2012

IMPROVED ESTIMATION STRATEGIES IN MULTIVARIATE MULTIPLE REGRESSION MODELS

Shabnam Chitsaz
University of Windsor

IMPROVED ESTIMATION STRATEGIES IN MULTIVARIATE MULTIPLE REGRESSION MODELS

by

Shabnam Chitsaz

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

Improved Estimation Strategies in Multivariate Multiple Regression Models

by

Shabnam Chitsaz

APPROVED BY

Y. P. Chaubey, External Examiner
Concordia University

J. Pathak
Odette School of Business

A.A. Hussein
Department of Mathematics and Statistics

M. Hlynka
Department of Mathematics and Statistics

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

L. Rueda, Chair of Defense
School of Computer Science

April 20, 2012

Declaration of Co-Authorship

I. Co-Authorship Declaration

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

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

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

II. Declaration of Previous Publication

This thesis includes two papers that have been published, and another accepted. It will be published on May 31, 2013.

Thesis Chapter	Publication title/ full citation	Publication Status
Chapter 2	Ahmed, S. E., Chitsaz, S., Data-Based Adaptive Estimation in an Investment Model. *Communications in Statistics-Theory and Methods* , 40(19-20), 2011	Published
Chapter 3	Chitsaz, S., Ahmed, S. E., Shrinkage estimation for the regression parameter matrix in multivariate regression model. *Journal of Statistical Computation & Simulation.*, 82(2), 2012	Published
Chapter 3	Chitsaz, S., Ahmed, S. E., An improved estimation in regression parameter matrix in multivariate regression model. *Communications in Statistics - Theory and Methods.*	Submitted

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

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

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

Abstract

The objective of this dissertation is to study properties of improved estimators of the parameters of interest in two different multivariate regression models, analogous to the fixed-X and random-X scenarios of multiple regression and compare the performance of these estimators with the usual least square estimator. In general, we study restricted versions of the multivariate regression problem based upon constraining the relationship between Y and X in some way where they may be known or unknown to the researcher prior to statistical analysis.

Chapter two contains a study of the properties of improved estimation strategies for the parameters of interest in a capital asset pricing model under a general linear constraint. Asymptotic results of the suggested estimators include derivation of asymptotic bias, asymptotic mean square error, and asymptotic distributional risk. The asymptotic results demonstrate the superiority of the suggested estimation technique. A simulation study is conducted to assess the performance of the suggested estimators for large samples. Both simulation study and data example corroborate with the theoretical result.

In Chapter three, we consider a multivariate multiple regression model when X is a fixed matrix. Here, we propose shrinkage and preliminary test estimation strategies for the matrix of regression parameters in the presence of a natural linear constraint. We examine the relative performances of the suggested estimators under the candidate subspace based on a quadratic risk function and the results are shown. A simulation study is conducted to compare the performance of the suggested estimators and two data examples are also presented. Our analytical and numerical results show that the suggested estimators perform better than the unrestricted estimator under the candidate subspace.

In Chapter four, we consider a multivariate reduced rank regression model when X is random and we propose preliminary test and shrinkage estimation strategies. We investigate the asymptotic properties of the shrinkage and pretest estimators under a quadratic loss function and compare the performance of the suggested estimators under the candidate subspace and beyond. The methods are applied on a real data set for illustrative purposes and a simulation study is also presented.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Ahmed, who guided me through the development of this research work. This dissertation would not have been completed without his continued academic advice.

I would like to thank the members of my Thesis Committee, Dr. Hlynka and Dr. Hussein for offering constructive suggestions, and to Dr. Pathak from Odette School of Business. Also I am grateful to Dr. Chaubey from University of Concordia for his valuable comments and advice.

I would also like to thank my wonderful fiance and my best friend Mohammadjavd. His support, encouragement, and understanding have turned the last year of my study into a pleasure. We proved that distance cannot, and will not hurt a bond between two people that is based on mutual respect, trust, commitment, and love.

Finally, my deepest gratitude goes to my mother Batoul Pirdel and my father Mahmood Chitsaz, for their love, encouragement, and continuous support. They helped me to believe in myself. Mom and dad, I love you, I am extremely grateful to what you have done for me and I am here because of you!

Contents

List of Figures

List of Tables

Abbreviations

ADB	Asymptotic distributional bias
ADR	Asymptotic distributional risk
AIC	Akaike information criterion
$AQDB$	Asymptotic quadratic distributional bias
$ANOVA$	Analysis of variance
$AMSE$	Asymptotic mean square error
B	Bias
BIC	Bayesian information criterion
$CAPM$	Capital asset pricing model
$GMANOVA$	Generalized multivariate analysis of variance
LS	Least square
$MANOVA$	Multivariate analysis of variance
$MMRM$	Multivariate multiple regression model
MSE	Mean square error
PT	pretest estimator
QB	Quadratic bias
R	Risk
RE	Restricted estimator
$R.E$	Relative efficiency
$RRRM$	Reduced rank regression model
UE	Unrestricted estimator
SCL	Security characteristic line
tr	Trace of matrix

List of Symbols

$\boldsymbol{\theta}$	Regression parameter vector, (intercept), asset's return
$\boldsymbol{\beta}$	Regression parameter vector, (slope), asset's return systematic risk
R_i	Return for stock i
R_m	Return on the market portfolio
R_f	return on treasury note
ε	Asset's nonsystematic risk
$\boldsymbol{H}$	$q \times p$ matrix
$\boldsymbol{h}$	$q \times 1$ vector
$\boldsymbol{K}$	$r \times m$ matrix
$\boldsymbol{L}$	$q \times n$ matrix
$\boldsymbol{A}$	$m \times r$ matrix
$\boldsymbol{B}$	$r \times q$ matrix
$\boldsymbol{\mathcal{F}}$	$r_1 \times r$ matrix
$\boldsymbol{\mathcal{G}}$	$q \times m$ matrix
$\boldsymbol{\mathcal{D}}$	$r_1 \times m$ matrix
$\boldsymbol{V}$	Variance covariance matrix
H_0	Null hypothesis
$\boldsymbol{I}$	Identity matrix
$\mathcal{L}$	Weighted quadratic loss function
$\boldsymbol{\Omega}$	Variance covariance matrix of ε
$\boldsymbol{\Sigma}^*$	Variance covariance matrix of restricted estimator
$\boldsymbol{\Omega}^*$	Variance covariance matrix of difference of unrestricted and restricted estimator

$\boldsymbol{\Sigma}_{12}$	Covariance matrix between UE and RE
$\boldsymbol{\Omega}_{12}$	Covariance matrix between UE and difference of UE and RE
$\hat{\boldsymbol{\beta}}$	Unrestricted estimator of regression parameter vector in CAPM
$\tilde{\boldsymbol{\beta}}$	Restricted estimator of regression parameter vector in CAPM
$\hat{\boldsymbol{\beta}}^{PT}$	Preliminary test estimator of regression parameter vector in CAPM
$\hat{\boldsymbol{\beta}}^{JS}$	Shrinkage estimator of regression parameter vector in CAPM
$\hat{\boldsymbol{\beta}}^{JS+}$	Positive shrinkage estimator of regression parameter vector in CAPM
$\hat{\boldsymbol{C}}$	Unrestricted estimator of regression parameter matrix in MMRM
$\tilde{\boldsymbol{C}}$	Restricted estimator of regression parameter matrix in MMRM
$\hat{\boldsymbol{C}}^{PT}$	Preliminary test estimator of regression parameter matrix in MMRM
$\hat{\boldsymbol{C}}^{JS}$	Shrinkage estimator of regression parameter matrix in MMRM
$\hat{\boldsymbol{C}}^{JS+}$	Positive shrinkage estimator of regression parameter matrix in MMRM
$\hat{\boldsymbol{B}}$	Unrestricted estimator of regression parameter matrix in RRRM
$\tilde{\boldsymbol{B}}$	Restricted estimator of regression parameter matrix in RRRM
$\hat{\boldsymbol{B}}^{PT}$	Preliminary test estimator of regression parameter matrix in RRRM
$\hat{\boldsymbol{B}}^{JS}$	Shrinkage estimator of regression parameter matrix in RRRM
$\hat{\boldsymbol{B}}^{JS+}$	Positive shrinkage estimator of regression parameter matrix in RRRM

Chapter 1

Introduction and Literature Review

1.1 Introduction

Multivariate multiple regression models (MMRMs) are generalizations of the usual multiple regression models when several response variables have to be predicted based on a set of predictor variables. MMRMs have recently found a wide range of applications in a variety of areas such as artificial intelligence, machine learning theory, education and psychology. (see for example Izenman (2008) and Timm (2002))

Regression analysis includes many techniques for modeling the relationships among variables and estimating the parameters of the model. When an estimator is obtained based on sample data only, it is well known that the maximum likelihood estimation leads to the best estimate among linear unbiased estimates. We call it an unrestricted maximum likelihood estimator. However, in problems of statistical inference, some-times we deal with uncertain prior information or some constraints on some of the parameters in a statistical model, which usually leads us to an improved inference based on alternative methods. Now the question arises as to how one can insert this uncertain prior information into the inference procedure. In this regard, Bancroft (1944) came up with the idea of testing the uncertainty of the prior information in the estimation procedure. It is reasonable to perform a pretest or pretest on the validity of the uncertain prior information and then analyze its development based on the result of the test. The new estimator that uses uncertain prior information to find improved estimates is called a restricted estimator. For examples on the results from many researchers, see Ahmed (2001), Ahmed et al. (2007), Ahmed and Chitsaz (2011), Chitsaz and Ahmed (2012b), Chitsaz and Ahmed (2012a), and others. We believe that the restricted estimator is more efficient than the unrestricted estimator after using prior information. Recent studies are mostly based on estimating the vector parameter.

The MMRM can be written as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{C}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}, \tag{1.1}$$

where, $\boldsymbol{\mathcal{Y}}$ is a full rank $n \times m$ matrix of response variables, $\boldsymbol{\mathcal{X}}$ is a full rank $q \times n$ matrix of predictor variables, $\boldsymbol{C}$ is a full rank $m \times q$ regression coefficients matrix, and $\boldsymbol{\mathcal{E}}$ is the $n \times m$ matrix of random errors. Linear restrictions on the regression coefficients are of such importance in estimation and testing that a special symbolism has been worked out. For example, in the analysis of variance, the analyst is concerned with whether treatment effects are equal, and the economist is often concerned about whether one or more parameters are zero. Similarly, researchers often wonder whether to pool data such as cross-sections over time, or whether linear combinations of coefficients are equal to a constant.

In this dissertation, we suggest some estimators for the parameter matrix in MMRM and we concentrate on estimating the parameter matrix, $\boldsymbol{C}$, under a very general set of linear constraints as prior information,

$$H_0 : \boldsymbol{KCL} = \mathbf{0}, \tag{1.2}$$

where $\boldsymbol{K}$ and $\boldsymbol{L}$ are known full-rank matrices of appropriate dimensions. Basically, we consider the estimation problem in two competing models, where one model includes all predictors and the other restricts variable coefficients to a linear constraint based on prior information. In this dissertation, we develop some improved estimation strategies such as pretest and shrinkage estimation methods for the matrix of a regression parameter in three different multivariate regression models.

Unrestricted and restricted estimator:

With a set of potential restrictions in mind as prior information , the researcher's attention is drawn to two estimators. Let $\hat{\boldsymbol{C}}$ be the ordinary (unrestricted) least square estimators of $\boldsymbol{C}$, and let $\tilde{\boldsymbol{C}}$ be the restricted estimator of $\boldsymbol{C}$ under a very general set of linear constraints as prior information named the candidate subspace (1.2). The estimator $\tilde{\boldsymbol{C}}$ has a smaller variance than the unrestricted estimator; if the restrictions are true, $\tilde{\boldsymbol{C}}$ is unbiased. Therefore, imposing false restrictions while reducing variance leads to bias, and the worse the restriction, the worse the bias. In many cases, researchers may have restrictions in mind such as pooling data, dropping variables, and so on. They may not be certain whether the restrictions are valid, or they may wish the data reveal something about the truth or falsity of the restrictions. A common practice in such situations is to test the restrictions as a statistical hypothesis.

Pretest estimator:

Let D be a test statistic for the null hypothesis in (1.2) and $d_{rn,\alpha}$ be the critical value of the distribution of D under the null hypothesis. We define the following pretest estimator:

$$\hat{\boldsymbol{C}}^{PT} = \hat{\boldsymbol{C}} - (\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})I(D < d_{rn,\alpha}),$$

where the $d_{rn,\alpha}$ is the upper α-level critical value of the χ^2 distribution with rn degrees of freedom, and $I(A)$ is an indicator function of a set A. There is always the chance of accepting a false hypothesis or rejecting a true hypothesis. When the analyst acts as if the restrictions are true, he runs a risk associated with a type two error; if the restrictions are false, he runs a risk associated with the other type of error involved in hypothesis testing. In general, the pretest estimator is biased, since $\tilde{\boldsymbol{C}}$ is biased if the restrictions are false, because there is a nonzero probability of accepting false restrictions via the test. However, the performance of this estimator is substantially better than the unrestricted estimator when uncertain prior information is nearly correct. Some useful literature about this estimator can be found in Bancroft (1944), Albertson (1991), and Ahmed (2001).

Estimators that are better in squared error loss than pretest estimators exist. Hence, James-Stein type estimators are defined and contrasted with pretest estimators.

Shrinkage and positive shrinkage estimator:

Following Ahmed and Krzanowski (2004), the shrinkage estimator of the regression parameters matrix based on a James-Stein type estimator is defined as

$$\hat{\boldsymbol{C}}^{JS} = \tilde{\boldsymbol{C}} + \{1 - cD^{-1}\}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}}), \quad c > 2,$$

where the optimal value of c is chosen in an interval in such a way that $\hat{\boldsymbol{C}}^{JS}$ dominates $\hat{\boldsymbol{C}}$. Note that the above estimator is derived by simply replacing the binary function of $I(A)$ by a continuous function cD^{-1}. Therefore, this estimator may have the opposite sign of $\hat{\boldsymbol{C}}$. To avoid that, we truncate $\hat{\boldsymbol{C}}^{JS}$ to obtain the positive shrinkage estimator which is defined as

$$\hat{\boldsymbol{C}}^{JS+} = \tilde{\boldsymbol{C}} + \{1 - cD^{-1}\}^{+}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}}), \quad c > 2.$$

If the candidate subspace as prior information is true, there is no issue since the imposition of a true candidate subspace reduces variances and does not cause bias. Imposition of a false candidate subspace introduces bias. Thus, the only way of making a judgment on listed estimators is to derive a risk function that assigns weights to bias and variance. We have studied the performance of the suggested estimators in terms of their risks. In an effort to provide the risk analysis, we considered the quadratic loss function of the form

$$\mathcal{L}(\boldsymbol{C}^*, \boldsymbol{C}) = [vec(\boldsymbol{C}^* - \boldsymbol{C})]' \boldsymbol{W} [vec(\boldsymbol{C}^* - \boldsymbol{C})],$$

where $\boldsymbol{W}$ is the positive semi-definite (p.s.d) matrix with an appropriate dimension. Then the risk of $\boldsymbol{C}^*$ or any estimator of $\boldsymbol{C}$ is

$$R(\boldsymbol{C}^*; \boldsymbol{W}) = tr[\boldsymbol{W} MSE(\boldsymbol{C}^*)], \tag{1.3}$$

where $MSE(\boldsymbol{C}^*) = E\{vec(\boldsymbol{C}^* - \boldsymbol{C})[vec(\boldsymbol{C}^* - \boldsymbol{C})]'\}$. For instance, if we get $MSE(\boldsymbol{C}^*) = (\boldsymbol{A} \otimes \boldsymbol{B})$ with a $\boldsymbol{A}$ and $\boldsymbol{B}$ nonsingular matrix, we define the quadratic risk as follows:

$$R(\boldsymbol{C}^*; \boldsymbol{W}) = tr(\boldsymbol{W}\boldsymbol{B})tr(\boldsymbol{A}).$$

1.2 Highlights of Contributions

The goal of this dissertation is to generalize estimation of the matrix parameter in MMRM. We describe two different multivariate regression scenarios, analogous to the fixed $\boldsymbol{\mathcal{X}}$ and random $\boldsymbol{\mathcal{X}}$ scenarios of multiple regression. We extend the concept of James-Stein type shrinkage estimation methods and pretest estimation in the context of three different linear regression models.

In Chapter two, we consider the simple multivariate regression model that includes basic investment models. For that, we have studied the capital asset pricing model which reflects how the expected return on an asset is a function of the expected returns on the market, the risk-free asset, and of the relevant risk of that asset. The goal of this model is to describe the properties of having an optimal portfolio given the best selection of stock for investors who like more return and less risk. Let us consider a system of regression models derived from a capital asset pricing model.

$$\boldsymbol{y}_t = \boldsymbol{\theta} + x_t\boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \qquad t = 1, \cdots, n, \tag{1.4}$$

where $\boldsymbol{y}_t$ is the $p \times 1$ vector of excess return on k assets; let x_t be the excess return on the market portfolio at time t. Here, the parameter $\boldsymbol{\beta}$ is the regression slope between the asset return and that of the market, which shows how a stock acts in relation to the market. Here, the goal is to maximize the performance of a portfolio when it is prior suspected that the asset's return, $\boldsymbol{\theta}$, may be restricted to a subspace. In this scenario, we are dealing with different estimation strategies for the parameters in a simple multivariate regression model. Here we consider alternative estimators of the slope parameter in a regression model with a non-normal error when uncertain prior information about the value of the intercept parameter is available and can be expressed in the general form of a null hypothesis, $\boldsymbol{H}_{q\times p}\boldsymbol{\theta}_{p\times 1} = \boldsymbol{h}_{q\times 1}$. We develop a large sample theory for the estimators that includes derivation of asymptotic bias and asymptotic distributional risk of the suggested estimators. The asymptotic results demonstrate the superiority of the suggested estimation technique. Also, a simulation study shows that the method suggested here has sound finite sample properties and strongly corroborates with the theoretical results of this chapter. A data example is also presented to illustrate the suggested estimation strategies.

In Chapter three, we generalize the estimation strategies for the matrix of a regression parameter in a multivariate multiple regression model in the presence of a natural linear constraint when the matrix of predictor variables $\boldsymbol{\mathcal{X}}$ is fixed and non stochastic. Also, we study the application of shrinkage and pretest estimation strategies in MMRM, which is the most important model for many practical situations. The goal is to critically examine the relative performances of the listed estimators in the direction of the subspace and candidate subspace restricted type estimators. In the case of multivariate multiple regression, we are dealing with the parameter matrix estimation. So, the fundamental results of Sclove et al. (1972) cannot be directly implemented to compute the expressions needed to check the validity and relative efficiency of proposed estimators under very general linear constraints. Therefore, we first generalized the results of Sclove et al. (1972) and then use them to derive those expressions for the suggested estimators. This chapter also addresses the pairwise comparisons of the proposed estimators. Our analytical and numerical results show that, overall, the proposed shrinkage estimators perform the best. The methods are also applied on a real data set for illustrative purposes.

However, there are many problems in multivariate statistical analysis that involve a test concerning regressions of reduced rank and restrictions on regressions. Therefore, the special feature that can be entered into the multivariate linear regression model case is that we admit the possibility that the rank of the regression coefficient matrix can be deficient. This implies that there are linear restrictions on the coefficient matrix, and these restrictions themselves are often not known a priori. Such a model is called a reduced rank regression model. The model structure and estimation strategies for this model will be explicitly discussed in Chapter four of this dissertation. Note that in this chapter we consider the predictor variables $\boldsymbol{\mathcal{X}}$ to be random.

In model (1.2) when $\boldsymbol{C}$ has reduced rank r, there exist two non-unique full rank matrices: an $(m \times r)$ matrix $\boldsymbol{A}$ and an $(r \times q)$ matrix $\boldsymbol{B}$, such that $\boldsymbol{C} = \boldsymbol{AB}$. We restate the model in (1.2) as a reduced rank regression model, such that

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{AB}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}. \tag{1.5}$$

The above mentioned features have practical implications. When we model with a large number of response and predictor variables, the implication in terms of restrictions serves a useful purpose. Certain linear combinations of response variables can, eventually, be ignored for regression modeling purposes, since these combinations will be found to be unrelated to the predictor variables. The alternative implication indicates that only certain linear combinations of the predictor variables need to be used in the regression model, since any remaining linear combinations may be found to have no influence on the response variables given the first set of linear combinations. Thus, the reduced rank regression procedure takes care of the dimension reduction aspect of a multivariate regression model when building through the assumption of a lower rank for the regression coefficient matrix. Statistical problems concerning reduced rank regression models have been studied in the statistical literature by Anderson (1951, 1984), Srivastava and Khatri (1979), (see Chapters 5 and 6), Muirhead and Koole (1982), (see Chapter 10), Reinsel (1998), Heinen and Rengifo (2007), Vounou et al. (2010), and others.

Therefore, in many practical situations, there is a need to reduce the number of parameters in model (1.2), and we approach this problem through the assumption of a lower rank of the matrix $\boldsymbol{B}$ in model (1.5) caused by linear constraints defined by

$$\boldsymbol{\mathcal{F}}\boldsymbol{B}\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}}. \tag{1.6}$$

In Chapter four, we consider shrinkage and pretest estimators in multivariate reduced rank regression model. We investigate the asymptotic properties of suggested estimators under a very general candidate subspace. In the support of our analytical results, we present a data example and simulation study.

Chapter five summarizes the results, and concludes the dissertation with a discussion on related research and the direction for future research.

Chapter 2

Data Based Adaptive Estimation in an Investment Model

2.1 Introduction

The capital asset pricing model (CAPM), or Sharpe-Lintner model, stands out among asset pricing models. This model reflects how the expected return on an asset is a function of the expected returns on the market, the risk-free asset, and the relevant risk of that asset. The goal of the CAPM is to describe the properties of having an optimal portfolio given the best selection of stocks for any investors who like more return and less risk. The Portfolio theory describes the process by which investors seek the best possible portfolio in terms of the tradeoff of risk for return. Portfolio management involves deciding what assets to include in the portfolio, how many to purchase, and when to purchase them. For this purpose, Jensen (1968) studied a regression model of the CAPM, given below:

$$R_{it} - R_f = \theta_i + \beta_i(R_{mt} - R_f) + \varepsilon_{it},$$

where R_{it} is the return for stock i in period t, R_f is the return of treasury note, θ_i is an asset's return in a access of it's risk adjusted, β_i is an asset's systematic risk for stock i, R_m is the return on the market portfolio in period t, and ε_{it} is the asset's nonsystematic risk in period t. In this chapter, we consider a system of regression models derived from a capital asset pricing model.

$$\boldsymbol{y}_t = \boldsymbol{\theta} + x_t\boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \qquad t = 1, \cdots, n. \tag{2.1}$$

Let $\boldsymbol{y}_t$ be the $p \times 1$ vector of excess return on k assets, and let x_t be the excess return on the market portfolio at time t. For the inference purpose, we assume that $E(\boldsymbol{\epsilon}_t) = 0$, $Cov(\boldsymbol{\epsilon}_t) = \boldsymbol{\Omega}$, and $E(x_t\boldsymbol{\epsilon}_t) = 0$. Here, the parameter $\boldsymbol{\beta}$ is the regression slope between the asset return and that of the market (security characteristic line, (SCL)), which shows how a stock acts in relation to the market. The measure of the sensitivity of the asset return to the market movement is given so that market variance is equal for all the assets. We will call a security an "aggressive security" if its beta exceeds 1,"$\beta_i > 1$", and "defensive" if its beta falls below 1, "$\beta_i < 1$". The factor $\boldsymbol{\theta}$ of the i^{th} risk asset represents the difference between the expected return according to the observed reality and the expected return according to the CAPM theory. Now, if the estimated $\boldsymbol{\theta}$ is significantly positive or negative, then the given risk asset produces returns that are over or below the appropriate values following the theory. Thus, in the market, the asset seems to be either underestimated or overestimated, respectively. As noted before, a portfolio is efficient when it yields a higher average return for a given risk, and a lower risk for a determined average return. It would be beneficial, if we have some preliminary information about $\boldsymbol{\theta}$, to have better estimation for the systematic risk and more efficient portfolio with more returns. The goal of this chapter is to maximize the performance of a portfolio when it is prior suspected that the asset's return, $\boldsymbol{\theta}$, may be restricted to a subspace. Ahmed and Krzanowski (2004) have considered estimation of the intercept vector in a simple multivariate normal regression model when it is a priori suspected that the slope vector may be restricted to a subspace. In this chapter, we investigate this problem when there is no assumption about the error term, and the prior information about the value of the intercept parameter can be expressed in the general form of a null hypothesis, $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{h}$.

2.2 Candidate Subspace

Let the candidate subspace be defined by $\boldsymbol{H}_{q\times p}\boldsymbol{\theta}_{p\times 1} = \boldsymbol{h}_{q\times 1}$ and $\boldsymbol{\Omega} = \sigma^2\boldsymbol{V}$. When $\boldsymbol{V}$ is known and nonsingular, then the weighted least square estimators (WLSE) of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are given by

$$\hat{\boldsymbol{\beta}} = \sum_{t=1}^{n}(x_t - \bar{x})\boldsymbol{y}_t / \sum_{t=1}^{n}(x_t - \bar{x})^2$$

and

$$\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{y}} - \bar{x}\hat{\boldsymbol{\beta}}.$$

However, even when $\boldsymbol{V}$ is unknown, the estimator of $\boldsymbol{\beta}$ does not depend on $\boldsymbol{V}$; $\boldsymbol{V}$ drops out of since the covariate is scalar. Now, considering the problem of finding $\tilde{\boldsymbol{\beta}}$ that minimizes the following expression subject to the constraint, we form the

Lagrangian function where $\boldsymbol{\lambda}$ is an $q \times 1$ vector of Lagrange multipliers:

$$\ell = \sum_t (\boldsymbol{y}_t - \boldsymbol{\theta} - x_t\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}_t - \boldsymbol{\theta} - x_t\boldsymbol{\beta}) + 2\boldsymbol{\lambda}'(\boldsymbol{H}\boldsymbol{\theta} - \boldsymbol{h}).$$

Differentiation with respect to $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ yields the following results:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}),$$

where

$$\boldsymbol{C} = d\boldsymbol{V}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}')^{-1}$$

and

$$d = \frac{n\bar{x}}{\sum_{t=1}^{n}(x_t - \bar{x})^2}.$$

Note that $\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'$ is invertible. We call $\tilde{\boldsymbol{\beta}}$ a candidate sub-model or a restricted estimator of $\boldsymbol{\beta}$. $\tilde{\boldsymbol{\beta}}$ will be equal to $\hat{\boldsymbol{\beta}}$, an unbiased estimator, if the subspace information is correct i.e. $\boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{h}$. Therefore, $\tilde{\boldsymbol{\beta}}$ will be a biased estimator if the subspace information is not correct. On the other hand, it will be relatively more efficient than the classical estimator $\hat{\boldsymbol{\beta}}$ when such subspace information represents the data.

A useful but compromising method for tackling the uncertainty regarding the subspace information is to implement estimation strategies based on shrinkage and pretest principles. For point estimation, we refer to Ahmed (2001), Khan and Ahmed (2003), and Ahmed et al. (2010), among others. For the purpose of statistical inference in such cases, one could employ an empirical Bayes approach to the computation of standard errors of these shrinkage estimators; for example, see Maddala et al. (1997). Alternatively, Kazimi and Brownstone (1999) proposed confidence bands for shrinkage estimators using a simple percentile bootstrapping method. Wan et al. (2003)

have proposed the use of mean squared error matrices with a class of shrinkage estimators for the purposes of constructing confidence ellipsoids.

The remainder of this chapter is organized as follows. In Section 3, pretest and shrinkage estimators are defined. In Section 4, we derive the expressions for asymptotic bias and risk of the proposed estimators, and provide their relative performances. Section 5 provides a simulation study and a real data example. Conclusions are offered in Section 6. Finally the proof of the main results are provided in Section 7.

2.3 Proposed Estimation Strategies

Here, we consider pretest and shrinkage estimations of the regression parameter vector $\boldsymbol{\beta}$. The pretest estimator is defined as

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}} I(P > \chi^2_{q,\alpha}) + \tilde{\boldsymbol{\beta}} I(P \leq \chi^2_{q,\alpha}),$$

where $I(A)$ is an indicator function of the set A, and P is the test statistic for testing $\boldsymbol{H\theta} = \boldsymbol{h}$ as given by

$$P = s_e^2(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' \boldsymbol{M}^{-1} (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$$

where

$$\boldsymbol{M} = (\frac{1}{n} + \frac{\bar{x}^2}{Q}) \boldsymbol{CHVH'C'}.$$

When the subspace information is true, ($\boldsymbol{H\theta} = \boldsymbol{h}$), the statistic P follows χ^2 distribution with q degrees of freedom as $n \to \infty$. The James-Stein or shrinkage estimator as a smooth function of P is given by

$$\hat{\boldsymbol{\beta}}^{JS} = \tilde{\boldsymbol{\beta}} + (1 - mP^{-1})(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}), \qquad q > 2.$$

Notice that $\hat{\boldsymbol{\beta}}^{JS}$ is similar to $\hat{\boldsymbol{\beta}}^{PT}$ where we have replaced the indicator function $I(P \leq \chi^2_{q,\alpha})$ by a continuous decreasing function mP^{-1} of P. Thus, instead of two extreme choices, namely $\hat{\boldsymbol{\beta}}$ or $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{JS}$ provides the choice for all values between $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ depending on the value of P for a given sample. Here, m is the shrinkage constant and is chosen in an interval in such a way that $\hat{\boldsymbol{\beta}}^{JS}$ dominates $\hat{\boldsymbol{\beta}}$ in terms of risk. m is allowed to vary over $[0, 2(q-2)), q > 2$, often set to $m = q - 2$; thus, we assume that $q \geq 3$. We can see that $\hat{\boldsymbol{\beta}}^{JS}$ tends to $\hat{\boldsymbol{\beta}}$ as P tends to infinity, and it tends to $\tilde{\boldsymbol{\beta}}$ as $P \to q - 2$. Finally, the positive-part shrinkage estimator is

$$\hat{\boldsymbol{\beta}}^{JS+} \quad = \quad \tilde{\boldsymbol{\beta}} + (1 - mP^{-1})^{+}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}),$$

where $z^{+} = max(0, z)$, or, equivalently, as

$$\hat{\boldsymbol{\beta}}^{JS+} = \tilde{\boldsymbol{\beta}} + (1 - mP^{-1})I(P > m)(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}).$$

Having defined all these estimators, we investigate their asymptotic properties in the following section.

2.4 Main Results

Let $\boldsymbol{\nu}_1 = \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, $\boldsymbol{\nu}_2 = \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$, and $\boldsymbol{\nu}_3 = \sqrt{n}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$. To establish the asymptotic properties of listed estimators, we consider the local alternatives to guarantee convergence and overcome the difficulty of identical asymptotic distribution of some listed estimators in large samples under fixed alternatives. To do so, we consider a sequence of local alternatives $\{L_n\}$ defined by $L_n : \boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{h} + \frac{\boldsymbol{\xi}}{\sqrt{n}}$, where $\boldsymbol{\xi}$ is a real fixed vector. Consider model (2.1), where $\boldsymbol{\epsilon}$ is not normally distributed. Therefore, we need the following three regularity conditions for the asymptotic normality of $(\boldsymbol{\theta}, \boldsymbol{\beta})$ as $n \to \infty$.

Theorem 2.4.1. Consider the simple regression model when the components of the error vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ are independent, $E(\boldsymbol{\epsilon}_t) = 0$, $Cov(\boldsymbol{\epsilon}_t) = \sigma^2 \boldsymbol{V}$, and the distribution of $\boldsymbol{\epsilon}$ is non-normal. Now assume the following regularity assumptions:

(i) $\lim_{n\to\infty} \bar{x} = \bar{x}_0, \qquad |\bar{x}_0| < \infty$

(ii) Let $q_t = x_t - \frac{\bar{x}}{\sqrt{Q}}$ then $\max_{1\le t\le n} q_t^2 \to 0$ as $n \to \infty$.

(iii) Let $Q = \sum_{t=1}^{n}(x_t - \bar{x})^2$. Then $\lim_{n\to\infty} n^{-1}Q = Q_0 < \infty$.

Then,

$$\begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix} \sim N_{2p}\left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma^2 \boldsymbol{V} \begin{pmatrix} (1 + \frac{\bar{x}^2}{Q_0}) & -\frac{\bar{x}}{Q_0} \\ -\frac{\bar{x}}{Q_0} & Q_0^{-1} \end{pmatrix} \right\}.$$

Theorem 2.4.2. Under assumed regularity conditions given in Theorem 2.4.1 and $\{L_n\}$, we have

$$\text{(i)} \quad \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix} \sim N_{2p}\left\{ \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\gamma} \end{pmatrix}, \sigma^2 \begin{pmatrix} Q_0^{-1}\boldsymbol{V} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}^* \end{pmatrix} \right\}$$

$$\text{(ii)} \quad \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_3 \end{pmatrix} \sim N_{2p}\left\{ \begin{pmatrix} \mathbf{0} \\ -\boldsymbol{\gamma} \end{pmatrix}, \sigma^2 \begin{pmatrix} Q_0^{-1}\boldsymbol{V} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}^* \end{pmatrix} \right\}$$

where $\boldsymbol{\gamma} = \boldsymbol{C\xi}$, $Q = \sum_{t=1}^{n}(x_t - \bar{x})^2$, $\boldsymbol{\Sigma}^* = Q_0^{-1}\boldsymbol{V} + (1 + \frac{\bar{x}^2}{Q_0})\boldsymbol{CHVH'C'} - 2\bar{x}Q_0^{-1}\boldsymbol{CHV}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = Q_0^{-1}\boldsymbol{V} - \bar{x}Q_0^{-1}\boldsymbol{CHV}$, $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}' = \bar{x}Q_0^{-1}\boldsymbol{CHV}$, and $\boldsymbol{\Omega}^* = (1 + \frac{\bar{x}^2}{Q_0})\boldsymbol{CHVH'C'}$.

Proof: See Appendix, Section 2.7.1.

2.4.1 Asymptotic Bias and Risk Analysis

In this section, we obtain expressions for the asymptotic distributional bias (ADB) and the risks (ADR) of the proposed estimators. Also, we compare the performance of the suggested estimators in terms of asymptotic bias and risk, respectively. Here, we present the expression for the asymptotic distribution bias (ADB) of the proposed estimators. The ADB of an estimator $\boldsymbol{\beta}^*$ is defined as

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

To study the asymptotic quadratic risks of the estimators, we define a quadratic loss function using a positive definite matrix (p.d.m.) $\boldsymbol{W}$, namely

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

where $\boldsymbol{\beta}^*$ is any one of $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{PT}$, $\hat{\boldsymbol{\beta}}^{JS}$, and $\hat{\boldsymbol{\beta}}^{JS+}$.

Now we assume that for the estimator $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$ the cumulative distribution function of $\boldsymbol{\beta}^*$ under $\{L_n\}$ exists and can be denoted as $F(\boldsymbol{x}) = \lim_{n\to\infty} P\{\sqrt{n}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \leq \boldsymbol{x}|L_n\}$, where $F(\boldsymbol{x})$ is nondegenerate. Then, the ADR of $\boldsymbol{\beta}^*$ is defined as

$$ADR(\boldsymbol{\beta}^*, \boldsymbol{W}) = \text{tr}\left\{ \boldsymbol{W} \int_{\mathcal{R}^{p_1}} \int \boldsymbol{x}\boldsymbol{x}' dF(\boldsymbol{x}) \right\} = \text{tr}(\boldsymbol{W}\boldsymbol{Z}),$$

where $\boldsymbol{Z}$ is the dispersion matrix for the asymptotic distribution $F(\boldsymbol{x})$. We say that $\hat{\boldsymbol{\beta}}$ dominates $\hat{\boldsymbol{\beta}}^\star$ for all $\boldsymbol{\beta}$, if $ADR(\hat{\boldsymbol{\beta}}; \boldsymbol{W}) < ADR(\hat{\boldsymbol{\beta}}^\star; \boldsymbol{W})$.

Theorem 2.4.3. Under $\{L_n\}$ the asymptotic distribution biases (ADB) of the proposed estimators are respectively

$$
\begin{aligned}
ADB(\tilde{\boldsymbol{\beta}}) &= \boldsymbol{\gamma} \\
ADB(\hat{\boldsymbol{\beta}}^{PT}) &= \boldsymbol{\gamma} H_{q+2}(\chi_q^2(\alpha), \Delta) \\
ADB(\hat{\boldsymbol{\beta}}^{JS}) &= m\boldsymbol{\gamma} E(\chi_{q+2}^{-2}(\Delta)) \\
ADB(\hat{\boldsymbol{\beta}}^{JS+}) &= ADB(\hat{\boldsymbol{\beta}}^{JS}) - \boldsymbol{\gamma} E\{(m\chi_{q+2}^{-2}(\Delta) - 1)I(\chi_{q+2}^2(\Delta) < m)\},
\end{aligned}
$$

where $\boldsymbol{\gamma} = \boldsymbol{C\xi}$, $m = q - 2$, and the notation $H_\nu(x; \Delta)$ is the distribution function of non-central chi-square distribution with ν degrees of freedom and non-centrality parameter $\Delta = Q_0 \sigma^{-2}(\boldsymbol{\gamma}' V^{-1} \boldsymbol{\gamma})$.

Proof: See Appendix, Section 2.7.2.

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

$$AQDB(\boldsymbol{\beta}^*) = Q_0 \sigma^{-2}[ADB(\boldsymbol{\beta}^*)]' \boldsymbol{V}^{-1}[ADB(\boldsymbol{\beta}^*)].$$

Based on the above, we can easily obtain the AQDB of the estimators.

$$
\begin{aligned}
AQDB(\tilde{\boldsymbol{\beta}}) &= \Delta \\
AQDB(\hat{\boldsymbol{\beta}}^{PT}) &= \Delta\{H_{q+2}(\chi_q^2(\alpha), \Delta)\}^2 \\
AQDB(\hat{\boldsymbol{\beta}}^{JS}) &= m^2 \Delta\{E(\chi_{q+2}^{-2}(\Delta))\}^2 \\
AQDB(\hat{\boldsymbol{\beta}}^{JS+}) &= \Delta\{mE(\chi_{q+2}^{-2}(\Delta)) - E\{(m\chi_{q+2}^{-2}(\Delta) - 1)I(\chi_{q+2}^2(\Delta) < m)\}^2.
\end{aligned}
$$

Clearly, the asymptotic bias of $\tilde{\boldsymbol{\beta}}$ is unbounded, and the bias of $\hat{\boldsymbol{\beta}}^{PT}$ depends on the size of α and Δ. The asymptotic bias of $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JS^+}$ depend on Δ alone. Thus, we can establish the following relation:

$$0 = AQDB(\hat{\boldsymbol{\beta}}) \leq AQDB(\hat{\boldsymbol{\beta}}^{JS^+}) \leq AQDB(\hat{\boldsymbol{\beta}}^{JS}) \leq AQDB(\hat{\boldsymbol{\beta}}^{PT}) \leq AQDB(\tilde{\boldsymbol{\beta}}).$$

Theorem 2.4.4. Under $\{L_n\}$, the asymptotic covariance matrices (AMSE) of the estimators are as follows:

$$\begin{aligned}
AMSE(\hat{\boldsymbol{\beta}}) &= Q_0^{-1}\boldsymbol{V} \\
AMSE(\tilde{\boldsymbol{\beta}}) &= Q_0^{-1}\boldsymbol{V} + \boldsymbol{G} - 2\boldsymbol{F} + \boldsymbol{\gamma\gamma}' \\
AMSE(\hat{\boldsymbol{\beta}}^{PT}) &= Q_0^{-1}\boldsymbol{V} + [\boldsymbol{G} - 2\boldsymbol{F}]H_{q+2}(\chi_q^2(\alpha), \Delta) \\
&+ \boldsymbol{\gamma\gamma}'\{-2\boldsymbol{A}H_{q+2}(\chi_q^2(\alpha), \Delta) \\
&- 2\boldsymbol{A}H_{q+4}(\chi_q^2(\alpha), \Delta) + H_{q+4}(\chi_q^2(\alpha), \Delta)\} \\
AMSE(\hat{\boldsymbol{\beta}}^{JS}) &= Q_0^{-1}\boldsymbol{V} + m^2\boldsymbol{G}E(\chi_{q+2}^{-4}(\Delta)) - 2m\boldsymbol{F}[E(\chi_{q+2}^{-2}(\Delta))] + m\boldsymbol{\gamma\gamma}' \\
& \quad [-2\boldsymbol{A}E(\chi_{q+4}^{-2}(\Delta)) - 2\boldsymbol{A}E(\chi_{q+2}^{-2}(\Delta)) + mE(\chi_{q+4}^{-4}(\Delta))] \\
AMSE(\hat{\boldsymbol{\beta}}^{JS+}) &= AMSE(\hat{\boldsymbol{\beta}}^{JS}) - \boldsymbol{G}[E(1 - m\chi_{q+2}^{-2}(\Delta))^2 I(\chi_{q+2}^2(\Delta) < m)] \\
&+ \boldsymbol{\gamma\gamma}'\{2E(1 - m\chi_{q+2}^{-2}(\Delta))I(\chi_{q+2}^2(\Delta) < m) - E(1 - m\chi_{q+4}^{-2}(\Delta))^2 \\
&\times I(\chi_{q+4}^2(\Delta) < m)\},
\end{aligned}$$

where $\boldsymbol{A} = \boldsymbol{G}^{-1}\boldsymbol{F}$, $\boldsymbol{G} = \boldsymbol{\Omega}^*$, and $\boldsymbol{F} = \boldsymbol{\Omega}_{12}$.

Proof: See Appendix, Section 2.7.3.

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

Theorem 2.4.5. Under $\{L_n\}$, the asymptotic distributional risks (ADR) are as follows:

$$
\begin{aligned}
ADR(\hat{\boldsymbol{\beta}};\boldsymbol{W}) &= Q_0^{-1}tr(\boldsymbol{WV}),\\
ADR(\tilde{\boldsymbol{\beta}};\boldsymbol{W}) &= Q_0^{-1}tr(\boldsymbol{WV}) + a\times tr(\boldsymbol{Z}_{11}) - 2b\times tr(\boldsymbol{WCHV}) + \eta_1'\boldsymbol{Z}_{11}\eta_1\\
ADR(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{W}) &= Q_0^{-1}tr(\boldsymbol{WV}) + [a\times tr(\boldsymbol{Z}_{11}) - 2b\times tr(\boldsymbol{WCHV})]H_{q+2}(\chi_q^2(\alpha),\Delta)\\
&- 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{A})[H_{q+2}(\chi_q^2(\alpha),\Delta) + H_{q+4}(\chi_q^2(\alpha),\Delta)]\\
&+ \eta_1'\boldsymbol{Z}_{11}\eta_1 H_{q+4}(\chi_q^2(\alpha),\Delta))\\
ADR(\hat{\boldsymbol{\beta}}^{JS};\boldsymbol{W}) &= Q_0^{-1}tr(\boldsymbol{WV}) - 2mb\times tr(\boldsymbol{WCHV})E(\chi_{q+2}^{-2}(\Delta))\\
&- 2m\times tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{A})[E(\chi_{q+4}^{-2}(\Delta)) + E(\chi_{q+2}^{-2}(\Delta))]\\
&+ am^2\times tr(\boldsymbol{Z}_{11})E(\chi_{q+2}^{-4}(\Delta)) + m^2\eta_1'\boldsymbol{Z}_{11}\eta_1 E(\chi_{q+4}^{-4}(\Delta))]\\
ADR(\hat{\boldsymbol{\beta}}^{JS+};\boldsymbol{W}) &= ADR(\hat{\boldsymbol{\beta}}^{JS};\boldsymbol{W}) - a\times tr(\boldsymbol{Z}_{11})[E(1-m\chi_{q+2}^{-2}(\Delta))^2 I(\chi_{q+2}^2(\Delta)<m)]\\
&+ \eta_1'\boldsymbol{Z}_{11}\eta_1\{2E(1-m\chi_{q+2}^{-2}(\Delta))I(\chi_{q+2}^2(\Delta)<m) - E(1-m\chi_{q+4}^{-2}(\Delta))^2\\
&\times I(\chi_{q+4}^2(\Delta)<m)\},
\end{aligned}
$$

where $a = (1+\frac{\bar{x}^2}{Q_0})$ and $b = \bar{x}Q_0^{-1}$.

Proof: See Appendix, Section 2.7.4.

2.4.2 Comparison of $\hat{\boldsymbol{\beta}}^{JS+}$ and $\hat{\boldsymbol{\beta}}$

Let us consider the risk of $\hat{\boldsymbol{\beta}}^{JS+}$ under a subspace, in terms of the risk of $\hat{\boldsymbol{\beta}}$:

$$
\begin{aligned}
ADR(\hat{\boldsymbol{\beta}}^{JS+};\boldsymbol{W}) &= ADR(\hat{\boldsymbol{\beta}};\boldsymbol{W}) - 2mb\times tr(\boldsymbol{WCHV})E(\chi_{q+2}^{-2}(0))\\
&+ a\times tr(\boldsymbol{Z}_{11})\{m^2\times E(\chi_{q+2}^{-4}(0)) - E[(1-m\chi_{q+2}^{-2}(0))^2]\\
&\quad I(\chi_{q+2}^2(0)<m)\}.
\end{aligned}
$$

Since $\{E(1-m\chi_{q+2}^{-2}(0))^2 I(\chi_{q+2}^2(0)<m)\} \leq E[(1-m\chi_{q+2}^{-2}(0))^2]$ and the expectation of a positive random variable is positive, then, for all m satisfying the condition

$$tr(\boldsymbol{Z}_{11}) > \frac{2mb \times tr(\boldsymbol{WCHV})E(\chi_{q+2}^{-2}(0))}{a[m^2 \times E(\chi_{q+2}^{-4}(0)) - E(1-m\chi_{q+2}^{-2}(0))^2]},$$

$\hat{\boldsymbol{\beta}}$ performs better than $\hat{\boldsymbol{\beta}}^{JS+}$. However, as Δ increases, $\hat{\boldsymbol{\beta}}^{JS+}$ dominates $\hat{\boldsymbol{\beta}}$ outside an interval near the origin.

2.4.3 Comparison of $\hat{\boldsymbol{\beta}}^{JS+}$ and $\hat{\boldsymbol{\beta}}^{JS}$

For comparing the asymptotic risk of $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JS+}$, we consider the risk difference of them

$$\begin{aligned} ADR(\hat{\boldsymbol{\beta}}^{JS};\boldsymbol{W}) - ADR(\hat{\boldsymbol{\beta}}^{JS+};\boldsymbol{W}) &= a \times tr(\boldsymbol{Z}_{11})[E(1-m\chi_{q+2}^{-2}(\Delta))^2 I(\chi_{q+2}^2(\Delta)<m)] \\ &- \eta_1' \boldsymbol{Z}_{11} \eta_1 2E(1-m\chi_{q+2}^{-2}(\Delta)) I(\chi_{q+2}^2(\Delta)<m) \\ &+ \eta_1' \boldsymbol{Z}_{11} \eta_1 E(1-m\chi_{q+4}^{-2}(\Delta))^2 I(\chi_{q+4}^2(\Delta)<m), \end{aligned}$$

since the expectation of a positive random variable is positive and, by the definition of an indicator function, the first and last terms are positive. For the second term in the equation, we observe that $(0<\chi_{q+2}^2(\Delta)<m) \Longleftrightarrow (m\chi_{q+2}^{-2}(\Delta)-1) \geq 0$, we get $E[(1-m\chi_{q+2}^{-2}(\Delta))I(\chi_{q+2}^2(\Delta)<m)] \leq 0$. Thus, the second term is nonnegative too. Therefore, the risk of $\hat{\boldsymbol{\beta}}^{JS+}$ will be smaller than $\hat{\boldsymbol{\beta}}^{JS}$ for all Δ in $(0,\infty)$.

2.4.4 Comparison of $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{JS}$

We investigate the risk-difference of $\hat{\boldsymbol{\beta}}^{JS}$ and $\tilde{\boldsymbol{\beta}}$ under subspace is

$$ADR(\tilde{\boldsymbol{\beta}};\boldsymbol{W})-ADR(\hat{\boldsymbol{\beta}}^{JS};\boldsymbol{W}) = 2b\times tr(\boldsymbol{WCHV})(m-1)-am^2\times tr(\boldsymbol{Z}_{11})E(\chi_{q+2}^{-4}(0))$$

the risk of $\hat{\boldsymbol{\beta}}^{JS}$ is smaller than $\tilde{\boldsymbol{\beta}}$ when the condition

$$tr(\boldsymbol{WCHV}) > \frac{am^2\times tr(\boldsymbol{Z}_{11})E(\chi_{q+2}^{-4}(0))}{2b(m-1)}.$$

Thus, $\hat{\boldsymbol{\beta}}^{JS}$ does not dominate $\tilde{\boldsymbol{\beta}}$ under a subspace. However, for the large values of Δ, the reverse conclusion holds.

2.4.5 Comparison of $\hat{\boldsymbol{\beta}}^{PT}$ and $\hat{\boldsymbol{\beta}}$

The risk difference is given by

$$\begin{aligned} ADR(\hat{\boldsymbol{\beta}};\boldsymbol{W}) - ADR(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{W}) &= -[a\times tr(\boldsymbol{Z}_{11}) - 2b\times tr(\boldsymbol{WCHV})]H_{q+2}(\chi_q^2(\alpha),\Delta) \\ &+ 2tr(\boldsymbol{W\gamma\gamma}'\boldsymbol{A})[(H_{q+2}(\chi_q^2(\alpha),\Delta)+H_{q+4}(\chi_q^2(\alpha),\Delta))] \\ &- \eta_1'\boldsymbol{Z}_{11}\eta_1 H_{q+4}(\chi_q^2(\alpha),\Delta)) \end{aligned}$$

The right hand side is nonnegative whenever

$$\eta_1'\boldsymbol{Z}_{11}\eta_1 < \frac{[2b\times tr(\boldsymbol{WCHV}) - a\times tr(\boldsymbol{Z}_{11})]H_{q+2}(\chi_q^2(\alpha),\Delta)}{H_{q+4}(\chi_q^2(\alpha),\Delta)},$$

In this range, $\hat{\boldsymbol{\beta}}^{PT}$ performs better than $\hat{\boldsymbol{\beta}}$ as well as under the null hypothesis $ADR(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{W}) \leq ADR(\hat{\boldsymbol{\beta}};\boldsymbol{W})$, since the risk difference for all α is positive.

After comparing the ADR of all the estimators, we can see that

$$ADR(\hat{\boldsymbol{\beta}}^{JS+}; \boldsymbol{W}) \leq ADR(\hat{\boldsymbol{\beta}}^{JS}; \boldsymbol{W}) \leq ADR(\hat{\boldsymbol{\beta}}; \boldsymbol{W}).$$

Also comparing the $\hat{\boldsymbol{\beta}}^{PT}$ with $\hat{\boldsymbol{\beta}}$, we see that $ADR(\hat{\boldsymbol{\beta}}^{PT}; \boldsymbol{W}) \leq ADR(\hat{\boldsymbol{\beta}}; \boldsymbol{W})$.

2.5 Numerical Study

2.5.1 Simulation Study

In this section, we use Monte Carlo simulation experiments to examine the performance of the proposed estimators based on a moderate and a large sample methodology. In this study, we simulate data from the following model:

$$\begin{pmatrix} y_{t1} \\ y_{t2} \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + x_t \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \end{pmatrix} \qquad t = 1, \cdots, n.$$

For simulation, we consider $\boldsymbol{\theta} = (1.5, 2.5)$, $\boldsymbol{H} = ((1,0)', (0,1)')'$, and $\boldsymbol{h} = (1.5, 2.5)$. Under the candidate subspace, we generate 5000 samples using the above model, which is adequate since a further increase in the number of replications did not significantly change the result. We define Δ as a departure parameter which is a function of the distance between the true value of $\boldsymbol{\theta}$ and that under the null hypothesis. In order to investigate the behavior of the proposed estimators, different values of $\boldsymbol{\theta}$ were chosen to produce the value of Δ between 0 and 4. The performance of an estimator of $\boldsymbol{\theta}$ will be reappraised using the mean square error criterion. All computations were conducted using the **R** statistical system. We numerically calculated the relative risk of $\tilde{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}^{PT}$, $\hat{\boldsymbol{\beta}}^{JS}$, and $\hat{\boldsymbol{\beta}}^{JS+}$ with respect to $\hat{\boldsymbol{\beta}}$ by simulation. The simulated relative efficiency of the estimator $\boldsymbol{\beta}^*$ to the unrestricted $\hat{\boldsymbol{\beta}}$ is defined by $R.E = risk(\hat{\boldsymbol{\beta}})/risk(\hat{\boldsymbol{\beta}}^*)$. We applied our method to several simulated data sets, and the results were similar. Since the result for different n were similar, here we only report the results for $n = 30$ and $n = 100$ in Tables 2.1 and 2.2, and Figures (2.1) and (2.2).

Table 2.1: R.E of estimators, $n = 30$.

Δ	$\tilde{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	3.198	2.449	1.323	1.467
0.2	2.199	1.410	1.020	1.320
0.4	1.104	0.85	1.065	1.115
0.6	0.621	0.898	1.051	1.054
0.8	0.381	0.883	1.029	1.029
1.2	0.181	0.894	1.015	1.015
1.6	0.105	1.000	1.009	1.009
2.0	0.067	1.000	1.006	1.006
4.0	0.017	1.000	1.001	1.001

Table 2.2: R.E of estimators, $n = 100$.

Δ	$\tilde{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
0.0	2.225	1.877	1.260	1.356
0.2	0.809	0.797	1.040	1.042
0.4	0.284	0.998	1.010	1.010
0.6	0.134	1.000	1.004	1.004
0.8	0.077	1.000	1.002	1.002
1.2	0.036	1.000	1.001	1.001
1.6	0.020	1.000	1.001	1.001
2.0	0.013	1.000	1.000	1.000
4.0	0.003	1.000	1.000	1.000

We can see the relative efficiencies of the estimators change with the change in the

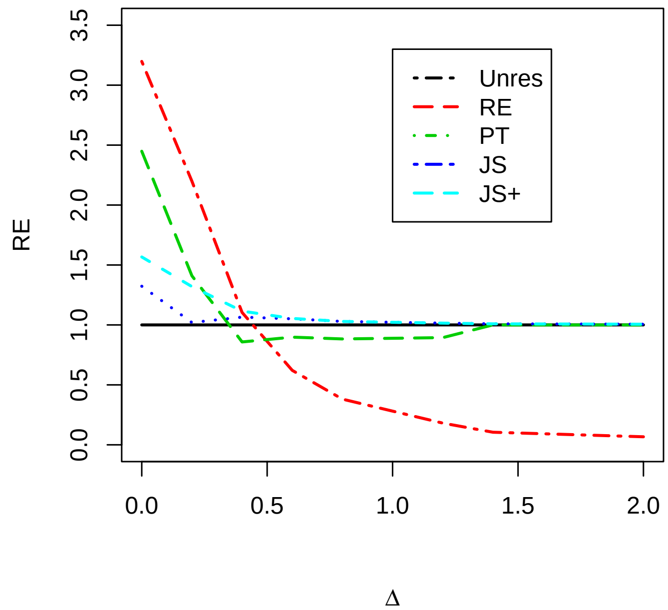

Figure 2.1: R.E of the estimators for $n = 30$.

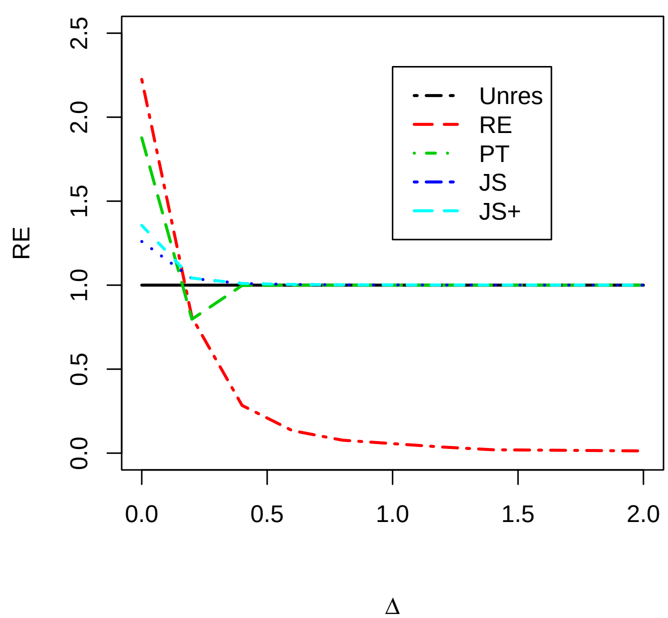

Figure 2.2: R.E of the estimators for $n = 100$.

value of the departure parameter Δ. The tables and figures reconfirm the typical characteristics of the listed estimators. We conclude that $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^{PT}$ dominate the usual $\hat{\boldsymbol{\beta}}$ at and near the candidate subspace. $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JS+}$ are more efficient than an unrestricted one in the unrestricted parameter space. If the candidate subspace is correctly specified, that is, $\Delta = 0$ or in the neighborhood of that, then the $\hat{\boldsymbol{\beta}}^{PT}$ is more efficient than $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JS+}$. However, for a larger value of α, the level of significance, $\hat{\boldsymbol{\beta}}^{JS+}$ dominates $\hat{\boldsymbol{\beta}}^{PT}$ uniformly. As the value of Δ grows, the $\hat{\boldsymbol{\beta}}^{PT}$ becomes more inefficient than the unrestricted one, and its efficiency value monotonically decreases, achieves a minimum after crossing the efficiency line at 1, and then monotonically increases and approaches to the $\hat{\boldsymbol{\beta}}$ efficiency. So $\tilde{\boldsymbol{\beta}}$ is more efficient than all the other estimators under the candidate subspace but, as Δ increases, its efficiency converges to zero since it is an unbounded function of Δ.

2.5.2 Real Data Example

A motivating example is the study of financial data taken from the Standard and Poors 500 (S&P) index. We consider nine of the largest mutual funds in the United States for the past thirty one years, from 1977 to 2007. Most data are cited from Chen and Wen (2004), while the information for 2005-2007 is gathered from Yahoo's finance website. We treat the nine funds' annual returns as response variables:
y_1-Washington Mutual Fund A (AWSHX), y_2-Fidelity Contra Fund (FCNTX), y_3-American Income Fund (AMECX), y_4-Dodqe and Cox Stock Fund(DODGX), y_5-New Perspective Fund A (ANWPX), y_6-Fidelity Puritan Fund (FPURX), y_7-Vauguard Windsor Fund (VWNDX), y_8-Janus family-janus fund (JANNSX), and y_9-Fidelity Equity Income Fund (FEQIX).

Table 2.3: Estimators of $\boldsymbol{\beta}$ for nine diversified funds

Fund	$\hat{\boldsymbol{\beta}}$	$\tilde{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}^{PT}$	$\hat{\boldsymbol{\beta}}^{JS}$	$\hat{\boldsymbol{\beta}}^{JS+}$
AWSHX	0.756583	1.1018080	0.9278829	0.8918164	0.8918164
FCNTX	0.93232	1.3925302	1.1918986	1.1125959	1.1125959
AMECX	0.4754531	0.6199427	0.5293787	0.5320533	0.5320533
DODGX	0.6753388	0.8373402	0.6385641	0.7387988	0.7387988
ANWPX	0.744376	0.8849010	0.8030956	0.7994232	0.7994232
FPURX	0.5409128	0.9847795	0.7452551	0.7147867	0.7147867
VWNDX	0.6280855	0.6637484	0.5257206	0.6420556	0.6420556
JANNSX	1.137417	1.0711357	1.1255026	1.1114528	1.1114528
FEQIX	0.7484044	1.0847322	0.8798584	0.8801525	0.8801525

Let the (S&P) index be a predictor variable, then we can construct the simple linear multivariate regression model as model (2.1). We consider the data from the first 10 years (1977-1986) to find the average of the asset's return as our preliminary information, which gives the following result:

$$\boldsymbol{\theta}_0 = (6.0060, -0.9894, 8.2210, 1.3112, 8.1563, 8.0703, 11.9730, 7.2887, 10.3300)'.$$

Using $\boldsymbol{\theta}_0$ as prior information, we estimate the systematic risks of $\boldsymbol{\beta}$ using suggested estimation strategies. The point estimation of the proposed estimators are presented in Table 2.3. We calculate the risk of the listed estimators, based on 1000 replicates from bootstrapping. We obtain the efficiency of estimators relative to $\hat{\boldsymbol{\beta}}$; the results are given in Table 2.4, which are in agreement with the findings of our theoretical and simulated work.

Table 2.4: The relative efficiency of estimators

Estimator	$R.E(\hat{\boldsymbol{\beta}} : \boldsymbol{\beta}^*)$
$\tilde{\boldsymbol{\beta}}$	2.4472
$\hat{\boldsymbol{\beta}}^{JS}$	1.5343
$\hat{\boldsymbol{\beta}}^{JS+}$	1.7750
$\hat{\boldsymbol{\beta}}^{PT}$	2.0853

2.6 Concluding Remarks

For a simple multivariate regression model that includes basic investment models, we have considered various estimation strategies based on a pretest and shrinkage estimation. In conclusion, the positive-part shrinkage estimator dominates the usual shrinkage estimator uniformly. Both shrinkage estimators perform well relative to the usual unrestricted least squares estimator in a wider range that the pretest estimator. The subspace candidate least squares estimator depends heavily on the quality of the subspace information. The ADR of the restricted least squares estimator is unbounded when the parameter moves far from the subspace of the restriction, while the pretest estimator provides good control on the magnitude of the ADR. It is exceedingly important to note that the shrinkage estimators have the smallest possible risk in most cases, as compared to other estimators except when the subspace information is nearly correct. Further, the application of shrinkage estimators are subject to the condition that $q \geq 3$, where q is the number of parameters in the unrestricted parameter vector.

The theoretical results in the chapter were verified based on a Monte Carlo simulation. Indeed, the simulation study shows that the method suggested has sound finite sample properties. The analysis of a motivating financial data example is also consistent with findings of the analytical and simulation results.

2.7 Appendix: Proof of Main Results

The following lemma is listed in Sclove et al. (1972), and is used to prove the theorem in this chapter.

Lemma 2.7.1. Let $\boldsymbol{y}$ be a q-dimensional normal vector distributed as $N_q(\boldsymbol{\mu_y}, \boldsymbol{I}_q)$. Then, for a measurable function of ϕ, we have

$$\begin{aligned} E[\boldsymbol{y}\phi(\boldsymbol{y}'\boldsymbol{y})] &= \boldsymbol{\mu_y} E[\phi(\chi^2_{q+2}(\Delta))] \\ E[\boldsymbol{y}\boldsymbol{y}'\phi(\boldsymbol{y}'\boldsymbol{y})] &= \boldsymbol{I}_q E[\phi(\chi^2_{q+2}(\Delta))] + \boldsymbol{\mu_y}\boldsymbol{\mu_y}' E[\phi(\chi^2_{q+4}(\Delta))], \end{aligned}$$

where $\Delta = \boldsymbol{\mu_y}'\boldsymbol{\mu_y}$

2.7.1 Proof of Theorem 2.4.2

Since $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2$, and $\boldsymbol{\nu}_3$ are asymptotically normal, the joint distribution of $(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$ and $(\boldsymbol{\nu}_2, \boldsymbol{\nu}_3)$ will be asymptotically normal as well.

$$\begin{aligned} E(\boldsymbol{\nu}_2) &= \lim_{n\to\infty} E[n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})]\} \\ &= \lim_{n\to\infty} E[n^{1/2}(\hat{\boldsymbol{\beta}} + \boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}) - \boldsymbol{\beta})] \qquad \text{under } L_n \\ &= 0 + \lim_{n\to\infty} E[n^{1/2}\boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h})] \\ &= \boldsymbol{C\xi} \\ &= \boldsymbol{\gamma} \end{aligned}$$

$$
\begin{aligned}
Cov(\boldsymbol{\nu}_2) &= Cov(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= Cov(\hat{\boldsymbol{\beta}} + \boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}) - \boldsymbol{\beta}) \\
&= Cov(\hat{\boldsymbol{\beta}}) + Cov(\boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h})) - 2Cov(\hat{\boldsymbol{\beta}}, \boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h})) \\
&= Q_0^{-1}\boldsymbol{V} + (1 + \frac{\bar{x}^2}{Q_0})\boldsymbol{CHVH'C'} - 2\bar{x}Q_0^{-1}\boldsymbol{CHV} \\
&= \boldsymbol{\Sigma}^* \\
E(\boldsymbol{\nu}_2) &= \lim_{n\to\infty} E[n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})]\} \\
&= \lim_{n\to\infty} E[n^{1/2}(\hat{\boldsymbol{\beta}} + \boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}) - \boldsymbol{\beta})] \qquad \text{under } L_n \\
&= 0 + \lim_{n\to\infty} E[n^{1/2}\boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h})] \\
&= \boldsymbol{C\xi} \\
&= \boldsymbol{\gamma} \\
E(\boldsymbol{\nu}_3) &= E(\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2) \\
&= \lim_{n\to\infty} E[n^{1/2}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})] \\
&= \lim_{n\to\infty} E\{n^{1/2}[-\boldsymbol{C}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h})]\} \qquad \text{under } L_n \\
&= -\boldsymbol{C\xi} \\
&= -\boldsymbol{\gamma} \\
Cov(\boldsymbol{\nu}_3) &= Cov(\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2) \\
&= Cov(\boldsymbol{\nu}_1) + Cov(\boldsymbol{\nu}_2) - 2Cov(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) \\
&= Q_0^{-1}\boldsymbol{V} + Q_0^{-1}\boldsymbol{V} + (1 + \frac{\bar{x}^2}{Q_0})\boldsymbol{CHVH'C'} - 2\bar{x}Q_0^{-1}\boldsymbol{CHV} \\
&- 2Q_0^{-1}\boldsymbol{V} + 2\bar{x}Q_0^{-1}\boldsymbol{CHV} \\
&= (1 + \frac{\bar{x}^2}{Q_0})\boldsymbol{CHVH'C'} \\
&= \boldsymbol{\Omega}^*
\end{aligned}
$$

2.7.2 Proof of Theorem 2.4.3

In this section we explicitly present a proof of Theorem 2.4.3. Clearly the ADB of $\tilde{\boldsymbol{\beta}}$ is equal to $\boldsymbol{\gamma}$.

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{PT}) &= \lim_{n\to\infty} \sqrt{n}E(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta}) \\
&= \lim_{n\to\infty} \sqrt{n}E(\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})I(P \leq \chi^2_{q,\alpha}) - \boldsymbol{\beta}) \\
&= \lim_{n\to\infty} E[\boldsymbol{\nu}_1 - \boldsymbol{\nu}_3 I(P \leq \chi^2_{q,\alpha})] \\
&= \boldsymbol{\gamma} H_{q+2}(\chi^2_q(\alpha), \Delta), \\
ADB(\hat{\boldsymbol{\beta}}^{JS}) &= \lim_{n\to\infty} \sqrt{n}E(\tilde{\boldsymbol{\beta}} + (1 - mP^{-1})(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}) \\
&= \lim_{n\to\infty} E(\boldsymbol{\nu}_1 - m\boldsymbol{\nu}_3 P^{-1}) \\
&= m\boldsymbol{\gamma} E(\chi^{-2}_{q+2}(\Delta)), \\
ADB(\hat{\boldsymbol{\beta}}^{JS+}) &= \lim_{n\to\infty} \sqrt{n}E\{(\hat{\boldsymbol{\beta}}^{JS} - \boldsymbol{\beta}) - (1 - mP^{-1})(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})I(P < m)\} \\
&= ADB(\hat{\boldsymbol{\beta}}^{JS}) - \boldsymbol{\gamma} E\{(m\chi^{-2}_{q+2}(\Delta) - 1)I(\chi^2_{q+2}(\Delta) < m)\}.
\end{aligned}$$

2.7.3 Proof of Theorem 2.4.4

Clearly the $AMSE(\tilde{\boldsymbol{\beta}})$ is equal to $\boldsymbol{\Sigma}^* + \boldsymbol{\gamma}\boldsymbol{\gamma}'$. $AMSE(\hat{\boldsymbol{\beta}}^{PT})$ can be written as

$$\begin{aligned}
&= \lim_{n\to\infty} E\{n(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{PT} - \boldsymbol{\beta})'\} \\
&= \lim_{n\to\infty} E\{[\boldsymbol{\nu}_1 - \boldsymbol{\nu}_3 I(P < \chi^2_q(\alpha))][\boldsymbol{\nu}_1 - \boldsymbol{\nu}_3 I(P < \chi^2_q(\alpha))]'\} \\
&= \lim_{n\to\infty} E\{\boldsymbol{\nu}_1\boldsymbol{\nu}_1{}' - \boldsymbol{\nu}_1\boldsymbol{\nu}_3{}' I(P < \chi^2_q(\alpha)) - \boldsymbol{\nu}_3\boldsymbol{\nu}_1{}' I(P < \chi^2_q(\alpha)) + \boldsymbol{\nu}_3\boldsymbol{\nu}_3{}' I^2(P < \chi^2_q(\alpha))\}.
\end{aligned}$$

Note that, by using Theorem 2.4.2 and Lemma 2.7.1, for $E(\boldsymbol{\nu}_3\boldsymbol{\nu}_1{}'I(P < \chi_q^2(\alpha)))$ we have

$$\begin{aligned} &= E(E(\boldsymbol{\nu}_3\boldsymbol{\nu}_1{}'I(P < \chi_q^2(\alpha))|\boldsymbol{\nu}_3)) \\ &= E(\boldsymbol{\nu}_3[E(\boldsymbol{\nu}_1) + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}^{*-1}(\boldsymbol{\nu}_3 - E(\boldsymbol{\nu}_3))]'I(P < \chi_q^2(\alpha))) \\ &= \boldsymbol{\Omega}_{12}H_{q+2}(\chi_q^2(\alpha),\Delta) + \boldsymbol{\gamma\gamma}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Omega}_{12}[H_{q+4}(\chi_q^2(\alpha),\Delta) + H_{q+2}(\chi_q^2(\alpha),\Delta)]. \end{aligned}$$

Therefore,

$$\begin{aligned} AMSE(\hat{\boldsymbol{\beta}}^{PT}) &= Q_0^{-1}\boldsymbol{V} - 2\{\boldsymbol{\Omega}_{12}H_{q+2}(\chi_q^2(\alpha),\Delta) + \boldsymbol{\gamma\gamma}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Omega}_{12}[H_{q+4}(\chi_q^2(\alpha),\Delta) + \\ &\quad H_{q+2}(\chi_q^2(\alpha),\Delta)]\} + \boldsymbol{\Omega}^*H_{q+2}(\chi_q^2(\alpha),\Delta) + \boldsymbol{\gamma\gamma}'H_{q+4}(\chi_q^2(\alpha),\Delta) \\ AMSE(\hat{\boldsymbol{\beta}}^{JS}) &= \lim_{n\to\infty} E\{n[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - mP^{-1}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})][(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - mP^{-1}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})]'\} \\ &= Var(\boldsymbol{\nu}_1) + E(\boldsymbol{\nu}_1)E(\boldsymbol{\nu}_1)' - 2E(\boldsymbol{\nu}_3\boldsymbol{\nu}_1{}'mP^{-1}) + E(\boldsymbol{\nu}_3\boldsymbol{\nu}_3{}'(mP^{-1})^2). \end{aligned}$$

Note that, by using Theorem 2.4.2 and Lemma 2.7.1, we have

$$\begin{aligned} E(\boldsymbol{\nu}_3\boldsymbol{\nu}_1{}'P^{-1}) &= E(E(\boldsymbol{\nu}_3\boldsymbol{\nu}_1{}'P^{-1}|\boldsymbol{\nu}_3)) \\ &= E(\boldsymbol{\nu}_3[E(\boldsymbol{\nu}_1) + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}^{*-1}(\boldsymbol{\nu}_3 - E(\boldsymbol{\nu}_3))]'P^{-1}) \\ &= E(\boldsymbol{\nu}_3\boldsymbol{\nu}_3{}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Omega}_{12}P^{-1} + \boldsymbol{\nu}_3\boldsymbol{\gamma}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Omega}_{12}P^{-1}) \\ &= \boldsymbol{\Omega}_{12}E(\chi_{q+2}^{-2}(\Delta)) + \boldsymbol{\gamma\gamma}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Omega}_{12}[E(\chi_{q+4}^{-2}(\Delta)) + E(\chi_{q+2}^{-2}(\Delta))]. \end{aligned}$$

Therefore,

$$\begin{aligned} AMSE(\hat{\boldsymbol{\beta}}^{JS}) &= Q_0^{-1}\boldsymbol{V} - 2m\{\boldsymbol{\Omega}_{12}E(\chi_{q+2}^{-2}(\Delta)) + \boldsymbol{\gamma\gamma}'\boldsymbol{\Omega}^{*-1}\boldsymbol{\Omega}_{12}[E(\chi_{q+4}^{-2}(\Delta)) + \\ &\quad E(\chi_{q+2}^{-2}(\Delta))]\} + m^2[\boldsymbol{\Omega}^*E(\chi_{q+2}^{-4}(\Delta)) + \boldsymbol{\gamma\gamma}'E(\chi_{q+4}^{-4}(\Delta))]. \end{aligned}$$

We got the result after some computation. Similarly,

$$\begin{aligned} AMSE(\hat{\boldsymbol{\beta}}^{JS+}) &= \lim_{n\to\infty} E\{n[(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})-(1-mP^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})I(P<m)] \\ &\times [(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})-(1-mP^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})I(P<m)]'\} \\ &= AMSE(\hat{\boldsymbol{\beta}}^{JS}) \\ &- 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^{JS}-\boldsymbol{\beta})'(1-mP^{-1})I(P<m)\} \\ &+ \lim_{n\to\infty} nE[(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})'(1-mP^{-1})^2 I(P<m)]. \end{aligned}$$

Note that by using the definition of $\hat{\boldsymbol{\beta}}^{JS}$ from Section 2.3 in the second term of the above equation and substituting $\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}$, we have

$$\begin{aligned} &- 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})[(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})+(1-mP^{-1})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})]' \times (1-mP^{-1})I(P<m)\} = \\ &- 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(1-mP^{-1})I(P<m)\} \qquad (1^*) \\ &+ 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})'(1-mP^{-1})I(P<m)\} \qquad (2^*) \\ &- 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})'(1-mP^{-1})^2 I(P<m)\} \qquad (3^*). \end{aligned}$$

Now, by substituting $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}+\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}$ in (1^*), we get

$$\begin{aligned} (1^*) = &- 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})'(1-mP^{-1})I(P<m)\} \qquad \text{same as } (2^*) \\ &- 2\lim_{n\to\infty} nE\{(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'(1-mP^{-1})I(P<m)\}. \end{aligned}$$

Therefore, the second term in $AMSE(\hat{\boldsymbol{\beta}}^{JS+})$ will be simplified as follows:

$$-2\lim_{n\to\infty} E[\boldsymbol{\nu}_3{\boldsymbol{\nu}_2}'(1-mP^{-1})I(P<m)] - 2\lim_{n\to\infty} E[\boldsymbol{\nu}_3{\boldsymbol{\nu}_3}'(1-mP^{-1})^2 I(P<m)].$$

As well the third term in $AMSE(\hat{\boldsymbol{\beta}}^{JS+})$ will be simplified as

$$
\begin{aligned}
&= E[\boldsymbol{\nu}_3\boldsymbol{\nu}_3{}'(1-mP^{-1})^2 I(P<m)] \\
&= \boldsymbol{\Omega}^* E(1-m\chi^{-2}_{q+2}(\Delta))^2 I(\chi^2_{q+2}(\Delta)<m) + \boldsymbol{\gamma}\boldsymbol{\gamma}' E(1-m\chi^{-2}_{q+4}(\Delta))^2 I(\chi^2_{q+4}(\Delta)<m).
\end{aligned}
$$

Finally, by using Theorem 2.4.2 and Lemma 2.7.1, the AMSE of a positive shrinkage estimator is

$$
\begin{aligned}
&= AMSE(\hat{\boldsymbol{\beta}}^{JS}) - \boldsymbol{\Omega}^* E(1-m\chi^{-2}_{q+2}(\Delta))^2 I(\chi^2_{q+2}(\Delta)<m) \\
&+ 2\boldsymbol{\gamma}\boldsymbol{\gamma}' E(1-m\chi^{-2}_{q+2}(\Delta)) I(\chi^2_{q+2}(\Delta)<m) - \boldsymbol{\gamma}\boldsymbol{\gamma}' E(1-m\chi^{-2}_{q+4}(\Delta))^2 I(\chi^2_{q+4}(\Delta)<m).
\end{aligned}
$$

2.7.4 Proof of Theorem 2.4.5

In an effort to prove Theorem 2.4.5, we need to show some useful preliminary results. Clearly, the asymptotic risk of $\hat{\boldsymbol{\beta}}$ is equal to $tr(\boldsymbol{W}Q_0^{-1}\boldsymbol{V}) = Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V})$. Also, we get the following expression for the asymptotic risk of $\tilde{\boldsymbol{\beta}}$:

$$
ADR(\tilde{\boldsymbol{\beta}};\boldsymbol{W}) = Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V}) + a\times tr(\boldsymbol{W}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}') - 2b\times tr(\boldsymbol{W}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}) + \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}.
$$

Since the risk of $\tilde{\boldsymbol{\beta}}$ depends on $\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = \boldsymbol{C}\boldsymbol{\xi}$, note that $\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}'\boldsymbol{V}^{-\frac{1}{2}}$ is symmetric and an idempotent matrix with rank q. Thus, there exists an orthogonal matrix $\mathbf{\Gamma}$ such that

$$\mathbf{\Gamma}\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}'\boldsymbol{V}^{-\frac{1}{2}}\mathbf{\Gamma}' = \begin{pmatrix} \boldsymbol{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$\mathbf{\Gamma}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{V}^{\frac{1}{2}}\mathbf{\Gamma}' = \begin{pmatrix} \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} \\ \boldsymbol{Z}_{21} & \boldsymbol{Z}_{22} \end{pmatrix}.$$

We need to show that

$$\begin{aligned} tr[\boldsymbol{W}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}'] &= tr[\{\mathbf{\Gamma}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{V}^{\frac{1}{2}}\mathbf{\Gamma}'\} \times \{\mathbf{\Gamma}\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}'\boldsymbol{V}^{-\frac{1}{2}}\mathbf{\Gamma}'\}] \\ &= tr[\begin{pmatrix} \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} \\ \boldsymbol{Z}_{21} & \boldsymbol{Z}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}] = tr(\boldsymbol{Z}_{11}), \end{aligned}$$

and we may write

$$\begin{aligned} \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma} &= \boldsymbol{\xi}'\boldsymbol{C}'\boldsymbol{W}\boldsymbol{C}\boldsymbol{\xi} \\ &= \boldsymbol{\xi}'\{\mathbf{\Gamma}\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}'\boldsymbol{V}^{-\frac{1}{2}}\mathbf{\Gamma}'\}\{\mathbf{\Gamma}\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{V}^{\frac{1}{2}}\mathbf{\Gamma}'\} \\ &\quad \{\mathbf{\Gamma}\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{C}'\boldsymbol{V}^{-\frac{1}{2}}\mathbf{\Gamma}'\}\boldsymbol{\xi} \\ &= \boldsymbol{\eta}' \begin{pmatrix} \boldsymbol{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} \\ \boldsymbol{Z}_{21} & \boldsymbol{Z}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{\eta} = \boldsymbol{\eta}_1'\boldsymbol{Z}_{11}\boldsymbol{\eta}_1, \end{aligned}$$

where $\boldsymbol{\eta} = \mathbf{\Gamma}\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}\boldsymbol{H}'\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix}$. Therefore,

$$ADR(\tilde{\boldsymbol{\beta}};\boldsymbol{W}) = Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V}) + a \times tr(\boldsymbol{Z}_{11}) - 2b \times tr(\boldsymbol{W}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V}) + \boldsymbol{\eta}_1'\boldsymbol{Z}_{11}\boldsymbol{\eta}_1.$$

Similarly for $ADR(\hat{\boldsymbol{\beta}}^{PT};\boldsymbol{W})$ we have

$$\begin{aligned}
&= Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V}) + [tr(\boldsymbol{W}\boldsymbol{G}) - 2tr(\boldsymbol{W}\boldsymbol{F})]H_{q+2}(\chi_q^2(\alpha),\Delta)\\
&- 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{A})(H_{q+2}(\chi_q^2(\alpha),\Delta) + H_{q+4}(\chi_q^2(\alpha),\Delta)) + tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')H_{q+4}(\chi_q^2(\alpha),\Delta))\\
&= Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V}) + [a\times tr(\boldsymbol{Z}_{11}) - 2b\times tr(\boldsymbol{W}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V})]H_{q+2}(\chi_q^2(\alpha),\Delta)\\
&- 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{A})[H_{q+2}(\chi_q^2(\alpha),\Delta) + H_{q+4}(\chi_q^2(\alpha),\Delta)] + \boldsymbol{\eta}_1'\boldsymbol{Z}_{11}\boldsymbol{\eta}_1 H_{q+4}(\chi_q^2(\alpha),\Delta)).
\end{aligned}$$

Finally, for $ADR(\hat{\boldsymbol{\beta}}^{JS};\boldsymbol{W})$ we have

$$\begin{aligned}
&= Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V}) - 2m\{tr(\boldsymbol{W}\boldsymbol{F})E(\chi_{q+2}^{-2}(\Delta))\\
&- tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{A})[E(\chi_{q+4}^{-2}(\Delta)) + E(\chi_{q+2}^{-2}(\Delta))]\}\\
&+ m^2[tr(\boldsymbol{W}\boldsymbol{G})E(\chi_{q+2}^{-4}(\Delta)) + tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')E(\chi_{q+4}^{-4}(\Delta))]\\
&= Q_0^{-1}tr(\boldsymbol{W}\boldsymbol{V}) - 2mb\times tr(\boldsymbol{W}\boldsymbol{C}\boldsymbol{H}\boldsymbol{V})E(\chi_{q+2}^{-2}(\Delta))\\
&- 2m\times tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{A})[E(\chi_{q+4}^{-2}(\Delta)) + E(\chi_{q+2}^{-2}(\Delta))]\\
&+ am^2\times tr(\boldsymbol{Z}_{11})E(\chi_{q+2}^{-4}(\Delta)) + m^2\times\boldsymbol{\eta}_1'\boldsymbol{Z}_{11}\boldsymbol{\eta}_1 E(\chi_{q+4}^{-4}(\Delta))],
\end{aligned}$$

and similarly for $ADR(\hat{\boldsymbol{\beta}}^{JS+};\boldsymbol{W})$ we have

$$\begin{aligned}
&= ADR(\hat{\boldsymbol{\beta}}^{JS}) - tr(\boldsymbol{W}\boldsymbol{G})[E(1-m\chi_{q+2}^{-2}(\Delta))^2 I(\chi_{q+2}^2(\Delta)<m)]\\
&+ tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')\{2E(1-m\chi_{q+2}^{-2}(\Delta))I(\chi_{q+2}^2(\Delta)<m)\\
&- E(1-m\chi_{q+4}^{-2}(\Delta))^2 I(\chi_{q+4}^2(\Delta)<m)\}\\
&= ADR(\hat{\boldsymbol{\beta}}^{JS}) - a\times tr(\boldsymbol{Z}_{11})[E(1-m\chi_{q+2}^{-2}(\Delta))^2 I(\chi_{q+2}^2(\Delta)<m)]\\
&+ \boldsymbol{\eta}_1'\boldsymbol{Z}_{11}\boldsymbol{\eta}_1\{2E(1-m\chi_{q+2}^{-2}(\Delta))I(\chi_{q+2}^2(\Delta)<m)\\
&- E(1-m\chi_{q+4}^{-2}(\Delta))^2 I(\chi_{q+4}^2(\Delta)<m)\}.
\end{aligned}$$

Chapter 3

Estimation Strategies for a Parameter Matrix in a Multivariate Regression Model

.

3.1 Introduction

In many areas of scientific research, the basic goal is to assess the simultaneous influence of several covariates on the response variable, and the quantity of interest. Multiple regression models provide an extremely powerful methodology to achieve this task. The multivariate multiple regression model (MMRM) generalizes the multiple regression model for the prediction of several response variables from the same set of explanatory variables. A common example of multivariate responses occur in classification and discrimination problems. Timm (2002) showcases a host of examples of applications in education and psychology. Some of the recent advances in multivariate analysis include artificial intelligence and machine learning theory (see for example Izenman (2008)).

The general multivariate regression model is defined as

$$Y_i = \boldsymbol{C} X_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \dots, n, \tag{3.1}$$

where $Y_i = (y_{1i}, \dots, y_{mi})'$ is a $m \times 1$ vector of response variables, $X_i = (x_{1i}, \dots, x_{qi})'$ is a $q \times 1$ vector of predictor variables, $\boldsymbol{C}$ is a full rank $m \times q$ regression coefficient matrix, and $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \dots, \epsilon_{mi})'$ is the $m \times 1$ vector of random errors with mean vector $E(\boldsymbol{\epsilon}_i) = 0$ and covariance matrix $Cov(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_{\epsilon\epsilon}$ is an $m \times m$ positive definite matrix. The ϵ_i are assumed to be independent for different i. We define the $m \times n$ and $q \times n$ data matrices, respectively, as $\boldsymbol{\mathcal{Y}} = [Y_1, \dots, Y_n]$ and $\boldsymbol{\mathcal{X}} = [X_1, \dots, X_n]$. We assume that $m + q \leq n$ and $\boldsymbol{\mathcal{X}}$ is a full rank matrix with rank $q = rank(\boldsymbol{\mathcal{X}}) < n$ to have a unique least square solution to the first order equations. Here we arrange the error vectors $\boldsymbol{\epsilon}_i$,$i = 1, \dots, n$, into an $m \times n$ matrix $\boldsymbol{\mathcal{E}} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_n]$. The MMRM in (3.1) can be rewritten as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{C}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}. \tag{3.2}$$

The model (3.2) is regarded as a candidate full model, the least squares (LS) estimate of $\boldsymbol{C}$ is given by

$$\hat{\boldsymbol{C}} = \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{X}}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}. \tag{3.3}$$

For the estimation problem at hand, for the sake of brevity, let us consider that the errors can be arranged into an $mn \times 1$ vector, $\boldsymbol{e} = vec(\boldsymbol{\mathcal{E}})$. Then,

$$E(\boldsymbol{e}) = \boldsymbol{0}, \qquad Cov(\boldsymbol{e}) = E\{\boldsymbol{e}\boldsymbol{e}'\} = \boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes \boldsymbol{I}_n, \qquad i = 1, \ldots, n.$$

For inference purposes we assume that the error terms are independently and identically distributed (iid) as multivariate normal, that is, $\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_n(\boldsymbol{0}, \boldsymbol{\Sigma}_{\epsilon\epsilon})$, and $\boldsymbol{\mathcal{X}}$ is a fixed matrix. Note that we are using the left Kronecker products in our expressions; see for example Izenman (2008). Here, we consider the $vec(\boldsymbol{\mathcal{Y}})$ where the "*vec*" operator transforms an $m \times n$ matrix into an nm-dimensional column vector by stacking the columns of the matrix below each other. Thus, $\boldsymbol{y} = vec(\boldsymbol{\mathcal{Y}}) = (\boldsymbol{I}_n \otimes \boldsymbol{\mathcal{X}}')vec(\boldsymbol{C}) + \boldsymbol{e}$, and $vec(\hat{\boldsymbol{C}}) = (\boldsymbol{I}_n \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{\mathcal{X}})vec(\boldsymbol{\mathcal{Y}})$. The distributional properties of $\hat{\boldsymbol{C}}$ follow easily from multivariate normality of the error terms ϵ_i; therefore,

$$vec(\hat{\boldsymbol{C}}) \sim N(vec(\boldsymbol{C}), \boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}).$$

Now, partition $\boldsymbol{\mathcal{X}} = [\boldsymbol{\mathcal{X}}_1', \boldsymbol{\mathcal{X}}_2']$ and corresponding $\boldsymbol{C} = [\boldsymbol{C}_1, \boldsymbol{C}_2]$ so that the model in (3.2) is written as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{C}_1\boldsymbol{\mathcal{X}}_1 + \boldsymbol{C}_2\boldsymbol{\mathcal{X}}_2 + \boldsymbol{\mathcal{E}}, \tag{3.4}$$

where $\boldsymbol{C}_1$ is a $m \times q_1$ and $\boldsymbol{C}_2$ is a $m \times q_2$ with $q_1 + q_2 = q$ dimensional matrix of unknown parameters. Model (3.4) may be regarded as a candidate full model, which is built at the initial stage of modeling and contains all possible relevant variables. Because of the high dimension of the regression parameter matrix, one usually uses a variable selection technique to remove less significant variables Li and Liang (2008). Without loss of generality, we suppose that $\boldsymbol{\mathcal{X}}_2$ is relatively insignificant and thus is removed from the model (3.4). Then, we obtain a candidate sub-model as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{C}_1 \boldsymbol{\mathcal{X}}_1 + \boldsymbol{\mathcal{E}}. \tag{3.5}$$

Nowadays a popularly used method is available to select variables and estimate parameters simultaneously; see for example Fan and Li (2001) and Li and Liang (2008). However, Leeb and Potscher (2005) reported that such an estimator is only point-wise consistent, and not globally well-working. On the other hand, some other estimation procedures, for example, the restricted-model estimation only depends on the sub-model (3.5) and the pretest (PT) estimation, which uses a test to decide that the estimator for $\boldsymbol{C}_1$ is based on a candidate full model, are used in the literature; see for example Ahmed (2001), Ahmed et al. (2007), and others. The regression coefficients obtained after model selection are biased. In other words, bias caused by a misspecified model should be accounted for. These issues are well summarized in the scientific literature. The main objective of this chapter is to consider the estimation problem of the parameter matrix $\boldsymbol{C}$ under a very general set of linear constraints, which includes the sub-model (3.5). To this end, we can write the subspace as

$$\text{candidate subspace} : \boldsymbol{KCL} = \boldsymbol{0}, \tag{3.6}$$

where $\boldsymbol{K}$ and $\boldsymbol{L}$ are known full-rank matrices of appropriate dimensions $r \times m$ and $q \times n$, respectively. In the candidate subspace, the matrix $\boldsymbol{K}$ allows for restrictions between the different columns of $\boldsymbol{C}$, whereas $\boldsymbol{L}$ generates possible relationships between the different responses. Let us consider the following example to motivate the problem at hand. Zapala et al. (2005) considers the multiple multivariate linear regression model to explain the relationships of gene expression patterns between different brain regions in the adult mouse. The data set involves gene expression data from multiple brain regions and multiple inbred mouse strains. They have built a gene expression-based brain map. However, the preliminary analysis indicates that the gene expression patterns of these brain regions could be related to each other based on adult anatomy, evolutionary relationships, or embryonic origin. The complete collection of extensively annotated gene expression data, along with data mining and visualization tools, have been made available on a publicly accessible web site (www.barlow-lockhart-brainmapnimhgrant.org).

The capital asset pricing model (CAPM) can be viewed as a special case of the MMRM as provided in Chapter 2. The candidate subspace includes many interesting hypotheses and a variety of applications that can be based upon this general set of linear constraints. Further, this constraint can be applied to tackle a variety of experimental design problems, including profile analysis. The similarity of a given number of profiles can be expressed as a set of linear constraints on $\boldsymbol{C}$. We are interested in establishing an optimal estimation strategy for the parameter matrix in multivariate multiple regression models when the parameter is suspected to satisfy a certain constraint. We let

$$\tilde{\boldsymbol{C}} = arg \min_{\text{KCL=0}} tr\{(\boldsymbol{\mathcal{Y}} - \boldsymbol{C}\boldsymbol{\mathcal{X}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{C}\boldsymbol{\mathcal{X}})'\}$$

denote the constrained or subspace candidate estimator. We get

$$\tilde{\boldsymbol{C}} = \hat{\boldsymbol{C}} - \boldsymbol{S}(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})\boldsymbol{T},$$

where $\boldsymbol{S} = \boldsymbol{K}'(\boldsymbol{K}\boldsymbol{K}')^{-1}$ and $\boldsymbol{T} = (\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}$.

Clearly, $\tilde{\boldsymbol{C}}$ will be a biased estimator if the candidate subspace does not hold. Alternatively, we can write the null hypothesis in (3.6) as

$$H_0 : (\boldsymbol{K} \otimes \boldsymbol{L}')vec(\boldsymbol{C}) = \boldsymbol{0}. \tag{3.7}$$

On the other hand, it can be easily verified that $\tilde{\boldsymbol{C}}$ is an inconsistent estimator because of the bias inherited by the sub-model. More precisely, the model bias is $B(\tilde{\boldsymbol{C}}) = -\boldsymbol{\gamma}$ where $\boldsymbol{\gamma} = (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C})$. The amount of bias can be reduced by shrinking the full model estimator towards the candidate sub-model estimator. A natural way to balance the potential bias of the estimator under the restriction against the classical estimator is to take a weighted average of $\hat{\boldsymbol{C}}$ and $\tilde{\boldsymbol{C}}$. Such an integrated or composite estimator commonly known as the shrinkage estimator is given by

$$\hat{\boldsymbol{C}}^S = \tau\hat{\boldsymbol{C}} + (1-\tau)\tilde{\boldsymbol{C}},$$

where $\tau \in [0, 1]$ denotes the shrinkage intensity. Note that, for $\tau = 1$, the shrinkage estimate equals the shrinkage target $\tilde{\boldsymbol{C}}$; whereas, for $\tau = 0$, the classical or full model estimate (FE) is recovered. The key advantage of this construction is that it outperforms the FE in some part of the parameter space. However, the key question in this type of estimator is how to select an optimal value for the shrinkage parameter τ. In some situations, it may suffice to fix the parameter τ at some given value. The second choice is to select the parameter τ in a data-driven fashion by explicitly minimizing a suitable risk function. A common but also computationally intensive approach to estimate the optimal τ is by using cross-validation. On the other hand, from a Bayesian perspective, one can employ the empirical Bayes technique to infer. In this case, τ is treated as a hyper-parameter and may be estimated from the data by optimizing the marginal likelihood. In this work, we treat τ as the degree of trust in the prior information in the null hypothesis. The value of τ may be assigned by the experimenter according to her/his prior belief in the prior value. Ahmed and Krzanowski (2004) among others pointed out that such an estimator yields a smaller mean squared error (MSE) when the constraint is correct or nearly correct. However, this is at the expense of poorer performance in the rest of the parameter space induced by the candidate subspace information. Here, we will demonstrate that $\tilde{\boldsymbol{C}}$ will have a smaller MSE than $\hat{\boldsymbol{C}}$ near the restriction given in (3.7). However, it becomes considerably biased and inefficient when the restriction may not be judiciously justified. Therefore, It can be easily verified that $\tilde{\boldsymbol{C}}$ is an inconsistent estimator because of the bias of the sub-model. The magnitude of the bias can be controlled by judiciously selecting the values of τ; in this sense, it has an edge over on $\hat{\boldsymbol{C}}$. In any event, it is also a function of $\boldsymbol{\gamma}$, so it will be an inconsistent estimator of $\boldsymbol{C}$, regardless. As such, when the linear constraint information is rather suspicious, it may be reasonable to construct pretest estimators. We use the following test statistic for defining a pretest estimator, which can be found based on a likelihood ratio method of the test construction.

$$\begin{aligned} D &= tr\{(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})'(\boldsymbol{K}\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}')^{-1}(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\} \\ &= [(\boldsymbol{K}\otimes\boldsymbol{L}')vec(\hat{\boldsymbol{C}})]'[(\boldsymbol{K}\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\otimes\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})]^{-1}[(\boldsymbol{K}\otimes\boldsymbol{L}')vec(\hat{\boldsymbol{C}})]. \end{aligned}$$

Under the null hypothesis in (3.7), D follows a central χ^2 distribution with rn degrees of freedom for known $\boldsymbol{\Sigma}_{\epsilon\epsilon}$. Consequently,

$$\hat{\boldsymbol{C}}^{PT} = \hat{\boldsymbol{C}} - (\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})I(D < d_{rn,\alpha}),$$

where the $d_{rn,\alpha}$ is the upper α-level critical value of the χ^2 distribution with rn degrees of freedom, and $I(A)$ is an indicator function of a set A. Ahmed (1992) suggested the following improved version of pretest estimator, namely the shrinkage pretest estimator.

$$\hat{\boldsymbol{C}}^{SPT} = \hat{\boldsymbol{C}}^{S} I(D < d_{rn,\alpha}) + \hat{\boldsymbol{C}} I(D > d_{rn,\alpha}).$$

Like $\hat{\boldsymbol{C}}^{S}$, it controls the magnitude of the bias of the estimator and it is less demanding with regards to the size of the pretest, α. Following Ahmed and Krzanowski (2004), we consider two shrinkage estimations of the regression parameters matrix based on a James-Stein type estimator. The positive part shrinkage estimator is given by

$$\hat{\boldsymbol{C}}^{JS+} = \tilde{\boldsymbol{C}} + \{1 - cD^{-1}\}^{+}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}}), \quad rn > 2,$$

where the optimal value of c is $c_{opt} = rn - 2$ and is chosen in an interval in such a way that $\hat{\boldsymbol{\beta}}^{JS}$ dominates $\hat{\boldsymbol{\beta}}$. c is allowed to vary over $[0, 2(rn - 2)), rn > 2$, often set to $c = rn - 2$; thus, we assume that $rn \geq 3$. Finally,

$$\hat{\boldsymbol{C}}^{JS} = \tilde{\boldsymbol{C}} + \{1 - cD^{-1}\}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}}), \quad rn > 2.$$

The remainder of this chapter is organized as follows. In Section 2, we showcase some important results which will be needed in deriving the expressions for the listed estimators. In Section 3, we obtain the expressions for bias and risk, and present the pairwise risk comparison of the listed estimators. To facilitate the risk expressions for estimators given in this Section, we first generalize the two important lemmas of Sclove et al. (1972). Section 4 provides two real data examples and a simulation study. Conclusions are offered in Section 5. The proof of the main results, including

Lemmas and Theorems, are given in Section 6.

3.2 Main Results

In an effort to establish some important properties of the estimators, let $\boldsymbol{p}_1 = vec(\hat{\boldsymbol{C}} - \boldsymbol{C})$, $\boldsymbol{p}_2 = vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})$, and $\boldsymbol{p}_3 = vec(\tilde{\boldsymbol{C}} - \boldsymbol{C})$. Assuming that the errors are distributed as (iid) gaussian random vectors, i.e., $\boldsymbol{\epsilon}_i \overset{iid}{\sim} N_n(\boldsymbol{0}, \boldsymbol{\Sigma}_{\epsilon\epsilon}), i = 1, \ldots, n$ and $\boldsymbol{\mathcal{X}}$ is a fixed matrix, then we have the following distributional results:

Theorem 3.2.1.

(i) $\boldsymbol{p}_1 \sim N(\boldsymbol{0}, (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}))$

(ii) $\boldsymbol{p}_2 \sim N((\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}), \boldsymbol{\Sigma}^*)$

where $\boldsymbol{\Sigma}^* = (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')(\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})$

(iii) $\boldsymbol{p}_3 \sim N(-(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}), \boldsymbol{\Omega}^*)$

where $\boldsymbol{\Omega}^* = (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21} + \boldsymbol{\Sigma}^*$,

$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})$

(iv) $\begin{pmatrix} \boldsymbol{p}_1 \\ \boldsymbol{p}_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{0} \\ (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}) \end{pmatrix}, \begin{pmatrix} (\boldsymbol{\Sigma}_{\boldsymbol{\epsilon\epsilon}} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}^* \end{pmatrix} \right\}$

(v) $\begin{pmatrix} \boldsymbol{p}_2 \\ \boldsymbol{p}_3 \end{pmatrix} \sim N\left\{ \begin{pmatrix} (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}) \\ -(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}) \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}^* & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}^* \end{pmatrix} \right\}$

where

$$\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}' = \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}^*.$$

Proof: See Appendix, Section 3.6.1.

3.2.1 Two Useful Lemmas

In this chapter, we are dealing with parameter matrix estimation in a multivariate multiple regression model. The fundamental results of Sclove et al. (1972) cannot be directly implemented to compute the expressions for bias and risk of the proposed estimators. However, they can be generalized as in the following Lemmas.

Lemma 3.2.1. Let $\boldsymbol{y}$ be an $(nm \times 1)$ vector that follows normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{y}}$ vector and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}}$, i.e. $\boldsymbol{y} \sim N(\boldsymbol{\mu}_{\boldsymbol{y}}, \boldsymbol{\Sigma}_{\boldsymbol{y}})$. Then, for a measurable function of ϕ, we have

$$E[\boldsymbol{y}\phi(\boldsymbol{y}'\boldsymbol{y})] = \boldsymbol{\mu}_{\boldsymbol{y}} E[\phi(\chi^2_{nm+2}(\Delta))],$$

where $\Delta = \boldsymbol{\mu}'_{\boldsymbol{y}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}}\boldsymbol{\mu}_{\boldsymbol{y}}$.

Proof: See Appendix, Section 3.6.2.

Lemma 3.2.2. Let $\boldsymbol{y}$ be an $(nm \times 1)$ vector that follows normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{y}}$ matrix and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}}$, i.e. $\boldsymbol{y} \sim N(\boldsymbol{\mu}_{\boldsymbol{y}}, \boldsymbol{\Sigma}_{\boldsymbol{y}})$. Then, for a measurable function of ϕ we have

$$E[\boldsymbol{y}\boldsymbol{y}'\phi(\boldsymbol{y}'\boldsymbol{y})] = \boldsymbol{\Sigma}_{\boldsymbol{y}} E[\phi(\chi^2_{nm+2}(\Delta))] + \boldsymbol{\mu}_{\boldsymbol{y}}\boldsymbol{\mu}'_{\boldsymbol{y}} E[\phi(\chi^2_{nm+4}(\Delta))],$$

where $\Delta = \boldsymbol{\mu}'_{\boldsymbol{y}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}}\boldsymbol{\mu}_{\boldsymbol{y}}$.

Proof: See Appendix, Section 3.6.3.

3.3 Bias and Risk Analysis

In this section, we compare the performance of the suggested estimators in terms of bias and risk, respectively. First, the bias expressions of the listed estimators of the regression coefficients are given in the following theorem.

Theorem 3.3.1. Under the assumptions for model (3.2), the bias of the listed estimators are given as follows.

(i) $B(\tilde{\boldsymbol{C}}) = -\boldsymbol{\gamma}$

(ii) $B(\hat{\boldsymbol{C}}^{S}) = -(1-\tau)\boldsymbol{\gamma}.$

(iii) $B(\hat{\boldsymbol{C}}^{PT}) = -\boldsymbol{\gamma} H_{rn+2}(\chi^2_{rn}(\alpha); \Delta)$

(iv) $B(\hat{\boldsymbol{C}}^{SPT}) = -(1-\tau)\boldsymbol{\gamma} H_{rn+2}(\chi^2_{rn}(\alpha); \Delta)$

(v) $B(\hat{\boldsymbol{C}}^{JS}) = -c\boldsymbol{\gamma} E[\chi^{-2}_{rn+2}(\Delta)]$

(vi) $B(\hat{\boldsymbol{C}}^{JS^+}) = B(\hat{\boldsymbol{C}}^{S}) - \boldsymbol{\gamma} E\{[1 - c\chi^{-2}_{rn+2}(\Delta)] I(\chi^2_{rn+2}(\Delta) < c)\},$

where

$$\boldsymbol{\gamma} = (\boldsymbol{S}\boldsymbol{K} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C})$$

and

$$\Delta = \boldsymbol{\gamma}'(\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})^{-1}\boldsymbol{\gamma}.$$

Proof: See Appendix, Section 3.6.4.

Since the bias expressions of all the estimators are not in scalar form, we therefore take the recourse of converting them into the quadratic form. Thus, let us define the quadratic bias (QB) of an estimator $\boldsymbol{C}^*$ of $\boldsymbol{C}$ by

$$QB = [B(\hat{\boldsymbol{C}}^*)]'(\boldsymbol{\Sigma}_{\epsilon\epsilon}^{-1} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}'))[B(\hat{\boldsymbol{C}}^*)].$$

Based on the above, we can easily obtain the QB of the listed estimators.

(i) $QB(\tilde{\boldsymbol{C}}) = \Delta$

(ii) $QB(\hat{\boldsymbol{C}}^S) = (1-\tau)^2\Delta.$

(iii) $QB(\hat{\boldsymbol{C}}^{PT}) = \Delta\{H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)\}^2$

(iv) $QB(\hat{\boldsymbol{C}}^{SPT}) = (1-\tau)^2\Delta\{H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)\}^2$

(v) $QB(\hat{\boldsymbol{C}}^{JS}) = c^2\Delta\{E[\chi^{-2}_{rn+2}(\Delta)]\}^2$

(vi) $QB(\hat{\boldsymbol{C}}^{JS^+}) = \Delta\{cE[\chi^{-2}_{rn+2}(\Delta)] - E[(1-c\chi^{-2}_{rn+2}(\Delta))I(\chi^2_{rn+2}(\Delta) < c)]\}^2.$

Clearly, for the quadratic bias of $\tilde{\boldsymbol{C}}$, $\hat{\boldsymbol{C}}^{JS}$, and $\hat{\boldsymbol{C}}^{JS^+}$ the component Δ is common and they differ only by scalar factors: it suffices to compare the scalar factors only. Therefore we have the following two results:

$$QB(\hat{\boldsymbol{C}}^{JS^+}) \leq QB(\hat{\boldsymbol{C}}^{JS}) \leq QB(\tilde{\boldsymbol{C}})$$

$$QB(\hat{\boldsymbol{C}}) \leq QB(\hat{\boldsymbol{C}}^S) \leq QB(\hat{\boldsymbol{C}}^{SPT}) \leq QB(\hat{\boldsymbol{C}}^{PT}) \leq QB(\tilde{\boldsymbol{C}}).$$

Now, we present some useful results in the following theorem which will be used in deriving the risk expressions for the estimators.

Lemma 3.3.1. Under the assumptions for model (3.2),

(i) $E(\boldsymbol{p}_1|\boldsymbol{p}_2) = (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')^{-1}(\boldsymbol{p}_2 - \boldsymbol{\gamma})$

(ii) $E(\boldsymbol{p}_2\boldsymbol{p}_1{}'I(D < \chi^2_{rn}(\alpha)) = \boldsymbol{\Sigma}^*(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})^{-1}H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})^{-1}\left[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)\right]$

(iii) $E(\boldsymbol{p}_2\boldsymbol{p}_1{}'D^{-1}) = \boldsymbol{\Sigma}^*(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})^{-1}E(\chi^{-2}_{rn+2}(\Delta)) + \boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})^{-1}\{E[\chi^{-2}_{rn+4}(\Delta)] - E[\chi^{-2}_{rn+2}(\Delta)]\}$,

Proof: See Appendix, Section 3.6.5.

3.3.1 Relative Performance of the Estimators

In this section, we compare the performance of the suggested estimators in terms of their risks. In an effort to provide the risk analysis, we consider the quadratic loss function of the form

$$\begin{aligned}\mathcal{L}(\boldsymbol{C}^*, \boldsymbol{C}) &= [vec(\boldsymbol{C}^* - \boldsymbol{C})]'\boldsymbol{W}[vec(\boldsymbol{C}^* - \boldsymbol{C})] \\ &= tr\{\boldsymbol{W}[vec(\boldsymbol{C}^* - \boldsymbol{C})][vec(\boldsymbol{C}^* - \boldsymbol{C})]'\},\end{aligned}$$

where $\boldsymbol{W}$ is the positive semi-definite (p.s.d) matrix with an appropriate dimension. Then the risk of $\boldsymbol{C}^*$ or any estimator of $\boldsymbol{C}$ is

$$R(\boldsymbol{C}^*; \boldsymbol{W}) = tr[\boldsymbol{W}MSE(\boldsymbol{C}^*)], \tag{3.8}$$

where $MSE(\boldsymbol{C}^*) = E\{vec(\boldsymbol{C}^* - \boldsymbol{C})[vec(\boldsymbol{C}^* - \boldsymbol{C})]'\}$. For instance, if we get $MSE(\boldsymbol{C}^*) = (\boldsymbol{A} \otimes \boldsymbol{B})$ with a $\boldsymbol{A}$ and $\boldsymbol{B}$ nonsingular matrix, we define the quadratic risk as follows:

$$R(\boldsymbol{C}^*; \boldsymbol{W}) = tr(\boldsymbol{WB})tr(\boldsymbol{A}).$$

The quadratic risk of the estimators are given in the following theorem.

Theorem 3.3.2. The quadratic risk of the listed estimators are given as follows.

$$
\begin{aligned}
R(\hat{\boldsymbol{C}};\boldsymbol{W}) &= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) \\
R(\tilde{\boldsymbol{C}};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}};\boldsymbol{W}) - \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11}) + (\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}) \\
R(\hat{\boldsymbol{C}}^{S};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}};\boldsymbol{W}) - (1-\tau^2)\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11}) + (1-\tau)^2(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}) \\
R(\hat{\boldsymbol{C}}^{PT};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}};\boldsymbol{W}) - \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11})H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) \\
&- 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M}) \times [H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)] \\
&+ (\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma})H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) \\
R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}};\boldsymbol{W}) - (1-\tau^2)\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11})H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) \\
&- 2(1-\tau)\operatorname{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M}) \times [H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)] \\
&+ (1-\tau)^2\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) \\
R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}};\boldsymbol{W}) - c \times \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11}) \\
&\quad \{2E(\chi^{-2}_{rn+2}(\Delta)) - cE(\chi^{-4}_{rn+2}(\Delta))\} + 2c \times \operatorname{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M}) \\
&\quad \{E(\chi^{-2}_{rn+2}(\Delta)) - E(\chi^{-2}_{rn+4}(\Delta))\} + c^2 \times (\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma})E(\chi^{-4}_{rn+4}(\Delta)), \\
R(\hat{\boldsymbol{C}}^{JS+};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W}) - \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11})E[(1-c\chi^{-2}_{rn+2}(\Delta))^2 \\
&\quad I(\chi^2_{rn+2}(\Delta) < c)] + \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}E[(1-c^2\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta) < c)] \\
&- 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta) < c)],
\end{aligned}
$$

where

$$\boldsymbol{M} = (\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{L}\boldsymbol{T})^{-1}$$

$$\boldsymbol{A} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'$$

$$\mathrm{tr}(\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}) = \mathrm{tr}(\boldsymbol{A}_{11})$$

$$\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*\boldsymbol{M}) = \mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11})$$

$$\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}) = \mathrm{tr}(\boldsymbol{A}_{11}).$$

Proof: See Appendix, Section 3.6.6.

Now, we provide the pairwise comparison of the estimators.

3.3.2 Comparison of $\hat{\boldsymbol{C}}^{SPT}$ and $\hat{\boldsymbol{C}}$

Consider the difference between two risks:

$$\begin{aligned} R(\hat{\boldsymbol{C}};\boldsymbol{W}) \quad - \quad & R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W}) = (1-\tau^2)\mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11})H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \\ & 2(1-\tau)\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)] - \\ & (1-\tau)^2\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}H_{rn+4}(\chi^2_{rn}(\alpha);\Delta)]. \end{aligned}$$

The right hand side is nonnegative whenever

$$tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M}) > \frac{(1-\tau)\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}H_{rn+4}(\chi^2_{rn}(\alpha);\Delta)}{2[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)]}.$$

In this range, $\hat{\boldsymbol{C}}^{SPT}$ performs better than $\hat{\boldsymbol{C}}$ and also under the null hypothesis $R(\hat{\boldsymbol{C}};\boldsymbol{W}) > R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W})$, since the difference between two risks for all α is positive. Furthermore, we get the same conclusion for $R(\hat{\boldsymbol{C}}^{PT})$ after taking $\tau = 0$ in $R(\hat{\boldsymbol{C}}^{SPT})$.

3.3.3 Comparison of $\hat{\boldsymbol{C}}^{SPT}$ and $\hat{\boldsymbol{C}}^{PT}$

For comparing the risk of $\boldsymbol{C}^{SPT}$ and $\boldsymbol{C}^{PT}$, we consider the difference between them:

$$\begin{aligned} R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W}) \quad - \quad & R(\hat{\boldsymbol{C}}^{PT};\boldsymbol{W}) = (\tau^2)\mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11})H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \\ & 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)] - \\ & (\tau^2)\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}H_{rn+4}(\chi^2_{rn}(\alpha);\Delta). \end{aligned}$$

The knowledge of Δ determines which estimator is to be chosen. From the difference between two risks, it is obvious that the risk of $\hat{\boldsymbol{C}}^{SPT}$ will be larger than the risk of $\hat{\boldsymbol{C}}^{PT}$ under the null hypothesis. Further, as Δ increases, the difference between two risks becomes negative and $\hat{\boldsymbol{C}}^{SPT}$ dominates $\hat{\boldsymbol{C}}^{PT}$ in the rest of the parameter space.

3.3.4 Comparison of $\hat{\boldsymbol{C}}^{SPT}$ and $\hat{\boldsymbol{C}}^{S}$

Let us consider the risk of $\hat{\boldsymbol{C}}^{SPT}$ and $\hat{\boldsymbol{C}}^{S}$:

$$\begin{aligned} R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W}) \quad = \quad & R(\hat{\boldsymbol{C}}^{S};\boldsymbol{W}) - (1-\tau^2)\mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11})[H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) - 1] + \\ & (1-\tau)^2\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - 1] + \\ & 2(\tau-1)\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)]. \end{aligned}$$

The risk of $\hat{\boldsymbol{C}}^{SPT}$ is smaller than $\hat{\boldsymbol{C}}^{S}$ for all Δ in $(0,\infty)$, whenever

$$\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma} < \frac{2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)]}{(1-\tau)[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - 1]}.$$

3.3.5 Comparison of $\tilde{C}$ and $\hat{C}^S$

We investigate the risk-difference of $\hat{\boldsymbol{C}}^S$ and $\tilde{\boldsymbol{C}}$

$$R(\tilde{\boldsymbol{C}};\boldsymbol{W}) - R(\hat{\boldsymbol{C}}^S;\boldsymbol{W}) = -\tau\{\tau \times \mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11}) - [\tau-2]\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}\}.$$

The risk of $\tilde{\boldsymbol{C}}$ is smaller than the risk of $\hat{\boldsymbol{C}}^S$ when the value of τ satisfies the following condition:

$$\tau\mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11}) > \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}(2-\tau).$$

Finally, based on the above finding we will suggest to use $\hat{\boldsymbol{C}}^{SPT}$. It provides a good control on the risk function unlike the sub-model estimators $\tilde{\boldsymbol{C}}$ and $\hat{\boldsymbol{C}}^S$. Under the candidate subspace, the risks of the estimators may be ordered as

$$R(\tilde{\boldsymbol{C}};\boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}^{PT};\boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}^{S};\boldsymbol{W}) \leq R(\hat{\boldsymbol{C}};\boldsymbol{W}).$$

3.3.6 Comparison of $\hat{C}^{JS}$ and $\hat{C}$

Let us consider the risk of $\hat{\boldsymbol{C}}^{JS}$ under the candidate subspace, in terms of the risk of $\hat{\boldsymbol{C}}$

$$\begin{aligned} R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W}) = {} & R(\hat{\boldsymbol{C}};\boldsymbol{W}) - 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*\boldsymbol{M})E(\chi^{-2}_{rn+2}(\Delta)) + \\ & c^2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*)E(\chi^{-4}_{rn+2}(\Delta)) + 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E(\chi^{-2}_{rn+2}(\Delta)) - \\ & 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E(\chi^{-2}_{rn+4}(\Delta)) + c^2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')E(\chi^{-4}_{rn+4}(\Delta)). \end{aligned}$$

The risk difference is positive whenever

$$\mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11}) < \frac{-\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}'\boldsymbol{\gamma}\boldsymbol{M})E(\chi^{-2}_{rn+4}(\Delta))}{E(\chi^{-2}_{rn+2}(\Delta))}.$$

The risk of $\hat{\boldsymbol{C}}^{JS}$ is smaller than $\hat{\boldsymbol{C}}$ when, for all Δ, the opposite of above inequality is satisfied.

3.3.7 Comparison of $\hat{\boldsymbol{C}}^{JS+}$ and $\hat{\boldsymbol{C}}^{JS}$

For comparing the risk of $\hat{\boldsymbol{C}}^{JS}$ and $\hat{\boldsymbol{C}}^{JS+}$, we consider the risk difference

$$\begin{aligned} R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W}) \quad - & \quad R(\hat{\boldsymbol{C}}^{JS+};\boldsymbol{W}) = \\ + & \quad \mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11})E[(1-c\chi^{-2}_{rn+2}(\Delta))^2 I(\chi^2_{rn+2}(\Delta)<c)] \\ - & \quad \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}E[(1-c^2\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta)<c)] \\ + & \quad 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+2}(\Delta)<c)]. \end{aligned}$$

The right hand side is positive, since the expectation of a positive random variable is positive by the definition of an indicator function,

$$(0<\chi^2_{rn+4}(\Delta)<c) \Longleftrightarrow (c\chi^{-2}_{rn+4}(\Delta)-1)\geq 0;$$

therefore,

$$E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+2}(\Delta)<c)] \leq 0.$$

Thus, for all Δ, $R(\hat{\boldsymbol{C}}^{JS+};\boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W})$ and $\hat{\boldsymbol{C}}^{JS+}$ not only confirms inadmissibility of $\hat{\boldsymbol{C}}^{JS}$ but also provides a simple superior estimator.

3.3.8 Comparison of $\tilde{\boldsymbol{C}}$ and $\hat{\boldsymbol{C}}^{JS}$

We investigate the risk-difference of $\hat{\boldsymbol{C}}^{JS}$ and $\tilde{\boldsymbol{C}}$ under candidate subspace is

$$\begin{aligned} R(\hat{\boldsymbol{C}}^{JS}; \boldsymbol{W}) &= R(\tilde{\boldsymbol{C}}; \boldsymbol{W}) + \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11})[1 - cE(\chi^{-2}_{rn+2}(0)]^2 \\ &\geq R(\tilde{\boldsymbol{C}}; \boldsymbol{W}). \end{aligned}$$

Thus, $\tilde{\boldsymbol{C}}$ performs better than $\hat{\boldsymbol{C}}^{JS}$ under candidate subspace. However, as Δ moves away from the origin, the risk of $\tilde{\boldsymbol{C}}$ becomes unbounded, and the risk of $\hat{\boldsymbol{C}}^{JS}$ remains below the risk of $\hat{\boldsymbol{C}}$ and merges with it as $\Delta \to \infty$. Therefore, $\hat{\boldsymbol{C}}^{JS}$ dominates $\tilde{\boldsymbol{C}}$ outside an interval around the origin.

Finally, based on the above finding, we can order the risk of the estimators under the candidate subspace as

$$R(\tilde{\boldsymbol{C}}; \boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}^{JS+}; \boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}^{JS}; \boldsymbol{W}) \leq R(\hat{\boldsymbol{C}}; \boldsymbol{W}).$$

3.4 Numerical Study

3.4.1 Simulation Study

In this section, we use Monte Carlo simulation experiments to examine the relative performance of the proposed estimators. In this study, we simulate the data from the following model:

$$Y_i = \mathbf{c}_0 + \mathbf{c}_1 x_{1i} + \mathbf{c}_2 x_{2i} + \mathbf{c}_3 x_{3i} + \mathbf{c}_4 x_{4i} + \boldsymbol{\epsilon}_i \equiv \boldsymbol{C} X_i + \boldsymbol{\epsilon}_i \qquad i = 1, \ldots, n,$$

where $Y_i = (y_{1i}, y_{2i}, y_{3i}, y_{4i})'$ and $X_i = (1, x_{1i}, x_{2i}, x_{3i}, x_{4i})'$ with $m = 1, 2, 3, 4$. Therefore, $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{X}}$, and $\boldsymbol{C}$ denote $4 \times n$, $5 \times n$, and 4×5 data matrices, respectively. For simulation, we consider

$$\boldsymbol{C} = ((0.5, 0.5, 0.25, 0.5, 0.25)', (0, 0.12, 0.1, 0.5, -0.5)', (-0.14, 0, -0.1, -0.5, 0)',$$

$$(-0.1, 0, -0.03, 0.4, 0.11)',$$

$\boldsymbol{K} = [I_4]$, and $\boldsymbol{L} = [0', I_1']'$, or more explicitly,

$$\boldsymbol{K} = ((1, 0, 0, 0)', (0, 1, 0, 0)', (0, 0, 1, 0)', (0, 0, 0, 1)'),$$

$$\boldsymbol{L} = ((0, 0, 0, 0, 0)', (0, 0, 0, 0, 0)', (0, 0, 0, 0, 0)',$$

$$(0, 0, 0, 0, 0)', (1, 0, 0, 0, 0)')'.$$

Under the candidate subspace, we generate 5000 samples using the above model. We define the Δ as a departure parameter which is a function of the distance between the true value of $\boldsymbol{C}$ and that under the null hypothesis. In order to investigate the behavior of the proposed estimators, different values of $\boldsymbol{C}$ were chosen to produce the value of Δ between 0 and 4. The performance of an estimator of $\boldsymbol{C}$ will be reappraised using the mean square error criterion. All computations were conducted using the **R** statistical system. We numerically calculated the relative risk of $\tilde{\boldsymbol{C}}$, $\hat{\boldsymbol{C}}^{PT}$, $\hat{\boldsymbol{C}}^{JS}$ and $\hat{\boldsymbol{C}}^{JS+}$ with respect to $\hat{\boldsymbol{C}}$ by simulation. The simulated relative efficiency of the estimator $\boldsymbol{C}^*$ to the unrestricted $\hat{\boldsymbol{C}}$ is defined by $R.E = risk(\hat{\boldsymbol{C}})/risk(\hat{\boldsymbol{C}}^*)$. Since the result for different n were similar, here we only report the results for $n = 20$ and $n = 40$ in Figures (3.1) and (3.2).

We can see the relative efficiencies of the estimators change with the value of the departure parameter Δ. The figures reconfirm the typical characteristics of the listed estimators. We conclude that $\tilde{\boldsymbol{C}}$ and $\hat{\boldsymbol{C}}^{PT}$ dominate the usual $\hat{\boldsymbol{C}}$ at or near the candidate subspace. $\hat{\boldsymbol{C}}^{JS}$ and $\hat{\boldsymbol{C}}^{JS+}$ are more efficient than an unrestricted estimator in the unrestricted parameter space; for $\Delta = 0$, or in a neighborhood of that, $\hat{\boldsymbol{C}}^{PT}$ is more efficient that $\hat{\boldsymbol{C}}^{JS}$ and $\hat{\boldsymbol{C}}^{JS+}$. Note that, for a larger value of α, the level of significance, $\hat{\boldsymbol{C}}^{JS+}$ dominates $\hat{\boldsymbol{C}}^{PT}$ uniformly. For the larger value of Δ, $\hat{\boldsymbol{C}}^{PT}$ becomes more inefficient than the unrestricted estimator, and its efficiency value monotonically decreases, achieves a minimum after crossing the efficiency line at 1, and then monotonically increases and approaches the $\hat{\boldsymbol{C}}$ efficiency. Under the candidate subspace, $\tilde{\boldsymbol{C}}$ is more efficient than all the other estimators, but, as Δ increases, its efficiency converges to zero since it is an unbounded function of Δ.

3.4.2 Real Data Example I

As a first data example, we consider multivariate regression analysis methods on data from a study by Rohwer (given in Timm (1975)) on kindergarten children. It was designed to determine how well a set of paired-associate (PA) tasks determine performance on some tests. The data involve a sample from an upper-class, white, residential school. The data considered in this example consist of 32 kindergarten students and three response variables: the peabody picture vocabulary test (y_1); the raven progressive matrices test (y_2); and a student achievement test (y_3). PA learning

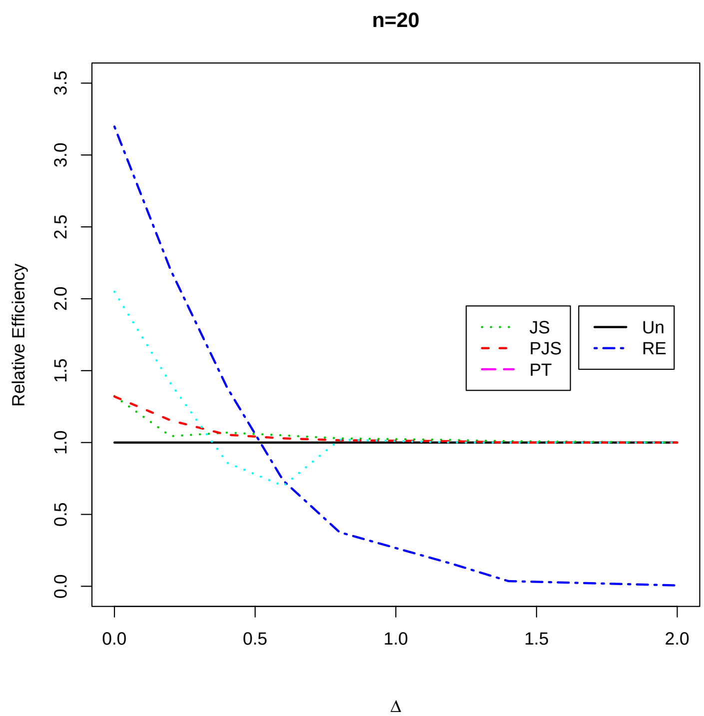

Figure 3.1: R.E of the estimators for $n = 20$.

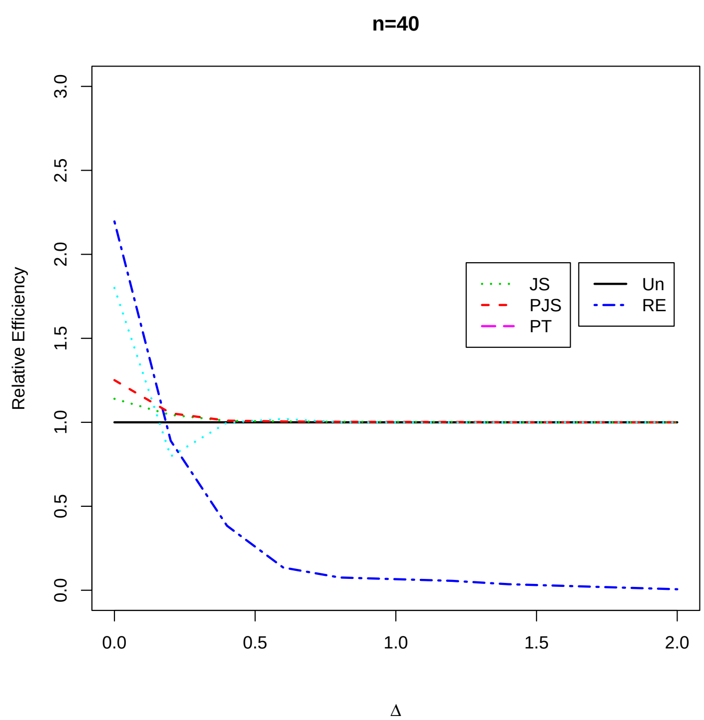

Figure 3.2: R.E of the estimators for $n = 40$.

proficiency tasks represent the sum of the number of items correct out of 20 (on two exposures). The tasks involved prompts that were used to facilitate learning. The five PA word prompts involved x_1-named (N), x_2-still (S), x_3-named action (NA), x_4-named still (NS), and x_5-sentence still (SS) prompts. We consider the following model for the i^{th} vector of responses of the form:

$$Y_i = \mathbf{c}_0 + \mathbf{c}_1 x_{1i} + \mathbf{c}_2 x_{2i} + \mathbf{c}_3 x_{3i} + \mathbf{c}_4 x_{4i} + \mathbf{c}_5 x_{5i} + \boldsymbol{\epsilon}_i \equiv \boldsymbol{C} X_i + \boldsymbol{\epsilon}_i \qquad i = 1, \ldots, n,$$

where $Y_i = (y_{1i}, y_{2i}, y_{3i})'$ and $X_i = (1, x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i})'$ with $m = 1, 2, 3$ and $n = 32$. Therefore $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{X}}$, and $\boldsymbol{C}$ denote 3×32, 6×32, and 3×6 data matrices, respectively. The closest literature to the model and moment selection results of this chapter is that concerning likelihood-based model selection criteria. We now review the criteria related to this chapter. The AIC criterion was introduced by Akaike (1969). The BIC criterion was introduced by Schwarz (1978), Rissanen (1978), and Akaike (1977). HQ and HQIC criterion were introduced by Hannan and Quinn (1979). The stepwise, C_p, and HQ procedures selected variables X_1, X_2, X_3, and X_4, while the corrected AIC selected only variables X_2 and X_4. The uncorrected criteria AIC, BIC, and $HQIC$ only selected one variable X_4. All methods exclude the fifth variable (SS). Based on all methods, the predictor variable X_5 does not enter into the linear regression model, so the restrictions $\{c_{15} = c_{25} = c_{35} = 0\}$ can be imposed on the model. The coefficient matrices of the subspace are selected as $\boldsymbol{K} = [I_3]$ and $\boldsymbol{L} = [0', I_1']'$ or, more explicitly,

$$\boldsymbol{K} = ((1,0,0)', (0,1,0)', (0,0,1)')$$

Table 3.1: Estimate of the regression coefficient matrix

	Estimator	*Intercept*	x_1	x_2	x_3	x_4	x_5
	$\hat{\boldsymbol{C}}$	39.76467	0.34929	0.40265	-1.05918	1.89365	0.39208
	$\tilde{\boldsymbol{C}}$	41.94818	0.33364	0.57819	-0.99751	2.06337	0
y_1	$\hat{\boldsymbol{C}}^S$	40.85643	0.34147	0.49042	-1.02834	1.97851	0.19604
	$\hat{\boldsymbol{C}}^{PT}$	39.79491	0.34915	0.41428	-1.25860	1.89523	0.38845
	$\hat{\boldsymbol{C}}^{SPT}$	39.77479	0.349225	0.40346	-1.05889	1.86444	0.36026
	$\hat{\boldsymbol{C}}$	12.97366	0.03742	0.50580	-0.18529	0.15699	-0.13105
	$\tilde{\boldsymbol{C}}$	12.24380	0.04265	0.44713	-0.20590	0.10026	0
y_2	$\hat{\boldsymbol{C}}^S$	12.60873	0.04004	0.47647	-0.19560	0.12863	-0.06552
	$\hat{\boldsymbol{C}}^{PT}$	12.94386	0.03763	0.50341	-0.18613	0.15467	-0.12570
	$\hat{\boldsymbol{C}}^{SPT}$	12.95876	0.03753	0.50461	-0.18571	0.15583	-0.12838
	$\hat{\boldsymbol{C}}$	-45.04718	2.14209	2.83837	-5.32059	6.22416	-0.49068
	$\tilde{\boldsymbol{C}}$	-47.77978	2.16168	2.61869	-5.39776	6.01177	0
y_3	$\hat{\boldsymbol{C}}^S$	-46.41348	2.15188	2.72853	-5.35917	6.11797	-0.24534
	$\hat{\boldsymbol{C}}^{PT}$	-45.23896	2.14203	2.83903	-5.14036	6.29480	-0.49215
	$\hat{\boldsymbol{C}}^{SPT}$	-45.04307	2.14206	2.83870	-5.32047	6.22448	-0.49142

and

$$\boldsymbol{L} = ((0,0,0,0,0,0)', (0,0,0,0,0,0)', (0,0,0,0,0,0)',$$

$$(0,0,0,0,0,0)', (0,0,0,0,0,0)', (1,0,0,0,0,0)')'.$$

The least squares estimates of the regression coefficient matrix $\hat{\boldsymbol{C}}$, $\tilde{\boldsymbol{C}}$, $\hat{\boldsymbol{C}}^S$, $\hat{\boldsymbol{C}}^{PT}$, and $\hat{\boldsymbol{C}}^{SPT}$ are given in Table 3.1.

In an effort to appraise the performance of the suggested estimators, we conduct a numerical study that implements bootstrapping. We conduct bootstrapping for 5000 replicates to evaluate the performance of suggested estimators in our data example. The performance of the estimators was evaluated in terms of the relative efficiency of estimators. The relative efficiencies are given in Table 3.2, assuming the candidate subspace is correct.

Table 3.2: The relative efficiency of estimators

Estimator	$R.E(\hat{\boldsymbol{C}} : \hat{\boldsymbol{C}}^*)$
$\tilde{\boldsymbol{C}}$	3.8894
$\hat{\boldsymbol{C}}^S$	2.3503
$\hat{\boldsymbol{C}}^{PT}$	2.7039
$\hat{\boldsymbol{C}}^{SPT}$	2.7806

3.4.3 Real Data Example II

As a second data example, we consider multivariate regression analysis methods and some biochemical data taken from a study by Smith et al. (1962). The data involve several biochemical measurements and characteristics of urine specimens from men classified under two weight groups: overweight and underweight. Each subject contributed two or three morning samples of urine, and the data considered in this example consist of the 33 individual samples; five response variables: pigment creatinine (y_1), concentrations of phosphate (y_2), phosphorous (y_3), creatinine (y_4), and choline (y_5); and three predictor variables: weight of each subject (x_1), volume (x_2), and specific gravity (x_3). The multivariate data set for this example is presented in Table 3.5 in Section 3.6. We consider the following model for the i^{th} vector of responses of the form

$$Y_i = \mathbf{c}_0 + \mathbf{c}_1 x_{1i} + \mathbf{c}_2 x_{2i} + \mathbf{c}_3 x_{3i} + \boldsymbol{\epsilon}_i \equiv \boldsymbol{C} X_i + \boldsymbol{\epsilon}_i \qquad i = 1, \ldots, 33,$$

where $Y_i = (y_{1i}, y_{2i}, y_{3i}, y_{4i}, y_{5i})'$ and $X_i = (1, x_{1i}, x_{2i}, x_{3i})'$ with $n = 5$. Therefore $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{X}}$, and $\boldsymbol{C}$ denote 5×33, 4×33, and 5×4 data matrices, respectively. Based on stepwise selection, the predictor variable X_3 does not enter into the linear regression model for the first four components of the response variables $\boldsymbol{\mathcal{Y}}$. So, the restrictions $\{c_{13} = c_{23} = c_{33} = c_{43} = 0\}$ can be imposed to the model. The

Table 3.3: Estimates of the regression coefficient matrix

	Estimator	*Intercept*	x_1	x_2	x_3
y_1	$\hat{\boldsymbol{C}}$	15.3469	-2.91975	1.9534	0.1992
	$\tilde{\boldsymbol{C}}$	15.7497	-2.8132	1.9364	0
	$\hat{\boldsymbol{C}}^{JS}$	15.8659	-3.7742	2.6142	0.1042
	$\hat{\boldsymbol{C}}^{JS+}$	15.8541	-3.5594	2.4659	0.0813
y_2	$\hat{\boldsymbol{C}}$	1.4159	0.6043	-0.4815	0.2666
	$\tilde{\boldsymbol{C}}$	2.2413	0.6530	-0.5702	0
	$\hat{\boldsymbol{C}}^{JS}$	3.7064	0.9704	-0.7264	0.6097
	$\hat{\boldsymbol{C}}^{JS+}$	3.3883	0.8994	-0.6928	0.4961
y_3	$\hat{\boldsymbol{C}}$	2.0187	0.5768	-0.4245	-0.0400
	$\tilde{\boldsymbol{C}}$	1.8842	0.5650	-0.4041	0
	$\hat{\boldsymbol{C}}^{JS}$	1.87203	0.6381	-1.0681	-0.4153
	$\hat{\boldsymbol{C}}^{JS+}$	1.8775	0.6205	-0.9281	-0.3187
y_4	$\hat{\boldsymbol{C}}$	1.8717	0.6159	-0.5780	0.3518
	$\tilde{\boldsymbol{C}}$	2.9710	0.6825	-0.7009	0
	$\hat{\boldsymbol{C}}^{JS}$	3.80438	0.2832	-1.0717	0.1340
	$\hat{\boldsymbol{C}}^{JS+}$	3.6220	0.3647	-0.9888	0.1007
y_5	$\hat{\boldsymbol{C}}$	-0.8902	1.3797	-0.6289	2.8907
	$\tilde{\boldsymbol{C}}$	-1.8248	1.5113	-0.5216	3.0692
	$\hat{\boldsymbol{C}}^{JS}$	-2.8369	1.7688	-0.4883	3.0692
	$\hat{\boldsymbol{C}}^{JS+}$	-2.6398	1.7085	-0.4911	3.0692

coefficient matrices of the subspace are selected as $\boldsymbol{K} = [I_4, 0]$ and $\boldsymbol{L} = [0', I_1']'$ or, more explicitly, $\boldsymbol{K} = ((1,0,0,0,0)', (0,1,0,0,0)', (0,0,1,0,0)', (0,0,0,1,0)')$ and $\boldsymbol{L} = ((0,0,0,0)', (0,0,0,0)', (0,0,0,0)', (1,0,0,0)')'$. The least squares estimate of the regression coefficient matrix $\hat{\boldsymbol{C}}$, $\hat{\boldsymbol{C}}^{JS}$, and $\hat{\boldsymbol{C}}^{JS}$ are given in Table 3.3.

We conduct a bootstrap with 1000 replicates to evaluate the performance of suggested estimators in our data example. The results for relative efficiency at $\Delta = 0$ are given in Table 3.4.

Table 3.4: The relative efficiency of estimators

$Estimator$	$R.E(\hat{\boldsymbol{C}} : \hat{\boldsymbol{C}}^{*})$
$\tilde{\boldsymbol{C}}$	6.07
$\hat{\boldsymbol{C}}^{JS}$	2.35
$\hat{\boldsymbol{C}}^{JS+}$	2.54

3.5 Concluding Remarks

The goal of this chapter was to examine the relative performance of estimators based on full model and submodels in the context of multivariate multiple regression models when X is fixed. However, studying pretest and shrinkage estimation strategies for a matrix parameter in a MMRM is not considered much in the reviewed literature. Mostly it was studied for a vector form of parameter. Here we extended our analysis to a matrix form. The fundamental results of Sclove et al. (1972) could not be directly implemented to derive the expressions. Therefore, we first generalize the results of Sclove et al. (1972).

We succinctly investigated the bias and risk properties of the suggested estimators. The shrinkage estimator provides a wider range than the restricted estimator in which it dominates the unrestricted one. The pretest estimators with data based weights outperform the full model estimator in a meaningful part of the parameter space induced by the candidate subspace. A shrinkage pretest estimator dominates the unrestricted estimator in a wider range that the pretest estimator. Further, the suggested approach is free from any tuning parameters, and calculations are not iterative. It would also be interesting to replace the known $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ by an estimated one. We leave these for future investigation. We conclude that the risk improvement of the submodel estimator over other estimators is substantial near the restriction. However, the improvement starts diminishing as the restriction moves further and further away from assumed subspaces. Thus, the performance of the submodel estimator heavily depends on the quality of the subspace information. In summary, we show large gains of a suggested shrinkage approach over an ordinary least-square. Finally, real data examples and the simulation study support the contention that the suggested method is superior to classical estimation.

3.6 Appendix: Proof of Main Results

3.6.1 Proof of Theorem 3.2.1

Since $\boldsymbol{p}_1, \boldsymbol{p}_2$, and $\boldsymbol{p}_3$ are normal, the joint distribution of $(\boldsymbol{p}_1, \boldsymbol{p}_2)$ and $(\boldsymbol{p}_2, \boldsymbol{p}_3)$ will also be normal.

$$\begin{aligned}
E(\boldsymbol{p}_2) &= E[vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})] \\
&= E[vec(\boldsymbol{S}(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})\boldsymbol{T})] \\
&= (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')E(vec(\hat{\boldsymbol{C}})) \\
&= (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}) \\
Cov(\boldsymbol{p}_2) &= Cov(vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})) \\
&= Cov(vec(\boldsymbol{S}(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})\boldsymbol{T})) \\
&= Cov[(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\hat{\boldsymbol{C}})] \\
&= (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')Cov[vec(\hat{\boldsymbol{C}})](\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')' \\
&= (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')(\Sigma_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT}) \\
&= \boldsymbol{\Sigma}^*
\end{aligned}$$

$$\begin{aligned}
E(\boldsymbol{p}_3) &= E(\boldsymbol{p}_1 - \boldsymbol{p}_2) \\
&= E[vec(\hat{\boldsymbol{C}} - \boldsymbol{C}) - vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})] \\
&= E[vec(\hat{\boldsymbol{C}} - \boldsymbol{C})] - E[vec(\boldsymbol{S}(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})\boldsymbol{T})] \\
&= 0 - E[(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')E(vec(\hat{\boldsymbol{C}}))] \\
&= -(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}) \\
Cov(\boldsymbol{p}_3) &= Cov(\boldsymbol{p}_1 - \boldsymbol{p}_2) \\
&= Cov(\boldsymbol{p}_1) + Cov(\boldsymbol{p}_1) - 2Cov(\boldsymbol{p}_1, \boldsymbol{p}_2) \\
&= (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) + \boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21} \\
&= \boldsymbol{\Omega}^*
\end{aligned}$$

3.6.2 Proof of Lemma 3.2.1

Let $\boldsymbol{\mathcal{Y}} = (Y_1, \ldots, Y_n)$ be an $m \times n$ matrix following a normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}}$, and the columns are each random m-vectors with covariance matrix $\boldsymbol{\Sigma}$. Note that pairs of column vectors, $(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_k)$, $j \neq k$, are uncorrelated with each other. Applying the vectoring operation, we get $E(\boldsymbol{y}) = E(vec(\boldsymbol{\mathcal{Y}})) = \boldsymbol{\mu}_{\boldsymbol{y}}$ and $cov(\boldsymbol{y}) = cov(vec(\boldsymbol{\mathcal{Y}})) = (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes \boldsymbol{I}) = \boldsymbol{\Sigma}_{\boldsymbol{y}}$. Hence, by arranging $\boldsymbol{\mathcal{Y}}$ to a vector form as $\boldsymbol{y} = (y_1, y_2, \ldots, y_{nm})'$, where $\boldsymbol{y}'\boldsymbol{y} = \sum_{j=1}^{nm} y_j^2$, we can rewrite the left hand side of Lemma 3.2.1 as

$$\begin{aligned}
E[\boldsymbol{y}\phi(\boldsymbol{y}'\boldsymbol{y})] &= \Bigg\{E\Bigg[E\Bigg[y_1\phi(y_1^2 + \sum_{j=2}^{nm} y_j^2)|y_j, j \neq 1\Bigg]\Bigg], \ldots, E\Bigg[E\Bigg[y_{nm}\phi(y_{nm}^2 \\
&+ \sum_{j=1}^{nm-1} y_j^2)|y_j, j \neq nm\Bigg]\Bigg]\Bigg\}'.
\end{aligned}$$

Now, using the result of Lemma 2 in Judge and Bock (1978), (p. 320), we have

$$E[y_i\phi(\boldsymbol{y}'\boldsymbol{y})] = \mu_i E\Big[E\Big[\phi(\chi^2_{(3,\mu_i^2/2)} + \sum_{j\neq i} y_j^2)|y_j, j \neq i\Big]\Big].$$

Using the fact that the distribution of Z_i is $\chi^2_{(1,\mu_i)}$, then $\sum_{i=1}^{p} Z_i$ is $\chi^2_{(p,\sum_{i=1}^{p}\mu_i)}$. Hence $E[\boldsymbol{y}\phi(\boldsymbol{y}'\boldsymbol{y})] = (\mu_1, \ldots, \mu_{nm})' E[\phi(\chi^2_{(nm+2,\mu'\mu/2)})]$.

3.6.3 Proof of Lemma 3.2.2

Consider the diagonal elements of $E[\boldsymbol{y}\boldsymbol{y}'\phi(\boldsymbol{y}'\boldsymbol{y})]$. By using the result of Lemma 1 in Judge and Bock (1978), (p. 320), we get

$$\begin{aligned}
E[y_i^2\phi(\sum_{i}^{nm} y_i^2)] &= E\Big[E\Big[y_i^2\phi(y_i^2 + \sum_{j\neq i} y_j^2)|y_j^2, j \neq i\Big]\Big] \\
&= E\Big[E\Big[\phi(\chi^2_{(3,\mu_i^2/2)} + \sum_{j\neq i} y_j^2)\Big]|y_j^2, j \neq i\Big] \\
&+ \mu_i^2 E\Big[\phi(\chi^2_{(5,\mu_i^2/2)} + \sum_{j\neq i} y_j^2)\Big]|y_j^2, j \neq i\Big] \\
&= E\Big[\phi(\chi^2_{(nm+2,\mu'\mu/2)})\Big] + \mu_i^2 E\Big[\phi(\chi^2_{(nm+4,\mu'\mu/2)})\Big].
\end{aligned}$$

Then, for the off-diagonal elements, we consider the result of Lemma 2 given in Judge and Bock (1978), (p. 320), for $i \neq j$, so that

$$
\begin{aligned}
E\Big[y_i y_j \phi(\sum_{k=1}^{nm} y_k^2)\Big] &= E\Big[y_j E\Big[y_i \phi(y_i^2 + \sum_{k\neq i} y_k^2)|y_k, k\neq i\Big]\Big] \\
&= E\Big[y_j \mu_i E\Big[\phi\Big(\chi^2_{(3,\mu_i^2/2)} + \sum_{k\neq i} y_k^2\Big)|y_k, k\neq i\Big]\Big] \\
&= E\Big[y_j \mu_i E\Big[\phi\Big(\chi^2_{(3,\mu_i^2/2)} + y_j^2 \\
&+ \sum_{k\neq i, k\neq j} y_k^2\Big)|\chi^2_{(3,\mu_i^2/2)}, y_k, k\neq i, k\neq j\Big]\Big].
\end{aligned}
$$

The unconditional expectations of the off-diagonal elements are

$$
\mu_i \mu_j E\Big[\phi\Big(\chi^2_{(3,\mu_j^2/2)} + \sum_{k\neq i, k\neq j} y_k^2 + \chi^2_{(3,\mu_i^2/2)}\Big)\Big] = \mu_i \mu_j E\Big[\phi(\chi^2_{(nm+4, \sum_{i=1}^{nm} \mu_i^2/2)})\Big].
$$

Now, combining the diagonal and off-diagonal components, we get the desired result.

3.6.4 Proof for Theorem 3.3.1

Here, we provide the proof of bias expressions for all proposed estimators. Clearly, the bias of $\tilde{\boldsymbol{C}}$ is equal to $-\boldsymbol{\gamma}$.

$$
\begin{aligned}
B(\hat{\boldsymbol{C}}^S) &= E[vec(\hat{\boldsymbol{C}}^S - \boldsymbol{C})] \\
&= E[vec(\tilde{\boldsymbol{C}} - \boldsymbol{C}) + \tau \times vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})],
\end{aligned}
$$

using Theorem 3.2.1, we get the desired result.

$$\begin{aligned} B(\hat{\boldsymbol{C}}^{PT}) &= E[vec(\hat{\boldsymbol{C}}^{PT} - \boldsymbol{C})] \\ &= E[vec(\hat{\boldsymbol{C}} - \boldsymbol{C} - (\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})I(D < \chi^2_{rn}(\alpha))] \\ &= E[-vec(\boldsymbol{S}(\boldsymbol{K}\hat{\boldsymbol{C}}\boldsymbol{L})\boldsymbol{T}I(D < \chi^2_{rn}(\alpha)))] \\ &= E[-(\boldsymbol{S}\boldsymbol{K} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\hat{\boldsymbol{C}})I(D < \chi^2_{rn}(\alpha))], \end{aligned}$$

using Theorem 3.2.1, Lemma 3.2.1, and some computations, we get the above result.

$$\begin{aligned} B(\hat{\boldsymbol{C}}^{SPT}) &= E[vec(\hat{\boldsymbol{C}}^{SPT} - \boldsymbol{C})] \\ &= E\{vec[\hat{\boldsymbol{C}}I(D > d_{rn,\alpha}) + \tau\hat{\boldsymbol{C}}I(D < d_{rn,\alpha}) + (1-\tau)\tilde{\boldsymbol{C}}I(D < d_{rn,\alpha})]\} \\ &= E\{vec[\hat{\boldsymbol{C}}I(D > d_{rn,\alpha}) + \hat{\boldsymbol{C}}I(D < d_{rn,\alpha}) - \hat{\boldsymbol{C}}I(D < d_{rn,\alpha}) + \\ &\quad \tau\hat{\boldsymbol{C}}I(D < d_{rn,\alpha}) + (1-\tau)\tilde{\boldsymbol{C}}I(D < d_{rn,\alpha})]\} \\ &= E\{vec[\hat{\boldsymbol{C}} - (1-\tau)(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})I(D < d_{rn,\alpha})]\}, \end{aligned}$$

finally, using Theorem 3.2.1 and Lemma 3.2.1 we obtained the desired result.

The proof of $B(\hat{\boldsymbol{C}}^{JS})$ is as follows,

$$\begin{aligned} B(\hat{\boldsymbol{C}}^{JS}) &= E[vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C})] \\ &= E[vec(\tilde{\boldsymbol{C}} + \{1 - cD^{-1}\}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}}) - \boldsymbol{C})] \\ &= E[vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})cD^{-1}] \end{aligned}$$

using Theorem 3.2.1 and Lemma 3.2.1, we get the desired result. Finally we have

$$
\begin{aligned}
B(\hat{\boldsymbol{C}}^{JS^+}) &= E[vec(\tilde{\boldsymbol{C}} + \{1 - cD^{-1}\}^{+}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}}) - \boldsymbol{C})] \\
&= E(vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C}) - E(vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})(1 - cD^{-1})I(D < c)) \\
&= B(\hat{\boldsymbol{C}}^{JS}) - \{E[vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})I(D < c)] + E[vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})cD^{-1}I(D < c)]\} \\
&= B(\hat{\boldsymbol{C}}^{JS}) - (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C})E\{I(\chi^2_{rn+2}(\Delta) < c)\} + \\
&\quad c(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C})E\{\chi^{-2}_{rn+2}(\Delta)I(\chi^2_{rn+2}(\Delta) < c)\}.
\end{aligned}
$$

By factoring the term $-(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C})$, we get the result.

3.6.5 Proof of Lemma 3.3.1:

For the proof of part (i), we use the result of Theorem 3.2.1 part (i),(ii), and (iv):

$$
\begin{aligned}
E(\boldsymbol{p}_1|\boldsymbol{p}_2) &= E(\boldsymbol{p}_1) + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{p}_2 - E(\boldsymbol{p}_2)) \\
&= (\boldsymbol{\Sigma}_{\boldsymbol{\epsilon\epsilon}} \otimes (\boldsymbol{\mathcal{X}\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT}) \\
&\quad [(\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon\epsilon}} \otimes (\boldsymbol{\mathcal{X}\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{LT})]^{-1} \\
&\quad [\boldsymbol{p}_2 - (\boldsymbol{SK} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C})].
\end{aligned}
$$

We got the result by using left Kronecker product rules . A proof of (ii) is similar to the proof of (iii), provided below. Based on Theorem 3.2.1 part (i),(ii), and (iv), we have

$$
\begin{aligned}
E(\boldsymbol{p}_2\boldsymbol{p}_1{}'I(D<\chi^2_{rn}(\alpha))) &= E(E(\boldsymbol{p}_2\boldsymbol{p}_1{}'I(D<\chi^2_{rn}(\alpha))|\boldsymbol{p}_2)) \\
&= E(\boldsymbol{p}_2E(\boldsymbol{p}_1{}'I(D<\chi^2_{rn}(\alpha))|\boldsymbol{p}_2)) \\
&= E(\boldsymbol{p}_2[E(\boldsymbol{p}_1)+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{p}_2-E(\boldsymbol{p}_2))]'I(D<\chi^2_{rn}(\alpha))) \\
&= E(\boldsymbol{p}_2[(\boldsymbol{p}_2-\boldsymbol{\gamma})'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}I(D<\chi^2_{rn}(\alpha))]) \\
&= E(\boldsymbol{p}_2\boldsymbol{p}_2{}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}I(D<\chi^2_{rn}(\alpha))) \\
&- E(\boldsymbol{p}_2)\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}I(D<\chi^2_{rn}(\alpha)) \\
&= [Var(\boldsymbol{p}_2)H_{rn+2}(\chi^2_{rn}(\alpha))+\boldsymbol{\gamma}\boldsymbol{\gamma}'H_{rn+4}(\chi^2_{rn}(\alpha))]\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21} \\
&- \boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}H_{rn+2}(\chi^2_{rn}(\alpha)) \\
&= \boldsymbol{\Sigma}^*(\boldsymbol{K}'\boldsymbol{S}'\otimes\boldsymbol{LT})^{-1}H_{rn+2}(\chi^2_{rn}(\alpha))+\boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{K}'\boldsymbol{S}'\otimes\boldsymbol{LT})^{-1} \\
&\quad [H_{rn+4}(\chi^2_{rn}(\alpha))-H_{rn+2}(\chi^2_{rn}(\alpha))], \\
E(\boldsymbol{p}_2\boldsymbol{p}_1{}'D^{-1}) &= E[E(\boldsymbol{p}_2\boldsymbol{p}_1{}'D^{-1})|\boldsymbol{p}_2] \\
&= E(\boldsymbol{p}_2E(\boldsymbol{p}_1{}'D^{-1}|\boldsymbol{p}_2)) \\
&= E(\boldsymbol{p}_2[E(\boldsymbol{p}_1)+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{p}_2-E(\boldsymbol{p}_2))]'D^{-1}) \\
&= E(\boldsymbol{p}_2[(\boldsymbol{p}_2-\boldsymbol{\gamma})'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}D^{-1}]) \\
&= E(\boldsymbol{p}_2\boldsymbol{p}_2{}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}D^{-1}-\boldsymbol{p}_2\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}D^{-1}) \\
&= \boldsymbol{\Sigma}_{21}E(\chi^{-2}_{rn+2}(\Delta))+\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}[E(\chi^{-2}_{rn+4}(\Delta))-E(\chi^{-2}_{rn+2}(\Delta))] \\
&= \boldsymbol{\Sigma}^*(\boldsymbol{K}'\boldsymbol{S}'\otimes\boldsymbol{LT})^{-1}E(\chi^{-2}_{rn+2}(\Delta)) \\
&+ \boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}[E(\chi^{-2}_{rn+4}(\Delta))-E(\chi^{-2}_{rn+2}(\Delta))].
\end{aligned}
$$

3.6.6 Proof of Theorem 3.3.2

Following the definition in (3.8), clearly the risk of $\hat{\boldsymbol{C}}$ is equal to $\mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})$. Note that

$$MSE(\tilde{\boldsymbol{C}}) = E\{vec(\tilde{\boldsymbol{C}} - \boldsymbol{C})(vec(\tilde{\boldsymbol{C}} - \boldsymbol{C}))'\}.$$

Using the definition of $\tilde{\boldsymbol{C}}$ and Theorem 3.2.1, we have

$$\begin{aligned} MSE(\tilde{\boldsymbol{C}}) &= E\{\boldsymbol{p}_3\boldsymbol{p}_3{}'\} \\ &= (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2(\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{L}\boldsymbol{T}) + \\ &\quad (\boldsymbol{S}\boldsymbol{K} \otimes \boldsymbol{T}'\boldsymbol{L}')(\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})(\boldsymbol{K}'\boldsymbol{S}' \otimes \boldsymbol{L}\boldsymbol{T}) + \gamma\gamma'. \end{aligned}$$

Using the definition in (3.8) we have,

$$\begin{aligned} R(\tilde{\boldsymbol{C}}; \boldsymbol{W}) &= \mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2tr(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}) \\ &+ \mathrm{tr}(\boldsymbol{S}\boldsymbol{K}\Sigma_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}) + \mathrm{tr}(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}). \end{aligned}$$

The risk of $\tilde{\boldsymbol{C}}$ depends on $\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}$, where

$$\begin{aligned} \boldsymbol{\gamma} &= (\boldsymbol{S}\boldsymbol{K} \otimes \boldsymbol{T}'\boldsymbol{L}')vec(\boldsymbol{C}) \\ \boldsymbol{T}'\boldsymbol{L}' &= (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'. \end{aligned}$$

Note that $(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}$ is symmetric and an idempotent matrix with rank r. Thus, there exists an orthogonal matrix $\boldsymbol{\Gamma}$ such that

$$\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{I}_{rn} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}.$$

Hence,

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}] &= \mathrm{tr}[\{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\
&\quad \{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\}] \\
&= \mathrm{tr}[\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_{rn} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}] \\
&= \mathrm{tr}[\boldsymbol{A}_{11}]
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}] &= \mathrm{tr}[(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'] \\
&= \mathrm{tr}[(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'] \\
&= \mathrm{tr}[\{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\
&\quad \{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\}] \\
&= \mathrm{tr}[\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_{rn} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}] \\
&= \mathrm{tr}[\boldsymbol{A}_{11}],
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}] &= \mathrm{tr}[\boldsymbol{L}\boldsymbol{T}\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}] \\
&= \mathrm{tr}[\{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\
&\quad \{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\
&\quad \{\boldsymbol{\Gamma}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{L}(\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L})^{-1}\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-\frac{1}{2}}\boldsymbol{\Gamma}'\}] \\
&= tr[\begin{pmatrix} \boldsymbol{I}_{rn} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{rn} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}] \\
&= \mathrm{tr}[\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{rn} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}] \\
&= \mathrm{tr}[\boldsymbol{A}_{11}],
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{S}\boldsymbol{K}\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}'] &= \mathrm{tr}[\boldsymbol{K}'(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}] \\
&= \mathrm{tr}[\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{K}'(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}] \\
&= \mathrm{tr}[\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'(\boldsymbol{K}'\boldsymbol{K})^{-1}\boldsymbol{K}] \\
&= \mathrm{tr}[\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}'].
\end{aligned}$$

By some computations, we obtain the expression for the risk of $R(\tilde{\boldsymbol{C}};\boldsymbol{W})$ as follows:

$$R(\tilde{\boldsymbol{C}};\boldsymbol{W}) = \mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - \mathrm{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\mathrm{tr}(\boldsymbol{A}_{11}) + \boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}.$$

Now, similarly, we get

$$\begin{aligned}
MSE(\hat{\boldsymbol{C}}^S) &= E\{(\boldsymbol{p}_3 + \tau\boldsymbol{p}_2)(\boldsymbol{p}_3 + \tau\boldsymbol{p}_2)'\} \\
&= (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2(1-\tau)\boldsymbol{\Sigma}_{12} + (1-\tau)^2\boldsymbol{\Sigma}^* + (1-\tau)^2\boldsymbol{\gamma}\boldsymbol{\gamma}'.
\end{aligned}$$

Hence,

$$\begin{aligned} R(\hat{\boldsymbol{C}}^S;\boldsymbol{W}) &= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2(1-\tau)\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}) + \\ & (1-\tau)^2\operatorname{tr}(\boldsymbol{S}\boldsymbol{K}\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{L}\boldsymbol{T}) + (1-\tau)^2\operatorname{tr}(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}) \\ &= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2(1-\tau)\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11}) + \\ & (1-\tau)^2\operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11}) + (1-\tau)^2\operatorname{tr}(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma}). \end{aligned}$$

After some simplification, we get the expression for the risk of $R(\tilde{\boldsymbol{C}}^S;\boldsymbol{W})$. For the mean square error of the pretest estimation, we have

$$MSE(\hat{\boldsymbol{C}}^{PT}) = E\{vec(\hat{\boldsymbol{C}}^{PT} - \boldsymbol{C})[vec(\hat{\boldsymbol{C}}^{PT} - \boldsymbol{C})]'\}.$$

Using the definition of $\hat{\boldsymbol{C}}^{PT}$ and Theorem 4.2.1, we have

$$\begin{aligned} MSE(\hat{\boldsymbol{C}}^{PT}) &= E\{\boldsymbol{p}_1\boldsymbol{p}_1' - \boldsymbol{p}_1\boldsymbol{p}_2' I(D < \chi^2_{rn}(\alpha)) - \boldsymbol{p}_2\boldsymbol{p}_1' I(D < \chi^2_{rn}(\alpha)) \\ &+ \boldsymbol{p}_2\boldsymbol{p}_2' I^2(D < \chi^2_{rn}(\alpha))\}. \end{aligned}$$

Hence,

$$\begin{aligned} MSE(\hat{\boldsymbol{C}}^{PT}) &= (\boldsymbol{\Sigma}_{\epsilon\epsilon} \otimes (\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2\boldsymbol{\Sigma}^*(\boldsymbol{S}\boldsymbol{K} \otimes \boldsymbol{T}'\boldsymbol{L}')^{-1} H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) \\ &- 2\boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{S}\boldsymbol{K} \otimes \boldsymbol{T}'\boldsymbol{L}')^{-1}(H_{rn+4}(\chi^2_{rn}(\alpha);\Delta^2) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)) \\ &+ \boldsymbol{\Sigma}^* H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \boldsymbol{\gamma}\boldsymbol{\gamma}' H_{rn+4}(\chi^2_{rn}(\alpha);\Delta). \end{aligned}$$

Finally for $R(\hat{\boldsymbol{C}}^{PT};\boldsymbol{W})$ we have

$$\begin{aligned} &= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2\operatorname{tr}[\boldsymbol{W}\boldsymbol{\Sigma}^*(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1}]H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) \\ &- 2\operatorname{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1})(H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)) \\ &+ \operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*)H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \operatorname{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')H_{rn+4}(\chi^2_{rn}(\alpha);\Delta). \end{aligned}$$

We obtain more simplified expressions for $\operatorname{tr}[\boldsymbol{W}\boldsymbol{\Sigma}^*(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1}]$ and $\operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*)$,

$$\begin{aligned} \operatorname{tr}[\boldsymbol{W}\boldsymbol{\Sigma}^*(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1}] &= \operatorname{tr}[\boldsymbol{W}(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')(\boldsymbol{\Sigma}_{\epsilon\epsilon}\otimes(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) \\ &\quad (\boldsymbol{K}'\boldsymbol{S}'\otimes\boldsymbol{LT})(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1}] \\ &= \operatorname{tr}[\boldsymbol{W}(\boldsymbol{SK}\boldsymbol{\Sigma}_{\epsilon\epsilon}\otimes\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1})] \\ &= \operatorname{tr}[\boldsymbol{SK}\boldsymbol{\Sigma}_{\epsilon\epsilon}]\operatorname{tr}[\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}] \\ &= \operatorname{tr}[\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}']\operatorname{tr}[\boldsymbol{A}_{11}], \end{aligned}$$

and

$$\begin{aligned} \operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*) &= \operatorname{tr}(\boldsymbol{SK}\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}[\boldsymbol{W}\boldsymbol{T}'\boldsymbol{L}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{LT}] \\ &= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon}\boldsymbol{K}'\boldsymbol{S}')\operatorname{tr}(\boldsymbol{A}_{11}). \end{aligned}$$

By some computations after substituting the above results, we get the expression for the risk of $R(\hat{\boldsymbol{C}}^{PT};\boldsymbol{W})$.

For the mean square error of the shrinkage pretest estimation, we have

$$\begin{aligned} MSE(\hat{\boldsymbol{C}}^{SPT}) &= E\{vec(\hat{\boldsymbol{C}}^{SPT}-\boldsymbol{C})[vec(\hat{\boldsymbol{C}}^{SPT}-\boldsymbol{C})]'\}, \\ &= E\{(\boldsymbol{p}_1-(1-\tau)\boldsymbol{p}_2 I(D<d_{rn,\alpha})(\boldsymbol{p}_1-(1-\tau)\boldsymbol{p}_2 I(D<d_{rn,\alpha})'\}. \end{aligned}$$

After some manipulations, we get

$$\begin{aligned}
MSE(\hat{\boldsymbol{C}}^{SPT}) &= (\boldsymbol{\Sigma}_{\epsilon\epsilon}\otimes(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2(1-\tau)\{\boldsymbol{\Sigma}^{*}(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1}H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) \\
&+ \boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1}[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)]\} \\
&+ (1-\tau)^2[\boldsymbol{\Sigma}^{*}H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \boldsymbol{\gamma}\boldsymbol{\gamma}'H_{rn+4}(\chi^2_{rn}(\alpha);\Delta)].
\end{aligned}$$

Hence, for $R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W})$ we have

$$\begin{aligned}
&= \operatorname{tr}(\boldsymbol{\Sigma}_{\epsilon\epsilon})\operatorname{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - 2(1-\tau)\{\operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^{*}(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1})H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) \\
&+ \operatorname{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{SK}\otimes\boldsymbol{T}'\boldsymbol{L}')^{-1})[H_{rn+4}(\chi^2_{rn}(\alpha);\Delta) - H_{rn+2}(\chi^2_{rn}(\alpha);\Delta)]\} \\
&+ (1-\tau)^2[\operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^{*})H_{rn+2}(\chi^2_{rn}(\alpha);\Delta) + \operatorname{tr}(\boldsymbol{\gamma}'\boldsymbol{W}\boldsymbol{\gamma})H_{rn+4}(\chi^2_{rn}(\alpha);\Delta)].
\end{aligned}$$

By some computations we get the expression for the risk of $R(\hat{\boldsymbol{C}}^{SPT};\boldsymbol{W})$. For the proof for $R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W})$, let us consider

$$MSE(\hat{\boldsymbol{C}}^{JS}) = E\{vec(\hat{\boldsymbol{C}}^{JS}-\boldsymbol{C})[vec(\hat{\boldsymbol{C}}^{JS}-\boldsymbol{C})]'\}.$$

Using the definition of $\hat{\boldsymbol{C}}^{JS}$ and Theorem 3.2.1 part (i) and (ii), we have

$$\begin{aligned}
MSE(\hat{\boldsymbol{C}}^{JS}) &= E\{(\boldsymbol{p}_1 - c\boldsymbol{p}_2D^{-1})(\boldsymbol{p}_1 - c\boldsymbol{p}_2D^{-1})'\} \\
&= E\{\boldsymbol{p}_1\boldsymbol{p}_1{}' - c\boldsymbol{p}_1\boldsymbol{p}_2{}'D^{-1} - c\boldsymbol{p}_2\boldsymbol{p}_1{}'D^{-1} + \boldsymbol{p}_2\boldsymbol{p}_2{}'(cD^{-1})^2\}.
\end{aligned}$$

By combining Theorem 3.2.1, Lemma 3.2.1, and 3.2.1, we get the following result,

$$\begin{aligned}
MSE(\hat{\boldsymbol{C}}^{JS}) &= (\boldsymbol{\Sigma}\otimes(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}) - c\boldsymbol{\Sigma}^{*}(2E(\chi^{-2}_{rn+2}(\Delta))\boldsymbol{M} - cE(\chi^{-4}_{rn+2}(\Delta))) \\
&+ c\boldsymbol{\gamma}\boldsymbol{\gamma}'[2E(\chi^{-2}_{rn+2}(\Delta))\boldsymbol{M} - 2E(\chi^{-2}_{rn+4}(\Delta))\boldsymbol{M} + cE(\chi^{-4}_{rn+4}(\Delta))].
\end{aligned}$$

Using the definition in (3.8), we have

$$\begin{aligned} R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W}) &= \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon\epsilon}})\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\mathcal{X}\mathcal{X}}')^{-1}) - 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*\boldsymbol{M})E(\chi^{-2}_{rn+2}(\Delta)) \\ &+ c^2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*)E(\chi^{-4}_{rn+2}(\Delta)) + 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E(\chi^{-2}_{rn+2}(\Delta)) \\ &- 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E(\chi^{-2}_{rn+4}(\Delta)) + c^2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')E(\chi^{-4}_{rn+4}(\Delta)). \end{aligned}$$

By some computation we get the risk expression for $\hat{\boldsymbol{C}}^{JS}$. Finally,

$$\begin{aligned} MSE(\hat{\boldsymbol{C}}^{JS+}) &= E\{vec(\hat{\boldsymbol{C}}^{JS+} - \boldsymbol{C})[vec(\hat{\boldsymbol{C}}^{JS+} - \boldsymbol{C})]'\} \\ &= E\{[vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C}) - vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})(1 - cD^{-1})I(D < c)] \\ &\quad [vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C}) - vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})(1 - cD^{-1})I(D < c)]'\} \\ &= E\{[vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C})][vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C})]'\} \\ &- 2E\{[vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})(1 - cD^{-1})I(D < c)][vec(\hat{\boldsymbol{C}}^{JS} - \boldsymbol{C})]'\} \\ &+ E\{[vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})(1 - cD^{-1})I(D < c)] \\ &\quad [vec(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})(1 - cD^{-1})I(D < c)]'\}. \end{aligned}$$

Using Theorem 3.2.1, Lemma 3.2.1, and 3.2.2, we have

$$\begin{aligned}
MSE(\hat{\boldsymbol{C}}^{JS+}) &= MSE(\hat{\boldsymbol{C}}^{JS}) - 2E\{\boldsymbol{p}_2\boldsymbol{p}_1{}'(1-cD^{-1})I(D<c)\} \\
&+ 2E\{\boldsymbol{p}_2\boldsymbol{p}_2{}'(1-cD^{-1})I(D<c)\} - E\{\boldsymbol{p}_2\boldsymbol{p}_2{}'(1-cD^{-1})^2 I(D<c)\} \\
&= MSE(\hat{\boldsymbol{C}}^{JS}) - 2(\boldsymbol{K}'\boldsymbol{S}'\otimes\boldsymbol{LT})^{-1}E[(1-c\chi^{-2}_{rn+2}(\Delta))I(\chi^2_{rn+2}(\Delta)<c)] \\
&- 2\boldsymbol{\gamma}\boldsymbol{\gamma}'(\boldsymbol{K}'\boldsymbol{S}'\otimes\boldsymbol{LT})^{-1}E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta)<c)] \\
&+ 2\boldsymbol{\Sigma}^* E[(1-c\chi^{-2}_{rn+2}(\Delta))I(\chi^2_{rn+2}(\Delta)<c)] \\
&+ 2\boldsymbol{\gamma}\boldsymbol{\gamma}' E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta)<c)] \\
&- \boldsymbol{\Sigma}^* E[(1-c\chi^{-2}_{rn+2}(\Delta))^2 I(\chi^2_{rn+2}(\Delta)<c)] \\
&- \boldsymbol{\gamma}\boldsymbol{\gamma}' E[(1-c\chi^{-2}_{rn+4}(\Delta))^2 I(\chi^2_{rn+4}(\Delta)<c)].
\end{aligned}$$

Using the definition in (3.8) we have

$$\begin{aligned}
R(\hat{\boldsymbol{C}}^{JS+};\boldsymbol{W}) &= R(\hat{\boldsymbol{C}}^{JS};\boldsymbol{W}) - 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*\boldsymbol{M})E[(1-c\chi^{-2}_{rn+2}(\Delta))I(\chi^2_{rn+2}(\Delta)<c)] \\
&- 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{M})E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta)<c)] \\
&+ 2tr(\boldsymbol{W}\boldsymbol{\Sigma}^*)E[(1-c\chi^{-2}_{rn+2}(\Delta))I(\chi^2_{rn+2}(\Delta)<c)] \\
&+ 2tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')E[(1-c\chi^{-2}_{rn+4}(\Delta))I(\chi^2_{rn+4}(\Delta)<c)] \\
&- tr(\boldsymbol{W}\boldsymbol{\Sigma}^*)E[(1-c\chi^{-2}_{rn+2}(\Delta))^2 I(\chi^2_{rn+2}(\Delta)<c)] \\
&- tr(\boldsymbol{W}\boldsymbol{\gamma}\boldsymbol{\gamma}')E[(1-c\chi^{-2}_{rn+4}(\Delta))^2 I(\chi^2_{rn+4}(\Delta)<c)].
\end{aligned}$$

Using some computation we get the risk expression for $\hat{\boldsymbol{C}}^{JS+}$.

Table 3.5: Biomedical data on urine samples of patients from a study by Smith et al. (1962).

y_1	y_2	y_3	y_4	y_5	x_1	x_2	x_3
17.6	1.50	1.50	1.88	7.5	0.98	2.05	2.4
13.4	1.65	1.32	2.24	7.1	0.98	1.60	3.2
20.3	0.90	0.89	1.28	2.3	1.15	4.80	1.7
22.3	1.75	1.50	2.24	4.0	1.28	2.30	3.0
18.5	1.20	1.03	1.84	2.0	1.28	2.15	2.7
12.1	1.90	1.87	2.40	16.8	1.22	2.15	2.5
10.1	2.30	2.08	2.68	0.9	1.22	1.90	2.8
14.7	2.35	2.55	3.00	2.0	1.30	1.75	2.4
14.7	2.50	2.38	2.84	3.8	1.30	1.55	2.7
18.1	1.50	1.20	2.60	14.5	1.35	2.20	3.1
16.9	1.40	1.15	1.72	8.0	1.35	3.05	3.2
23.7	1.65	1.58	1.60	4.9	1.35	2.75	2.0
18.0	1.60	1.68	2.00	3.6	1.35	2.10	2.3
14.8	2.45	2.15	3.12	12.0	1.40	1.70	3.1
15.6	1.65	1.42	2.56	5.2	1.40	2.35	2.8
16.2	1.65	1.62	2.04	10.2	1.41	1.85	2.1
17.5	1.05	1.56	1.48	9.6	1.41	2.65	1.5
14.1	2.70	2.77	2.56	6.9	1.62	3.05	2.6
22.5	0.85	1.65	1.20	3.5	1.62	4.30	1.6
17.0	0.70	0.97	1.24	1.9	2.05	3.50	1.8
12.5	0.80	0.80	0.64	0.7	2.05	4.75	1.0
21.5	1.80	1.77	2.60	8.3	2.30	1.95	3.3
13.0	2.20	1.85	3.84	13.0	2.30	1.60	3.5
13.0	3.55	3.18	3.48	18.3	2.15	2.40	3.3
12.0	3.65	2.40	3.00	14.5	2.15	2.70	3.4
22.8	0.55	1.00	1.14	3.3	2.30	4.75	1.6
18.4	1.05	1.17	1.36	4.9	2.30	4.90	2.8
8.7	4.25	3.62	3.84	19.5	2.62	1.15	2.5
9.4	3.85	3.36	5.12	1.3	2.62	0.97	2.8
15.0	2.45	2.38	2.40	20.0	2.55	3.25	2.7
12.9	1.70	1.74	2.48	1.0	2.55	3.10	2.3
12.1	1.80	2.00	2.24	5.0	2.70	2.45	2.5
11.5	2.25	2.25	3.12	5.1	2.70	2.20	3.4

Chapter 4

Estimation Strategies for a Parameter Matrix in a Multivariate Reduced Rank Regression Model

4.1 Introduction

We consider the multivariate multiple regression models (MMRMs) that were presented in Chapter 3, when the number of parameters in the regression matrix is large, which happens in financial and economic analysis. Thus, in many practical problems, there is a need to reduce the number of parameters in model (3.2). We study this problem through the possibility that the rank of the regression coefficient matrix $\boldsymbol{C}$ is deficient. Therefore, there may be a number of linear constraints on the set of regression coefficients in the model. The resulting model is called a multivariate reduced rank regression model (Izenman 1975, 2008). Anderson (1951) was the first to consider in detail the RRR problem when X_i is fixed.

Statistical problems concerning reduced rank regression models and their properties have been studied in the statistical literature by Anderson (1984), Robinson (1973), Robinson (1974), Tso (1981), Davies and Tso (1982), Zhou (1994), Geweke (1996), Reinsel and Velu (1998), Heinen and Rengifo (2007), Vounou et al. (2010), and Yee and Hastie (2003), also, from a Bayesian point of view, Schmidli (1995), Geweke (1996), and others. Most applications of RRR have been directed toward problems in time series (time domain and frequency domain) and econometrics: see Velu and Reinsel (1987), Reinsel (1983), Johansen (1988, 1991) among others.

A physical interpretation of the reduced rank model was offered by Brillinger (1969). Suppose we wish to send a message based upon the q components of a vector X that represents information which is to be used to send a message Y having m components ($m \leq q$), but such a message can only be transmitted through r channels ($r \leq m$). Thus, first we would need to encode X into a r vector $\boldsymbol{\zeta} = \boldsymbol{B}X$, where $\boldsymbol{B}$ is a $(r \times q)$ matrix. After receiving the coded message, we would need to decode it using an $(m \times r)$ matrix $\boldsymbol{A}$ to form the m vector $\boldsymbol{A\zeta}$, which we hope to be as close as possible to the desired $\boldsymbol{\mathcal{Y}}$.

When $\boldsymbol{C}$ has reduced rank r, then there exist two nonunique full rank matrices, an $(m \times r)$ matrix $\boldsymbol{A}$ and a $(r \times q)$ matrix $\boldsymbol{B}$, such that $\boldsymbol{C} = \boldsymbol{AB}$. The nonuniqueness happens because we can always find a nonsingular $(r \times r)$ matrix $\boldsymbol{\mathcal{R}}$ such that

$$\boldsymbol{C} = (\boldsymbol{A}\boldsymbol{\mathcal{R}})(\boldsymbol{\mathcal{R}}^{-1}\boldsymbol{B}) = \boldsymbol{\mathcal{J}}\boldsymbol{\mathcal{T}},$$

which gives a different decomposition of $\boldsymbol{C}$.

Hence, we restate the model in (3.2) as RRR model

$$Y_i = \boldsymbol{AB}X_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, n, \tag{4.1}$$

where vectors $\boldsymbol{\epsilon}_i$ are assumed to be independent with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\epsilon\epsilon}$, and where $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ is a positive definite matrix.

In this chapter, we are interested in how to place linear constraints in a RRR model on $\boldsymbol{C}$ when $\boldsymbol{\mathcal{X}}$ is considered to be random. Therefore, we describe the RRR scenario in which $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are jointly distributed.

As we will see shortly, reduced rank regression estimation will be obtained as a certain reduced rank approximation of the full rank least squares estimate of the regression coefficient matrix. Therefore, we need an essential matrix result which represents how to approximate a full rank matrix by a matrix of lower rank from Eckart and Young (1936).

Theorem 4.1.1. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both $(J \times K)$ matrices, and we plan on using $\boldsymbol{B}$ with reduced rank $r(\boldsymbol{B}) = b$ to approximate $\boldsymbol{A}$ with full rank $r(\boldsymbol{A}) = min(J, K)$, then we have

$$\lambda_j((\boldsymbol{A} - \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})') \geq \lambda_{j+b}(\boldsymbol{AA}')$$

with equality if

$$\boldsymbol{B} = \sum_{i=1}^{b} \lambda_i^{1/2} \boldsymbol{u}_i \boldsymbol{v}_i',$$

where $\lambda_i = \lambda_i(\boldsymbol{AA}')$, $\boldsymbol{u}_i = \boldsymbol{v}_i(\boldsymbol{AA}')$, and $\boldsymbol{v}_i = \boldsymbol{v}_i(\boldsymbol{A}'\boldsymbol{A})$. Because the above choice of $\boldsymbol{B}$ provides a simultaneous minimization for all eigenvalues λ_j, it follows that the minimum is achieved for different functions of those eigenvalues, say, the trace or the determinant of $(\boldsymbol{A} - \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})'$.

Estimation of $\boldsymbol{A}$ and $\boldsymbol{B}$ in (4.1) is based on the following Theorem (Brillinger, 1981, Section 10.2), which essentially uses Theorem 4.1.1.

Theorem 4.1.2. Suppose the $(m+q)$ dimensional random vector $(Y', X')'$ has mean vector zero and covariance matrix $\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy} = Cov(Y, X)$, and $\boldsymbol{\Sigma}_{xx} = Cov(X)$ nonsingular. Then for any positive definite matrix $\boldsymbol{\Gamma}$, an $(m \times r)$ matrix $\boldsymbol{A}$ and $(r \times q)$ matrix $\boldsymbol{B}$, for $r \leq min(m, q)$, which minimize

$$tr\{E[\boldsymbol{\Gamma}^{1/2}(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{\mathcal{X}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{\mathcal{X}})'\boldsymbol{\Gamma}^{1/2}]\}$$

are given by

$$\boldsymbol{A}^{(r)} = \boldsymbol{\Gamma}^{-1/2}[V_1, \ldots, V_r] = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{V}, \qquad \boldsymbol{B}^{(r)} = \boldsymbol{V}'\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1},$$

where $\boldsymbol{V} = [V_1, \ldots, V_r]$ and V_j is the (normalized) eigenvector that corresponds to the j^{th} largest eigenvalue λ_j^2 of the matrix $\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Gamma}^{1/2}$, $(j = 1, \ldots, r)$.

We also observed earlier that the decomposition $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ is not unique, hence to determine $\boldsymbol{A}$ and $\boldsymbol{B}$ uniquely, we must impose some normalization conditions. The eigenvectors V_j in Theorem 4.1.2 are normalized to satisfy $V_j'V_j = 1$, and this is equivalent to normalization for $\boldsymbol{A}$ and $\boldsymbol{B}$ as follows:

$$\boldsymbol{B}\boldsymbol{\Sigma}_{xx}\boldsymbol{B}' = \Lambda^2, \qquad \boldsymbol{A}'\boldsymbol{\Gamma}\boldsymbol{A} = \boldsymbol{I}_r,$$

where $\boldsymbol{\Lambda}^2 = diag(\lambda_1^2, \ldots, \lambda_r^2)$ and $\mathbf{I}_r$ is an $r \times r$ identity matrix. Thus, the number of independent regression parameters in the reduced rank model (4.1) is $(m + q - r)$ compared to (mq) parameters in the full rank model. The elements of the reduced rank approximation of the matrix $\boldsymbol{C}$ are given as

$$\boldsymbol{C}^{(r)} = \boldsymbol{A}^{(r)}\boldsymbol{B}^{(r)} = \boldsymbol{\Gamma}^{-1/2}\left(\sum_{j=1}^{r} V_j V_j'\right)\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} = \boldsymbol{\mathcal{T}}_{\Gamma}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}$$

where $\boldsymbol{\mathcal{T}}_{\Gamma}$ is an idempotent matrix for any $\boldsymbol{\Gamma}$, but it does not need to be symmetric. Observe that, $\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}$ is the usual full rank (population) regression coefficient matrix. When $r = m$, $\sum_{j=1}^{r} V_j V_j' = \boldsymbol{I}_m$ and, therefore, $\boldsymbol{C}^{(r)}$ reduces to a full rank coefficient matrix. Robinson (1974) has shown that the solution for $\boldsymbol{A}$ and $\boldsymbol{B}$ in Theorem 4.1.2 are as follows when sample quantities are substituted, namely

$$\hat{\boldsymbol{A}}^{(r)} = \boldsymbol{\Gamma}^{-1/2}[\hat{V}_1, \ldots, \hat{V}_r] = \boldsymbol{\Gamma}^{-1/2}\hat{\boldsymbol{V}}, \qquad \hat{\boldsymbol{B}}^{(r)} = [\hat{V}_1, \ldots, \hat{V}_r]'\boldsymbol{\Gamma}^{1/2}\hat{\boldsymbol{\Sigma}}_{yx}\hat{\boldsymbol{\Sigma}}_{xx}^{-1},$$

$\hat{V}_j$ is the eigenvector that corresponds to the j^{th} largest eigenvalues $\hat{\lambda}_j^2$ of $\boldsymbol{\Gamma}^{1/2}\hat{\boldsymbol{\Sigma}}_{yx}\hat{\boldsymbol{\Sigma}}_{xx}^{-1}$ $\hat{\boldsymbol{\Sigma}}_{yx}\boldsymbol{\Gamma}^{1/2}$, with the choice $\boldsymbol{\Gamma} = \hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}^{-1}$, where $\hat{\boldsymbol{\Sigma}}_{yx} = (1/n)\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{X}}'$, $\hat{\boldsymbol{\Sigma}}_{xx} = (1/n)\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}'$, and $\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon} = (1/n)(\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}' - \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{X}}'(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}')^{-1}\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{Y}}') \equiv (1/n)(\boldsymbol{\mathcal{Y}} - \hat{\boldsymbol{C}}\boldsymbol{\mathcal{X}})(\boldsymbol{\mathcal{Y}} - \hat{\boldsymbol{C}}\boldsymbol{\mathcal{X}})'$.

The small sample distribution of the reduced rank estimators is somewhat difficult to work with. Therefore, we focus on the large sample behavior of the estimators. The asymptotic results follow from Robinson (1973).

The main results on asymptotic distribution of $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ when $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}_{\epsilon\epsilon}^{-1}$ is assumed to be known is contained in the next Theorem.

Theorem 4.1.3. For the model (4.1), let $(vec(\boldsymbol{A}), vec(\boldsymbol{B})) \in \Theta$, which is a compact set defined by the normalization condition in Theorem 4.1.2. Then, with $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{B}}$ and $\boldsymbol{A}$, $\boldsymbol{B}$ given by

$$\hat{\boldsymbol{A}} = \boldsymbol{\Gamma}^{-1/2}[\hat{V}_1, \ldots, \hat{V}_r] = \boldsymbol{\Gamma}^{-1/2}\hat{\boldsymbol{V}}, \qquad \hat{\boldsymbol{B}} = \hat{\boldsymbol{V}}'\boldsymbol{\Gamma}^{1/2}\hat{\boldsymbol{\Sigma}}_{yx}\hat{\boldsymbol{\Sigma}}_{xx}^{-1},$$

$$\boldsymbol{A} = \boldsymbol{\Gamma}^{-1/2}[V_1, \ldots, V_r] = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{V}, \qquad \boldsymbol{B} = \boldsymbol{V}'\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1},$$

respectively, $vec(\hat{\boldsymbol{A}})$ and $vec(\hat{\boldsymbol{B}})$ converge in probability to $vec(\boldsymbol{A})$ and $vec(\boldsymbol{B})$, respectively, as $n \longrightarrow \infty$.

In this chapter, we assume $\boldsymbol{A}$ to be known and based on results by Robinson (1974, Theorem 1 and Section 2.3). Then $n^{1/2}vec(\hat{\boldsymbol{B}} - \boldsymbol{B})$ converges in distribution to $N(\mathbf{0}, (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1}))$.

The main objective of this chapter is to consider the estimation problem of the parameter matrix $\boldsymbol{B}$ under a very general set of linear constraints,

$$H_0 : \boldsymbol{\mathcal{F}}\boldsymbol{B}\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}}, \tag{4.2}$$

where $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{G}}$ are known full rank matrices of appropriate dimensions $r_1 \times r$ and $q \times m$, respectively. Let $\tilde{\boldsymbol{B}}$ be a restricted estimate of $\boldsymbol{B}$ when it is suspected but not known that $\boldsymbol{B}$ may be restricted to the subspace defined by $\boldsymbol{\mathcal{F}}\boldsymbol{B}\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}}$. Alternatively, we can write the null hypothesis in (4.2) by using the left Kronecker product as

$$H_0 : (\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{G}}')vec(\boldsymbol{B}) = vec(\boldsymbol{\mathcal{D}}). \tag{4.3}$$

We wish to find an $(r \times q)$ matrix $\boldsymbol{B}$ to minimize the following expression

$$\tilde{\boldsymbol{B}} = arg \min_{\boldsymbol{\mathcal{F}}\mathrm{B}\boldsymbol{\mathcal{G}}=\boldsymbol{\mathcal{D}}} tr\{(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{\mathcal{X}})\boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{\mathcal{X}})'\}$$

where $\boldsymbol{\Gamma}$ is a positive definite symmetric matrix. The candidate sub-model estimator is

$$\tilde{\boldsymbol{B}} = \hat{\boldsymbol{B}} - \boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}},$$

where $\boldsymbol{\mathcal{N}} = (\boldsymbol{A}'\Gamma\boldsymbol{A})^{-1}\boldsymbol{\mathcal{F}}'(\boldsymbol{\mathcal{F}}(\boldsymbol{A}'\Gamma\boldsymbol{A})^{-1}\boldsymbol{\mathcal{F}}')^{-1}$ and $\boldsymbol{\mathcal{Q}} = (\boldsymbol{\mathcal{G}}'\Sigma_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\Sigma_{xx}^{-1}$. Clearly, $\tilde{\boldsymbol{B}}$ will be a unbiased estimator under the candidate subspace in (4.2). On the other hand, it can be easily verified that $\tilde{\boldsymbol{B}}$ is an inconsistent estimator because of the bias inherited by the submodel. It is obviously important to be able to find a test statistic for testing $H_0 : \boldsymbol{\mathcal{F}}\boldsymbol{B}\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}}$ which can be used for further research. Based on the asymptotic results given in Robinson (1974), we define that the test statistic follows from the likelihood ratio method of test construction [see Anderson (1951)], as we now indicate. For model (4.1) under the asymptotically normality assumption on the $\boldsymbol{\epsilon}_i$, the likelihood ratio test statistic for testing $H_0 : \boldsymbol{\mathcal{F}}\boldsymbol{B}\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}}$ is $\nu = U^{n/2}$, where $U = |\boldsymbol{\mathcal{S}}|/|\boldsymbol{\mathcal{S}}_1|$, $\boldsymbol{\mathcal{S}} = (\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{X}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{X}})'$, and

$$
\begin{aligned}
\boldsymbol{\mathcal{S}}_1 &= (\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\tilde{\boldsymbol{B}}\boldsymbol{\mathcal{X}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\tilde{\boldsymbol{B}}\boldsymbol{\mathcal{X}})' \\
&= (\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{X}} + \boldsymbol{A}\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{X}}) \\
&\quad (\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{X}} + \boldsymbol{A}\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{X}})' \\
&= (\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\hat{\boldsymbol{B}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{A}\hat{\boldsymbol{B}})' \\
&+ \boldsymbol{A}\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}}\hat{\Sigma}_{xx}\boldsymbol{\mathcal{Q}}'(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})'\boldsymbol{\mathcal{N}}'\boldsymbol{A}' \\
&= \boldsymbol{\mathcal{S}} + \boldsymbol{A}\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})(\boldsymbol{\mathcal{G}}'\hat{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xx} \\
&\times \hat{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\hat{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})'\boldsymbol{\mathcal{N}}'\boldsymbol{A}' \\
&= \boldsymbol{\mathcal{S}} + \boldsymbol{A}\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})(\boldsymbol{\mathcal{G}}'\hat{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})'\boldsymbol{\mathcal{N}}'\boldsymbol{A}' \\
&= \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{H}}.
\end{aligned}
$$

It has been shown in Anderson (1984) that the test statistic $L = -[n - q + \frac{q-m-1}{2}]log(U)$ under H_0 follows asymptotically $\chi^2_{r_1 m}$.

As mentioned before, when the linear constraint as a prior information is rather suspicious, it may be reasonable to construct pretest estimators. Therefore, we use

$$\hat{\boldsymbol{B}}^{PT} = \hat{\boldsymbol{B}} - (\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})I(L < l_{r_1 m,\alpha}),$$

where the $l_{r_1 m,\alpha}$ is the upper α-level critical value of the χ^2 distribution with $r_1 m$ degrees of freedom, and $I(A)$ is an indicator function of a set A. Also, the shrinkage and positive shrinkage estimators are defined as

$$\hat{\boldsymbol{B}}^{JS} = \tilde{\boldsymbol{B}} + \{1 - sL^{-1}\}(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}}), \quad r_1 m > 2$$

$$\hat{\boldsymbol{B}}^{JS+} = \tilde{\boldsymbol{B}} + \{1 - sL^{-1}\}^{+}(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}}), \quad r_1 m > 2,$$

where the optimal value of s is $s_{opt} = r_1 m - 2$ and is chosen in an interval in such a way that $\hat{\boldsymbol{B}}^{JS}$ dominates $\hat{\boldsymbol{B}}$. s is allowed to vary over $[0, 2(r_1 m - 2)), r_1 m > 2$, often set to $s = r_1 m - 2$; thus, we assume that $r_1 m \geq 3$.

The remainder of this chapter is organized as follows. In Section 2, we showcase some important results which will be needed in deriving the expressions for the suggested estimators. In Section 3, we present the expressions for asymptotic bias and risk with their analysis. A data example and simulation study are presented in Section 4. Conclusions are offered in Section 5. Finally, the proofs of the main results are given in Section 6.

4.2 Main Results

In an effort to establish some important properties of the estimators, we consider the asymptotic distribution of $n^{1/2} vec(\hat{\boldsymbol{B}} - \boldsymbol{B}) \sim N(\boldsymbol{0}, (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1}))$. To obtain a meaningful asymptotic distribution for this, we consider the class of local alternatives K_n defined by

$$K_n : \boldsymbol{\mathcal{F}}\boldsymbol{B}\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{D}} + n^{-1/2}\boldsymbol{\Upsilon},$$

where $\boldsymbol{\Upsilon}$ is a non zero matrix. Let $\boldsymbol{q}_1 = n^{1/2}vec(\hat{\boldsymbol{B}} - \boldsymbol{B})$, $\boldsymbol{q}_2 = n^{1/2}vec(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})$, and $\boldsymbol{q}_3 = n^{1/2}vec(\tilde{\boldsymbol{B}} - \boldsymbol{B})$. Then we have the following asymptotic distributional results:

Theorem 4.2.1.

(i) $\boldsymbol{q}_1 \sim N(\mathbf{0}, (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1}))$

(ii) $\boldsymbol{q}_2 \sim N(\boldsymbol{\delta}, \boldsymbol{\Sigma}^*)$

(iii) $\boldsymbol{q}_3 \sim N(-\boldsymbol{\delta}, \boldsymbol{\Omega}^*)$

(iv) $\begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1}) & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}^* \end{pmatrix} \right\}$

(v) $\begin{pmatrix} \boldsymbol{q}_2 \\ \boldsymbol{q}_3 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{\delta} \\ -\boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}^* & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}^* \end{pmatrix} \right\}$

where $\boldsymbol{\delta} = (\boldsymbol{\mathcal{N}} \otimes \boldsymbol{\mathcal{Q}}')vec(\boldsymbol{\Upsilon})$,

$\boldsymbol{\Sigma}^* = (\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})$,

$\boldsymbol{\Omega}^* = (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1}) - \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21} + \boldsymbol{\Sigma}^*$,

$\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})$,

$\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{21}' = \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}^*$.

Proof: See Appendix, Section 4.6.1.

4.3 Asymptotic Bias and Risk Analysis

In this section, we obtain expressions for the asymptotic distributional bias (ADB) and the risks (ADR) of the proposed estimators. Also, we compare the performance of the suggested estimators in terms of asymptotic risk. First we present the expression for the asymptotic distribution bias (ADB) of the proposed estimators. The ADB of an estimator $\boldsymbol{B}^*$ is defined as

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

Theorem 4.3.1. Under $\{K_n\}$ the asymptotic distribution biases (ADB) of the proposed estimators are, respectively,

(i) $ADB(\tilde{\boldsymbol{B}}) = -\boldsymbol{\delta}$

(ii) $ADB(\hat{\boldsymbol{B}}^{PT}) = -\boldsymbol{\delta} H_{r_1m+2}(\chi^2_{r_1m}(\alpha), \Delta)$

(iii) $ADB(\hat{\boldsymbol{B}}^{JS}) = -s\boldsymbol{\delta} E[\chi^{-2}_{r_1m+2}(\Delta)]$

(iv) $ADB(\hat{\boldsymbol{B}}^{JS^+}) = ADB(\hat{\boldsymbol{B}}^{S}) - \boldsymbol{\delta} E\{[1 - s\chi^{-2}_{r_1m+2}(\Delta)]I(\chi^2_{r_1m+2}(\Delta) < s)\}.$

The notation $H_\nu(x, \Delta)$ is the distribution function of non-central chi-square distribution with ν degrees of freedom and non-centrality parameter $\Delta = \boldsymbol{\delta}'(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}^{-1}_{xx})^{-1}\boldsymbol{\delta}$.

Proof: See Appendix, Section 4.6.2.

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

$\boldsymbol{B}$ by

$$AQDB(\boldsymbol{B}^*) = [ADB(\hat{\boldsymbol{B}}^*)]'(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})[ADB(\hat{\boldsymbol{B}}^*)].$$

Based on the Theorem 4.3.1, we can easily obtain the AQDB of the estimators.

$$\begin{aligned}
AQDB(\tilde{\boldsymbol{B}}) &= \Delta \\
AQDB(\hat{\boldsymbol{B}}^{PT}) &= \Delta\{H_{r_1m+2}(\chi^2_{r_1m}(\alpha), \Delta)\}^2 \\
AQDB(\hat{\boldsymbol{B}}^{JS}) &= s^2\Delta\{E(\chi^{-2}_{r_1m+2}(\Delta))\}^2 \\
AQDB(\hat{\boldsymbol{B}}^{JS+}) &= \Delta\{sE(\chi^{-2}_{r_1m+2}(\Delta)) - E\{(s\chi^{-2}_{r_1m+2}(\Delta) - 1)I(\chi^2_{r_1m+2}(\Delta) < s)\}^2.
\end{aligned}$$

Clearly, the asymptotic bias of $\tilde{\boldsymbol{B}}$ is unbounded, and the bias of $\hat{\boldsymbol{B}}^{PT}$ depends on the size of α and Δ. The asymptotic bias of $\hat{\boldsymbol{B}}^{JS}$ and $\hat{\boldsymbol{B}}^{JS^+}$ depends on Δ alone. Thus, we can establish the following two results:

$$AQDB(\hat{\boldsymbol{B}}^{JS^+}) \le AQDB(\hat{\boldsymbol{B}}^{JS}) \le AQDB(\tilde{\boldsymbol{B}}),$$

$$0 = AQDB(\hat{\boldsymbol{B}}) \le AQDB(\hat{\boldsymbol{B}}^{PT}) \le AQDB(\tilde{\boldsymbol{B}}).$$

4.3.1 Relative Performance of the Estimators

Now, we present some useful results in the following Theorem, which will be used in deriving the risk expressions for the estimators.

Lemma 4.3.1. Under the assumed conditions for model (4.1),

(i) $E(\boldsymbol{q}_1|\boldsymbol{q}_2) = (\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')^{-1}(\boldsymbol{q}_2 - \boldsymbol{\delta})$

(ii) $E(\boldsymbol{q}_2\boldsymbol{q}_1{}'I(L < \chi^2_{r_1m}(\alpha))) = \boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) + \boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}[H_{r_1m+4}(\chi^2_{r_1m}(\alpha)) - H_{r_1m+2}(\chi^2_{r_1m}(\alpha))]$, where $\boldsymbol{\delta} = E(\boldsymbol{q}_2)$

(iii) $E(\boldsymbol{q}_2\boldsymbol{q}_1{}'L^{-1}) = (\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{N}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}^{-1}_{xx})E(\chi^{-2}_{r_1m+2}(\Delta)) + \boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}$ $[E(\chi^{-2}_{r_1m+4}(\Delta)) - E(\chi^{-2}_{r_1m+2}(\Delta))]$.

Proof: See Appendix, Section 4.6.3.

Theorem 4.3.2. Under $\{K_n\}$, the asymptotic mean square errors (AMSE) of the listed estimators are as follows:

$$\begin{aligned}
AMSE(\tilde{\boldsymbol{B}}) &= \boldsymbol{\Omega}^* + \boldsymbol{\delta}\boldsymbol{\delta}' \\
AMSE(\hat{\boldsymbol{B}}^{PT}) &= (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}^{-1}_{xx}) - 2\boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) \\
&- 2\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}[H_{r_1m+4}(\chi^2_{r_1m}(\alpha)) - H_{r_1m+2}(\chi^2_{r_1m}(\alpha))] \\
&+ \boldsymbol{\delta}\boldsymbol{\delta}'H_{r_1m+4}(\chi^2_{r_1m}(\alpha)) + \boldsymbol{\Sigma}^*H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) \\
AMSE(\hat{\boldsymbol{B}}^{JS}) &= (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}^{-1}_{xx}) - s\boldsymbol{\Sigma}^*[2E(\chi^{-2}_{r_1m+2}(\Delta))(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1} \\
&- sE(\chi^{-4}_{r_1m+2}(\Delta))] + s\boldsymbol{\delta}\boldsymbol{\delta}'\{2[E(\chi^{-2}_{r_1m+2}(\Delta)) - E(\chi^{-2}_{r_1m+4}(\Delta))] \\
&\times (\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1} + sE(\chi^{-4}_{r_1m+4}(\Delta))\} \\
AMSE(\hat{\boldsymbol{B}}^{JS+}) &= AMSE(\hat{\boldsymbol{B}}^{JS}) - 2\boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1} \\
&\times E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta) < s)] - 2\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1} \\
&\times E[(1 - s\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta) < s)] \\
&+ 2\boldsymbol{\Sigma}^*E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&+ 2\boldsymbol{\delta}\boldsymbol{\delta}'E[(1 - s\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta) < s)] \\
&- \boldsymbol{\Sigma}^*E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))^2I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&- \boldsymbol{\delta}\boldsymbol{\delta}'E[(1 - s\chi^{-2}_{r_1m+4}(\Delta))^2I(\chi^2_{r_1m+4}(\Delta) < s)].
\end{aligned}$$

Proof: See Appendix, Section 4.6.4.

Theorem 4.3.3. The asymptotic quadratic risks of estimators are

$$\begin{aligned}
ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) &= r\text{tr}(\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-1}) \\
ADR(\tilde{\boldsymbol{B}};\boldsymbol{W}) &= ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) - 2\text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}) + \text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}') \\
&+ \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta} \\
ADR(\hat{\boldsymbol{B}}^{PT};\boldsymbol{W}) &= ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) - 2\text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) \\
&- 2\text{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})[H_{r_1m+4}(\chi^2_{r_1m}(\alpha)) - H_{r_1m+2}(\chi^2_{r_1m}(\alpha))] \\
&+ \text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) \\
&+ \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}H_{r_1m+4}(\chi^2_{r_1m}(\alpha)) \\
ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W}) &= ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) - 2s\text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})E(\chi^{-2}_{r_1m+2}(\Delta)) + s^2\text{tr}(\boldsymbol{\mathcal{D}}_{11}) \\
&\quad \text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')E(\chi^{-4}_{r_1m+2}(\Delta)) - 2s\text{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})[E(\chi^{-2}_{r_1m+4}(\Delta)) \\
&- E(\chi^{-2}_{r_1m+2}(\Delta))] + s^2\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}E(\chi^{-4}_{r_1m+4}(\Delta)) \\
ADR(\hat{\boldsymbol{B}}^{JS+};\boldsymbol{W}) &= ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W}) \\
&- 2\text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&- 2\text{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})E[(1 - s\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta) < s)] \\
&+ \text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')E[(1 - s^2\chi^{-4}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&+ \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}E[(1 - s^2\chi^{-4}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta) < s)],
\end{aligned}$$

where

$$\boldsymbol{\mathcal{U}} = (\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1},$$

$$\boldsymbol{\mathcal{D}} = \boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}',$$

$$\text{tr}(\boldsymbol{W}\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}'\Sigma_{xx}^{-1}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}) = \text{tr}(\boldsymbol{\mathcal{D}}_{11}),$$

$$\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}) = \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11}),$$

$$\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*\boldsymbol{\mathcal{U}}) = \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}),$$

$$\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*) = \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')$$

.

Proof: See Appendix, Section 4.6.5.

4.3.2 Comparison of $\hat{\boldsymbol{B}}^{PT}$ and $\hat{\boldsymbol{B}}$

Consider the risk-difference:

$$\begin{aligned} ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) &- ADR(\hat{\boldsymbol{B}}^{PT};\boldsymbol{W}) = 2\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta) \\ &+ 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})[H_{r_1m+4}(\chi^2_{r_1m}(\alpha),\Delta) - H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta)] \\ &- \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}H_{r_1m+4}(\chi^2_{r_1m}(\alpha),\Delta) \\ &- tr(\boldsymbol{\mathcal{D}}_{11})tr(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta). \end{aligned}$$

The right hand side is nonnegative whenever

$$\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta} \leq \frac{-\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta)}{H_{r_1m+4}(\chi^2_{r_1m}(\alpha),\Delta)}.$$

In this range, $\hat{\boldsymbol{B}}^{PT}$ performs better than $\hat{\boldsymbol{B}}$ for all Δ.

4.3.3 Comparison of $\hat{\boldsymbol{B}}^{PT}$ and $\tilde{\boldsymbol{B}}$

For comparing the risk of $\hat{\boldsymbol{B}}^{PT}$ and $\tilde{\boldsymbol{B}}$, we consider the differences between them:

$$\begin{aligned} ADR(\tilde{\boldsymbol{B}};\boldsymbol{W}) &- ADR(\hat{\boldsymbol{B}}^{PT};\boldsymbol{W}) = (H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta)-1)[2\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}) \\ &- \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')] + \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}(1-H_{r_1m+4}(\chi^2_{r_1m}(\alpha),\Delta)) \\ &+ 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})[H_{r_1m+4}(\chi^2_{r_1m}(\alpha),\Delta) - H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta)]. \end{aligned}$$

Thus, the risk of $\hat{\boldsymbol{B}}^{PT}$ is smaller than $\tilde{\boldsymbol{B}}$ whenever

$$\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta} \geq \frac{(H_{r_1m+2}(\chi^2_{r_1m}(\alpha),\Delta)-1)2\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})[\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}) - \mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')]}{(H_{r_1m+4}(\chi^2_{r_1m}(\alpha),\Delta)-1)}.$$

Note that both $\hat{\boldsymbol{B}}^{PT}$ and $\tilde{\boldsymbol{B}}$ are superior to $\hat{\boldsymbol{B}}$ under H_0. Therefore, under H_0 the risk of the three estimators may be ordered as

$$ADR(\tilde{\boldsymbol{B}};\boldsymbol{W}) \leq ADR(\hat{\boldsymbol{B}}^{PT};\boldsymbol{W}) \leq ADR(\hat{\boldsymbol{B}};\boldsymbol{W}).$$

4.3.4 Comparison of $\hat{\boldsymbol{B}}^{JS}$ and $\hat{\boldsymbol{B}}$

The risk-differences are given by

$$\begin{aligned} ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W}) &- ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) = -2c\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})E(\chi^{-2}_{r_1m+2}(\Delta)) + c^2\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11}) \\ &\quad \mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')E(\chi^{-4}_{r_1m+2}(\Delta)) - 2c\mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})[E(\chi^{-2}_{r_1m+4}(\Delta)) \\ &- E(\chi^{-2}_{r_1m+2}(\Delta))] + c^2\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}E(\chi^{-4}_{r_1m+4}(\Delta)) \end{aligned}$$

Note that $\Delta E(\chi^{-4}_{r_1m+4}(\Delta)) = E(\chi^{-2}_{r_1m+2}(\Delta)) - cE(\chi^{-4}_{r_1m+2}(\Delta))$. Thus, for all Δ, the $\hat{\boldsymbol{B}}^{JS}$ uniformly dominates $\hat{\boldsymbol{B}}$ whenever

$$\Delta < \frac{\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}[cE(\chi^{-4}_{r_1m+2}(\Delta)) - E(\chi^{-2}_{r_1m+4}(\Delta))]}{\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')E(\chi^{-4}_{r_1m+2}(\Delta))}$$

4.3.5 Comparison of $\hat{\boldsymbol{B}}^{JS}$ and $\tilde{\boldsymbol{B}}$

Let us consider the risk of $\hat{\boldsymbol{B}}^{JS}$ under a candidate subspace in terms of the risk of $\tilde{\boldsymbol{B}}$:

$$\begin{aligned} ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W}) &= ADR(\tilde{\boldsymbol{B}};\boldsymbol{W}) - 2tr(\boldsymbol{\mathcal{D}}_{11})tr(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}) \\ &\quad [cE(\chi^{-2}_{r_1m+2}(0)) - 1] + \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')[c^2E(\chi^{-4}_{r_1m+2}(0)) - 1]. \end{aligned}$$

Under H_0, the risk of $\tilde{\boldsymbol{B}}$ is smaller than that of the risk of $\hat{\boldsymbol{B}}^{JS}$ when the value of c satisfies the following condition

$$\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})[c^2E(\chi^{-4}_{r_1m+2}(0)) - 1] > 2[cE(\chi^{-2}_{r_1m+2}(0)) - 1].$$

However, as Δ increases, the risk of $\tilde{\boldsymbol{B}}$ becomes unbounded, and the risk of $\hat{\boldsymbol{B}}^{JS}$ remains below the risk of $\tilde{\boldsymbol{B}}$ and merges with it as $\Delta \longrightarrow \infty$. Thus, $\hat{\boldsymbol{B}}^{JS}$ dominates $\tilde{\boldsymbol{B}}$ outside an interval around the origin.

4.3.6 Comparison of $\hat{\boldsymbol{B}}^{JS+}$ and $\hat{\boldsymbol{B}}^{JS}$

For comparing the risk of $\hat{\boldsymbol{B}}^{JS}$ and $\hat{\boldsymbol{B}}^{JS+}$, we consider the risk-difference between them:

$$\begin{aligned} ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W}) \; - \; & ADR(\hat{\boldsymbol{B}}^{JS+};\boldsymbol{W}) = \\ + \; & 2\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})E[(1-c\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<c)] \\ + \; & 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})E[(1-c\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<c)] \\ - \; & \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')E[(1-c^2\chi^{-4}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<c)] \\ - \; & \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}E[(1-c^2\chi^{-4}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta)<c)]. \end{aligned}$$

The first two terms on the right hand side are positive, and we have

$$(0<\chi^2_{r_1m+4}(\Delta)<c^2) \Longleftrightarrow (c^2\chi^{-2}_{r_1m+4}(\Delta)-1)\geq 0$$

so that

$$E[(1-c^2\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta)<c)]\leq 0.$$

Thus, for all Δ, $ADR(\hat{\boldsymbol{B}}^{JS+})\leq ADR(\hat{\boldsymbol{B}}^{JS})$ and $\hat{\boldsymbol{B}}^{JS+}$ not only confirms inadmissibility of $\hat{\boldsymbol{B}}^{JS}$ but also provides a simple superior estimator.

Under H_0, we can order the risk of estimators

$$ADR(\tilde{\boldsymbol{B}};\boldsymbol{W})\leq ADR(\hat{\boldsymbol{B}}^{JS+};\boldsymbol{W})\leq ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W})\leq ADR(\hat{\boldsymbol{B}};\boldsymbol{W}).$$

4.4 Numerical Study

4.4.1 Simulation study

In this section, we use Monte Carlo simulation experiments to examine the relative performance of the proposed estimators. In this study, we simulate data from the following model: $Y_i = \mathbf{c}_0 + \mathbf{c}_1 x_{1i} + \mathbf{c}_2 x_{2i} + \mathbf{c}_3 x_{3i} + \mathbf{c}_4 x_{4i} + \boldsymbol{\epsilon}_i \equiv \boldsymbol{C} X_i + \boldsymbol{\epsilon}_i,\ i = 1, \ldots, n$, where $Y_i = (y_{1i}, y_{2i}, y_{3i}, y_{4i}, y_{5i})'$ and $X_i = (1, x_{1i}, x_{2i}, x_{3i}, x_{4i})'$ with $m = 1, 2, 3, 4, 5$. Therefore $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{X}}$, and $\boldsymbol{C}$ denote $5 \times n$, $5 \times n$, and 5×5 data matrices, respectively. For simulation we consider

$$\hat{\boldsymbol{C}} = \begin{bmatrix} 5.42 & -0.90 & -0.06 & -0.42 & 1.90 \\ 1.24 & -0.10 & 0.13 & -0.32 & 0.56 \\ -5.92 & 0.29 & -0.37 & 1.21 & -1.05 \\ 3.13 & -0.51 & -0.11 & -0.13 & 1.38 \\ 9.46 & -0.37 & 0.71 & -0.86 & 0.39 \end{bmatrix}$$

$$\hat{\boldsymbol{A}}' = \begin{bmatrix} 2.70 & 1.09 & -1.69 & 2.54 & 2.84 \\ 0.21 & 0.31 & 0.15 & -0.86 & 3.80 \end{bmatrix}$$

$$\hat{\boldsymbol{B}} = \begin{bmatrix} 0.81 & -0.13 & 0.08 & -0.24 & 0.61 \\ 1.57 & 0.19 & 0.28 & -0.24 & -0.49 \end{bmatrix}$$

$\boldsymbol{\mathcal{F}} = [I_2]$, and $\boldsymbol{\mathcal{G}} = [0', I_1']'$ or, more explicitly,

$$\boldsymbol{\mathcal{F}} = ((1,0)', (0,1)'),$$

$$\boldsymbol{\mathcal{G}} = ((0,0,0,0,0)', (0,0,0,0,1)')'.$$

We generate 1000 samples using the above model. We define Δ as a departure parameter which is a function of the distance between the true value of $\boldsymbol{C}$ and that under the candidate subspace. In order to investigate the behavior of the proposed estimators, different values of $\boldsymbol{B}$ are chosen to produce the value of Δ between 0 and 4. The performance of an estimator of $\boldsymbol{C}$ will be reappraised using the mean squared error criterion. All computations are conducted using the **R** statistical system. We numerically calculate the relative risk of $\tilde{\boldsymbol{B}}$, $\hat{\boldsymbol{B}}^{PT}$, $\hat{\boldsymbol{B}}^{JS}$, and $\hat{\boldsymbol{B}}^{JS+}$ with respect to $\hat{\boldsymbol{B}}$ by simulation. Since the result for different n were similar, here we only report the results for $n = 40$ and $n = 100$ in Figures (4.1) and (4.2).

Table 4.1: R.E of estimators, $n = 40$.

Δ	$\tilde{\boldsymbol{B}}$	$\hat{\boldsymbol{B}}^{JS}$	$\hat{\boldsymbol{B}}^{JS+}$
0.0	2.251	1.239	1.439
0.2	2.056	1.204	1.400
0.4	1.738	1.152	1.260
0.6	1.355	1.094	1.112
0.8	1.077	1.077	1.100
1.2	0.636	1.024	1.068
1.6	0.386	1.013	1.015
2.0	0.267	1.007	1.007
4.0	0.059	0.999	0.999

Table 4.2: R.E of estimators, $n = 100$.

Δ	$\tilde{\boldsymbol{B}}$	$\hat{\boldsymbol{B}}^{JS}$	$\hat{\boldsymbol{B}}^{JS+}$
0.0	3.227	2.544	2.027
0.2	1.589	1.051	1.341
0.4	0.687	0.969	1.105
0.6	0.341	1.000	1.048
0.8	0.197	1.000	1.025
1.2	0.094	1.000	1.011
1.6	0.052	1.000	1.007
2.0	0.033	1.000	1.004
4.0	0.009	1.000	1.000

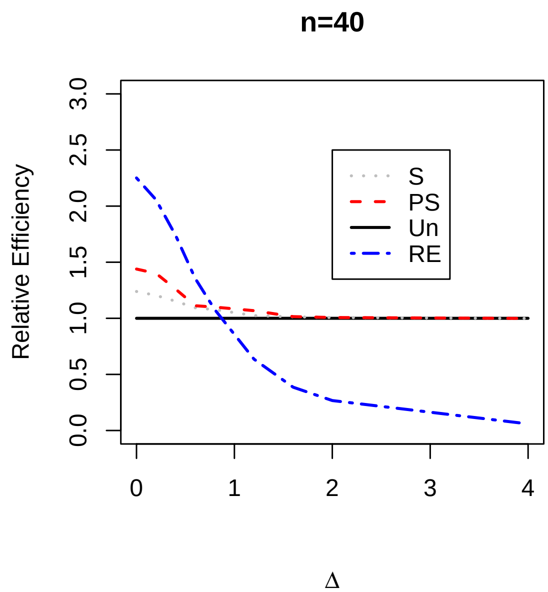

Figure 4.1: R.E of the estimators for $n = 40$.

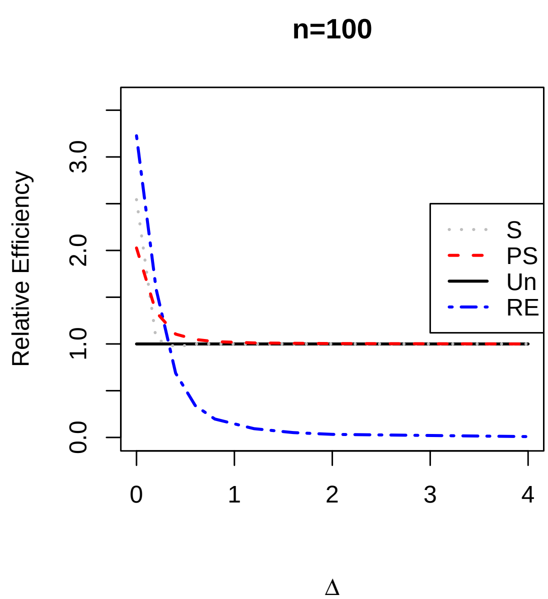

Figure 4.2: R.E of the estimators for $n = 100$.

4.4.2 Real Data Example

In this section, we consider reduced rank regression methods using the example on biochemical data that was examined in Section 3.4.2. It would appear from the results of the likelihood ratio test statistic that the possibility that the rank of the matrix $\boldsymbol{C}$ is either one or two could be entertained. Only estimation results for the rank two situation will be presented in detail here. Therefore, there is a reduced rank structure for the regression coefficient matrix (excluding the constant term) of the predictor variables $X_i = (x_{1i}, x_{2i}, x_{3i})'$. The normalized eigenvectors are $\hat{V}_1' = (-0.290, 0.234, 0.025, 0.906, 0.199)$ and $\hat{V}_2' = (0.425, 0.269, -0.841, 0.047, 0.194)$

Therefore, $\hat{\boldsymbol{A}}^{(2)} = \hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}^{1/2}\hat{V}_{(2)}'$ and $\hat{\boldsymbol{B}}^{(2)} = \hat{V}_{(2)}'\hat{\boldsymbol{\Sigma}}_{\epsilon\epsilon}^{-1/2}\hat{\boldsymbol{\Sigma}}_{yx}\hat{\boldsymbol{\Sigma}}_{xx}^{-1}$, with $\hat{V}_{(2)}' = [\hat{V}_1', \hat{V}_2']$. The least squares estimates of the $\boldsymbol{A}$ and $\boldsymbol{B}$ are given below:

$$\hat{\boldsymbol{A}}' = \begin{bmatrix} -1.085 & 0.333 & 0.231 & 0.397 & 1.109 \\ 1.335 & -0.126 & -0.266 & -0.092 & 0.875 \end{bmatrix}$$

$$\hat{\boldsymbol{B}} = \begin{bmatrix} 1.361 & -1.326 & 1.190 \\ -0.908 & 0.494 & 1.240 \end{bmatrix}.$$

The reduced rank estimate of the regression coefficient matrix is

$$\hat{\boldsymbol{C}} = \begin{bmatrix} -2.6893 & 2.0981 & 0.3649 \\ 0.5679 & -0.5040 & 0.2400 \\ 0.5558 & -0.4374 & -0.0555 \\ 0.6248 & -0.5726 & 0.3583 \\ 0.7142 & -1.0379 & 2.4044 \end{bmatrix}.$$

Notice again that we are excluding the first column of intercepts from the regression coefficient matrix. Most of the coefficients of $\hat{\boldsymbol{C}}$ were found to be statistically significant in the reduced rank estimate of $\boldsymbol{C}$. Now, based on stepwise selection, the predictor variable X_3 does not enter into the linear regression model. So, the restriction $\{b_{13} = b_{23} = 0\}$ can be imposed on the model. The coefficient matrices of the subspace are selected as $\boldsymbol{\mathcal{F}} = [I_2]$ and $\boldsymbol{\mathcal{G}} = [0', I_1']'$ or, more explicitly,

$$\boldsymbol{\mathcal{F}} = ((1,0)', (0,1)')$$

and

$$\boldsymbol{\mathcal{G}} = ((0,0)', (0,0)', (0,1)')'.$$

The least squares estimate of the matrix $\hat{\boldsymbol{B}}$, $\tilde{\boldsymbol{B}}$, $\hat{\boldsymbol{B}}^{JS}$, and $\hat{\boldsymbol{B}}^{JS+}$ are given below:

$$\hat{\boldsymbol{B}}^{JS+} = \begin{bmatrix} 2.1318964 & -1.2726333 & 1.587769 \\ -0.8517842 & 0.9462463 & 1.433390 \end{bmatrix},$$

$$\hat{\boldsymbol{B}} = \begin{bmatrix} 1.3609628 & -1.3258840 & 1.189601 \\ -0.9084776 & 0.4941074 & 1.240191 \end{bmatrix},$$

$$\tilde{\boldsymbol{B}} = \begin{bmatrix} 1.8218686 & -1.0393488 & 0 \\ -0.6059243 & 0.6546354 & 0 \end{bmatrix},$$

$$\hat{\boldsymbol{B}}^{JS} = \begin{bmatrix} 2.2244241 & -1.340690 & 1.592792 \\ -0.9210026 & 1.024746 & 1.449279 \end{bmatrix}.$$

We conduct a bootstrap with 1000 replicates to evaluate the performance of the suggested estimators in our data example. The performance of the estimators is evaluated in terms of the relative efficiency of the estimators, where relative efficiency of the estimator $\hat{\boldsymbol{B}}^*$ to the unrestricted least square estimator $\hat{\boldsymbol{B}}$ is denoted by $R.E(\hat{\boldsymbol{B}} : \hat{\boldsymbol{B}}^*) = \frac{risk(\hat{\boldsymbol{B}})}{risk(\hat{\boldsymbol{B}}^*)}$. The estimated relative efficiency at $\Delta = 0$ is given below:

$Estimator$	$R.E(\hat{\boldsymbol{B}} : \boldsymbol{B}^*)$
$\tilde{\boldsymbol{B}}$	4.11
$\hat{\boldsymbol{B}}^{JS}$	2.09
$\hat{\boldsymbol{B}}^{JS+}$	2.24

4.5 Concluding Remarks

We consider shrinkage and pretest estimators in a reduced rank regression model. We investigate the asymptotic properties of listed estimators under a very general candidate subspace. The relative performance of the estimators is examined using asymptotic analysis of quadratic risk functions; it is found that the shrinkage estimator outperforms the full model estimator uniformly. On the other hand, the pretest estimator dominates the least squares estimator only in a small part of the parameter space. Also, the risk performance of the listed estimators is investigated through an asymptotic distributional risk. A data example and our analytical results show that the suggested estimators perform better than the classical estimator under a candidate subspace and beyond. We conclude that a positive shrinkage estimator dominates the usual shrinkage estimator, and they both perform well relative to the classical full model generalized least squares estimator of the reduced rank regression parameter matrix in the entire parameter space. Note that the performance of the restricted and pretest estimators heavily depend on the quality of prior information.

4.6 Appendix: Proof of Main Results

4.6.1 Proof of Theorem 4.2.1

Since $\boldsymbol{q}_1, \boldsymbol{q}_2$, and $\boldsymbol{q}_3$ are asymptotically normal, the joint distribution of $(\boldsymbol{q}_1, \boldsymbol{q}_2)$ and $(\boldsymbol{q}_2, \boldsymbol{q}_3)$ will be asymptotically normal as well.

$$\begin{aligned} E(\boldsymbol{q}_2) &= \lim_{n\to\infty} E[n^{1/2} vec(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})] \\ &= \lim_{n\to\infty} E[n^{1/2} vec(\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}})] \qquad \text{under } K_n \\ &= \lim_{n\to\infty} E[n^{1/2} vec(\boldsymbol{\mathcal{N}} n^{-1/2} \boldsymbol{\Upsilon} \boldsymbol{\mathcal{Q}})] \\ &= (\boldsymbol{\mathcal{N}} \otimes \boldsymbol{\mathcal{Q}}') vec(\boldsymbol{\Upsilon}) \\ &= \boldsymbol{\delta} \end{aligned}$$

$$
\begin{aligned}
Cov(\boldsymbol{q}_2) &= Cov(vec(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})) \\
&= Cov(vec(\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}})) \\
&= Cov[(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')vec(\hat{\boldsymbol{B}}) - (\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')vec(\boldsymbol{\mathcal{D}})] \\
&= (\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')Cov[vec(\hat{\boldsymbol{B}})](\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')' \\
&= (\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}) \\
&= \boldsymbol{\Sigma}^*
\end{aligned}
$$

$$
\begin{aligned}
E(\boldsymbol{q}_3) &= E(\boldsymbol{q}_1 - \boldsymbol{q}_2) \\
&= \lim_{n\to\infty} E\{n^{1/2}[vec(\hat{\boldsymbol{B}} - \boldsymbol{B}) - vec(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})]\} \\
&= \lim_{n\to\infty} E[n^{1/2}vec(\hat{\boldsymbol{B}} - \boldsymbol{B})] - E[n^{1/2}vec(\boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}})] \qquad \text{under } K_n \\
&= 0 - \lim_{n\to\infty} E[n^{1/2}vec(\boldsymbol{\mathcal{N}}n^{-1/2}\boldsymbol{\Upsilon}\boldsymbol{\mathcal{Q}})] \\
&= -(\boldsymbol{\mathcal{N}} \otimes \boldsymbol{\mathcal{Q}}')vec(\boldsymbol{\Upsilon}) \\
&= -\boldsymbol{\delta}
\end{aligned}
$$

$$
\begin{aligned}
Cov(\boldsymbol{q}_3) &= Cov(\boldsymbol{q}_1 - \boldsymbol{q}_2) \\
&= Cov(\boldsymbol{q}_1) + Cov(\boldsymbol{q}_1) - 2Cov(\boldsymbol{q}_1, \boldsymbol{q}_2) \\
&= (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1}) + \boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21} \\
&= \boldsymbol{\Omega}^*
\end{aligned}
$$

4.6.2 Proof of Theorem 4.3.1

Here, we provide the proof of asymptotic bias expressions for the proposed estimators.

$$\begin{aligned}
ADB(\tilde{\boldsymbol{B}}) &= \lim_{n\to\infty} \sqrt{n}E[vec(\tilde{\boldsymbol{B}} - \boldsymbol{B})] \\
&= \lim_{n\to\infty} \sqrt{n}E[vec(\hat{\boldsymbol{B}} - \boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}} - \boldsymbol{B})] \\
&= \lim_{n\to\infty} \sqrt{n}E[vec(\hat{\boldsymbol{B}} - \boldsymbol{B} - \boldsymbol{\mathcal{N}}(\boldsymbol{\mathcal{F}}\hat{\boldsymbol{B}}\boldsymbol{\mathcal{G}} - \boldsymbol{\mathcal{D}})\boldsymbol{\mathcal{Q}})] \\
&= -(\boldsymbol{\mathcal{N}} \otimes \boldsymbol{\mathcal{Q}}')vec(\boldsymbol{\Upsilon}) \\
&= -\boldsymbol{\delta},
\end{aligned}$$

$$\begin{aligned}
ADB(\hat{\boldsymbol{B}}^{PT}) &= \lim_{n\to\infty} \sqrt{n}E[vec(\hat{\boldsymbol{B}}^{PT} - \boldsymbol{B})] \\
&= \lim_{n\to\infty} \sqrt{n}E[vec(\hat{\boldsymbol{B}} - \boldsymbol{B} - (\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})I(L < \chi^2_{r_1 m}(\Delta)))] \\
&= \lim_{n\to\infty} \sqrt{n}E[\boldsymbol{q}_1 - \boldsymbol{q}_3 I(L < \chi^2_{r_1 m}(\Delta))] \\
&= -\boldsymbol{\delta} H_{r_1 m+2}(\chi^2_{r_1 m}(\Delta)),
\end{aligned}$$

$$\begin{aligned}
ADB(\hat{\boldsymbol{B}}^{JS}) &= \lim_{n\to\infty} \sqrt{n}E[vec(\hat{\boldsymbol{B}}^{JS} - \boldsymbol{B})] \\
&= \lim_{n\to\infty} \sqrt{n}E[vec(\tilde{\boldsymbol{B}} - \boldsymbol{B} + (\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})(1 - sL^{-1}))] \\
&= \lim_{n\to\infty} \sqrt{n}E[\boldsymbol{q}_1 - s\boldsymbol{q}_3 L^{-1}] \\
&= -s\boldsymbol{\delta} E(\chi^{-2}_{r_1 m+2}(\Delta)),
\end{aligned}$$

$$\begin{aligned}
ADB(\hat{\boldsymbol{B}}^{JS+}) &= \lim_{n\to\infty} \sqrt{n}E[vec(\hat{\boldsymbol{B}}^{JS+} - \boldsymbol{B})] \\
&= \lim_{n\to\infty} \sqrt{n}E\{vec[(\hat{\boldsymbol{B}}^{JS} - \boldsymbol{B}) - (1 - sL^{-1})(\hat{\boldsymbol{B}} - \tilde{\boldsymbol{B}})I(L < s)]\} \\
&= ADB(\hat{\boldsymbol{B}}^{S}) - \boldsymbol{\delta} E\{[1 - s\chi^{-2}_{r_1 m+2}(\Delta)]I(\chi^2_{r_1 m+2}(\Delta) < s)\}.
\end{aligned}$$

4.6.3 Proof of Lemma 4.3.1

For the proof of part (i), we use the result of Theorem 4.2.1 part (i),(ii), and (iv)

$$\begin{aligned} E(\boldsymbol{q}_1|\boldsymbol{q}_2) &= E(\boldsymbol{q}_1) + \boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}^*)^{-1}(\boldsymbol{q}_2 - E(\boldsymbol{q}_2)) \\ &= (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}) \\ &\quad [(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})]^{-1} \\ &\quad [\boldsymbol{q}_2 - \boldsymbol{\delta}]. \end{aligned}$$

We got the result by using left Kronecker products. Proof of (ii) is similar to the proof of (iii), provided below, based on Theorem 4.2.1 part (i),(ii), and (iv). We have

$$\begin{aligned} E(\boldsymbol{q}_2\boldsymbol{q}_1{}'I(L < \chi^2_{r_1m}(\alpha))) &= E(E(\boldsymbol{q}_2\boldsymbol{q}_1{}'I(L < \chi^2_{r_1m}(\alpha))|\boldsymbol{q}_2)) \\ &= E(\boldsymbol{q}_2E(\boldsymbol{q}_1{}'I(L < \chi^2_{r_1m}(\alpha))|\boldsymbol{q}_2)) \\ &= E(\boldsymbol{q}_2[E(\boldsymbol{q}_1) + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{q}_2 - E(\boldsymbol{q}_2))]'I(L < \chi^2_{r_1m}(\alpha))) \\ &= E(\boldsymbol{q}_2[(\boldsymbol{q}_2 - \boldsymbol{\delta})'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}I(L < \chi^2_{r_1m}(\alpha))]) \\ &= E(\boldsymbol{q}_2\boldsymbol{q}_2{}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}I(L < \chi^2_{r_1m}(\alpha))) \\ &- E(\boldsymbol{q}_2)\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}I(L < \chi^2_{r_1m}(\alpha)) \\ &= [Var(\boldsymbol{q}_2)H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) + \boldsymbol{\delta}\boldsymbol{\delta}'H_{r_1m+4}(\chi^2_{r_1m}(\alpha))]\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21} \\ &- \boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) \\ &= \boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) + \boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1} \\ &\quad [H_{r_1m+4}(\chi^2_{r_1m}(\alpha)) - H_{r_1m+2}(\chi^2_{r_1m}(\alpha))], \end{aligned}$$

and

$$\begin{aligned}
E(\boldsymbol{q}_2\boldsymbol{q}_1{}'L^{-1}) &= E(E(\boldsymbol{q}_2\boldsymbol{q}_1{}'L^{-1}|\boldsymbol{q}_2))\\
&= E(\boldsymbol{q}_2E(\boldsymbol{q}_1{}'L^{-1}|\boldsymbol{q}_2))\\
&= E(\boldsymbol{q}_2[E(\boldsymbol{q}_1)+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{q}_2-E(\boldsymbol{q}_2))]'L^{-1})\\
&= E(\boldsymbol{q}_2[(\boldsymbol{q}_2-\boldsymbol{\delta})'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}L^{-1}])\\
&= E(\boldsymbol{q}_2\boldsymbol{q}_2{}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}L^{-1}-\boldsymbol{q}_2\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}L^{-1})\\
&= \boldsymbol{\Sigma}_{21}E(\chi^{-2}_{r_1m+2}(\Delta))+\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}[E(\chi^{-2}_{r_1m+4}(\Delta))-E(\chi^{-2}_{r_1m+2}(\Delta))]\\
&= (\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\otimes\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}^{-1}_{xx})E(\chi^{-2}_{r_1m+2}(\Delta))\\
&+ \boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}[E(\chi^{-2}_{r_1m+4}(\Delta))-E(\chi^{-2}_{r_1m+2}(\Delta))].
\end{aligned}$$

4.6.4 Proof of Theorem 4.3.2

Clearly, the $AMSE$ of $\hat{\boldsymbol{B}}$ is equal to $\boldsymbol{\Omega}^*+\boldsymbol{\delta}\boldsymbol{\delta}'$. Using Theorem 4.2.1,

$$\begin{aligned}
AMSE(\tilde{\boldsymbol{B}}) &= \lim_{n\to\infty} E\{n(\tilde{\boldsymbol{B}}-\boldsymbol{B})(\tilde{\boldsymbol{B}}-\boldsymbol{B})'\}\\
&= \lim_{n\to\infty} E(\boldsymbol{q}_3\boldsymbol{q}_3{}')\\
&= Var(\boldsymbol{q}_3)+E(\boldsymbol{q}_3)E(\boldsymbol{q}_3)'\\
&= \boldsymbol{\Omega}^*+\boldsymbol{\delta}\boldsymbol{\delta}'.
\end{aligned}$$

Note that

$$\begin{aligned}
AMSE(\hat{\boldsymbol{B}}^{PT}) &= \lim_{n\to\infty} E\{n(\hat{\boldsymbol{B}}^{PT}-\boldsymbol{B})(\hat{\boldsymbol{B}}^{PT}-\boldsymbol{B})'\} \\
&= \lim_{n\to\infty} E\{n[(\hat{\boldsymbol{B}}-(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})I(L<\chi^2_{r_1m}(\alpha))-\boldsymbol{B}) \\
&\quad (\hat{\boldsymbol{B}}-(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})I(L<\chi^2_{r_1m}(\alpha))-\boldsymbol{B})']\} \\
&= \lim_{n\to\infty} E\{[\boldsymbol{q}_1-\boldsymbol{q}_2I(L<\chi^2_{r_1m}(\alpha))][\boldsymbol{q}_1-\boldsymbol{q}_2I(L<\chi^2_{r_1m}(\alpha))]'\} \\
&= \lim_{n\to\infty} E\{\boldsymbol{q}_1\boldsymbol{q}_1{}'-\boldsymbol{q}_1\boldsymbol{q}_2{}'I(L<\chi^2_{r_1m}(\alpha))-\boldsymbol{q}_2\boldsymbol{q}_1{}'I(L<\chi^2_{r_1m}(\alpha)) \\
&+ \boldsymbol{q}_2\boldsymbol{q}_2{}'I^2(L<\chi^2_{r_1m}(\alpha))\} \\
&= Var(\boldsymbol{q}_1)+E(\boldsymbol{q}_1)E(\boldsymbol{q}_1)'-2E(\boldsymbol{q}_2\boldsymbol{q}_1{}'I(L<\chi^2_{r_1m}(\alpha))) \\
&+ E(\boldsymbol{q}_2\boldsymbol{q}_2{}'I^2(L<\chi^2_{r_1m}(\alpha))) \\
&= (\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_{xx}^{-1})-2\boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}'\otimes\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}H_{r_1m+2}(\chi^2_{r_1m}(\alpha)) \\
&- 2\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}'\otimes\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}[H_{r_1m+4}(\chi^2_{r_1m}(\alpha))-H_{r_1m+2}(\chi^2_{r_1m}(\alpha))] \\
&+ \boldsymbol{\delta}\boldsymbol{\delta}'H_{r_1m+4}(\chi^2_{r_1m}(\alpha))+\boldsymbol{\Sigma}^*H_{r_1m+2}(\chi^2_{r_1m}(\alpha)).
\end{aligned}$$

$AMSE(\hat{\boldsymbol{B}}^{JS})$ can be written as

$$\begin{aligned}
&= \lim_{n\to\infty} E\{n(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B})(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B})'\} \\
&= \lim_{n\to\infty} E\{n[\tilde{\boldsymbol{B}}+(1-sL^{-1})(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})-\boldsymbol{B}][\tilde{\boldsymbol{B}}+(1-sL^{-1})(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})-\boldsymbol{B}]'\} \\
&= \lim_{n\to\infty} E\{n[(\tilde{\boldsymbol{B}}-\boldsymbol{B})+(1-sL^{-1})(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})][(\tilde{\boldsymbol{B}}-\boldsymbol{B})+(1-sL^{-1})(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})]'\} \\
&= \lim_{n\to\infty} E\{n[(\hat{\boldsymbol{B}}-\boldsymbol{B})-sL^{-1}(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})][(\hat{\boldsymbol{B}}-\boldsymbol{B})-sL^{-1}(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})]'\} \\
&= E\{[\boldsymbol{q}_1-sL^{-1}\boldsymbol{q}_2][\boldsymbol{q}_1-sL^{-1}\boldsymbol{q}_2]'\} \\
&= E[\boldsymbol{q}_1\boldsymbol{q}_1{}'-\boldsymbol{q}_1\boldsymbol{q}_2{}'sL^{-1}-\boldsymbol{q}_2\boldsymbol{q}_1{}'sL^{-1}+\boldsymbol{q}_2\boldsymbol{q}_2{}'(sL^{-1})^2] \\
&= (\boldsymbol{I}_r\otimes\boldsymbol{\Sigma}_{xx}^{-1})-2s\{\boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}'\otimes\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}E(\chi^{-2}_{r_1m+2}(\Delta))+\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}'\otimes\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1} \\
&\quad (E[\chi^{-2}_{r_1m+4}(\Delta)]-E[\chi^{-2}_{r_1m+2}(\Delta)])\}+s^2[\boldsymbol{\Sigma}^*E(\chi^{-4}_{r_1m+2}(\Delta))+\boldsymbol{\delta}\boldsymbol{\delta}'E(\chi^{-4}_{r_1m+4}(\Delta))].
\end{aligned}$$

$AMSE(\hat{\boldsymbol{B}}^{JS+})$ can be written as

$$
\begin{aligned}
&= \lim_{n\to\infty} nE\{vec(\hat{\boldsymbol{B}}^{JS+}-\boldsymbol{B})[vec(\hat{\boldsymbol{B}}^{JS+}-\boldsymbol{B})]'\} \\
&= \lim_{n\to\infty} nE\{[vec(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B}) - vec(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})(1-sL^{-1})I(L<s)] \\
&\quad [vec(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B}) - vec(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})(1-sL^{-1})I(L<s)]'\} \\
&= \lim_{n\to\infty} nE\{[vec(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B})][vec(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B})]'\} - \\
&\quad 2E\{[vec(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})(1-sL^{-1})I(L<s)][vec(\hat{\boldsymbol{B}}^{JS}-\boldsymbol{B})]'\} + \\
&\quad E\{[vec(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})(1-sL^{-1})I(L<s)][vec(\hat{\boldsymbol{B}}-\tilde{\boldsymbol{B}})(1-sL^{-1})I(L<s)]'\} \\
&= AMSE(\hat{\boldsymbol{B}}^{JS}) - 2E\{\boldsymbol{q}_2\boldsymbol{q}_1{}'(1-sL^{-1})I(L<s)\} + \\
&\quad 2E\{\boldsymbol{q}_2\boldsymbol{q}_2{}'(1-sL^{-1})I(L<s)\} - E\{\boldsymbol{q}_2\boldsymbol{q}_2{}'(1-sL^{-1})^2 I(L<s)\} \\
&= AMSE(\hat{\boldsymbol{B}}^{JS}) - 2\boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}'\otimes\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}E[(1-s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<s)] - \\
&\quad 2\boldsymbol{\delta}\boldsymbol{\delta}'(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}'\otimes\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}E[(1-s\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<s)] + \\
&\quad 2\boldsymbol{\Sigma}^*E[(1-s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<s)] + \\
&\quad 2\boldsymbol{\delta}\boldsymbol{\delta}'E[(1-s\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+2}(\Delta)<s)] - \\
&\quad \boldsymbol{\Sigma}^*E[(1-s\chi^{-2}_{r_1m+2}(\Delta))^2 I(\chi^2_{r_1m+2}(\Delta)<s)] - \\
&\quad \boldsymbol{\delta}\boldsymbol{\delta}'E[(1-s\chi^{-2}_{r_1m+4}(\Delta))^2 I(\chi^2_{r_1m+4}(\Delta)<s)].
\end{aligned}
$$

4.6.5 Proof of Theorem 4.3.3

Following the definition in (4.4), the risk of $\hat{\boldsymbol{B}}$ is clearly equal to $r\text{tr}(\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-1})$. Note that

$$\begin{aligned} ADR(\tilde{\boldsymbol{B}};\boldsymbol{W}) &= \text{tr}(\boldsymbol{W}AMSE(\tilde{\boldsymbol{B}})) \\ &= \text{tr}(\boldsymbol{W}\boldsymbol{\Omega}^*) + \text{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}') \\ &= r\text{tr}(\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-1}) - 2\text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}') \\ &+ \text{tr}(\boldsymbol{\mathcal{D}}_{11})\text{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}') + \boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}. \end{aligned}$$

The risk of $\tilde{\boldsymbol{B}}$ depends on $\boldsymbol{\delta}'\boldsymbol{W}\boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\boldsymbol{\mathcal{N}} \otimes \boldsymbol{\mathcal{Q}}')vec(\boldsymbol{\Upsilon})$. Also, $\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}$. Note that $\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}$ is symmetric and idempotent with rank r_1. Thus, there exists an orthogonal matrix $\boldsymbol{\Gamma}$ such that

$$\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{I}_{r_1 m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$
$$\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}' = \begin{pmatrix} \boldsymbol{\mathcal{D}}_{11} & \boldsymbol{\mathcal{D}}_{12} \\ \boldsymbol{\mathcal{D}}_{21} & \boldsymbol{\mathcal{D}}_{22} \end{pmatrix}.$$

Hence,

$$\begin{aligned} \text{tr}[\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}] &= \text{tr}[\{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\ &\quad \{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}'\}] \\ &= \text{tr}[\begin{pmatrix} \boldsymbol{\mathcal{D}}_{11} & \boldsymbol{\mathcal{D}}_{12} \\ \boldsymbol{\mathcal{D}}_{21} & \boldsymbol{\mathcal{D}}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_{r_1 m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}] \\ &= \text{tr}[\boldsymbol{\mathcal{D}}_{11}], \end{aligned}$$

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{W}\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}] &= \mathrm{tr}[\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}\boldsymbol{W}\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}'] \\
&= \mathrm{tr}[\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}\boldsymbol{W}\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}] \\
&= \mathrm{tr}[\{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\
&\quad \{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}'\} \\
&\quad \{\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}})^{-1}\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-\frac{1}{2}}\boldsymbol{\Gamma}'\}] \\
&= \mathrm{tr}[\begin{pmatrix} \boldsymbol{I}_{r_1 m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{D}}_{11} & \boldsymbol{\mathcal{D}}_{12} \\ \boldsymbol{\mathcal{D}}_{21} & \boldsymbol{\mathcal{D}}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{r_1 m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}] \\
&= \mathrm{tr}[\begin{pmatrix} \boldsymbol{\mathcal{D}}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{r_1 m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}] \\
&= \mathrm{tr}[\boldsymbol{\mathcal{D}}_{11}],
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{W}\boldsymbol{\Sigma}^*] &= \mathrm{tr}[\boldsymbol{W}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})] \\
&= \mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')\mathrm{tr}(\boldsymbol{W}\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}}) \\
&= \mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}')\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11}),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{W}\boldsymbol{\Sigma}^*\boldsymbol{\mathcal{U}}] &= \mathrm{tr}[\boldsymbol{W}\boldsymbol{\Sigma}^*(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}}\boldsymbol{\mathcal{Q}})^{-1}] \\
&= \mathrm{tr}[\boldsymbol{W}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}')(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})] \\
&= \mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})\mathrm{tr}(\boldsymbol{W}\boldsymbol{\mathcal{Q}}'\boldsymbol{\mathcal{G}}'\boldsymbol{\Sigma}_{xx}^{-1}) \\
&= \mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11}).
\end{aligned}$$

After some computation, we obtain the expression for the risk of $ADR(\tilde{\boldsymbol{B}};\boldsymbol{W})$ as follows:

$$
\begin{aligned}
ADR(\tilde{\boldsymbol{B}};\boldsymbol{W}) &= \mathrm{tr}(\boldsymbol{W\Omega}^*) + \mathrm{tr}(\boldsymbol{W\delta\delta}') \\
&= \mathrm{tr}(\boldsymbol{W}(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})) - 2\mathrm{tr}(\boldsymbol{W}(\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}_{xx}^{-1})(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}' \otimes \boldsymbol{\mathcal{G}\mathcal{Q}})) \\
&+ \mathrm{tr}(\boldsymbol{W\Sigma}^*) + \boldsymbol{\delta}'\boldsymbol{W\delta} \\
&= ADR(\hat{\boldsymbol{B}};\boldsymbol{W}) - 2tr(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{N}}') + \mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}\mathcal{F}\mathcal{F}}'\boldsymbol{\mathcal{N}}') \\
&+ \boldsymbol{\delta}'\boldsymbol{W\delta}.
\end{aligned}
$$

Now, similarly, we get

$$
\begin{aligned}
ADR(\hat{\boldsymbol{B}}^{PT};\boldsymbol{W}) &= r\ \mathrm{tr}(\boldsymbol{W\Sigma}_{xx}^{-1}) - 2\mathrm{tr}(\boldsymbol{W\Sigma}^*\boldsymbol{\mathcal{U}})H_{r_1m+2}(\chi^2_{r_1m}(\Delta)) - \\
&\quad 2\mathrm{tr}(\boldsymbol{W\delta\delta}'\boldsymbol{\mathcal{U}})[H_{r_1m+4}(\chi^2_{r_1m}(\Delta)) - H_{r_1m+2}(\chi^2_{r_1m}(\Delta))] + \\
&\quad \mathrm{tr}(\boldsymbol{W\Sigma}^*)H_{r_1m+2}(\chi^2_{q+2}(\Delta)) + \mathrm{tr}(\boldsymbol{W\delta\delta}')H_{r_1m+4}(\chi^2_{r_1m}(\Delta)).
\end{aligned}
$$

After some computation, we get the expression for the risk of $ADR(\hat{\boldsymbol{B}}^{PT};\boldsymbol{W})$. Also,

$$
\begin{aligned}
ADR(\hat{\boldsymbol{B}}^{JS};\boldsymbol{W}) &= r\mathrm{tr}(\boldsymbol{W\Sigma}_{xx}^{-1}) - 2s\mathrm{tr}(\boldsymbol{W\Sigma}^*\boldsymbol{\mathcal{U}})E(\chi^{-2}_{r_1m+2}(\Delta)) \\
&+ 2s\mathrm{tr}(\boldsymbol{W\delta\delta}'\boldsymbol{\mathcal{U}})\{[E(\chi^{-2}_{r_1m+2}(\Delta)) - E(\chi^{-2}_{r_1m+4}(\Delta))]\} \\
&+ s^2\mathrm{tr}(\boldsymbol{W\Sigma}^*)E(\chi^{-4}_{r_1m+2}(\Delta))] + s^2\boldsymbol{\delta}'\boldsymbol{W\delta}E(\chi^{-4}_{r_1m+4}(\Delta)).
\end{aligned}
$$

Finally,

$$\begin{aligned}
ADR(\hat{\boldsymbol{B}}^{JS+}) &= ADR(\hat{\boldsymbol{B}}^{JS}) - 2\mathrm{tr}(\boldsymbol{\mathcal{D}}_{11})\mathrm{tr}(\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{F}})E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&- 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\mathcal{U}})E[(1 - s\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta) < s)] \\
&+ 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*)E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&+ 2\mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}')E[(1 - c\chi^{-2}_{r_1m+4}(\Delta))I(\chi^2_{r_1m+4}(\Delta) < s)] \\
&- \mathrm{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^*)E[(1 - s\chi^{-2}_{r_1m+2}(\Delta))^2 I(\chi^2_{r_1m+2}(\Delta) < s)] \\
&- \mathrm{tr}(\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\delta}')E[(1 - s\chi^{-2}_{r_1m+4}(\Delta))^2 I(\chi^2_{r_1m+4}(\Delta) < s)].
\end{aligned}$$

After using some computation, we get the expression for the $ADR(\hat{\boldsymbol{B}}^{JS+})$.

Chapter 5

Concluding Remarks and Future Research

In this dissertation, we study different estimation strategies for multivariate regression models. The following estimation procedures are discussed in this dissertation.

- Unrestricted and restricted estimation;
- Shrinkage and positive shrinkage estimation;
- Pretest estimation.

We apply the above estimation procedures in some multivariate multiple regression models to improve the performance of existing estimators when non-sample information is available. The positive-part shrinkage estimator dominates the usual shrinkage estimator. At any rate, both shrinkage estimators perform well relative to the usual unrestricted least squares estimator of the parameters in the entire parameter space.

In Chapter 2, we study the simple multivariate regression model that includes basic investment models. We consider various estimation strategies based on the pretest and shrinkage estimation. The subspace candidate least squares estimator depends heavily on the quality of the subspace information. The asymptotic distributional risk of the restricted least squares estimator is unbounded when the parameter moves far from the subspace of the restriction, while the pretest estimator provides good control on the magnitude of the asymptotic distributional risk. It is exceedingly important to note that the shrinkage estimators have the smallest possible asymptotic risk in most cases, as compared to other estimators except when the subspace information is nearly correct. Finally, a numerical study based on a real data set demonstrates how to implement and use the proposed estimation procedure. The statistical properties of the estimators are investigated analytically and numerically. Also, the simulation study supports our theoretical findings.

In Chapter 3, we generalize the estimation strategies for the matrix of a regression parameter in a multivariate multiple regression model. Also, we study the application of shrinkage and pretest estimation strategies and discuss the relative performance of the full model LSE, a candidate subspace estimator, a pretest, and shrinkage estimators in MMRMs in the presence of a natural candidate subspace when the matrix of explanatory variables is fixed. We succinctly investigate the bias and risk properties of the suggested estimators. We conclude that the risk improvement of the submodel estimator over other estimators is substantial near the restriction. However, the improvement starts diminishing as the restriction moves further and further away from the assumed subspaces. Thus, the performance of the submodel estimator heavily depends on the quality of the subspace information. The shrinkage estimators with data based weights outperform the full model estimator which outperforms the full model LSE in the most parameter space induced by the candidate subspace. Based on the risk comparisons and relative efficiencies among the suggested estimators, the weight of the shrinkage estimator has an appealing intuitive property. In summary, we show large gains of a suggested shrinkage approach over an ordinary least-squares approach. Finally, the real data example supports the contention that the suggested method is superior to classical estimation.

In Chapter 4, we consider shrinkage and pretest estimators in a multivariate reduced rank regression model. We investigate the asymptotic properties of listed estimators under a very general candidate subspace. The relative performance of the estimators is examined using asymptotic analysis of quadratic risk functions. It is found that the pretest estimator dominates the least squares estimator in some part of the parameter space. Also, the risk performance of the listed estimators are investigated through asymptotic distributional risk. A data example and our analytical results show that all suggested estimators perform better than the classical estimator under a candidate subspace and beyond. We conclude that a positive shrinkage estimator dominates the usual shrinkage estimator uniformly, and they both perform well relative to the classical full model weighted least squares estimator of the reduced rank regression parameter matrix. Further, the performance of the restricted and pretest estimators heavily depend on the quality of prior information.

5.1 Future Research

For future research, we can consider shrinkage and pretest estimations in an extension of the basic growth curve (or generalized MANOVA) model. In this extended model, the mean of the response vector is represented by a sum of the two components, a growth curve portion, and a standard MANOVA portion. The model could be viewed as an analysis of a covariance model that adjusts the growth curve structure for the influence of additional covariates for the response vector. Specifically, the model is considered as

$$\boldsymbol{Y}_i = \boldsymbol{A}\boldsymbol{B}\boldsymbol{X}_i + \boldsymbol{D}\boldsymbol{Z}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, n.$$

Notice that the form of this model is similar to the reduced rank model in Chapter 4, however, in the presence of a growth curve model context, the matrix $\boldsymbol{A}$ is known. The error terms $\boldsymbol{\epsilon}_i$ are assumed to follow a multivariate normal distribution with zero mean vector and positive definite covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$.

Also, we can consider shrinkage and pretest strategies in linear simultaneous equation models. The model can be written in the form

$$\boldsymbol{Y}\boldsymbol{B} = \boldsymbol{Z}\boldsymbol{\Gamma} + \boldsymbol{U}.$$

A simultaneous equation model (SEM) relates a set of endogenous or dependent variables to a set of exogenous, independent, or predetermined variables with error variables. In contrast to many statistical studies, the interest in simultaneous equation models is in linear restrictions on the regression of the dependent variables on the independent variables. In order to have nontrivial linear restrictions on the regression coefficients, the regression matrix has to be a reduced rank.

Bibliography

Ahmed, S. E. (1992). Shrinkage preliminary test estimation in multivariate normal distributions. *Statistical Computation and Simulation*, 43:177–195.

Ahmed, S. E. (2001). Shrinkage estimation of regression coefficients from censored data with multiple observations. empirical bayes and likelihood inference. *Lecture Notes in Statistics, Springer-Verlag: New York*, 148:103–120.

Ahmed, S. E. and Chitsaz, S. (2011). Data-based adaptive estimation in an investment model. *Communications in Statistics - Theory and Methods*, 40:3540–3554.

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

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

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

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21:243–247.

Akaike, H. (1977). On entropy maximization principle in: P.r. krishnaiah, (ed.). *Applications of Statistics. North-Holland, Amsterdam.*

Albertson, K. V. (1991). Pre-test estimation in a regression model with a misspecified error covariance matrix. *Department of Economics, University of Canterbury*, 9115.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 22:327–351.

Anderson, T. W. (1984). An introduction to multivariate statistical analysis. *Second edition, New York, Wiley.*

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.*, 15:190–204.

Brillinger, D. R. (1969). The canonical analysis of stationary time series. *Multivariate Analysis*, 2:331–350.

Chitsaz, S. and Ahmed, S. E. (2012a). An improved estimation in regression parameter matrix in multivariate regression model. *Communications in Statistics, Theory and Methods.*

Chitsaz, S. and Ahmed, S. E. (2012b). Shrinkage estimation for the regression parameter matrix in multivariate regression model. *Statistical Computation and Simulation*, 82(2):309–323.

Davies, P. T. and Tso, M. K. S. (1982). On the asymptotic properties of ldu-based tests of the rank of a matrix. *Journal of the American Statistical Association*, 31:244–255.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218.

Geweke, J. F. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75:121–146.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41:190–195.

Heinen, A. and Rengifo, E. (2007). Multivariate reduced rank regression in non-gaussian contexts, using copulas. *Computational Statistics and Data Analysis*, 52(6):2931–2944.

Izenman, A. J. (2008). Modern multivariate statistical techniques: Regression, classification and manifold learning. *Springer*.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *Journal of Finance*, 23:389–461.

Judge, G. G. and Bock, M. E. (1978). The statistical implications of pretest and stein-rule estimators in econometrics. *Amsterdam: North Holand Publishing Company.*

Kazimi, C. and Brownstone, D. (1999). Bootstrap confidence bands for shrinkage estimators. *Journal of Econometrics*, 90:99–127.

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

Leeb, H. and Potscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21:21–59.

Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *Annals of Statistics*, 36(1):261–286.

Maddala, G. S., Li, R. P., and Joutz, H. (1997). Estimation of sort-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics*, 15:90–100.

Muirhead, R. J. and Koole, G. (1982). Aspects of multivarite statistical theory. *New York: Wiley.*

Reinsel, G. C. (1983). Some results on multivariate autoregressive index models. *Biometrika*, 70:145–156.

Reinsel, G. C. (1998). On multivariate linear regression shrinkage and reduced rank procedures. *Technical report, Department of Statistics, University of Wisconsin, Madison, WI.*

Reinsel, G. C. and Velu, R. P. (1998). Multivariate reduced-rank regression: Theory and applications. *New York: Springer.*

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14:465–471.

Robinson, P. M. (1973). Generalized canonical analysis for time series. *Journal of multivariate Analysis*, 3:141–160.

Robinson, P. M. (1974). Identification, estimation and large-sample theory for regressions containing unobservable variables. *International Economic Review*, 15:680–692.

Schmidli, H. (1995). Reduced rank regression: With applications to quantitative structure-activity relationships. *Heidelberg:Physica.*

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

Sclove, S. L., Morris, C., and Radhakrishnan, R. (1972). Optimality of preliminary test estimation for the multinormal mean. *The Annals of Mathematical Statistics*, 43:1481–1490.

Srivastava, M. S. and Khatri, C. G. (1979). An introduction to multivariate statistics. *New York: North Holand*.

Timm, N. H. (1975). Spanish-english code-switching: el porque y how-not-to. *Romance Philology*, 28:473–482.

Timm, N. H. (2002). Applied multivariate analysis. *Springer*.

Tso, M. K. S. (1981). Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society, B*, 43:183–189.

Velu, R. P. and Reinsel, G. C. (1987). Reduced-rank regression with autoregressive errors. *Journal of Econometrics*, 35:317–335.

Vounou, M., Nichols, T. E., and Montana, G. (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Computational Statistics and Data Analysis*, 53(3):1147–1159.

Wan, A., Chaturvedi, A., and Zou, G. (2003). Unbiased estimation of the mse matrices of improved estimators in linear regression. *Journal of Applied Statistics*, 30(2):173–189.

Yee, T. W. and Hastie, T. J. (2003). Reduced-rank vector generalized linear models. *Statistical Modelling*, 3:1541.

Zhou, G. (1994). Analytical gmm tests: Asset pricing with time varying risk premiums. *Review of Financial Studies*, 7:687–709.

Vita Auctoris

NAME:	Shabnam Chitsaz
PLACE OF BIRTH:	Arak, Iran
EDUCATION	Shahid Beheshti University, Tehran, Iran
	1999-2003 B.Sc.
	Allameh Tabatabai University, Tehran, Iran
	2004-2007 M.Sc.
	University of Windsor, Windsor, Canada
	2008-2012 Ph.D.