

Optimal Generalized Shrunk Least Squares Estimator in Linear Models

by

Zhiming Zhang

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Statistics
University of Manitoba
Winnipeg

Copyright © 2025 by Zhiming Zhang

Abstract

It is widely acknowledged that, in a linear model, when the predictor variables are highly correlated so that the design matrix is nearly singular, the Ordinary Least Squares (OLS) estimator may have poor performance due to its large variance and numerical instability. The problem is particularly acute in modern times where models with large number of variables are routinely used. To address the issue of multicollinearity, linear biased estimators have received significant attention in the literature. This study focuses on the class of Generalized Shrunk Least Square (GSLS) estimators that contain various popular biased estimators as special cases. In particular, we derive the optimal GSLS estimator and develop an effective algorithm for its numerical calculation. Comprehensive simulation studies are conducted to evaluate and compare the performance of the proposed optimal estimator with other estimators across varying multicollinearity scenarios under the mean square error (MSE) criterion. Real data applications are also presented to demonstrate the usefulness of the proposed method.

Acknowledgment

First, I express my deepest gratitude to my supervisor, Dr. Liquan Wang, whose continuous support has been instrumental throughout my research journey. I remember the first time I came to our department, problems such as language barrier, feeling of loneliness, challenges in my studies, really confused me. I did not know what to do in the future, almost lost confidence in my research process. Without my supervisor's patience and insightful guidance, I would not be able to continue my program. Without his expertise and encouragement, this thesis would not have been possible.

I also sincerely thank my advisory committee members, Drs. Saumen Mandal and Po Yang, for their valuable suggestions and constructive feedback, which greatly improved the quality of my thesis.

Special thanks go to the Department of Statistics, the Faculty of Graduate Studies at the University of Manitoba, and the Natural Sciences and Engineering Research Council of Canada (NSERC), for the financial supports throughout my study.

I would like to express my heartfelt appreciation to my workmates, Bahram and Avanthi, for their kindness, understanding, and unwavering moral support during the ups and downs of my graduate life at the UM.

Finally, I am forever grateful to my parents for their love, sacrifices, and constant encouragement. This achievement would not have been possible without their belief in me.

Dedication Page

To the memory of those who inspired me, and to the pursuit of knowledge that has guided my path throughout this research.

Contents

Contents	v
List of Tables	ix
List of Figures	xi
1 Introduction	1
1.1 Overview	1
1.2 Outline of the Thesis	3
2 Generalized Shrunk Least Square Estimators	5
2.1 Introduction	5
2.2 Linear biased estimators	7
2.2.1 Ridge Estimator	7
2.2.2 Liu Estimator	8
2.2.3 Liu Type Estimator	10

2.2.4	Two-Parameter Estimator	11
2.2.5	New Two-Parameter Estimator	12
2.2.6	Modified Two-Parameter Estimator	13
2.2.7	Modified New Two-Parameter Estimator	14
3	Admissibility of GSLS Estimators	17
3.1	Introduction	17
3.2	Admissibility	19
3.2.1	Admissibility of Ridge Estimator	19
3.2.2	Admissibility of Liu Estimator	20
3.2.3	Admissibility of Liu Type Estimator	21
3.2.4	Admissibility of Two-Parameter Estimator	23
3.2.5	Admissibility of Modified Two-Parameter Estimator	24
3.2.6	Admissibility of Modified New Two-Parameter Estimator	25
4	Optimal Estimators	27
4.1	Optimal GSLS estimator	27
4.2	Optimal estimators	30
4.2.1	Optimal ridge estimator	31
4.2.2	Optimal Liu Estimator	32

4.2.3	Optimal Liu Type Estimator	34
4.2.4	Optimal Two-Parameter Estimator	36
4.2.5	Optimal New Two-Parameter Estimator	38
4.2.6	Optimal Modified New Two-Parameter Estimator . . .	39
5	Numerical Calculation of the Optimal Estimators	43
5.1	Comparisons between GOLS estimators	43
5.2	Iteration algorithms	45
5.2.1	Old iteration method	45
5.2.2	New iteration method	46
5.2.3	Bootstrap iteration method	48
6	Monte Carlo Simulations	53
6.1	Simulation study 1	56
6.2	Simulation study 2	60
7	Real Data Application	65
7.1	Portland cement dataset	65
7.2	French economy dataset	68
8	Conclusion	71

A Simulation R Code	73
Bibliography	83
Index	90

List of Tables

2.1	Summary of GOLS Estimators	15
5.1	Results Comparison	51
6.1	AMSE under simulation study 1 when $n = 50$	57
6.2	AMSE under simulation study 1 when $n = 100$	57
6.3	AMSE under simulation study 1 when $n = 200$	58
6.4	AMSE under simulation study 2 when $n = 50$	60
6.5	AMSE under simulation study 2 when $n = 100$	61
6.6	AMSE under simulation study 2 when $n = 200$	61
7.1	Homogeneous model for cement data	66
7.2	Inhomogeneous model for cement data	67
7.3	Multicollinearity diagnostics	69
7.4	Estimated coefficients of French economy data	69

List of Figures

5.1	New iteration method	50
6.1	AMSE under simulation study 1	58
6.2	AMSE under simulation study 2	62

Chapter 1

Introduction

1.1 Overview

Regression analysis is one of the most widely used statistical techniques for modeling relationships between a dependent variable and one or more independent variables. Galton (1886) introduced the concept of regression towards the mean and laid the foundations for later developments in statistical regression, particularly in biology and social sciences. The modern formulation of linear regression was formalized by Pearson (1896) and Fisher (1921), whose contributions were vital in developing the least squares method to estimate regression coefficients. The introduction of multiple regression by Wright (1934) expanded the method to allow multiple independent variables, allowing researchers to control for confounding variables and examine the unique contribution of each predictor.

More recently, the development of regularized regression techniques, such as

Tibshirani (1996), has addressed the issue of multicollinearity and overfitting in high-dimensional data. Then many alternative methods were proposed in the literature, such as the Stein estimator (Stein et al. (1956)), principle components estimator (Massy (1965)), the ordinary ridge regression estimator (Hoerl and Kennard (1970)), the modified ridge regression estimator (Swindel (1976)) etc. These estimators are all biased estimators, and the ordinary ridge estimator (ORRE) is one of the most widely used estimators. It intends to overcome the problem of multicollinearity by adding a positive value k to the diagonal elements of the ill-conditioned matrix $\mathbf{X}'\mathbf{X}$, which is known as the biasing parameter or the shrinkage parameter. A huge amount of literature is available on the selection of the biasing parameter k , but the choice of a suitable k is difficult since the ridge estimator is a non-linear function of it. Liu (1993) proposed another estimator, whose advantage over the ridge estimator is that the Liu estimator (LE) is a linear function of the bias estimator d . Kaçiranlar et al. (1999) amended the LE in the form of a new estimator, known as the restricted LE. This restricted LE is superior under the MSE matrix criterion as compared to the usual LE and the restricted least squared estimator.

Later, Özkale and Kaciranlar (2007) proposed the two-parameter (TP) estimator to reduce OLSE, ORRE, and LE. The estimator is based on the biasing parameter k and d . proposed a new two-parameter estimator. Sakallıoğlu and Kaçiranlar (2008) proposed another new biased estimator, called the $(k - d)$ class estimator. The basic idea is to substitute $\hat{\beta}_R$ in place of $\hat{\beta}_O$ in the LE

and obtain a new biased estimator, which is more reliable than the LE. It is a general estimator which includes the OLSE, ORRE and LE. Recently, Dorugade (2014a) proposed a modified TPE. This estimator was a sort of almost unbiased estimator, following the work of Wu and Yang (2013). Moreover, this estimator was proposed after two-step estimation in which the initial estimator was given following LE, TPE.

1.2 Outline of the Thesis

The remainder of this thesis is organized as follows. In Chapter 2, we investigate a class of generalized shrunken least squares (GSLS) estimators, originally introduced by Wang (1990). It is shown that many of these estimators can be represented as special cases within the framework of the GSLS estimators. Chapter 3 examines the admissibility conditions of these GSLS estimators, providing the necessary and sufficient criteria for the admissibility of each subclass, including Ridge, Liu, Liu-type, and various two-parameter estimators. In Chapter 4, we derive the locally optimal GSLS estimator by minimizing the mean square error (MSE) and discuss how this optimal estimator relates to other biased estimators through parameter comparisons. Chapter 5 outlines an iterative algorithm and bootstrap method to estimate the optimal parameters practically, and presents a comprehensive Monte Carlo simulation study to compare the performance of the proposed optimal estimator with conventional biased estimators under various multicollinearity scenarios. Finally, Chapter 6

concludes the study, summarizing the main findings and providing information for future research.

Chapter 2

Generalized Shrunk Least Square Estimators

2.1 Introduction

Consider the linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{Y}$ is a $n \times 1$ vector of observations of the response variable, $\mathbf{X}$ is the $n \times p$ design matrix of the predictor variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of random errors. Throughout this thesis, we make the following assumptions:

Assumption 1: $\mathbf{X}$ is a matrix of fixed constants and has full rank $2 \leq p \leq n$.

Assumption 2: The random error $\boldsymbol{\varepsilon}$ has zero mean and covariance matrix $\sigma^2 \mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix.

6CHAPTER 2. GENERALIZED SHRUNKEN LEAST SQUARE ESTIMATORS

Under these assumptions, the ordinary least squares (OLS) estimator is

$$\hat{\beta}_O = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (2.2)$$

which is the best linear unbiased estimator (BLUE) for β . Moreover, the variance of $\hat{\beta}_{OLS}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Let the eigenvalue of $\mathbf{X}'\mathbf{X}$ be denoted by $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = \lambda_{min} > 0$. In general,

$$MSE(\hat{\beta}) = \mathbb{E}((\hat{\beta} - \beta)'(\hat{\beta} - \beta)) = Trace(Cov(\hat{\beta})) + \|\mathbb{E}(\hat{\beta}) - \beta\|^2$$

Since $\mathbf{X}'\mathbf{X}$ is a symmetric matrix, the MSE of $\hat{\beta}_{OLS}$ is $\sigma^2 \sum_{j=1}^p \lambda_j^{-1}$.

When one of λ_j is extremely small, the MSE of $\hat{\beta}_{OLS}$ is very large, which means that when multicollinearity is present $\hat{\beta}_{OLS}$ is not the best choice. So, it is possible to significantly reduce the MSE by increasing a little bias but decrease the variance by a large amount, which means that we can use biased estimators to deal with this problem.

Wang (1990) introduced a class of generalized shrunk least squares (GSLS) estimators defined by

$$\hat{\beta}_{GS}(\mathbf{A}) = \mathbf{P}\mathbf{A}\mathbf{P}'\hat{\beta}_O, \quad (2.3)$$

where $\mathbf{A} = diag(a_1, a_2, \dots, a_p)$, $0 \leq a_i \leq 1$, $i = 1, \dots, p$ is called the id matrix to identify different biased estimators. $\mathbf{P}$ is the orthogonal matrix such that $\mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P} = diag(\lambda_1, \lambda_2, \dots, \lambda_p) = \mathbf{\Lambda}$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. $\mathbf{\Lambda}$ is called the eigenvalue matrix that simplifies certain calculation.

We can write equation (2.3) in the following form using $\mathbf{X}'\mathbf{X} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$,

$$\begin{aligned}\hat{\beta}_{GS}(\mathbf{A}) &= \mathbf{P}\mathbf{A}\mathbf{P}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{P}\mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{P}'\mathbf{X}'\mathbf{Y} \\ &= (\mathbf{P}\mathbf{\Lambda}\mathbf{A}^{-1}\mathbf{P}')^{-1}\mathbf{X}'\mathbf{Y},\end{aligned}\tag{2.4}$$

provided $\mathbf{A} > \mathbf{0}$. In the subsequent section, it will be demonstrated that this class includes many other linearly biased estimators proposed in the literature.

2.2 Linear biased estimators

In this section, we present a range of commonly known and some recently published biased estimators in the literature. These are summarized in Table 2.1.

2.2.1 Ridge Estimator

Hoerl and Kennard (1970) proposed an estimation procedure based on adding small positive quantities to the diagonal of $\mathbf{X}'\mathbf{X}$, the following class of Ridge estimators (RE) is

$$\hat{\beta}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \quad k > 0.\tag{2.5}$$

It is straightforward to show that a RE can be written in the form of GSLS as $\hat{\beta}_R(k) = \hat{\beta}_{GS}(\mathbf{A})$, Let $(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} = (\mathbf{P}\mathbf{\Lambda}\mathbf{A}^{-1}\mathbf{P}')^{-1}$ using equation (2.4), then

$$\begin{aligned}
\mathbf{A}_R &= (\mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P} + \mathbf{P}'k\mathbf{I}\mathbf{P})\mathbf{\Lambda} \\
&= \mathbf{\Lambda}(\mathbf{\Lambda} + k\mathbf{I})^{-1}
\end{aligned} \tag{2.6}$$

The ridge estimator can be expressed in the format of a GSLS estimator as

$$\hat{\boldsymbol{\beta}}_R = \mathbf{P}\mathbf{A}_R\mathbf{P}'\hat{\boldsymbol{\beta}}_O. \tag{2.7}$$

Ridge regression has been extensively studied and refined over time. [Swindel \(1976\)](#) re-examined ridge regression and introduced ridge estimators based on prior information, while [Marquardt and Snee \(1975\)](#) explored its connection to generalized inverse regression, presenting simulation results and three practical examples. Building on this, [McDonald and Galarneau \(1975\)](#) proposed two analytic methods to specify k , and [Kibria \(2003\)](#) further advanced the field by proposing new estimators based on the generalized ridge regression approach.

2.2.2 Liu Estimator

Quite a lot of literature is available on the selection of the biasing parameter k of the ridge estimator. However, it is a non-linear function of the biasing parameter and for this reason the choice of a suitable k remains difficult. [Liu \(1993\)](#) proposed another estimator, known as the Liu estimator (LE), which combined the advantages of the Stein estimator and RE. The Stein estimator was defined by [Stein et al. \(1956\)](#) as follows $\hat{\boldsymbol{\beta}}_S = c\hat{\boldsymbol{\beta}}_{OLS}$, where $0 < c < 1$.

The LE is defined as follows:

$$\hat{\beta}_L = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} + d\hat{\beta}_{OLS}), \quad 0 < d < 1 \quad (2.8)$$

It can be written in the form

$$\hat{\beta}_L = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{I} + d(\mathbf{X}'\mathbf{X})^{-1}) \mathbf{X}'\mathbf{Y}. \quad (2.9)$$

Using equation (2.4), we have the following.

$$\mathbf{A}_L = (\mathbf{\Lambda} + \mathbf{I})^{-1} (\mathbf{\Lambda} + d\mathbf{I}), \quad (2.10)$$

where d is the biasing parameter. The main advantage of LE over RE is that LE is a linear function of the biasing parameter d .

The Liu estimator has been further developed and refined in subsequent studies, showcasing its versatility in addressing multicollinearity. Akdeniz and Kaçiranlar (1995) proposed the almost unbiased generalized Liu estimator, providing an exact unbiased evaluation of bias and mean squared error. Building on this, Akdeniz and Erol (2003) compared the generalized Liu estimator with the generalized ridge regression estimator in terms of matrix mean squared error, highlighting its relative efficiency.

Later, Hubert and Wijekoon (2006) introduced a stochastic restricted Liu estimator, which demonstrated improved performance under certain constraints. More recently, Månsson et al. (2012) recommended the Liu estimator with the shrinkage parameter d to estimate logistic regression models when multicollinearity is present. These developments underscore the adaptability and robustness of the Liu estimator across diverse statistical frameworks.

2.2.3 Liu Type Estimator

Later, [Liu \(2003\)](#) explains that the RE estimator can be obtained by increasing $\mathbf{0} = k^{1/2}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ to (2.1) and then using the OLS estimator. He says that as k becomes larger, the distance between $k^{1/2}\boldsymbol{\beta}$ and $\mathbf{0}$ increases; therefore, the RR estimator becomes more biased. For this reason he increases $(-d/k^{1/2})\boldsymbol{\beta} = k^{1/2}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ to (2.1) and obtains the Liu-type estimator.

$$\hat{\boldsymbol{\beta}}_{LT} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} - d\tilde{\boldsymbol{\beta}}), \quad k > 0, \quad -\infty < d < \infty, \quad (2.11)$$

where $\tilde{\boldsymbol{\beta}}$ is any estimator of $\boldsymbol{\beta}$. We can choose a large k because we can adjust another parameter d to make the equation still fit.

1. When $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_O$, then $\hat{\boldsymbol{\beta}}_{LT} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{I} - d(\mathbf{X}'\mathbf{X})^{-1})\mathbf{X}'\mathbf{Y}$, use equation (2.4) and we can get

$$\mathbf{A}_{LTO} = \boldsymbol{\Lambda}(\boldsymbol{\Lambda} + k\mathbf{I})^{-1}(\mathbf{I} - d\boldsymbol{\Lambda}^{-1}). \quad (2.12)$$

2. When $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_R$, then $\hat{\boldsymbol{\beta}}_{LT} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{I} - d(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1})\mathbf{X}'\mathbf{Y}$, use equation (2.4) and we can get

$$\mathbf{A}_{LTR} = \boldsymbol{\Lambda}(\boldsymbol{\Lambda} + k\mathbf{I})^{-1}(\mathbf{I} - d(\boldsymbol{\Lambda} + k\mathbf{I})^{-1}). \quad (2.13)$$

[Liu \(2004\)](#) generalize to the random k case, showing that the Liu-type estimator improves ridge regression with k chosen by the Hoerl-Kennard formula

or the condition number for severe collinearity. [Li and Yang \(2012\)](#) introduce a new Liu-type estimator called a modified Liu estimator based on prior information for the vector of parameters in a linear regression model and discuss its properties.

2.2.4 Two-Parameter Estimator

When the collinearity is not severe, the Liu-type estimator produces negative k values. For this reason, when collinearity is moderate, [Liu \(2004\)](#) suggested using the estimate of k that [Hoerl and Kennard \(1970\)](#) proposed. However, he did not explain how to select the biasing parameter k and suggest the interval for d . With positive k values produced and the choice of d suggested by the form of the contraction estimator, [Özkale and Kaciranlar \(2007\)](#) introduced a Two-Parameter Estimator (TPE) as follows:

$$\hat{\beta}_{TP} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} + kd\hat{\beta}_O), \quad k > 0, 0 < d < 1. \quad (2.14)$$

It can be simplified to the following form,

$$\hat{\beta}_{TP} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1} (\mathbf{I} + kd(\mathbf{X}'\mathbf{X})^{-1})\mathbf{X}'\mathbf{Y}, \quad (2.15)$$

using equation [\(2.4\)](#), then

$$\mathbf{A}_{TP} = (\mathbf{\Lambda} + k\mathbf{I})^{-1} (\mathbf{\Lambda} + kd\mathbf{I}) \quad (2.16)$$

Özkale and Abbasi (2022) presents an iterative restricted OK estimator for generalized linear models, designed to address multicollinearity by incorporating exact linear restrictions on the parameters.

2.2.5 New Two-Parameter Estimator

Given the prior information about β in the form of linear stochastic restriction as follows: $(d-k)\hat{\beta}(k) = \beta + \varepsilon'$, where ε' is a $p \times 1$ random vector of disturbance with $E(\varepsilon') = 0$, $D(\varepsilon') = \sigma^2 \mathbf{I}$, and $E(\varepsilon \varepsilon') = 0$. Using the mixed estimation technique introduced by Theil (1963) and Theil and Goldberger (1992), Yang and Chang (2010) proposed a new biased estimator based on RE as follows:

$$\hat{\beta}_{NTP} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\beta}_R, \quad k > 0, \quad 0 < d < 1 \quad (2.17)$$

which is just replacing $\hat{\beta}_O$ in the Liu estimator with $\hat{\beta}_R$ if we write equation (2.9) as $\hat{\beta}_L = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\beta}_O$. It can be simplified to the following form,

$$\hat{\beta}_{NTP} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y} \quad (2.18)$$

using equation (2.4), then

$$\mathbf{A}_{NTP} = \mathbf{\Lambda}(\mathbf{\Lambda} + \mathbf{I})^{-1}(\mathbf{\Lambda} + d\mathbf{I})(\mathbf{\Lambda} + k\mathbf{I})^{-1}. \quad (2.19)$$

The OK and YC estimators, obtained by shrinking the OLS estimator using R and FT matrices, are sensitive to outliers in the y direction. To address this

issue, the OKM and YCM estimators, $\hat{\beta}_M(\text{OK})$ and $\hat{\beta}_M(\text{YC})$, are proposed to handle both multicollinearity and outliers in the y direction by [Ahmad and Aslam \(2022\)](#).

2.2.6 Modified Two-Parameter Estimator

[Dorugade \(2014b\)](#) proposed a modified TPE:

$$\hat{\beta}_{\text{MTP}} = \hat{\beta}^* + k(1 - d)(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1}\hat{\beta}^*, \quad k > 0, 0 < d < 1 \quad (2.20)$$

where

$$\hat{\beta}^* = (\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}. \quad (2.21)$$

This estimator was a sort of almost unbiased estimator (AUE), following the work of [Kadiyala \(1984\)](#), [Ohtani \(1986\)](#) and [Wu and Yang \(2013\)](#). Moreover, this estimator proposed after two-step estimation in which the initial estimator (2.21) was given following [Liu \(1993\)](#), [Kaçiranlar et al. \(1999\)](#), and [Yang and Chang \(2010\)](#). It can be written to the following form,

$$\hat{\beta}_{\text{MTP}} = (\mathbf{I} + k(1 - d)(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1})(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \quad (2.22)$$

using equation (2.4), then

$$\begin{aligned} \mathbf{A}_{\text{MTP}} &= \mathbf{\Lambda}(\mathbf{\Lambda} + kd\mathbf{I})^{-1}(\mathbf{I} + k(1 - d)(\mathbf{\Lambda} + kd\mathbf{I})^{-1}) \\ &= \mathbf{\Lambda}((\mathbf{\Lambda} + kd\mathbf{I})^{-1} + k(1 - d)(\mathbf{\Lambda} + kd\mathbf{I})^{-2}) \end{aligned} \quad (2.23)$$

2.2.7 Modified New Two-Parameter Estimator

Ahmad and Aslam (2022) proposed another biased estimator similar to the New Two-Parameter estimator. Instead of $\hat{\beta}_R$ in (2.17), they used the estimator (2.21) proposed by Dorugade (2014b). This estimator, called Modified New Two-Parameter Estimator (MNTPE), is given as follows:

$$\hat{\beta}_{\text{MNTP}} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})\hat{\beta}^*, \quad k > 0, \quad 0 < d < 1. \quad (2.24)$$

where k and d are two biasing parameters. It can be written to the following form,

$$\hat{\beta}_{\text{MNTP}} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}, \quad (2.25)$$

using equation (2.4), then

$$\mathbf{A}_{\text{MNTP}} = \mathbf{\Lambda}(\mathbf{\Lambda} + \mathbf{I})^{-1}(\mathbf{\Lambda} + d\mathbf{I})(\mathbf{\Lambda} + kd\mathbf{I})^{-1} \quad (2.26)$$

Akay et al. (2024) propose to tailor the Aslam-Ahmad estimator for the Cox proportional hazards model and define its properties.

Biased Estimators	Equations	GSLS Estimators
Ridge Estimator (1970)	$\hat{\beta}_R(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$ where $k > 0$	$\hat{\beta}_R(k) = \mathbf{P}\mathbf{A}_R\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_R = \mathbf{\Lambda}(\mathbf{\Lambda} + k\mathbf{I})^{-1}$
Liu Estimator (1993)	$\hat{\beta}_L = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} + d\hat{\beta}_{OLS})$ where $0 < d < 1$	$\hat{\beta}_L = \mathbf{P}\mathbf{A}_L\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_L = (\mathbf{\Lambda} + \mathbf{I})^{-1}(\mathbf{\Lambda} + d\mathbf{I})$
Liu Type OLS Estimator (2003)	$\hat{\beta}_{LTO} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{I} - d(\mathbf{X}'\mathbf{X})^{-1})\mathbf{X}'\mathbf{Y}$ where $k > 0, -\infty < d < \infty$	$\hat{\beta}_{LTO} = \mathbf{P}\mathbf{A}_{LTO}\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_{LTO} = \mathbf{\Lambda}(\mathbf{\Lambda} + k\mathbf{I})^{-1}(\mathbf{I} - d\mathbf{\Lambda}^{-1})$
Liu Type Ridge Estimator (2003)	$\hat{\beta}_{LTR} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{I} - d(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1})\mathbf{X}'\mathbf{Y}$ where $k > 0, -\infty < d < \infty$	$\hat{\beta}_{LTR} = \mathbf{P}\mathbf{A}_{LTR}\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_{LTR} = \mathbf{\Lambda}(\mathbf{\Lambda} + k\mathbf{I})^{-1}(\mathbf{I} - d(\mathbf{\Lambda} + k\mathbf{I})^{-1})$
Two-Parameter Estimator (2007)	$\hat{\beta}_{TP} = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} + kd\hat{\beta}_O)$ where $k > 0, 0 < d < 1$	$\hat{\beta}_{TP} = \mathbf{P}\mathbf{A}_{TP}\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_{TP} = (\mathbf{\Lambda} + k\mathbf{I})^{-1}(\mathbf{\Lambda} + kd\mathbf{I})$
New Two-Parameter Estimator (2010)	$\hat{\beta}_{NTP} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$ where $k > 0, 0 < d < 1$	$\hat{\beta}_{NTP} = \mathbf{P}\mathbf{A}_{NTP}\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_{NTP} = \mathbf{\Lambda}(\mathbf{\Lambda} + \mathbf{I})^{-1}(\mathbf{\Lambda} + d\mathbf{I})(\mathbf{\Lambda} + k\mathbf{I})^{-1}$
Modified Two-Parameter Estimator (2014b)	$\hat{\beta}_{MTP} = (\mathbf{I} + k(1 - d)(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1})(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$ where $k > 0, 0 < d < 1$	$\hat{\beta}_{MTP} = \mathbf{P}\mathbf{A}_{MTP}\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_{MTP} = \mathbf{\Lambda}((\mathbf{\Lambda} + kd\mathbf{I})^{-1} + k(1 - d)(\mathbf{\Lambda} + kd\mathbf{I})^{-2})$
Modified New Two-Parameter Estimator (2022)	$\hat{\beta}_{MNTP} = (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{X} + d\mathbf{I})(\mathbf{X}'\mathbf{X} + kd\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$ where $k > 0, 0 < d < 1$	$\hat{\beta}_{MNTP} = \mathbf{P}\mathbf{A}_{MNTP}\mathbf{P}'\hat{\beta}_O$ where $\mathbf{A}_{MNTP} = \mathbf{\Lambda}(\mathbf{\Lambda} + \mathbf{I})^{-1}(\mathbf{\Lambda} + d\mathbf{I})(\mathbf{\Lambda} + kd\mathbf{I})^{-1}$

Table 2.1: Summary of GSLS Estimators

Chapter 3

Admissibility of GSLS Estimators

3.1 Introduction

The concept of admissibility in statistics was first introduced by [Wald \(1949\)](#) as part of his foundational work in statistical decision theory. Admissibility was defined as a property of an estimator where no other estimator exists that has a uniformly better risk function under a given loss function. [Rao et al. \(1973\)](#) states that the Ordinary Least Squares (OLS) estimator is the Best Linear Unbiased Estimator (BLUE) under the MSE criterion. This result implicitly connects MSE with admissibility within the class of linear and unbiased estimators. However, it only applies to unbiased estimators and does not address the broader admissibility question, including biased estimators.

The focus on MSE-based admissibility became central in the work of [Stein \(1992\)](#) in the groundbreaking paper. [Marquardt \(1970\)](#) discusses the use of

biased estimators, specifically ridge regression, to address multicollinearity issues in linear regression models.

Let $\mathbf{b}_1$ and $\mathbf{b}_2$ be two estimators of $\boldsymbol{\beta}$. Then $\mathbf{b}_1$ is said to be better than $\mathbf{b}_2$, if for all $\boldsymbol{\beta}$ and σ^2 ,

$$\mathbb{E}[(\mathbf{b}_1 - \boldsymbol{\beta})'(\mathbf{b}_1 - \boldsymbol{\beta})] \leq \mathbb{E}[(\mathbf{b}_2 - \boldsymbol{\beta})'(\mathbf{b}_2 - \boldsymbol{\beta})].$$

An estimator $\mathbf{b}$ of $\boldsymbol{\beta}$ is said to be admissible in the class of all estimators of $\boldsymbol{\beta}$, if there exists no other estimator that is better than $\mathbf{b}$. To consider the admissibility of any GSLS estimator $\mathbf{b}_{GS}(\mathbf{A})$ in the class of all estimators of $\boldsymbol{\beta}$, we have the following theorem.

Theorem 3.1.1 (Wang (1990)). *A GSLS estimator $\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A})$ is admissible if and only if*

$$\text{rank}(\mathbf{I} - \mathbf{A}) \geq p - 2.$$

From Theorem 3.1.1, we know that the necessary and sufficient condition is as follows:

1. All eigenvalues of $\mathbf{A}$ are between 0 and 1;
2. At most two of them are equal to 1.

3.2 Admissibility

3.2.1 Admissibility of Ridge Estimator

The diagonal matrix for the ridge estimator is

$$\begin{aligned} \mathbf{A}_R &= \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_p+k} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p}{\lambda_p+k} \end{bmatrix}. \end{aligned}$$

For all $i \in \{1, 2, \dots, p\}$, we have

$$0 < \frac{\lambda_i}{\lambda_i + k} < 1,$$

since $k > 0, \lambda_i > 0$ and hence none of these elements is 1. Now we have the main result.

Corollary 1. *All Ridge estimators $\hat{\beta}_R$ are admissible.*

3.2.2 Admissibility of Liu Estimator

The diagonal matrix for Liu estimator is

$$\begin{aligned} \mathbf{A}_L &= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_p+1} \end{bmatrix} \begin{bmatrix} 1 + \frac{d}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 + \frac{d}{\lambda_p} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1+d}{\lambda_1+1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\lambda_p+d}{\lambda_p+1} \end{bmatrix}. \end{aligned}$$

For all $i \in \{1, 2, \dots, p\}$, we have

$$0 < \frac{\lambda_i + d}{\lambda_i + 1} < 1,$$

since $0 < d < 1$ and $\lambda_i > 0$. Hence, none of these elements is equal to 1.

Therefore, we obtain the main result.

Corollary 2. *All Ridge estimators $\hat{\beta}_L$ are admissible.*

3.2.3 Admissibility of Liu Type Estimator

The diagonal matrix of Liu type OLS estimator is

$$\begin{aligned} \mathbf{A}_{LTO} &= \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_p+k} \end{bmatrix} \begin{bmatrix} 1 - \frac{d}{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \frac{d}{\lambda_p} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1-d}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p-d}{\lambda_p+k} \end{bmatrix}. \end{aligned}$$

For the first condition of admissibility, we know that

$$0 \leq \frac{\lambda_i - d}{\lambda_i + k} \leq 1, \quad \forall i \in \{1, 2, \dots, p\} \quad (3.1)$$

The condition $-k \leq d \leq \lambda_i$ is simplified for all $i \in \{1, 2, \dots, p\}$, we obtain $d \leq \lambda_p = \lambda_{\min}$ if $d > 0$ and $k \geq -d$ if $d \leq 0$.

For the second condition, if $\frac{\lambda_i-d}{\lambda_i+k} = 1, \forall i \in \{1, 2, \dots, p\}$, then it implies $k = -d$.

However, if p , the number of coefficients is greater than 2, all these diagonal elements can equal 1, resulting in Condition 2 not being satisfied. Thus, we obtain the result.

Corollary 3. *A Liu-type OLS estimator $\hat{\beta}_{LTO}$ is admissible if and only if $-k \leq d \leq \lambda_p$ when $p \leq 2$ or $-k < d \leq \lambda_p$ when $p > 2$.*

The diagonal matrix of Liu type Ridge estimator is

$$\begin{aligned} \mathbf{A}_{LTR} &= \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_p+k} \end{bmatrix} \begin{bmatrix} 1 - \frac{d}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \frac{d}{\lambda_p+k} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1(\lambda_1+k-d)}{(\lambda_1+k)^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p(\lambda_p+k-d)}{(\lambda_p+k)^2} \end{bmatrix}. \end{aligned}$$

For the first condition of admissibility, we know that $\forall i \in \{1, 2, \dots, p\}$,

$$0 \leq \frac{\lambda_i(\lambda_i + k - d)}{(\lambda_i + k)^2} \leq 1 \quad (3.2)$$

For d , the inequality reduces to $d \leq k + \lambda$ and $d \geq \frac{-k^2}{\lambda} - k$. As $k + \lambda > 0$ and $\frac{-k^2}{\lambda} - k < 0$, combining this we deduce that $-k(1 + \frac{k}{\lambda_1}) \leq d \leq k + \lambda_p$.

For the second condition, if $\frac{\lambda(\lambda+k-d)}{(\lambda+k)^2} = 1$, then it implies $\lambda = \frac{-k^2}{k+d}$. We can get the result.

Corollary 4. *A Liu type Ridge estimator $\hat{\beta}_{LTR}$ is admissible if and only if $-k(1 + \frac{k}{\lambda_1}) \leq d \leq k + \lambda_p$ and at most 2 λ make $d = -k(1 + \frac{k}{\lambda})$.*

3.2.4 Admissibility of Two-Parameter Estimator

The diagonal matrix of Two-Parameter Estimator is

$$\begin{aligned} \mathbf{A}_{TP} &= \begin{bmatrix} \frac{1}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_p+k} \end{bmatrix} \begin{bmatrix} \lambda_1 + kd & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p + kd \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1+kd}{\lambda_1+k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p+kd}{\lambda_p+k} \end{bmatrix}. \end{aligned}$$

For the first condition of admissibility, we know that $\forall i \in \{1, 2, \dots, p\}$,

$$0 \leq \frac{\lambda_i + kd}{\lambda_i + k} \leq 1 \quad (3.3)$$

For k , the inequality simplifies to $k \geq \frac{-\lambda}{d}$, which is satisfied because $k > 0$.

For d , the inequality reduces to $\frac{-\lambda}{k} \leq d \leq 1$. It also satisfies since $0 < d < 1$.

For the second condition, if $\frac{\lambda+kd}{\lambda+k} = 1$, then it implies $d = 1$. This is impossible since $0 < d < 1$. So none of these diagonal elements can equal 1 and we can get the result.

Corollary 5. *All Two-Parameter Estimators are admissible.*

3.2.5 Admissibility of Modified Two-Parameter Estimator

The diagonal matrix of Modified Two-Parameter Estimator is

$$\begin{aligned} \mathbf{A}_{MTP} &= \begin{bmatrix} \frac{\lambda_1}{\lambda_1+kd} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p}{\lambda_p+kd} \end{bmatrix} + \begin{bmatrix} \frac{k(1-d)}{(\lambda_1+kd)^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{k(1-d)}{(\lambda_p+kd)^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1(\lambda_1+kd)+k(1-d)}{(\lambda_1+kd)^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p(\lambda_p+kd)+k(1-d)}{(\lambda_p+kd)^2} \end{bmatrix}. \end{aligned}$$

For the first condition of admissibility, we know that $\forall i \in \{1, 2, \dots, p\}$,

$$0 \leq \frac{\lambda_i(\lambda_i + kd) + k(1-d)}{(\lambda_i + kd)^2} \leq 1 \quad (3.4)$$

For k , the inequality simplifies to $k \geq \frac{1}{d^2}(1 - d(1 + \lambda_i))$, $\forall i \in \{1, 2, \dots, p\}$, we can get

1. If $d \geq \frac{1}{1+\lambda_p}$, then $k > 0$.
2. If $0 < d < \frac{1}{1+\lambda_p}$, then $k \geq \frac{1-(1+\lambda_p)d}{d^2}$.

For the second condition, if $\frac{\lambda(\lambda+kd)+k(1-d)}{(\lambda+kd)^2} = 1$, then it implies $\lambda = \frac{1-d-kd^2}{d}$. we

can get the result.

Corollary 6. *A Modified Two-Parameter estimator $\hat{\beta}_{MTP}$ is admissible if and only if $k \geq \frac{1-(1+\lambda_p)d}{d^2}$ when $0 < d < \frac{1}{1+\lambda_p}$ or $d \geq \frac{1}{1+\lambda_p}$ and at most 2 λ make $k = \frac{1-(1+\lambda)d}{d^2}$.*

3.2.6 Admissibility of Modified New Two-Parameter Estimator

The diagonal matrix of Modified New Two-Parameter Estimator is

$$\begin{aligned} \mathbf{A}_{MNTPE} &= \begin{bmatrix} \frac{\lambda_1}{\lambda_1+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p}{\lambda_p+1} \end{bmatrix} \begin{bmatrix} \frac{\lambda_1+d}{\lambda_1+kd} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p+d}{\lambda_p+kd} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1(\lambda_1+d)}{(\lambda_1+1)(\lambda_1+kd)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_p(\lambda_p+d)}{(\lambda_p+1)(\lambda_p+kd)} \end{bmatrix}. \end{aligned}$$

For the first condition of admissibility, we know that $\forall i \in \{1, 2, \dots, p\}$,

$$0 \leq \frac{\lambda_i(\lambda_i + d)}{(\lambda_i + 1)(\lambda_i + kd)} \leq 1 \quad (3.5)$$

For k , the inequality simplifies to $k \geq \frac{\lambda(d-1)}{d(\lambda+1)}$, which is satisfied because $\frac{\lambda(d-1)}{d(\lambda+1)} < 0$, so $k > 0$.

For d , the inequality reduces to $d \leq \frac{\lambda}{\lambda-k(1+\lambda)}$. It also satisfies since $\frac{\lambda}{\lambda-k(1+\lambda)} > 1$, so $0 < d < 1$.

For the second condition, if $\frac{\lambda(\lambda+d)}{(\lambda+1)(\lambda+kd)} = 1$, then it implies $\lambda = \frac{kd}{d-kd-1}$. This is a contradiction, as $\lambda > 0$, $\frac{kd}{d-kd-1} < 0$. So none of these diagonal elements can equal 1 and the result follows.

Corollary 7. *All Modified New Two-Parameter Estimators are admissible.*

Chapter 4

Optimal Estimators

4.1 Optimal GSLS estimator

Numerous biased estimators have been proposed in the literature; however, the issue of calculating the optimal estimators remains a significant challenge. Although admissibility criteria can help narrow the range of possible parameter values, determining the optimal parameters remains an open problem. To address these limitations, we introduce a novel estimator, termed the Optimal Shrunk Least Squares estimator (OSLS). This estimator possesses several key advantages. First, it achieves optimal performance by using parameter values determined under specific optimality conditions. Second, it is a member of the Generalized Shrunk Least Squares (GSLS) class, which ensures theoretical consistency and robustness. Finally, it eliminates the need for manual parameter selection, as the parameters are directly derived from the design matrix and the observed data.

The primary objective of this section is to explore the locally optimal estimator within the class of GSLS estimators and subsequently compare the elements of our estimator, corresponding to the optimal values of k , d or both k and d , with those obtained from previously proposed biasing parameters.

The mean square error (MSE) criterion is a widely used measure to evaluate the quality of an estimator. For an estimator $\hat{\beta}$ of β , the MSE of β is formally defined as:

$$\begin{aligned} MSE(\hat{\beta}) &= \mathbb{E}[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] \\ &= \text{tr}(\text{Cov}(\hat{\beta})) + \text{Bias}(\hat{\beta})'\text{Bias}(\hat{\beta}) \end{aligned} \tag{4.1}$$

If $\hat{\beta}_{GS}(\mathbf{A})$ is a GSLS estimator, the covariance matrix and bias of $\hat{\beta}_{GS}(\mathbf{A})$ are given by, respectively:

$$\begin{aligned} \text{Cov}(\hat{\beta}_{GS}(\mathbf{A})) &= \text{Cov}(\mathbf{PAP}'\hat{\beta}_{OLS}) \\ &= \mathbf{PAP}'\text{Cov}(\hat{\beta}_{OLS})\mathbf{PAP}' \\ &= \sigma^2\mathbf{PAP}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{PAP}' \\ &= \sigma^2\mathbf{PA}^2\mathbf{\Lambda}^{-1}\mathbf{P}', \end{aligned} \tag{4.2}$$

and

$$\begin{aligned}
\text{Bias}(\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A})) &= \mathbb{E}[\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A})] - \boldsymbol{\beta} \\
&= \mathbf{P}\mathbf{A}\mathbf{P}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{Y}] - \boldsymbol{\beta} \\
&= \mathbf{P}\mathbf{A}\mathbf{P}'\boldsymbol{\beta} - \boldsymbol{\beta} \\
&= \mathbf{P}(\mathbf{A} - \mathbf{I})\mathbf{P}'\boldsymbol{\beta}.
\end{aligned} \tag{4.3}$$

Combing (4.2) and (4.3),

$$\begin{aligned}
MSE(\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A})) &= \text{tr}(\text{Cov}(\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A}))) + (\text{Bias}(\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A})))'(\text{Bias}(\hat{\boldsymbol{\beta}}_{GS}(\mathbf{A}))) \\
&= \sigma^2 \text{tr}(\mathbf{A}^2 \boldsymbol{\Lambda}^{-1}) + \boldsymbol{\beta}' \mathbf{P}(\mathbf{I} - \mathbf{A})^2 \mathbf{P}' \boldsymbol{\beta} \\
&= \sum_{i=1}^p \left[\frac{\sigma^2 a_i^2}{\lambda_i} + \delta_i^2 (1 - a_i)^2 \right] \\
&= \sum_{i=1}^p D_i(a_i),
\end{aligned} \tag{4.4}$$

where δ_i is the i th entry of the vector $\boldsymbol{\delta} = \mathbf{P}'\boldsymbol{\beta}$, $i = 1, 2, \dots, p$.

Taking $\frac{\partial}{\partial a_i} \sum_{i=1}^p D_i(a_i) = 0$, we have

$$a_i^{\text{opt}} = \frac{\lambda_i \delta_i^2}{\lambda_i \delta_i^2 + \sigma^2}. \tag{4.5}$$

Note that for $i \neq j$, we have

$$\frac{\partial^2}{\partial a_i \partial a_j} \sum_{i=1}^p D_i(a_i) = 0, \tag{4.6}$$

and,

$$\frac{\partial^2}{\partial a_i^2} \sum_{i=1}^p D_i(a_i) = \frac{2\sigma^2}{\lambda_i} + 2\delta_i^2 > 0. \quad (4.7)$$

Thus, its Hessian matrix is positive definite, that is, $(a_1^{opt}, a_2^{opt}, \dots, a_p^{opt})$ is the point of optimal value that makes $MSE(\hat{\beta}_{GS}(\mathbf{A}))$ reach the minimum value.

4.2 Optimal estimators

In practical applications, the determination of the optimal values of the biasing parameters k and d plays a crucial role in improving estimator performance. The choice of appropriate parameter values is essential for achieving desirable statistical properties. Various alternative estimators have been proposed in the literature, each involving different forms of biasing parameters. In this study, we aim to compare these estimators with our proposed parameter a_i , and further explore the possibility of incorporating their optimal parameters into our framework through the lens of a_i .

4.2.1 Optimal ridge estimator

Hoerl et al. (1975) defined $(\mathbf{L}_i^*)^2 = (\hat{\boldsymbol{\delta}}^* - \boldsymbol{\delta})'(\hat{\boldsymbol{\delta}}^* - \boldsymbol{\delta})$, it follows that the proposed optimal values of k is given by taking derivative of $E(\mathbf{L}_i^*)^2$, known as MSE:

$$\hat{k}_{opt} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\delta}}' \hat{\boldsymbol{\delta}}} = \frac{\hat{\sigma}^2}{\sum_{i=1}^p \hat{\delta}_i^2 / p} = \hat{\sigma}^2 / \overline{\hat{\delta}^2}.$$

From (2.6), we obtain the elements of $\mathbf{A}_R$ and equate them to (5.4):

$$\frac{\lambda_i}{\lambda_i + k} = \frac{\lambda_i}{\lambda_i + \hat{\sigma}^2 / \hat{\delta}_i^2}. \quad (4.8)$$

Solving for k , we obtain the optimal value:

$$\hat{k}_{opt} = \frac{\hat{\sigma}^2}{\hat{\delta}_i^2}.$$

This result is unusual and incorrect because k is a constant, while $\hat{\delta}_i$ varies with different i . The issue arises from taking the partial derivative in an unconstrained problem with respect to a_i , which is a function of k . In the ridge estimator, a_i is expressed as $\frac{\lambda_i}{\lambda_i + k}$. To determine k_{opt} , we must impose the constraint $a_i = \frac{\lambda_i}{\lambda_i + k}$, substitute it in the MSE, and then take the derivative.

Alternatively, we can rewrite the proposed parameter k in the form

$$\hat{a}_i^{opt} = \frac{\lambda_i}{\lambda_i + \hat{\sigma}^2 / \hat{\delta}^2}.$$

Here, the denominator uses the mean, which implies a lack of detailed information to determine the optimal a_i^{opt} for each individual i .

4.2.2 Optimal Liu Estimator

From (2.10), we obtain the elements of $\mathbf{A}_L$:

$$a_i = \frac{\lambda_i + d}{\lambda_i + 1}.$$

From Liu (1993) , we can see that $MSE(\hat{\delta}_d)$ is minimized at

$$d_{MM} = \frac{\sum_{i=1}^p \frac{\delta_i^2 - \sigma^2}{(\lambda_i + 1)^2}}{\sum_{i=1}^p \frac{\sigma^2 + \lambda_i \delta_i^2}{\lambda_i (\lambda_i + 1)^2}}.$$

By using *CL* and *GCV* criterion, Liu also gets

$$\hat{d}_{CL} = 1 - \hat{\sigma}^2 \left(\frac{\sum_{i=1}^p \frac{1}{\lambda_i + 1}}{\sum_{i=1}^p \frac{\lambda_i \delta_i^2}{(\lambda_i + 1)^2}} \right),$$

and

$$\hat{d}_{MC} = 1 - \frac{\sum_{i=1}^p \frac{\hat{\delta}_i^2}{\lambda_i + 1} - \sqrt{\left(\sum_{i=1}^p \frac{\hat{\delta}_i}{\lambda_i + 1} \right)^2 - \hat{\sigma}^2 \sum_{i=1}^p \frac{\hat{\delta}_i^2}{(\lambda_i + 1)^2} \cdot \sum_{i=1}^p \frac{1}{\lambda_i}}}{\sum_{i=1}^p \frac{\hat{\delta}_i^2}{(\lambda_i + 1)^2}},$$

respectively.

Since $\boldsymbol{\delta}_L = \mathbf{P}'\boldsymbol{\beta}_L = \mathbf{P}'(\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} + d\hat{\boldsymbol{\beta}}_O)$, so d and $\boldsymbol{\delta}$ depend on each other. Liu applies the iterative method developed in [Hoerl and Kennard \(1976\)](#) to obtain the selection of d_{MM} , d_{CL} , d_{MC} . He found that the CL and GCV criterion gives the rather larger d whereas the minimum MSE estimate results in the smallest d . Here we just show the example for d_{MM} to estimate d_{opt} .

$$E(\hat{\boldsymbol{\delta}}) = E\left(\mathbf{P}'\hat{\boldsymbol{\beta}}_O\right) = \mathbf{P}'\boldsymbol{\beta} = \boldsymbol{\delta},$$

and

$$\text{Var}(\hat{\boldsymbol{\delta}}) = \text{Var}\left(\mathbf{P}'\hat{\boldsymbol{\beta}}_O\right) = \mathbf{P}'\text{Cov}(\hat{\boldsymbol{\beta}}_O)\mathbf{P} = \mathbf{P}'\left(\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)\mathbf{P} = \sigma^2\boldsymbol{\Lambda}^{-1}.$$

So $\hat{\delta}_i \sim N\left(\delta_i, \frac{\sigma^2}{\lambda_i}\right)$ and $E(\hat{\delta}_i^2 - \frac{\hat{\sigma}^2}{\lambda_i}) = \delta_i^2$.

- **Step 1:** Substituting δ_i^2 and σ^2 by their unbiased estimates $\hat{\delta}_i^2 - \frac{\hat{\sigma}^2}{\lambda_i}$ and $\hat{\sigma}^2$, get the estimate of d :

$$\hat{d}_0 = 1 - \hat{\sigma}^2 \left(\frac{\sum_{i=1}^p \frac{1}{\lambda_i(\lambda_i+1)}}{\sum_{i=1}^p \frac{\hat{\delta}_i^2}{(\lambda_i+1)^2}} \right).$$

- **Step 2:** Estimate $\boldsymbol{\delta}_{L1}$ by using $\hat{d}_0$, then estimate $\hat{d}_1$ by using $\boldsymbol{\delta}_{L1} \dots$

- **Step 3:** The procedure ends if $\frac{\hat{d}_j - \hat{d}_{j+1}}{\hat{d}_j} < 20T^{-1.3}$, where $T =$

$\text{tr}(\mathbf{X}'\mathbf{X})^{-1}/p$. Then we use $\hat{\boldsymbol{\delta}}(d_j)$ to estimate $\boldsymbol{\delta}$. So $d_{opt} = \hat{d}_j$.

So we can get

$$\hat{a}_i^{opt} = \frac{\lambda_i + \hat{d}_j}{\lambda_i + 1}.$$

4.2.3 Optimal Liu Type Estimator

From (2.11), we obtain the elements of $\mathbf{A}_{LT}$:

$$a_i = \frac{\lambda_i - d}{\lambda_i + k}.$$

The selection of k is based on the recommendation of [Belsley et al. \(1980\)](#).

The condition number is defined as $\kappa = \left(\frac{\text{largest eigenvalue}}{\text{smallest eigenvalue}} \right)^{1/2}$, κ around 10 indicate weak dependence, condition numbers between 30 and 100 indicate moderate to strong dependence, and numbers greater than 100 suggest serious collinearity problems. To address this, Liu chooses k such that the condition number of $\mathbf{X}'\mathbf{X} + k\mathbf{I}$ is reduced to 10. After selecting

$$\hat{k} = \frac{\lambda_1 - 100 \cdot \lambda_p}{99},$$

we treat k as a fixed value. Then, using the Mean Squared Error (MSE) formula, which is a quadratic equation, we aim to find the minimizer d while focusing on two choices: RE and OLS. The expression for $\hat{d}$ is given by:

$$\hat{d} = \frac{\sum_{i=1}^p \frac{(\hat{\sigma}^2 - \hat{k}\hat{\delta}_i^2)}{(\lambda_i + \hat{k})^2}}{\sum_{i=1}^p \frac{(\lambda_i\hat{\delta}_i^2 + \hat{\sigma}^2)}{\lambda_i(\lambda_i + \hat{k})^2}} \quad (\tilde{\beta} = \beta_R),$$

or

$$\hat{d} = \frac{\sum_{i=1}^p \frac{\lambda_i(\hat{\sigma}^2 - \hat{k}\hat{\delta}_i^2)}{(\lambda_i + \hat{k})^3}}{\sum_{i=1}^p \frac{\lambda_i(\lambda_i\hat{\delta}_i^2 + \hat{\sigma}^2)}{(\lambda_i + \hat{k})^4}} \quad (\tilde{\beta} = \beta_O).$$

Following the same way from the Liu estimator, the iterative procedure is

- **Step 1:** Substituting δ_i^2 and σ^2 by ridge estimates $\boldsymbol{\delta}_R = \mathbf{P}'\boldsymbol{\beta}_R = \mathbf{P}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$ and $\hat{\sigma}_R^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)}{n - p}$ get the estimate of d : $\hat{d}_0$.
- **Step 2:** Estimate $\boldsymbol{\delta}_{kd} = \mathbf{P}'\boldsymbol{\beta}_{LT} = \mathbf{P}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{Y} - d\tilde{\boldsymbol{\beta}})$ by using $\hat{d}_0, \hat{k}$; then estimate $\hat{d}_1$ by using $\hat{\boldsymbol{\delta}}_{kd1} \dots$
- **Step 3:** The procedure terminates if $\frac{\hat{d}_j - \hat{d}_{j+1}}{\hat{d}_j} < 20 T^{-1.3}$, where $T = \text{tr}(\mathbf{X}'\mathbf{X})^{-1}/p$. Then we use $\hat{\boldsymbol{\delta}}(d_{kdj})$ to estimate $\boldsymbol{\delta}$. So $d_{opt} = \hat{d}_{kdj}$.

We can get

$$\hat{a}_i^{kdj} = \frac{\lambda_i - \hat{d}_{opt}}{\lambda_i + \hat{k}}$$

where

$$\hat{d}_{opt} = \frac{\sum_{i=1}^p \frac{(\hat{\sigma}_R^2 - \hat{k}\hat{\delta}_{Ri}(d_{kdj})^2)}{(\lambda_i + \hat{k})^2}}{\sum_{i=1}^p \frac{(\lambda_i\hat{\delta}_i(d_{kdj})^2 + \hat{\sigma}^2)}{\lambda_i(\lambda_i + \hat{k})^2}} (\tilde{\beta} = \beta_R),$$

or

$$\frac{\sum_{i=1}^p \frac{\lambda_i(\hat{\sigma}_R^2 - \hat{k}\hat{\delta}_{Ri}(d_{kdj})^2)}{(\lambda_i + \hat{k})^3}}{\sum_{i=1}^p \frac{\lambda_i(\lambda_i\hat{\delta}_i(d_{kdj})^2 + \hat{\sigma}^2)}{(\lambda_i + \hat{k})^4}} (\tilde{\beta} = \beta_O).$$

4.2.4 Optimal Two-Parameter Estimator

From (2.16), we obtain the elements of $\mathbf{A}_{TP}$:

$$a_i = \frac{\lambda_i + kd}{\lambda_i + k}.$$

Hoerl et al. (1975) proposed an estimator of k by employing the harmonic mean of k values derived from Hoerl and Kennard (1970). Furthermore, Kibria (2003) extended this work by proposing estimators of k based on the arithmetic and geometric means of the values of k . Following this methodology, Özkale and Kaciranlar (2007) propose the estimators of k as follows:

$$\hat{k}_{HM} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \left[\hat{\delta}_i^2 - d \left(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\delta}_i^2 \right) \right]},$$

$$\hat{k}_{\text{AM}} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2}{\hat{\delta}_i^2 - d \left(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\delta}_i^2 \right)},$$

$$\hat{k}_{\text{GM}} = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^p \left[\hat{\delta}_i^2 - d \left(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\delta}_i^2 \right) \right] \right)^{1/p}}.$$

After minimizing MSE and replacing it by the unbiased estimators, we obtain the optimal estimator of d for a fixed value of k by as

$$\hat{d} = \frac{\sum_{i=1}^p \frac{k \hat{\delta}_i^2 - \hat{\sigma}^2}{(\lambda_i + k)^2}}{\sum_{i=1}^p \frac{k(\hat{\sigma}^2 + \hat{\delta}_i^2 \lambda_i)}{\lambda_i(\lambda_i + k)^2}}. \quad (4.9)$$

To make k always positive, they get the range for

$$d < \min \left\{ \frac{\delta_i^2}{\frac{\sigma^2}{\lambda_i} + \delta_i^2} \right\}. \quad (4.10)$$

The selection of the estimators of the parameters d and k in $\hat{\beta}(k, d)$ can be obtained by applying the following iterative method:

- **Step 1.** Calculate $\hat{d}$ from (4.10) using unbiased estimates $\hat{\delta}_i^2 - \frac{\hat{\sigma}^2}{\lambda_i}$ and $\hat{\sigma}^2$,

- **Step 2.** Estimate $\hat{k}_{\text{HM}}$, $\hat{k}_{\text{AM}}$, or $\hat{k}_{\text{GM}}$ by using $\hat{d}$ in Step 1.
- **Step 3.** Obtain $\hat{d}_{\text{opt}}$ from (4.9) by using the estimators $\hat{k}_{\text{HM}}$, $\hat{k}_{\text{AM}}$, or $\hat{k}_{\text{GM}}$ in Step 2.
- **Step 4.** If $\hat{d}_{\text{opt}}$ is negative, use $\hat{d}_{\text{opt}} = \hat{d}$. Note that $\hat{d}_{\text{opt}}$ in (4.9) is always less than one, but it can be smaller than zero. Further, d in (4.10) is always less than one and bigger than zero.

We can get

$$\hat{a}_i^{\text{opt}} = \frac{\lambda_i + \hat{k}\hat{d}_{\text{opt}}}{\lambda_i + \hat{k}}$$

4.2.5 Optimal New Two-Parameter Estimator

From (2.19), we obtain the elements of $\mathbf{A}_{MTP}$:

$$a_i = \frac{\lambda_i(\lambda_i + d)}{(\lambda_i + 1)(\lambda_i + k)}.$$

Using the MSE method, differentiate it with respect to k and d , replace σ^2 and δ_i with unbiased estimates, Yang and Chang (2010) obtain

$$\hat{k} = \frac{\hat{\sigma}^2(\lambda_i + d) - (1 - d)\lambda_i\hat{\delta}_i^2}{(\lambda_i + 1)\hat{\delta}_i^2}$$

and

$$\hat{d}_{\text{opt}} = \frac{\sum_{i=1}^p \frac{((k+1)\lambda_i + k)\lambda_i \hat{\delta}_i^2 - \lambda_i^2 \hat{\sigma}^2}{(\lambda_i + 1)^2 (\lambda_i + k)^2}}{\sum_{i=1}^p \frac{(\hat{\sigma}^2 + \lambda_i \hat{\delta}_i^2) \lambda_i}{(\lambda_i + 1)^2 (\lambda_i + k)^2}}.$$

For k , they use the arithmetic and geometric means value of k , which are

$$\hat{k}_{\text{AM}} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2 (\lambda_i + d) - (1 - d) \lambda_i \hat{\delta}_i^2}{(\lambda_i + 1) \hat{\delta}_i^2},$$

$$\hat{k}_{\text{GM}} = \left(\prod_{i=1}^p \frac{\hat{\sigma}^2 (\lambda_i + d) - (1 - d) \lambda_i \hat{\delta}_i^2}{(\lambda_i + 1) \hat{\delta}_i^2} \right)^{\frac{1}{p}}.$$

We observe that k and d are interdependent. However, in the practical examples presented in the paper, there is no explicit discussion on how to select d ; instead, it is simply fixed at 0.95, then use it to calculate $\hat{k}_{\text{AM}}$ and $\hat{k}_{\text{GM}}$.

4.2.6 Optimal Modified New Two-Parameter Estimator

From (2.26), we obtain the elements of $\mathbf{A}_{MNTP}$:

$$a_i = \frac{\lambda_i (\lambda_i + d)}{(\lambda_i + 1) (\lambda_i + kd)}.$$

Using the MSE method, differentiate it with respect to k and d , replace σ^2 and δ_i with unbiased estimates, [Ahmad and Aslam \(2022\)](#) obtain

$$\hat{k} = \frac{\hat{\sigma}^2(\lambda_i + d) - (1 - d)\lambda_i\hat{\delta}_i^2}{d(\lambda_i + 1)\hat{\delta}_i^2}$$

and

$$\hat{d} = \frac{\sum_{i=1}^p \lambda_i(\hat{\delta}_i^2 - \hat{\sigma}^2)}{\sum_{i=1}^p (\lambda_i\hat{\delta}_i^2 + \hat{\sigma}^2 - k\lambda_i\hat{\delta}_i^2)}.$$

For k , they use the arithmetic, geometric and harmonic means value of k , which are

$$\hat{k}_{\text{AM}} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2(\lambda_i + d) - (1 - d)\lambda_i\hat{\delta}_i^2}{d(\lambda_i + 1)\hat{\delta}_i^2};$$

$$\hat{k}_{\text{GM}} = \left(\prod_{i=1}^p \frac{\hat{\sigma}^2(\lambda_i + d) - (1 - d)\lambda_i\hat{\delta}_i^2}{d(\lambda_i + 1)\hat{\delta}_i^2} \right)^{\frac{1}{p}};$$

$$\hat{k}_{\text{HM}} = p \sum_{i=1}^p \frac{\hat{\sigma}^2(\lambda_i + d) - (1 - d)\lambda_i\hat{\delta}_i^2}{d(\lambda_i + 1)\hat{\delta}_i^2}.$$

Follow the same approach as used by [Özkale and Kaciranlar \(2007\)](#) to make k always positive, they get the range for d :

$$d > \max \left[\frac{(\delta_i^2 - \sigma^2)\lambda_i}{(\sigma^2 + \lambda_i\delta_i^2)} \right].$$

Now present the following algorithm:

- **Step 1:** Obtain $\hat{d}$ after replacing σ^2 and δ_i with their unbiased estimates as stated earlier.
- **Step 2:** Compute $\hat{k}_{\text{opt}}$ using the above computed $\hat{d}$.
- **Step 3:** Compute $\hat{d}_{\text{opt}}$ using $\hat{k}_{\text{opt}}$ from Step 2.
- **Step 4:** Finally, if $\hat{d}_{\text{opt}} < 0$, consider $\hat{d}_{\text{opt}} = \hat{d}$.

Chapter 5

Numerical Calculation of the Optimal Estimators

5.1 Comparisons between GSLS estimators

In this subsection, we analyze and compare two GSLS estimators, $\hat{\beta}_{GS}(\mathbf{A})$ and $\hat{\beta}_{GS}(\mathbf{B})$, under the Mean Squared Error (MSE) criterion.

For two GSLS estimators $\hat{\beta}_{GS}(\mathbf{A})$ and $\hat{\beta}_{GS}(\mathbf{B})$, the MSEs are given by:

$$\begin{aligned} \text{MSE}(\hat{\beta}_{GS}(\mathbf{A})) &= \sigma^2 \text{trace}(\mathbf{A}^2 \mathbf{\Lambda}^{-1}) + \boldsymbol{\beta}' \mathbf{P} (\mathbf{I} - \mathbf{A})^2 \mathbf{P}' \boldsymbol{\beta} \\ &= \sigma^2 \sum_{i=1}^p \frac{a_i^2}{\lambda_i} + \sum_{i=1}^p \delta_i^2 (1 - a_i)^2 \end{aligned} \tag{5.1}$$

where a_i represents a specific parameter associated with the id matrix, and λ_i and δ_i are respective quantities related to eigenvalue matrix and estimator.

Similarly, for the estimator $\hat{\beta}_{GS}(\mathbf{B})$, we have:

$$\begin{aligned} \text{MSE}(\hat{\beta}_{GS}(\mathbf{B})) &= \sigma^2 \text{trace}(\mathbf{B}^2 \mathbf{\Lambda}^{-1}) + \beta' \mathbf{P}(\mathbf{I} - \mathbf{B})^2 \mathbf{P}' \beta \\ &= \sigma^2 \sum_{i=1}^p \frac{b_i^2}{\lambda_i} + \sum_{i=1}^p \delta_i^2 (1 - b_i)^2 \end{aligned} \quad (5.2)$$

Using equations (5.1) and (5.2), we can find the difference between the MSEs:

$$\begin{aligned} \text{MSE}(\hat{\beta}_{GS}(\mathbf{B})) - \text{MSE}(\hat{\beta}_{GS}(\mathbf{A})) &= \sigma^2 \sum_{i=1}^p \left[\frac{b_i^2 - a_i^2}{\lambda_i} \right] + \sum_{i=1}^p \delta_i^2 [(1 - b_i)^2 - (1 - a_i)^2] \\ &= \sum_{i=1}^p \left[\frac{\sigma^2}{\lambda_i} (b_i + a_i)(b_i - a_i) + \delta_i^2 (a_i - b_i)^2 \right] \\ &= \sum_{i=1}^p \left[\left(\frac{\lambda_i \delta_i^2 + \sigma^2}{\lambda_i} \right) (b_i - a_i)(a_i + b_i - 2a_i^{\text{opt}}) \right]. \end{aligned} \quad (5.3)$$

Thus, we obtain the following result.

Theorem 5.1.1 (Comparison Theorem). *For any two GSLS estimators $\hat{\beta}_{GS}(\mathbf{A})$ and $\hat{\beta}_{GS}(\mathbf{B})$, the MSE of $\hat{\beta}_{GS}(\mathbf{A})$ is less than or equal to the MSE of $\hat{\beta}_{GS}(\mathbf{B})$, i.e.,*

$$\text{MSE}(\hat{\beta}_{GS}(\mathbf{A})) \leq \text{MSE}(\hat{\beta}_{GS}(\mathbf{B})),$$

if either of the following conditions is satisfied:

$$(a) \quad 2a_i^{opt} - b_i \leq a_i \leq b_i, \quad i = 1, 2, \dots, p;$$

$$(b) \quad b_i \leq a_i \leq 2a_i^{opt} - b_i, \quad i = 1, 2, \dots, p.$$

If we make $\mathbf{B} = \mathbf{I}$, since a_i is between 0 and 1, using condition (b) we can get the comparison between any GSLS estimator and the OLS estimator.

Corollary 8. *Let $\hat{\beta}_{GS}(\mathbf{A})$ be a GSLS estimator and $\mathbf{A} \neq \mathbf{I}$. If*

$$\frac{\lambda_i \delta_i^2 - \sigma^2}{\lambda_i \delta_i^2 + \sigma^2} \leq a_i \leq 1, \quad i = 1, 2, \dots, p,$$

, then

$$MSE(\hat{\beta}_{GS}(\mathbf{A})) \leq MSE(\hat{\beta}_O).$$

5.2 Iteration algorithms

5.2.1 Old iteration method

Usually, we do not know the true value of β and σ^2 , so we use the least squares

estimator $\hat{\beta}_O$ and the unbiased estimator $\hat{\sigma}^2 = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2}{n - p}$ to estimate β and

σ^2 , respectively. Finally, we obtain

$$\hat{a}_i^{opt} = \frac{\lambda_i \hat{\sigma}_i^2}{\lambda_i \hat{\sigma}_i^2 + \hat{\sigma}^2} = \frac{\lambda_i}{\lambda_i + \hat{\sigma}^2 / \hat{\sigma}_i^2}. \quad (5.4)$$

But it's far away from our real a_i^{opt} , so we decide to use iterative methods. We employ an iterative procedure to obtain the optimal id matrix $\mathbf{A}^{opt}$. We follow the way from the Liu-type iteration: after assuming $\mathbf{A}_0$, using $\hat{\beta}_{GS} = \mathbf{P}\mathbf{A}_0\mathbf{P}'\hat{\beta}_O$ then $\hat{\sigma}_{GS}^2$ to get $\mathbf{A}_1$, stop until $\|\mathbf{A}_{n+1}^{opt} - \mathbf{A}_n^{opt}\| < 10^{-5}$, so that our $\mathbf{A}^{opt} = \mathbf{A}_{n+1}^{opt}$.

5.2.2 New iteration method

The previous iteration method relies heavily on an initial matrix $\mathbf{A}_0$, which is often obtained through a heuristic or random-chosen approach. However, such an initial estimate can be unreliable. Empirically, we observe that when the optimal shrinkage parameter a_i^{opt} is small, the iterative process tends to converge rapidly but undesirably to zero, resulting in suboptimal performance.

According to the *Comparison Theorem* (Theorem 5.1.1), a reduction in the mean squared error (MSE) can be achieved under the following two conditions if $\text{MSE}(\hat{\beta}_{GS}(\mathbf{B})) \leq \text{MSE}(\hat{\beta}_{GS}(\mathbf{A}))$:

1. If $a_i < b_i$, then $a_i^{opt} > \frac{a_i + b_i}{2}$;

2. If $a_i > b_i$, then $a_i^{\text{opt}} < \frac{a_i + b_i}{2}$.

In practical scenarios, we propose randomly generating initial values $0 < a_i < 1$, which form the diagonal of the initial matrix $\mathbf{A}_0 = \text{diag}(a_i)$. Using this initial matrix, we compute a preliminary approximation $\mathbf{A}^* \approx \mathbf{A}^{\text{opt}}$, which serves as an intermediate estimator of the optimal shrinkage matrix.

We then apply the following selection rule for constructing a comparison matrix $\mathbf{B}$ based on the relationship between a_i , a_i^* , and the a_i^{opt} :

- If $a_i^* > \frac{a_i + 1}{2} > \frac{a_i + b_i}{2}$ and $a_i < \frac{a_i + 1}{2}$, then set $b_i = \frac{a_i + 1}{2}$.
- If $a_i^* < \frac{a_i}{2} < \frac{a_i + b_i}{2}$ and $a_i > \frac{a_i}{2}$, then set $b_i = \frac{a_i}{2}$.
- If $\frac{a_i}{2} < a_i^* < \frac{a_i + 1}{2}$, set $b_i = a_i^*$.

Using this construction, we define the new matrix $\mathbf{B}^{\text{opt}} = \text{diag}(b_i)$, which then replaces the original $\mathbf{A}_0$ as the updated starting point for the iterative process.

The iteration continues by updating $\mathbf{B}^{\text{opt}}$ until convergence, i.e., until the stopping criterion

$$\|\mathbf{B}_{n+1}^{\text{opt}} - \mathbf{B}_n^{\text{opt}}\| < 10^{-5}$$

is satisfied, ensuring numerical stability and sufficient accuracy in the approximation of the optimal shrinkage matrix.

5.2.3 Bootstrap iteration method

However, it has been observed that the new iteration method relies on an unreliable initial estimate A_0 , we can't make sure the following calculation of $A^* \approx A^{\text{opt}}$, but ensure that MSE is getting smaller.

The search for an appropriate A_0 is crucial. We keep the coefficients β the same, but generate different datasets. It has been observed that the components $\hat{a}_i^{\text{opt}}$ are not always stable and can occasionally exhibit considerable skewness. To address this issue and improve the robustness of the initial estimates, we adopt a bootstrap median approach. The procedure is detailed as follows:

(1) The bootstrap method involves following steps:

- Re-sample with replacement from $(\mathbf{X}_i, y_i)$ to generate N sub-datasets (each of size n).
- Compute the statistic $\hat{\beta}, \hat{\sigma}^2$ for each sub-dataset.
- Calculate $\hat{a}_i^{\text{opt}}$ by the N statistics $\hat{\beta}, \hat{\sigma}^2$.
- Take the median of all $\hat{a}_i^{\text{opt}}$ to get initial value A_0 .

Based on the new iteration method, we can apply our $\mathbf{A}_0$ into iterative algorithm:

1. Set $\mathbf{A}_0 = \text{diag}(a_{0i}^{\text{opt}})$ from (1) to calculate $\hat{\beta}_{GS} = \mathbf{P}\mathbf{A}_0\mathbf{P}'\hat{\beta}_O$ and $\hat{\sigma}_{GS}^2 =$

$$\frac{\|\mathbf{Y} - \mathbf{X}\hat{\beta}_{GS}\|^2}{n - p}.$$

2. Calculate the estimated $\mathbf{A}^*$ with optimal weights $a_i^* = \frac{\lambda_i \hat{\delta}_{GSi}^2}{\lambda_i \hat{\delta}_{GSi}^2 + \hat{\sigma}_{GS}^2}$
3. For each $i = 1, 2, \dots, p$, calculate

$$b_i^{\text{opt}} = \begin{cases} \frac{a_{0i}^{\text{opt}}}{2}, & \text{if } a_i^* < \frac{a_{0i}^{\text{opt}}}{2}, \\ \frac{a_{0i}^{\text{opt}} + 1}{2}, & \text{if } a_i^* > \frac{a_{0i}^{\text{opt}} + 1}{2}, \\ a_i^*, & \text{otherwise.} \end{cases}$$

4. Update $\mathbf{A}_0 = \mathbf{B}_0^{\text{opt}}$ or stop if $\|\mathbf{B}_{n+1}^{\text{opt}} - \mathbf{B}_n^{\text{opt}}\| < 10^{-5}$.

There are three methods to be compared. First, the original iteration method, which is derived from the optimality of the Liu-type estimator and was first introduced in [Liu \(2010\)](#). This approach skips Step 3 and proceeds directly to Step 4 by setting $\mathbf{A}_0 = \mathbf{A}^*$.

The second method, referred to as the new iteration method, initializes Step 1 in procedure (2) by arbitrarily selecting an initial value and then performing the iterative updates.

The third method, the bootstrap iteration method, selects an initial matrix $\mathbf{A}_0$ that is relatively close to the true value. This approach effectively combines elements of both methods (1) and (2). The flowchart is as shown in the following figure.

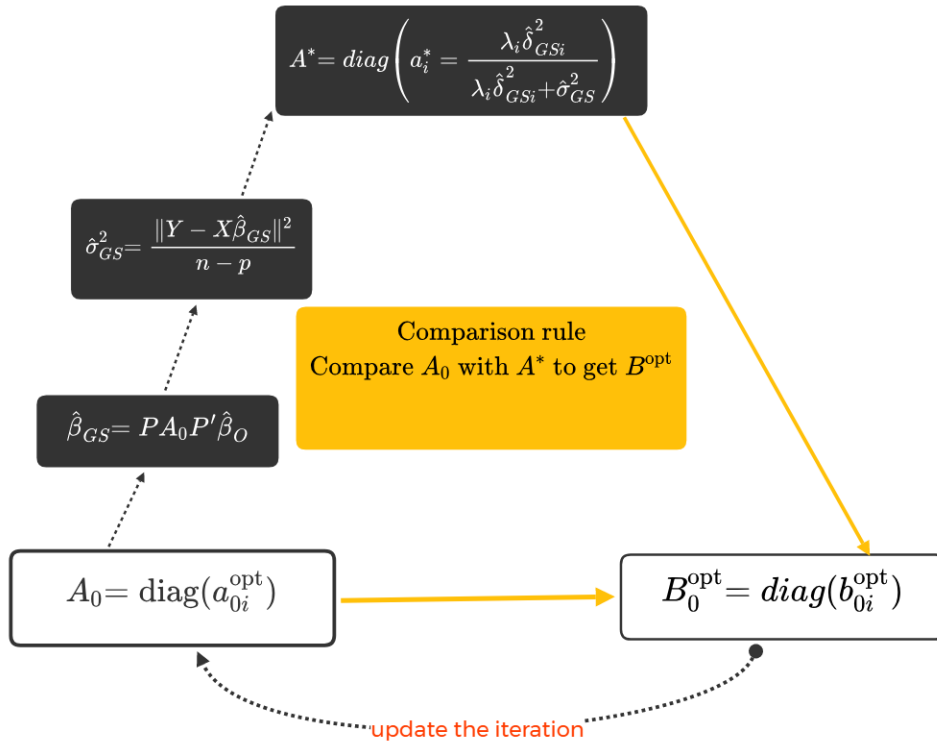


Figure 5.1: New iteration method

To introduce multicollinearity into the design matrix, we set $X_0 = \mathbf{1}$ and generate X_1 and X_2 independently from the standard normal distribution, i.e., $X_1, X_2 \sim \mathcal{N}(0, 1)$. The variable X_3 is constructed as a linear combination of X_1 and X_2 with an additional small random perturbation:

$$X_3 = X_1 + 2X_2 + e,$$

where $e \sim \mathcal{N}(0, 0.03^2)$.

The regression coefficients are specified as $\beta = (1, 2, 3, 4)$, and the error term is drawn from a normal distribution, $\epsilon \sim \mathcal{N}(0, 3^2)$. The response variable is generated according to the model:

$$y = X\beta + \epsilon.$$

From Table 5.1, we can see that using the bootstrap procedure to obtain the initial values and then applying the iterative algorithm (BootIter) produces estimates that are very close to the true values. At the same time, the resulting MSE is also very close to the true mean squared error, demonstrating the effectiveness and reliability of this method.

Table 5.1: Results Comparison

Method	aopt1	aopt2	aopt3	aopt4	MSE
True	0.9995	0.7949	0.9327	0.0035	2.8153
BootIter	0.9949	0.8030	0.9343	0.0051	2.8178
OldIter	0.9995	0.7986	0.9504	0.0000	2.8250
NewIter	0.9994	0.7981	0.9503	0.0006	2.8220

Chapter 6

Monte Carlo Simulations

In this section, we will carry out simulations to study the performance of the optimal GSLS and compare it with other linear biased estimators with various choices of k and d proposed in the literature. Observations on responses are then determined by

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n \quad (6.1)$$

In this simulation study, nine estimators are compared:

1. *LS*: Ordinary least squares estimator
2. *RH*: Ridge regression with $k_{\text{HKB}} = \frac{p\hat{\sigma}_O^2}{\hat{\beta}_O' \hat{\beta}_O}$
3. *LIU*: Liu estimator with iterated $d_{MM} = \frac{\sum_{i=1}^p \frac{\delta_i^2 - \sigma^2}{(\lambda_i + 1)^2}}{\sum_{i=1}^p \frac{\sigma^2 + \lambda_i \delta_i^2}{\lambda_i (\lambda_i + 1)^2}}$

4. LIU_{RH} : Liu Type estimator with

$$\hat{k} = \frac{\lambda_1 - 100 \cdot \lambda_p}{99}$$

and

$$\hat{d} = \frac{\sum_{i=1}^p \frac{\hat{\sigma}^2 - \hat{k}\hat{\delta}_i^2}{(\lambda_i + \hat{k})^2}}{\sum_{i=1}^p \frac{\lambda_i \hat{\delta}_i^2 + \hat{\sigma}^2}{\lambda_i(\lambda_i + \hat{k})^2}}.$$

5. TP : Two-Parameter Estimator with iterated

$$\hat{k}_{HM} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \left[\hat{\delta}_i^2 - d \left(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\delta}_i^2 \right) \right]}$$

and

$$\hat{d} = \frac{\sum_{i=1}^p \frac{k\hat{\delta}_i^2 - \hat{\sigma}^2}{(\lambda_i + k)^2}}{\sum_{i=1}^p \frac{k(\hat{\sigma}^2 + \hat{\delta}_i^2 \lambda_i)}{\lambda_i(\lambda_i + k)^2}}.$$

6. $MNTP$: Modified New Two-Parameter Estimator with iterated

$$\hat{k}_{AM} = \frac{1}{p} \sum_{i=1}^p \frac{\hat{\sigma}^2(\lambda_i + d) - (1 - d)\lambda_i \hat{\delta}_i^2}{d(\lambda_i + 1)\hat{\delta}_i^2}$$

and

$$\hat{d} = \frac{\sum_{i=1}^p \lambda_i (\hat{\delta}_i^2 - \hat{\sigma}^2)}{\sum_{i=1}^p (\lambda_i \hat{\delta}_i^2 + \hat{\sigma}^2 - k \lambda_i \hat{\delta}_i^2)}$$

7. OPT_{LS} : Optimal GSLS estimator with initial value $\hat{a}_i = \frac{\lambda_i}{\lambda_i + \hat{\sigma}_O^2 / \hat{\delta}_{Oi}^2}$,

and then use algorithm from Section 5.

8. OPT_{RH} : Optimal GSLS estimator with initial value $\hat{a}_i = \frac{\lambda_i}{\lambda_i + \hat{\sigma}_{RH}^2 / \hat{\delta}_{RH i}^2}$,

and then use algorithm from Section 5.

For each choice of n, p, ρ, β , and σ , variables are generated. The experiment is then repeated 5,000 times by generating new error terms. After the sample is generated, the average mean square error (AMSE) for $\hat{\beta}$ is calculated as follows:

$$AMSE(\hat{\beta}) = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\beta}_{(r)} - \beta)' (\hat{\beta}_{(r)} - \beta) \quad (6.2)$$

where $\hat{\beta}_{(r)}$ is the estimator for each replicate r ($r = 1, 2, \dots, 5,000$) and β is true parameter vector.

Our computations here and below are all performed using Rstudio 2024.12.0 + 467 and with R version 4.4.2.

6.1 Simulation study 1

The regressor variables are computed as McDonald and Galarneau (1975), Kibria (2003), Liu (2003), and Liu (2004) by

$$x_{ij} = (1 - \rho^2)^{1/2} z_{ij} + \rho z_{i(p+1)}, \quad i = 1, \dots, n, \quad j = 1, \dots, p \quad (6.3)$$

where z_{ij} are independent, standard, normal, pseudo-random numbers and ρ is specified so that the correlation between two regressors is given by ρ^2 . These variables are then standardized so that $X'X$ is in correlation form. Four different sets of correlations are considered that correspond to $\rho = 0.9, 0.99, 0.999, 0.9999$ for strong multicollinearity. And the conditional numbers are 4.26, 17.34, 45.62, 167.32 respectively. Regressors are used as $p = 3$ and $n = 50, 100, 200$. Three values of σ are investigated as 0.1, 1, and 10.

For the model, ϵ_i are independent normal pseudo-random numbers $(0, \sigma^2)$, and we will assume zero intercept without loss of any generality. Newhouse and Oman (1971) have noted that if the mean square error is a function of β , σ^2 , and k , and if the regressors are fixed, then subject to the constraint $\beta'\beta = 1$, the mean square error is minimized when β is the normalized eigenvector corresponding to the largest eigenvalue of the $X'X$ matrix. Hence, we selected the coefficients $\beta_1, \dots, \beta_p$ as the normalized eigenvector corresponding to the largest eigenvalue of the matrix $X'X$.

Table 6.1: AMSE under simulation study 1 when $n = 50$

ρ	σ	LS	RH	LIU	LIU _{RH}	TP	MNTP	OPT_LS	OPT_RH
0.9	0.1	0.0372	0.0371	0.0342	0.0339	0.0351	0.0337	0.0329	0.0329
	1	0.1092	0.0777	0.0746	0.0391	0.0365	0.0346	0.0336	0.0335
	10	42.0671	8.5717	6.9232	6.1509	5.1500	4.8275	2.7142	2.1624
0.99	0.1	0.0513	0.0505	0.0485	0.0487	0.0476	0.0411	0.0406	0.0403
	1	3.5805	0.6948	0.5605	0.5603	0.4495	0.3243	0.1409	0.1386
	10	163.3247	93.0349	86.2093	83.0732	80.0631	8.5088	5.7194	5.0276
0.999	0.1	0.4864	0.3336	0.3234	0.2745	0.2502	0.2152	0.1273	0.1262
	1	31.9792	9.2507	8.2505	7.2497	0.5253	0.3744	0.1433	0.1437
	10	3557.3968	720.5873	481.9412	454.7409	50.1587	41.4294	5.7432	5.4936
0.9999	0.1	7.2026	5.5855	5.0467	4.3918	4.3437	4.3436	0.1491	0.1383
	1	99.0769	38.4543	9.7552	7.3481	4.7648	4.5252	3.4540	1.1541
	10	27884.6853	6565.1778	4138.3532	3753.3777	625.7631	586.7946	6.9336	6.0981

Table 6.2: AMSE under simulation study 1 when $n = 100$

ρ	σ	LS	RH	LIU	LIU _{RH}	TP	MNTP	OPT_LS	OPT_RH
0.9	0.1	0.1277	0.1268	0.1266	0.1267	0.1247	0.1245	0.1237	0.1235
	1	0.1656	0.1608	0.1606	0.1404	0.1388	0.1387	0.1385	0.1345
	10	8.6595	5.5335	4.3692	3.7685	3.2683	3.2141	0.2559	0.2370
0.99	0.1	0.1289	0.1285	0.1282	0.1279	0.1259	0.1257	0.1249	0.1247
	1	0.4932	0.2966	0.2964	0.2458	0.1940	0.1825	0.1680	0.1665
	10	41.1718	16.0251	14.1948	13.3225	13.3223	6.1778	5.1331	5.1249
0.999	0.1	0.6529	0.6493	0.5491	0.5312	0.5302	0.1283	0.1339	0.1337
	1	6.2126	0.9531	0.8569	0.6793	0.6791	0.2655	0.2340	0.2253
	10	802.1643	636.9747	602.7605	550.7076	85.2343	85.2111	5.3811	5.4641
0.9999	0.1	5.5790	3.3260	3.3258	2.2452	2.2320	1.1721	0.1539	0.1536
	1	57.9121	10.6524	8.4000	5.9118	5.9107	1.4017	1.1320	1.1250
	10	5000.5342	1036.1791	679.6537	547.6940	457.3884	454.4785	5.6646	5.6603

Table 6.3: AMSE under simulation study 1 when $n = 200$

ρ	σ	LS	RH	LIU	LIU _{RH}	TP	MNTP	OPT_LS	OPT_RH
0.9	0.1	0.0021	0.0020	0.0015	0.0013	0.0011	0.0011	0.0002	0.0001
	1	0.4783	0.4752	0.4734	0.4582	0.4575	0.4575	0.4405	0.4403
	10	4.0612	1.0534	0.8326	0.7133	0.7068	0.4751	0.4605	0.4381
0.99	0.1	0.0029	0.0026	0.0018	0.0018	0.0010	0.0008	0.0006	0.0006
	1	0.7352	0.6524	0.6515	0.5384	0.4732	0.4609	0.4532	0.4530
	10	20.3234	4.4728	4.4298	4.2346	4.2343	0.7433	0.5352	0.5142
0.999	0.1	0.5345	0.5342	0.4236	0.4123	0.4023	0.4014	0.0012	0.0009
	1	4.1372	0.7714	0.7232	0.6342	0.6234	0.5294	0.5113	0.5030
	10	513.3902	494.2314	471.0324	394.3123	331.1424	57.1249	0.6213	0.5950
0.9999	0.1	3.4366	2.2924	2.2544	1.1636	1.1546	0.5253	0.1032	0.0924
	1	42.4543	8.3452	6.2453	3.5492	3.2542	1.0385	0.9389	0.9323
	10	2473.3244	624.2459	343.5796	143.9806	92.5490	61.3985	3.7352	3.2509

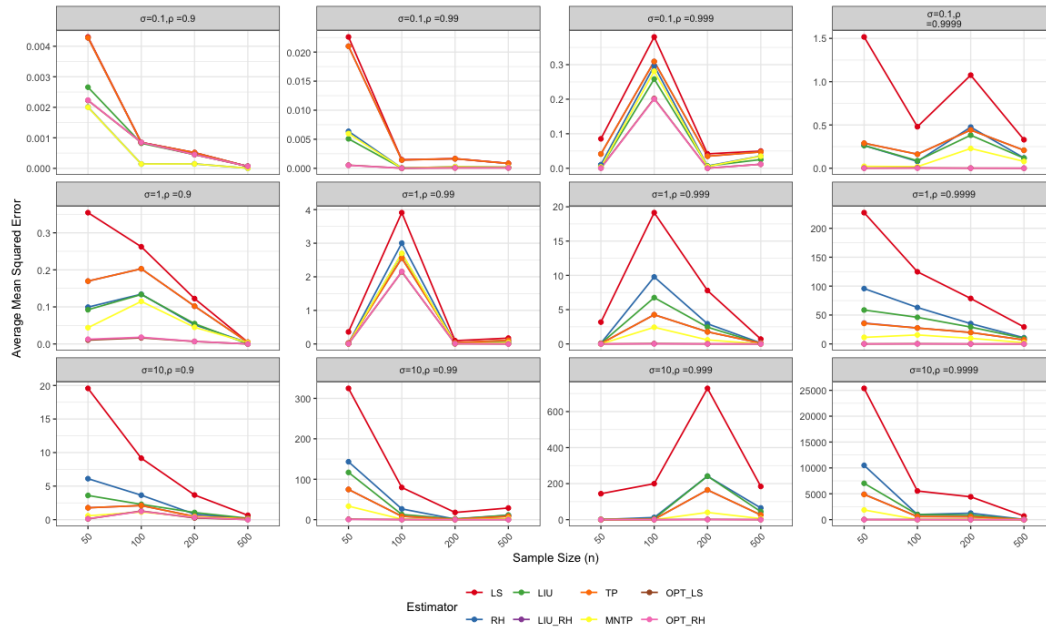


Figure 6.1: AMSE under simulation study 1

From the results shown in [Table 6.1](#), [Table 6.2](#), [Table 6.3](#) it can generally be concluded that for the error variance (σ): as the error variance (σ) increases, the AMSE increases for all estimators. This is particularly evident at higher values of σ (e.g., $\sigma = 10$), where the differences in AMSE between estimators become more pronounced, especially for small sample sizes ($n = 50$). OPT_LS and OPT_RH maintain relatively lower AMSE values even at higher levels of error variance, indicating their robustness in handling error variability. For degree of multicollinearity (ρ): as the degree of multicollinearity increases (that is, as ρ approaches 1), the AMSE increases for all estimators, which is a classic sign of ill conditioning in regression models. The impact is most noticeable when ρ reaches 0.9, where the AMSE for LS and other estimators drastically increases, indicating the difficulty of accurately estimating the parameters under high multicollinearity. OPT_LS and OPT_RH show relatively more stable performance compared to other estimators when ρ is near 1, suggesting that they are better equipped to handle multicollinearity. For sample size (n): as the sample size increases from $n = 50$ to $n = 200$, the AMSE decreases, which is expected in any well behaved estimator due to the law of large numbers. However, LS still shows a higher AMSE at larger sample sizes compared to OPT_LS and OPT_RH. OPT_LS and OPT_RH consistently achieve the smallest AMSE across all sample sizes, particularly at $n = 200$, where the error reduction is the most significant.

6.2 Simulation study 2

To simulate multicollinearity within the design matrix, we define the predictors X_1 and X_2 are independently sampled from a standard normal distribution:

$$X_1, X_2 \sim \mathcal{N}(0, 1).$$

The variable X_3 is generated by combining X_1 and X_2 linearly with an additional noise term:

$$X_3 = 3X_1 - 5X_2 + \omega,$$

where $\omega \sim \mathcal{N}(0, \eta^2)$ and we assume η as 0.03, 0.3, 3. For this simulation, our condition numbers are 1217.41, 110.13, 15.95 respectively. The coefficient vectors are specified randomly as

$$\beta = (-5, 2, 10, 4),$$

and the error term is ϵ that is drawn from a standard normal distribution, and σ keeps the same as in simulation 1.

Table 6.4: AMSE under simulation study 2 when $n = 50$

η	σ	LS	RH	LIU	LIU _{RH}	TP	MNTP	OPT_LS	OPT_RH
0.03	0.1	0.4074	0.0804	0.0682	0.0655	0.0628	0.0573	0.0569	0.0566
	1	11.0898	8.8623	8.1600	8.4876	6.2234	4.2030	0.0535	0.0528
	10	3513.0025	810.7465	909.4873	915.3756	809.5196	768.9396	659.7102	334.4588
0.3	0.1	0.0499	0.0482	0.0406	0.0414	0.0476	0.0421	0.0409	0.0403
	1	3.5110	0.4625	0.3517	0.5976	0.4659	0.2295	0.0449	0.0438
	10	1155.9341	457.2018	479.3425	721.1564	455.5874	42.0049	11.4075	2.6541
3	0.1	0.0452	0.0442	0.0364	0.0361	0.0449	0.0366	0.0285	0.0280
	1	0.1187	0.0860	0.0462	0.0418	0.0847	0.0471	0.0416	0.0432
	10	41.2547	7.6011	5.9180	5.1473	5.1655	3.2978	3.8431	3.5922

Table 6.5: AMSE under simulation study 2 when $n = 100$

η	σ	LS	RH	LIU	LIU _{RH}	TP	MNTP	OPT_LS	OPT_RH
0.03	0.1	0.3748	0.3076	0.2275	0.2506	0.3097	0.2352	0.0269	0.0264
	1.0	35.4913	5.7618	7.1362	7.3951	5.7421	3.2936	1.0431	1.0338
	10.0	3709.1149	573.9540	713.9066	850.0327	574.9922	483.4120	116.2842	88.5957
0.3	0.1	0.0419	0.0414	0.0332	0.0355	0.0427	0.0356	0.0287	0.0008
	1.0	0.3706	0.1218	0.0702	0.1065	0.1233	0.1479	0.0621	0.0370
	10.0	675.1129	469.2755	523.8776	570.3992	468.2141	107.7490	3.3872	1.1138
3	0.1	0.0285	0.0280	0.0281	0.0278	0.0286	0.0281	0.0209	0.0007
	1.0	0.1269	0.1047	0.0709	0.0622	0.1039	0.0553	0.0434	0.0141
	10.0	36.4711	5.8619	7.4238	9.8021	5.8472	4.3257	2.2985	1.5329

Table 6.6: AMSE under simulation study 2 when $n = 200$

η	σ	LS	RH	LIU	LIU _{RH}	TP	MNTP	OPT_LS	OPT_RH
0.03	0.1	0.0551	0.0514	0.0052	0.0076	0.0510	0.0069	0.0014	0.0012
	1.0	6.0944	0.5570	0.7341	0.8277	0.5566	0.0428	0.0109	0.0018
	10.0	3918.3376	990.5234	1103.2190	3452.9355	990.4798	1050.1789	672.9835	358.1902
0.3	0.1	0.0273	0.0232	0.0245	0.0232	0.0193	0.0102	0.0020	0.0019
	1.0	0.3689	0.1736	0.0710	0.1229	0.1723	0.0594	0.0047	0.0033
	10.0	40.8877	3.6781	4.2655	5.8976	3.6714	0.9923	0.1394	0.1572
3	0.1	0.0184	0.0132	0.0125	0.0101	0.0003	0.0086	0.0002	0.0001
	1.0	0.0419	0.0374	0.0152	0.0172	0.0371	0.0155	0.0019	0.0013
	10.0	8.5428	3.2017	4.0742	5.4328	3.1994	0.0546	0.0096	0.0018

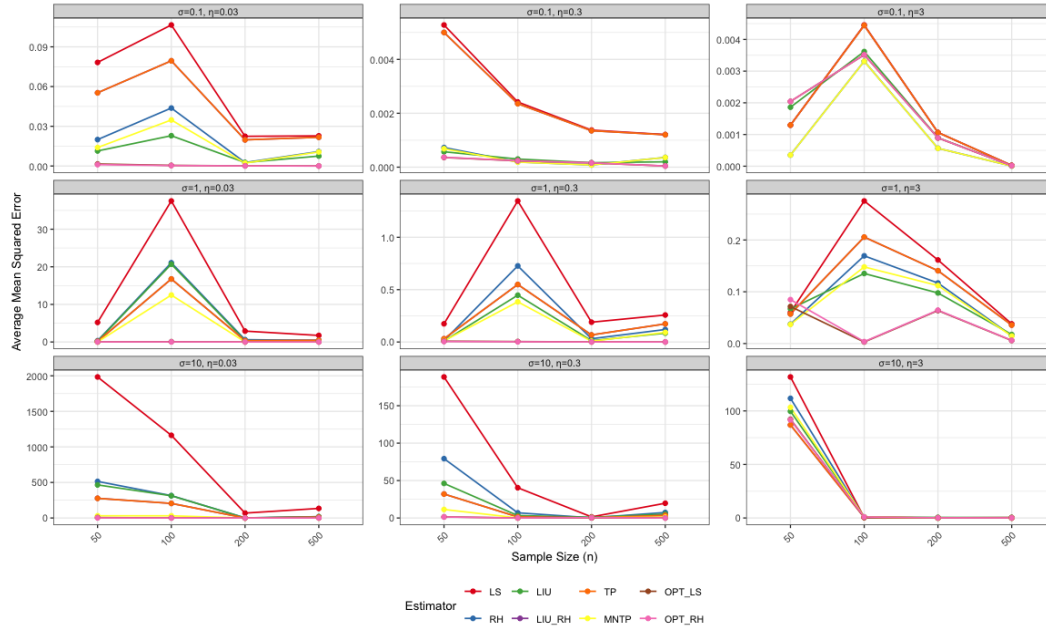


Figure 6.2: AMSE under simulation study 2

From the results shown in [Table 6.4](#), [Table 6.5](#), [Table 6.6](#) it can generally be concluded that under the typical statistical linear model, OPT_LS and OPT_RH are very close, but OPT_RH is more significant across most the conditions tested. There is a exception that when $n = 50$, $\eta = 3, \sigma = 10$, the MSE of MNTP are smaller than our optimal GSLS OLS estimator.

From [Figure 6.1](#) and [Figure 6.2](#), it is evident that the Mean Squared Error (MSE) of all our estimators is smaller than that of the Ordinary Least Squares (OLS) estimator, with our optimal estimator consistently demonstrating superior performance in most cases. When the sample size is small, the MSE differences in LS, RH, LIU, LIU_RH, TP and MNTP are relatively minor. But

our optimal GSLS are far less than them. However, as sample sizes increases, the gap between the best and worst MSEs becomes no clear-cut distinction.

Chapter 7

Real Data Application

7.1 Portland cement dataset

To demonstrate the validity of our theoretical results, we use an example that is commonly referenced in the literature for comparing our proposed estimator with ridge, Liu, Liu-RH, TP, and MNTP estimators.

Following [Woods et al. \(1932\)](#), we fit a homogeneous linear model without an intercept to the data. In this model, we have $n = 13$ observations and $p = 4$ unknown regression coefficients. The eigenvalues of $X'X$ are found to be 44676.206, 5965.422, 809.952, and 105.419. The condition number of $X'X$ is 20.5846, suggesting that $X'X$ is well conditioned with only weak multicollinearity present. For practical purposes, we replace the unknown parameters β and σ^2 with suitable estimators. The results of this data under the homogeneous linear regression model are presented as follows.

To assess the robustness of our theoretical framework, we utilize a cement heat released data set frequently cited in previous research to compare the proposed estimator against conventional alternatives, namely ridge, Liu, Liu-RH, TP, and MNTP estimators. The dataset contains 4 predictors, X_1 is the tricalcium aluminate, X_2 is tricalcium silicate, X_3 is tetracalcium aluminoferrite, and lastly X_4 is dicalcium silicate. These four variables are the percentage of components that make up the cement. Y is the heat per gram cement released.

Following the experimental setup in [Woods et al. \(1932\)](#), we implement an homogeneous linear regression model. The dataset comprises $n = 13$ observations with $p = 4$ regression coefficients. The multicollinear analysis of VIFs is 283, 254, 47, and 39. The condition number, calculated as 21, suggests moderate multicollinearity, confirming the instability of the matrix. For unknown true parameters β and σ^2 , we employ our proposed estimators. The regression results under this configuration are summarized as follows.

Table 7.1: Homogeneous model for cement data

Estimators	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
LS	2.1930	1.1533	0.7585	0.4863
RH	2.1655	1.1563	0.7463	0.4931
LIU	2.1646	1.1564	0.7458	0.4934
LIU_RH	2.1651	1.1564	0.7461	0.4932
TP	2.1641	1.1565	0.7455	0.4936
MNTP	2.1478	1.1638	0.7261	0.4931
OPT_LS	2.1645	1.1564	0.7458	0.4934
OPT_RH	2.1180	1.1671	0.7125	0.5005

Then we use the strategy adopted by Hald (1952, pp. 635–652) and fit

a linear model with intercept (inhomogeneous model) to the data. Now, the condition number increases to 6057, suggesting the presence of severe collinearity. The reason for this dramatic change in condition number is that there exists a near linear relationship among the four independent variables: $x_1 + x_2 + x_3 + x_4 \approx 100$. Given the presence of intercept, there would be stronger multicollinearity, the OLS estimator may not be reliable.

Table 7.2: Inhomogeneous model for cement data

Estimators	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
LS	62.4054	1.5511	0.5102	0.1019	-0.1441
RH	8.5870	2.1046	1.0648	0.6681	0.3996
LIU	0.8981	2.1837	1.1441	0.7490	0.4773
LIU_LS	0.0307	0.8574	1.4055	0.0789	0.6439
LIU_RH	0.0329	1.0676	1.3884	0.0899	0.6366
TP	0.2356	2.1851	1.1522	0.7520	0.4848
OPT_LS	0.0475	2.1604	1.1564	0.7436	0.4939
OPT_RH	0.0470	2.1603	1.1565	0.7436	0.4939

Compared with other estimators, the OLS estimator under the inhomogeneous model exhibits two significant issues. First, the intercept of the OLS estimator, $\hat{\beta}_0 = 62.4054$, is excessively large. This is impossible since, except for our four components, the rest material we could regard as impurities. They could not release much heat. Second, the sign of the correlation coefficient $\hat{\beta}_4 = -0.1441$ for X_4 is negative. It is not reasonable because our cement will generate heat when hardening. The numerical results present a comparison between the GSLS estimator and other estimators; we offer the following observations.

Our proposed estimator resolves both issues. Specifically, the correlation coefficient $\hat{\beta}_4 = 0.4939$ for X_4 is positive and the intercept $\hat{\beta}_0 = 0.0475$ for OPT_LS of the GSLS estimator is small. Moreover, our proposed estimator under the inhomogeneous linear regression model closely matches the OLS estimator under the homogeneous linear regression model, consistent with the original analysis by [Woods et al. \(1932\)](#). This means that even though we fit a different linear model with or without intercept, our result is still robust.

Due to the large condition number, the intercept of the ridge estimator's starting value $\hat{\beta}_0 = 8.5870$ is still considerably large. In this case, we recommend using the iterated OPT_RH estimator, which results in a much smaller intercept $\hat{\beta}_0 = 0.0470$.

7.2 French economy dataset

This study examines France's import patterns using data adapted from [Lukman et al. \(2020\)](#), where the dependent variable Y represents total imports (IMPORT) for developing years. The dataset comprises $n = 18$ quarterly observations from the national economic accounts. The import volume is modeled through three macroeconomic predictors:

- x_1 (DOPROD): Domestic industrial production index
- x_2 (STOCK): Inventory-to-sales ratio

- x_3 (CONSUM): Household final consumption expenditure

The design matrix exhibits severe multicollinearity, as evidenced by:

Table 7.3: Multicollinearity diagnostics

Diagnostic Measure	Value
Maximum VIF	58.7
Condition Index	89.27
Minimum Eigenvalue ($\lambda_{\min}$)	0.0043

The condition number is computed as:

$$\kappa(\mathbf{X}) = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = 89.27 \quad (7.1)$$

that exceeds the critical threshold of 30 suggested by [Belsley et al. \(1980\)](#).

The OLSE is no longer considered an appropriate estimation method in this context.

Table 7.4: Estimated coefficients of French economy data

Estimators	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	-19.7251	0.0322	0.4142	0.2427
RH	-16.7888	0.1410	0.3666	0.0722
LIU	42.5846	2.3400	-0.5961	-3.3760
LIU_RH	-18.7697	0.1466	0.4098	0.0746
TP	5.8502	0.9760	-0.0042	-1.2376
MNTP	-4.5908	0.5893	0.1651	-0.6312
OPT_LS	-18.7288	0.0118	0.0667	0.2736
OPT_RH	-15.5065	0.1300	0.0153	0.0880

Based on the table above, we make the following observations: We recommend using the optimal GSLS estimator due to the severe large condition

number. Given the presence of strong multicollinearity, the OLS estimator cannot be trusted. The sign of the correlation coefficient $\hat{\beta}_0 = -19.7251$ from the OLS estimator is negative. However, our proposed estimator Liu and TP, with the sign of the correlation coefficient being positive, is not correct since France's imports are influenced by policies that often set minimum domestic production quota and provide subsidies that reduce import dependence under normal conditions. Our proposed estimator corrects this issue and reduces the intercept of the OLS estimator without overshrinking it. Additionally, the sign of $\hat{\beta}_3$ should be negative, which is based on stocks that are higher causing less imports. But our optimal GSLS estimator provides shrunken estimator making it have small influence on the import, this makes the Domestic industrial production index account for the most in our shrunken estimators.

Chapter 8

Conclusion

In this thesis, we introduce a generalized shrunk least-square (GSLS) estimator, which includes special cases such as ridge, Liu, Liu-type, two-parameter, new two-parameter, modified two-parameter, and modified new two-parameter estimators. We then apply the theorem from Wang (1990) to derive the necessary and sufficient conditions for the GSLS estimator and conclude the conditions for each specific estimator.

Through numerical calculations, we determine the optimal GSLS estimator using the Mean Squared Error (MSE) method. After comparing two GSLS estimators, we propose that the optimality of $\mathbf{A}$ is based on β and σ^2 . The old iteration method yields quick results, and the optimal a_i^{opt} tends to be zero when small. Using the comparison theorem, we found that the new iteration method reduces the MSE, though it may not approach the true value closely. Combining both approaches, we adopt a bootstrap median method to obtain

the initial $\mathbf{A}_0$, which is then optimized using the new iteration method for the GSLS estimator.

Simulation studies were conducted to compare the optimal GSLS estimator with the OLS and other linear biased estimators. The MSE of the optimal GSLS estimator is smaller than that of the OLS and other biased estimators under various scenarios. When both σ and ρ are small, the differences in MSE are relatively minor. However, as either σ or ρ increases, the gap becomes more pronounced, with all estimators outperforming the OLS estimator under all conditions. This pattern also holds true in simulation study 2, which utilizes η to represent multicollinearity. In the real data case, our GSLS estimator under an inhomogeneous model can correct the signs of some coefficients and closely match the OLS estimator under the homogeneous model.

Appendix A

Simulation R Code

The following R code was used to perform the simulation study described.

```
1 n_values <- c(50, 100, 200) # Sample size loop
2 p <- 3 # Number of variables
3 rhos <- c(0.9, 0.99, 0.999) # Correlation coefficients
4 sigmas <- c(0.1, 1, 10) # Noise levels
5 reps <- 100 # Number of repetitions
6
7 # Initialize result storage
8 results <- data.frame()
9
10 for (n in n_values) {
11   for (rho in rhos) {
12
13     # Generate fixed design matrix X
14     w <- matrix(rnorm(n*(p+1)), nrow = n)
15     X <- matrix(nrow = n, ncol = p)
16     for (i in 1:n) {
17       for (j in 1:p) {
18         X[i, j] <- sqrt(1 - rho^2) * w[i, j] + rho * w[i,
19           p+1]
20       }
21     }
22     XtX <- t(X) %*% X
23     eigen_XtX <- eigen(XtX)
```

```

23 P <- eigen_XtX$eigenvectors
24 lambda <- eigen_XtX$values
25
26 # Generate true coefficients      (principal
    eigenvector)
27 beta_true <- P[, 1]
28
29 for (sigma in sigmas) {
30   # For each sigma combination (fixed X,y), repeat '
    reps' times
31   # Reinitialize this matrix at the beginning of each
    sigma loop
32   # to clear previous results and ensure independence
    between sigma results
33   emse_matrix <- matrix(0, nrow = reps, ncol = 8)
34   cond_count <- 0 # HKB condition-number exceedance
    count
35
36   # Generate response variable
37   epsilon <- rnorm(n, sd = sigma)
38   y <- X %*% beta_true + epsilon
39
40   for (r in 1:reps) {
41     # ---- Calculate estimators ----
42     # LS estimator
43     beta_LS <- solve(XtX) %*% t(X) %*% y
44     delta_LS <- t(P) %*% beta_LS
45     # Sigma^2 OLS estimator
46     res_LS <- y - X %*% beta_LS
47     sigma_sq <- sum(res_LS^2) / (n - p)
48
49     # Ridge with k_HKB
50     k_HKB <- (p * sigma_sq) / sum(beta_LS^2)
51     XtX_kHKB <- XtX + k_HKB * diag(p)
52     beta_RH <- solve(XtX_kHKB) %*% t(X) %*% y
53     res_RH <- y - X %*% beta_RH
54
55     # Liu estimator
56     delta_sq <- (t(P) %*% beta_LS)^2

```

```

57 T_val <- sum(1/lambda)/p # Condition-number
    baseline
58 d_old <- 0 # Initial value
59 max_iter <- 100
60 tol <- 20*T_val^(-1.3) # Convergence threshold
61 # Initialize iteration
62 iter <- 0
63 converged <- FALSE
64 while(iter < max_iter && !converged){
65   # New parameters
66   numerator <- sum((delta_sq-sigma_sq)/(lambda+1)
67 )
68   denominator <- sum((sigma_sq^2+lambda*delta_sq)
69 / lambda*(lambda + 1)^2)
70   d_new <- numerator / denominator
71
72   # Check convergence
73   if(iter > 0 && abs((d_old - d_new)/d_old) < tol
74 ){
75     converged <- TRUE
76   }
77
78   # Update parameters
79   d_old <- d_new
80   delta_sq <- (t(P) %*% solve(XtX +diag(p)) %*% (
81     t(X)%*%y + d_new*beta_LS))^2
82   iter <- iter + 1
83 }
84 # Final Liu estimator
85 beta_Liu <- solve(XtX + k_HKB*diag(p)) %*% (t(X)%
86   %*%y - d_old*beta_LS)
87
88 # Liu-type Ridge
89 delta_rh <- t(P) %*% beta_RH # Using ridge
    estimator's delta
90 T_val <- sum(1/lambda)/p # Condition-number
    baseline
91 d_old <- 0.1 # Initial value
92 max_iter <- 30

```

```

88   tol <- 20*T_val^(-1.3) # Convergence threshold
89   sigma_rh <- sum(res_RH^2) / (n - p)
90   lambda_max <- max(lambda) # Maximum eigenvalue
91   lambda_min <- min(lambda) # Minimum eigenvalue
92   k_liu <- (lambda_max - 100 * lambda_min) / 99
93
94   # Initialize iteration
95   iter <- 0
96   converged <- FALSE
97   while(iter < max_iter && !converged){
98     # New parameters
99     numerator <- sum(lambda*(sigma_rh - k_liu*delta
100       _rh^2)/(lambda + k_HKB)^3)
101     denominator <- sum(lambda*(lambda*delta_rh^2 +
102       sigma_rh)/(lambda + k_HKB)^4)
103     d_new <- numerator / denominator
104     # Check convergence
105     if(iter > 0 && abs((d_old - d_new)/d_old) < tol
106       ){
107       converged <- TRUE
108     }
109     # Update parameters
110     d_old <- d_new
111     delta_rh <- t(P) %*% solve(XtX + k_liu*diag(p))
112       %*% (t(X)%*%y - d_new*beta_RH)
113     iter <- iter + 1
114   }
115
116   # Final Liu-type Ridge estimator
117   beta_Liu_R <- solve(XtX + k_liu*diag(p)) %*% (t(X)
118     )%*%y - d_old*beta_RH)
119
120   # TP estimator
121   # Initialize iteration
122   iter <- 0
123   converged <- FALSE
124   d_old <- min(delta_LS^2/(sigma_sq/lambda+delta_LS
125     ^2)) # Initial value
126   max_iter <- 100

```

```

121
122 while(iter < max_iter && !converged){
123   # Compute new K-HM
124   denominator_k_vals <- delta_LS^2 - d_old*(sigma
      _sq/lambda + delta_LS^2)
125   # Skip if denominator is zero or negative
126   if (any(denominator_k_vals <= 0)) {
127     converged <- TRUE
128
129   }
130   k_vals <- sigma_sq^2 / denominator_k_vals
131   k_HM <- p / sum(1/k_vals)
132   numerator <- sum((k_HM*delta_LS^2 - sigma_sq)/(
      lambda + k_HM)^2)
133   denominator <- sum(k_HM*(sigma_sq + delta_LS^2*
      lambda)/(lambda*(lambda + k_HM)^2))
134   d_new <- -numerator / denominator
135   # check negativity
136   if(iter > 0 && (!is.finite(d_new) || d_new < 0)
      ){
137     converged <- TRUE
138   }
139   # Update
140   d_old <- d_new
141   iter <- iter + 1
142 }
143 # Final TP estimator
144 beta_TP <- solve(XtX + k_HM*diag(p)) %*% (t(X)%*%
      y + k_HM*d_old*beta_LS)
145
146 # MNTP estimator
147 # Initialize iteration
148 iter <- 0
149 converged <- FALSE
150 d_old <- max((delta_LS^2-sigma_sq)*lambda/(sigma_
      sq+lambda*delta_LS^2)) # Initial value
151 max_iter <- 100
152
153 while(iter < max_iter && !converged){

```

```

154 # Compute new K-AM
155 k_AM <- mean((sigma_sq*(lambda)-(d_old)*lambda*
156             delta_LS^2)/(d_old*delta_LS*(lambda+1)))
157 numerator <- sum(lambda*(delta_LS^2-sigma_sq))
158 denominator <- sum(lambda*delta_LS^2+sigma_sq-k
159                 _AM*lambda*delta_LS^2)
160 d_new <- numerator / denominator
161 # Check negativity
162 if(iter > 0 && (!is.finite(d_new) || d_new < 0)
163     ){
164     converged <- TRUE
165 }
166 # Update
167 d_old <- d_new
168 iter <- iter + 1
169 }
170 # Final MNTP estimator
171 beta_MNTP <- solve(XtX + diag(p)) %*% (XtX + d_
172     old*diag(p)) %*%
173     solve(XtX + d_old*diag(p)) %*% t(X)%*%y
174
175 # Bootstrap
176 num_boot <- 100L
177 aopls <- matrix(NA_real_, nrow = num_boot, ncol =
178     p)
179 aoprh <- matrix(NA_real_, nrow = num_boot, ncol =
180     p)
181
182 for (i in 1:num_boot) {
183     idx <- sample.int(n, size = n, replace = TRUE)
184     Xi <- X[idx, , drop = FALSE]
185     yi <- y[idx]
186
187     # OLS on bootstrap sample
188     fit_i <- lm.fit(x = Xi, y = yi)
189     beta_ols_i <- as.vector(fit_i$coefficients)
190     res_i <- yi - Xi %*% beta_ols_i
191     df_i <- nrow(Xi) - p
192     sigma2_ols_i <- sum(res_i^2) / df_i

```

```

187
188   delta_ols_i <- as.vector(t(P) %*% beta_ols_i)
189   aopls[i, ] <- (lambda * delta_ols_i^2) / (
190       lambda * delta_ols_i^2 + sigma2_ols_i)
191
192   # HKB ridge on bootstrap sample
193   k_HKB_i <- (p * sigma2_ols_i) / sum(beta_ols_i
194       ^2)
195   XtXi <- crossprod(Xi) # Xi'Xi
196   beta_rh_i <- as.vector(solve(XtXi + k_HKB_i *
197       diag(p), crossprod(Xi, yi))) # (Xi'Xi+kI)
198       ^{-1}Xi'yi
199
200   res_rh_i <- yi - Xi %*% beta_rh_i
201   sigma2_rh_i <- sum(res_rh_i^2) / df_i
202
203   delta_rh_i <- as.vector(t(P) %*% beta_rh_i)
204   aoprh[i, ] <- (lambda * delta_rh_i^2) / (lambda
205       * delta_rh_i^2 + sigma2_rh_i)
206 }
207
208 a0_ls <- apply(aopls, 2, median, na.rm = TRUE)
209 a0_rh <- apply(aoprh, 2, median, na.rm = TRUE)
210
211 # ----- Iteration function (clamping +
212   improved convergence check) -----
213 run_iteration <- function(a0_initial, tol = 1e-5,
214   max_iter = 100L) {
215   a <- as.numeric(a0_initial)
216   for (iter in 1:max_iter) {
217       # compute beta_gs given current a
218       beta_gs <- as.vector(P %*% (a * (t(P) %*%
219           beta_LS))) # equal to P diag(a) P' beta_
220           LS
221       # compute new a*
222       res_gs <- y - X %*% beta_gs
223       sigma2_gs <- sum(res_gs^2) / (n - p)
224       delta_gs <- as.vector(t(P) %*% beta_gs)
225       a_star <- (lambda * delta_gs^2) / (lambda *

```

```

217         delta_gs^2 + sigma2_gs)
218
219     # clamp into [a/2, (a+1)/2]
220     lower <- a / 2
221     upper <- (a + 1) / 2
222     a_new <- pmin(pmax(a_star, lower), upper)
223
224     if (sqrt(sum((a_new - a)^2)) < tol) {
225         a <- a_new
226         break
227     }
228     a <- a_new
229 }
230 # return final beta
231 as.vector(P %*% (a * (t(P) %*% beta_LS)))
232
233 beta_bt_ls <- run_iteration(a0_ls)
234 beta_bt_rh <- run_iteration(a0_rh)
235
236 # EMSE
237 betas <- list(beta_LS, beta_TP, beta_RH, beta_Liu,
238               beta_MNTP, beta_Liu_R, beta_bt_ls, beta_bt_rh)
239 emse_matrix[r, ] <- sapply(betas, function(b) sum
240                             ((b - beta_true)^2))
241
242 # Aggregate results
243 avg_emse <- colMeans(emse_matrix)
244 estimator_names <- c("LS", "RH", "LIU", "LIU_RH", "TP",
245                      "MNTP", "OPTLS", "OPTRH")
246 results <- rbind(results, data.frame(
247     n = n,
248     rho = rho,
249     sigma = sigma,
250     estimator = factor(estimator_names, levels =
251                         estimator_names),
252     EMSE = avg_emse
253 ))

```

```

251   }
252 }
253 }
254
255 # Generate table
256 library(reshape2)
257 table_data <- dcast(results, n + rho + sigma ~ estimator,
258   value.var = "EMSE")
259 View(round(table_data, 6))
260
261 library(ggplot2)
262 # Combined labels
263 results$condition <- paste0("  =", results$sigma, ",  ="
264   , results$rho)
265
266 # Integrated facet plot
267 results$n <- factor(results$n) # Convert n to factor
268 combined_plot <- ggplot(results, aes(x = n, y = EMSE,
269   color = estimator, group = estimator)) +
270   geom_line(linewidth = 0.6) +
271   geom_point(size = 1.5) +
272   facet_wrap(~ condition,
273     ncol = 3, # 3 combinations
274     per row (for 3 rho values)
275     scales = "free_y", # Independent y-
276     axis scaling
277     labeller = label_wrap_gen(width = 15)) + #
278     Label wrapping
279   labs(title = "AMSE Across All Experimental Conditions",
280     x = "Sample Size (n)",
281     y = "Average Mean Squared Error",
282     color = "Estimator") +
283   theme_bw(base_size = 10) +
284   theme(
285     axis.text.x = element_text(angle = 45, hjust = 1),
286     panel.spacing = unit(0.4, "lines"),
287     strip.text = element_text(size = 8, margin = margin
288       (2,0,2,0)),
289     legend.position = "bottom"

```

```
283   ) +  
284   scale_color_brewer(palette = "Set1")  
285  
286 # Display plot  
287 print(combined_plot)
```

Bibliography

Ahmad, S. and M. Aslam (2022). Another proposal about the new two-parameter estimator for linear regression model with correlated regressors. *Communications in Statistics-Simulation and Computation* 51(6), 3054–3072. (Cited on pages 13, 14, 15 and 40.)

Akay, K. U., E. Ertan, A. Erkoç, and F. Taş (2024). A new general ridge-type estimator in linear regression models: Accepted-january 2024. *REVSTAT-Statistical Journal*. (Cited on page 14.)

Akdeniz, F. and H. Erol (2003). Mean squared error matrix comparisons of some biased estimators in linear regression. *Communications in Statistics-Theory and Methods* 32(12), 2389–2413. (Cited on page 9.)

Akdeniz, F. and S. Kaçiranlar (1995). On the almost unbiased generalized liu estimator and unbiased estimation of the bias and mse. *Communications in Statistics-Theory and Methods* 24(7), 1789–1797. (Cited on page 9.)

Belsley, D. A., E. Kuh, and R. E. Welsch (1980). Regression diagnostics new york. *NY Wiley*. (Cited on pages 34 and 69.)

- Dorugade, A. (2014a, 03). A modified two-parameter estimator in linear regression. *Statistics in Transition new series* 15, 23–36. (Cited on page 3.)
- Dorugade, A. V. (2014b). A modified two-parameter estimator in linear regression. *Statistics in Transition. New Series* 15(1), 23–36. (Cited on pages 13, 14 and 15.)
- Fisher, R. A. (1921). Studies in crop variation. i. an examination of the yield of dressed grain from broadbalk. *The Journal of Agricultural Science* 11(2), 107–135. (Cited on page 1.)
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 15, 246–263. (Cited on page 1.)
- Hoerl, A. E., R. W. Kannard, and K. F. Baldwin (1975). Ridge regression: some simulations. *Communications in Statistics-Theory and Methods* 4(2), 105–123. (Cited on pages 31 and 36.)
- Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55–67. (Cited on pages 2, 7, 11, 15 and 36.)
- Hoerl, A. E. and R. W. Kennard (1976). Ridge regression iterative estimation of the biasing parameter. *Communications in Statistics-Theory and Methods* 5(1), 77–88. (Cited on page 33.)

Hubert, M. and P. Wijekoon (2006). Improvement of the liu estimator in linear regression model. *Statistical Papers* 47(3), 471–479. (Cited on page 9.)

Kaçiranlar, S., S. Sakallioğlu, F. Akdeniz, G. P. Styan, and H. J. Werner (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on portland cement. *Sankhyā: The Indian Journal of Statistics, Series B*, 443–459. (Cited on pages 2 and 13.)

Kadiyala, K. (1984). A class of almost unbiased and efficient estimators of regression coefficients. *Economics Letters* 16(3-4), 293–296. (Cited on page 13.)

Kibria, B. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation* 32(2), 419–435. (Cited on pages 8, 36 and 56.)

Li, Y. and H. Yang (2012). A new liu-type estimator in linear regression model. *Statistical Papers* 53, 427–437. (Cited on page 11.)

Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods* 22(2), 393–402. (Cited on pages 2, 8, 13, 15 and 32.)

Liu, K. (2003). Using liu-type estimator to combat collinearity. *Communications in Statistics-Theory and Methods* 32(5), 1009–1020. (Cited on pages 10, 15 and 56.)

- Liu, K. (2004). More on liu-type estimator in linear regression. *Communications in Statistics-Theory and Methods* 33(11), 2723–2733. (Cited on pages 10, 11 and 56.)
- Liu, X. (2010). A class of generalized shrunken least squares estimators in linear model. Msc thesis, University of Manitoba, Canada. (Cited on page 49.)
- Lukman, A. F., B. G. Kibria, K. Ayinde, and S. L. Jegede (2020). Modified one-parameter liu estimator for the linear regression model. *Modelling and Simulation in Engineering* 2020(1), 9574304. (Cited on page 68.)
- Månsson, K., B. G. Kibria, and G. Shukur (2012). On liu estimators for the logit regression model. *Economic Modelling* 29(4), 1483–1488. (Cited on page 9.)
- Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12(3), 591–612. (Cited on page 17.)
- Marquardt, D. W. and R. D. Snee (1975). Ridge regression in practice. *The American Statistician* 29(1), 3–20. (Cited on page 8.)
- Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association* 60(309), 234–256. (Cited on page 2.)

- McDonald, G. C. and D. I. Galarneau (1975). A monte carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association* 70(350), 407–416. (Cited on pages 8 and 56.)
- Newhouse, J. P. and S. D. Oman (1971). An evaluation of ridge estimators. Technical Report R-716-PR, Rand Corporation. A report prepared for United States Air Force Project RAND. (Cited on page 56.)
- Ohtani, K. (1986). On small sample properties of the almost unbiased generalized ridge estimator. *Communications in Statistics-Theory and Methods* 15(5), 1571–1578. (Cited on page 13.)
- Özkale, M. R. and A. Abbasi (2022). Iterative restricted ok estimator in generalized linear models and the selection of tuning parameters via mse and genetic algorithm. *Statistical Papers* 63(6), 1979–2040. (Cited on page 11.)
- Özkale, M. R. and S. Kaciranlar (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics—Theory and Methods* 36(15), 2707–2725. (Cited on pages 2, 11, 15, 36 and 40.)
- Pearson, K. (1896). Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character* 187(187), 253–318. (Cited on page 1.)
- Rao, C. R., C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao (1973). *Linear*

- statistical inference and its applications*, Volume 2. Wiley New York. (Cited on page 17.)
- Sakallıoğlu, S. and S. Kaçıranlar (2008). A new biased estimator based on ridge estimation. *Statistical papers* 49, 669–689. (Cited on page 2.)
- Stein, C. (1992). Estimation with quadratic loss. In S. Kotz and N. L. Johnson (Eds.), *Breakthroughs in Statistics*, pp. 443–460. Springer. (Cited on page 17.)
- Stein, C. et al. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 197–206. (Cited on pages 2 and 8.)
- Swindel, B. F. (1976). Good ridge estimators based on prior information. *Communications in Statistics-Theory and Methods* 5(11), 1065–1075. (Cited on pages 2 and 8.)
- Theil, H. (1963). On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association* 58(302), 401–414. (Cited on page 12.)
- Theil, H. and A. S. Goldberger (1992). On pure and mixed statistical estimation in economics. In *Henri Theil's Contributions to Economics and Econometrics: Econometric Theory and Methodology*, pp. 317–332. Springer. (Cited on page 12.)

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58(1), 267–288. (Cited on page 2.)
- Wald, A. (1949). Statistical decision functions. *The Annals of Mathematical Statistics*, 165–205. (Cited on page 17.)
- Wang, L. (1990). Generalized shrunken least squares estimators. *Chinese Journal of Applied Probability and Statistics* 6(3), 225–232. (Cited on pages 3, 6, 18 and 71.)
- Woods, H., H. H. Steinour, and H. R. Starke (1932). Effect of composition of portland cement on heat evolved during hardening. *Industrial & Engineering Chemistry* 24(11), 1207–1214. (Cited on pages 65, 66 and 68.)
- Wright, S. (1934). The method of path coefficients. *The annals of mathematical statistics* 5(3), 161–215. (Cited on page 1.)
- Wu, J. and H. Yang (2013). Efficiency of an almost unbiased two-parameter estimator in linear regression model. *Statistics* 47(3), 535–545. (Cited on pages 3 and 13.)
- Yang, H. and X. Chang (2010). A new two-parameter estimator in linear regression. *Communications in Statistics—Theory and Methods* 39(6), 923–934. (Cited on pages 12, 13, 15 and 38.)

Index

acknowledgment, [i](#)

dedication, [iii](#)

introduction, [1](#)

thesis regulations, [i](#), [iii](#)

 acknowledgment, [i](#)

 dedication, [iii](#)