COMPARISON OF RESTRICTED AND UNRESTRICTED ESTIMATORS IN THE MULTIPLE LINEAR REGRESSION ANALYSIS

Israa SHALTOOT

Master Dissertation

Department of Statistics

Programme in Applied Statistics

Supervisor: Prof. Dr. Sevil ŞENTÜRK

Eskişehir

Eskişehir Technical University

Institute of Graduate Programs

September 2021

ABSTRACT

COMPARISON OF RESTRICTED AND UNRESTRICTED ESTIMATORS IN THE MULTIPLE LINEAR REGRESSION ANALYSIS

Israa SHALTOOT

Department of Statistics

Programme in Applied Statistics

Eskişehir Technical University, Institute of Graduate Programs, September 2021

Supervisor: Prof. Dr. Sevil ŞENTÜRK

This thesis deals with the problem of multicollinearity in the linear regression model. In both classical and restricted linear regression models, when the least squares (LS) method is employed in the presence of multicollinearity, parameter estimates are unstable and have a high variance. In this case, to avoid the negative effects of multicollinearity over the LS it is recommended to use alternative biased estimators instead. In this thesis, Ridge, Contraction, Liu, Two-parameter, Restricted ridge, Restricted contraction, Restricted Liu, and Restricted two-parameter estimators were chosen among biased estimators to be studied and compared as two corresponding groups, with the aim of identifying which group gives better parameter estimates in the case of multicollinearity. Estimators' performance was compared according to matrix mean square error and scalar mean square error. In this study, applications were made with two real data sets known in the literature in order to show which of the estimators had better performance in the light of the theories previously discussed in the literature. These analyses are based on Portland cement data and total national research and development expenditures data using the MATLAB program. In addition to these applications with real data sets, the results obtained are also detailed with a Monte-Carlo simulation study. As a result, it was decided that the most effective estimators are the restricted biased estimators when it comes to the state of multicollinearity.

Keywords: Multicollinearity, Mean square error, Biased estimators, Monte Carlo simulation, Linear restrictions.

ÖZET

ÇOKLU DOĞRUSAL REGRESYON ANALIZINDE KISITLI VE KISITSIZ TAHMİN EDİCİLERİN KARŞILAŞTIRILMASI

Israa SHALTOOT

İstatistik Anabilim Dalı

Uygulamalı İstatistik Bilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Eylül 2021

Danışman: Prof. Dr. Sevil ŞENTÜRK

Bu tez, doğrusal regresyon modelindeki çoklu doğrusal bağlantı problemi ile ilgilidir. Hem klasik doğrusal regresyon hem de kısıtlı doğrusal regresyon modellerinde, çoklu doğrusal bağlantının varlığı durumunda, en küçük kareler (EKK) yöntemi kullanıldığında parametre tahminleri kararsız ve yüksek varyanslı olarak elde edilmektedirler. Bu durumda, çoklu doğrusal bağlantının EKK üzerindeki olumsuz etkilerinden kaçınmak için bunun yerine alternatif yanlı tahmin edicilerin kullanılması önerilmektedir. Bu tez çalışmasında, çoklu doğrusal bağlantı probleminin varlığı durumunda hangi grubun daha iyi parametre tahminleri verdiğini belirlemek amacıyla, Ridge, Contraction, Liu, İki parametreli, Kısıtlı ridge, Kısıtlı contraction, Kısıtlı Liu ve Kısıtlı iki parametreli tahmin edicileri incelenmek üzere yanlı tahmin ediciler arasından seçilmiş ve karşılık gelen iki grup olarak karşılaştırılmıştır. Tahmin edicilerin performansları matris hata kareler ortalaması ve skaler hata kareler ortalamasına göre karşılaştırılmıştır. Bu çalışmada literatürde önceden tartışılan teoriler ışığında tahmin edicilerin hangisinin daha iyi performansa sahip olduğunu göstermek amacıyla literatürde bilinen gerçek iki veri setiyle uygulamalar yapılmıştır. Söz konusu analizler, MATLAB programı kullanılarak, Portland çimento verileri ve toplam ulusal araştırma ve geliştirme harcamaları verileri üzerinedir. Gerçek veri setleri ile yapılan bu uygulamalara ek olarak elde edilen bulgular bir Monte-Carlo simülasyon çalışmasıyla da ayrıntılı bir şekilde verilmiştir. Sonuç olarak, çoklu doğrusal bağlantı durumu söz konusu iken en etkin tahmin edicinin kısıtlı yanlı tahmin ediciler olduğuna karar verilmiştir.

Anahtar Sözcükler: Çoklu doğrusal bağlantı, Ortalama kare hatası, Yanlı tahmin ediciler, Monte-Carlo simülasyonu, Doğrusal kısıtlamalar.

ACKNOWLEDGEMENTS

This dissertation is dedicated to my supervisor Prof. Dr. Sevil Şentürk for her support, guidance, and positive, compassionate words at all times, words can not express my endless respect and gratitude. To whom I owe her a lot, Res. Asst. Nihal İnce for her great assistance throughout the practical implementation, there is always a lot to be learned from her determination. Also, to Ass. Pro. Mohammad Arashi who shared his knowledge and experience very generously. Eventually to whom their presence, even in being far away, is a tremendous power for me, my Mum, Dad, and siblings. To whom that without him, this dissertation would not have been destined to be completed, Wael..

Israa SHALTOOT

03/09/2021

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with "scientific plagiarism detection program" used by Eskişehir Technical University, and that "it does not have any plagiarism" whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

Israa SHALTOOT

CONTENTS

Page

LIST OF TABLES

Page

GLOSSARY OF SYMBOLS AND ABBREVIATIONS

CN	:	Condition Number
LE	:	Liu Estimator
MRL	:	Modified Restricted Liu
MRR	:	Modified Ridge Regression
MMSE	:	Matrix Mean Square Error
SMSE	:	Scalar Mean Square Error
nnd	:	Nonnegative Defined
OLS	:	Ordinary Least Squares
PCR	:	Principal Components Regression
pd	:	Positive Defined
PMMSE	:	Prediction Matrix Mean Square Error
PSMSE	:	Prediction Scalar Mean Square Error
RL	:	Restricted Liu
RLM	:	Restricted Linear Model
RLS	:	Restricted Least Squares
RR	:	Ridge Regression
RRR	:	Restricted Ridge Regression
RTP	:	Restricted Two-Parameter
SRTP	:	Stochastic Restricted Two-Parameter
TP	:	Two-Parameter
VIF	:	Variance Inflation Factor
RC	:	Restricted Contraction
$\boldsymbol{S}$	:	The Correlation Matrix ($\boldsymbol{X'X}$)
WMMSE	:	Weighted Matrix Mean Square Error
WSMSE	:	Weighted Scalar Mean Square Error
EMSE	:	Estimated Mean Square Error
$E(\widetilde{\boldsymbol{\beta}})$	:	The Expected Value of $\widetilde{\boldsymbol{\beta}}$ Estimator
$Var(\widetilde{\boldsymbol{\beta}})$	:	The Variance-Covariance Matrix of $\widetilde{\boldsymbol{\beta}}$ Estimator
$Bias(\widetilde{\boldsymbol{\beta}})$	:	The Bias of $\widetilde{\boldsymbol{\beta}}$ Estimator
$\boldsymbol{\beta}$	:	The Regression Coefficients Vector

$\widetilde{\boldsymbol{\beta}}$	:	Any Estimator of $\boldsymbol{\beta}$ Parameter
$\boldsymbol{\alpha}$	:	The OLS Estimator of $\boldsymbol{\beta}$ Parameter
$\widetilde{\boldsymbol{\alpha}}$	:	The Regression Coefficients in the Canonical Form
$\widehat{\boldsymbol{\alpha}}$	:	Any Estimator of $\boldsymbol{\alpha}$ Parameter
$\boldsymbol{Q}$	:	An Orthogonal Matrix
k	:	Ridge Bias Parameter
d	:	Liu Bias Parameter
λ_i	:	The Eigenvalues of $\boldsymbol{X'X}$ Matrix
$\boldsymbol{\Lambda}$	:	A Diagonal Matrix Consisting of the Eigenvalues of $\boldsymbol{X'X}$ Matrix
$\widehat{\boldsymbol{\beta}}(k)$	:	Ridge Regression Estimator
$\widehat{\boldsymbol{\beta}}(k,b)$	:	: Modified Ridge Regression Estimator
$\widehat{\boldsymbol{\beta}}(d)$	:	: Contraction Estimator
$\widehat{\boldsymbol{\beta}}_d$	:	Liu Estimator
$\widehat{\boldsymbol{\beta}}(k,d)$	:	Two Parameter Estimator
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}$	:	Restricted Least Square Estimator
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(k)$	:	Restricted Ridge Regression Estimator
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}(d)$	:	Restricted Contraction Estimator
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(d)$	:	Restricted Liu Estimator
$\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(k,d)$	:	Restricted Two-Parameter Estimator

(The characters in bold indicate a vector or a matrix)

1. INTRODUCTION

1.1. Overview

Most scientific studies shed the light on the relationships that occur between the variable of interest and other related variables to provide a better understanding for them. In order to do that different statistical procedures are required to take place. The regression analysis method, which was presented first in the 1880s by Sir Francis Galton through his studies of inheritance and eugenics, is one of the most famous methods used to explain the relationship between different variables. As with many statistical analyses, this analysis method aims to summarize the observed data in as simplified and useful way as possible, except for when the regression analysis's key objective is to correctly approximate the values of regression coefficients to match a good model. Regression analysis is frequently used in almost all scientific fields to provide more sense of the relationships between variables, fields such as psychology, economics, physics, management, genetics, and sociology. Mainly, the construction of the regression model is based on a response variable and one or more observed explanatory variables. Moreover, unknown regression parameters are the basis that the functional relationship between the response variable and descriptive variable(s). Therefore, the model can be referred to as a linear model, when the parameters included in the model are the first-order polynomial. This is basically the most fundamental type of linear regression, as it only focuses on two variables, one is dependent and the other is independent. However, when more than one variable has a rule in explaining the dependent variable the result is another type called multiple linear regression model, and this is the model that is focused upon in this study. The matrix that contains all the explanatory variables is called the regressor matrix also known as the design matrix or model matrix and often denoted by $\boldsymbol{X}$. The methods used to estimate the unknown parameters of a regression model are based on the assumptions of regression analysis. The basic assumptions of linear regression analysis are:

I. The regressor matrix is non-stochastic matrix.
II. The rank of the regressor matrix is full column rank, i.e., explanatory variables are linear independent.
III. The error vector is a random variable that its elements cannot be observed.
IV. The response variable is a linear function of the explanatory variables.

V. The error term has a normal distribution with zero mean and constant variance.

VI. Error terms are uncorrelated to each other.

When these assumptions are fulfilled, usually the ordinary least squares (OLS) method is applied to the model, the unknown parameters are estimated and the analyses of interest are performed.

A linear statistical model fitted with the OLS without any doubt is the most commonly used statistical method in the literature (Chatterjee and Hadi, 2009). In order for the OLS to be the best-unbiased estimator, basic assumptions must be held. However, in practice, the assumptions often are encountered with a violation, and in that case, when the OLS method is applied the variances are going to increase although the OLS estimators of regression coefficients are still unbiased linear estimators, at the same time, the covariance will vary. In this case, the estimator's reliability will be reduced, the greater variance and covariance will cause the confidence intervals to be wider, and as a result, the validity of the statistical hypothesis tests will be lost.

In this study, the case when the assumption (II) given on page (1) does not hold true will be discussed. In practical applications, the problem encountered in case of violation of the assumption (II) is called Multicollinearity. Multicollinearity which is defined as the existence of high intercorrelation between regressors is a common problem in the multiple regression models, that creates undesirable effects on the estimator of the least squares. To get around this problem, the need of finding methods that are less sensitive to multicollinearity than the OLS method has arisen. That led to the emergence of biased estimators as estimation alternatives of the OLS. Another method is to use additional sample information with exact or stochastic linear constraints. So, the restricted biased estimation emerged by sitting the restrictions on the form $\boldsymbol{r} = \boldsymbol{R\beta}$ or $\boldsymbol{r} = \boldsymbol{R\beta} + \boldsymbol{\phi}$ to the model's unknown parameters.

Despite the fact that several comparison studies between biased estimators have been done in the literature, this study effort takes a new route in terms of practical application based on theoretical approach. To the best of our knowledge, no simulation studies have been conducted previously comparing the exact linear restricted estimators and the unrestricted biased estimators. Therefore, what distinguishes this work is the

concept of categorizing some biased estimators into two fundamentally separated groups, which shifts the comparison to a higher level of comparison between two linear regression models; the restricted linear regression model and the classical linear regression model, rather than only comparing the performance of a random group of estimators, as is always done.

1.2. Literature Review

Multicollinearity between regressors has several undesirable effects on the least square regression method. Such a problem has been discussed extensively in the literature, and several alternative approaches and remedial biased estimators have been presented to address the problem of multicollinearity. One of the oldest estimators was presented by Stein (1956). Then, another well-known technique was designed by Massy (1965) called principal components regression (PCR) estimator. This estimator's technique is based on selecting and using a subset of the column space of explanatory variables and projecting the response variable onto this subset. One of the most popular estimators is the ridge regression (RR) estimator proposed by Hoerl and Kennard (1970). The RR estimator is a well-known estimator of adding bias to an unbiased estimator in order to reduce the mean square error of parameter estimates. Swindel (1976) introduced modified ridge (MR) estimator by adding prior information to the ridge regression estimator. Another popular estimator is Liu (1993), Liu's biased estimator considers as an adjustment obtained by applying the Stein estimator to a particular state of the RR estimator. Since the bias parameter d of Liu estimator (LE) is a linear function, it is easier to select than ridge's bias parameter k, which is a complex function, then as a result it considers superior to the RR estimator. This estimator was called the Liu estimator for the first time by Akdeniz and Kaçiranlar (1995). Similar to modified ridge, Li and Yang (2012) proposed a new estimator called the modified Liu estimator based on prior information.

Baye and Parker (1984) developed the $r-k$ class estimator by merging the previously mentioned PCR and ridge estimators. This new estimator is considered as a general estimator which involves OLS, PCR and ridge estimators as special cases.

Kaçıranlar and Sakallıoğlu (2001) took the superiority of the Liu estimator over the ridge estimator into consideration and proposed the $r-d$ class estimator as a counterpart to the $r-k$ class estimator by combining the PCR estimator with the Liu estimator. Also,

as the former combination this estimator involves OLS, PCR and Liu estimators as special cases.

Liu (2003) added a second bias parameter to the Liu estimator creating what is known as the Liu type estimator in order to improve the performance of the Liu estimator. In this estimator, one of the bias parameters is used to eliminate the ill-condition of the design matrix, while the other bias parameter is used to increase the goodness of fit of the model. Then later by grafting the contraction estimator into Swindel's modified ridge estimator, Özkale and Kaciranlar (2007) introduced the two-parameter (TP) estimator, which considers a general case of the OLS, the ridge, the Liu, and the contraction estimators. Under the same logic few years later Yang and Chang (2010) developed another two-parameter alternative estimator based on two bias parameters, and gave its necessary and sufficient conditions for the superiority over the two-parameter estimator proposed by Özkale and Kaciranlar (2007) in terms of the matrix mean square error (MMSE). This estimator includes the OLS, Ridge and Liu estimators as special cases too.

Considering that the combination of two distinct estimators might inherit both estimators' advantages, Sarkar (1992) led to another biased estimator called the restricted ridge regression (RRR) estimator by grafting the RR estimator into the restricted least square (RLS) estimator. Years later, Groß (2003) stated that Sarkar's restricted estimator does not satisfy the exact restriction $\boldsymbol{R\beta} = \boldsymbol{r}$, then he introduced another restricted ridge regression estimator and derived the necessary and sufficient conditions for his new estimator's superiority over the restricted least-squares estimator. By developing the approach of Liu, Kaçiranlar et al. (1999) managed to find the restricted Liu estimator.

Comparisons in the literature

The diversity of the estimates that the deficiency of the OLS estimator in the presence of the multicollinearity have led to, opened the door to many comparisons to be conducted to identify the conditions of superiority either over the OLS estimator or any other estimator have been suggested earlier in the literature, some of the most relative comparisons to my research are as follows:

Kaciranlar (1998) compared the modified ridge regression (MRR) estimator, with the restricted ridge regression estimator presented by Sarkar (1992) in the sense of the

MMSE and gave sufficient and necessary conditions for the MMSE of the MRR estimator to exceed the MMSE of the RRR estimator.

Sakallioğlu et al. (2001) compared the performance of Liu ($\widehat{\boldsymbol{\beta}}_d$) and ridge regression ($\widehat{\boldsymbol{\beta}}_k$) estimators with respect to the MMSE where Liu (1993) himself did not compare his estimator's mean square error properties with that of RR estimator.

Akdeniz and Erol (2003) compared the almost unbiased generalized ridge regression estimator with the almost unbiased generalized Liu estimator in sense of the MMSE.

Özkale and Kaciranlar (2007) introduced a two-parameter estimator that is considered a general case for the OLS estimator, the contraction estimator, the RR estimator, and the Liu estimator. Then they gave a linear restricted version of it, and compared it theoretically in the sense of the MMSE criterion, once with the OLS estimator another with its restricted cases.

Yang and Chang (2010) introduced a new two parameter (NTP) estimator that includes the OLS estimator, the RR estimator, and the Liu estimator as a special cases and gave the necessary and sufficient conditions for the superiority of their estimator over the OLS, RR, Liu estimators, and the two parameter of Özkale and Kaciranlar (2007).

Li and Yang (2011) presented two sorts of restricted estimators based on prior information for the vector of parameters in a linear regression model with linear restrictions: the restricted modified Liu (RML) estimator and the restricted modified ridge (RMR) estimator and compared their performance in the terms of the MMSE.

Yang and Cui (2011) introduced a new two parameter estimator for the unknown parameter vector in the linear regression model. This new estimator which is called the stochastic restricted two-parameter (SRTP) estimator is obtained by combining the mixed estimator and the two-parameter estimator of Özkale and Kaciranlar (2007). The SRTP estimator simulates the RTP estimator; as the RTP estimator was implemented when exact linear restrictions are assumed to hold, the SRTP estimator was implemented when stochastic linear restrictions are assumed to hold. Furthermore, as Özkale and Kaciranlar (2007) compared between the RTP and the RLS estimators and gave the superiority's necessary and sufficient conditions of the RTP estimator over the RLS estimator, the founders of SRTP estimator compared it with the mixed estimator (ME) and the TP

estimator and gave the necessary and sufficient conditions of the superiority of their estimator over both the ME and the TP estimators.

Dorugade (2014) suggested a modified two-parameter estimator. The proposed estimator's properties are discussed, as well as its performance in terms of the matrix mean square error criterion over the OLS estimator, the NTP of Yang and Chang (2010), an almost unbiased two parameter estimator (AUTP), and other well-known estimators.

Recently, after Lukman et al. (2019) introduced the modified ridge-type (MRT) estimator, they compare it with the OLS, the RR, the MRR, and Dorugade's two parameter estimator. Lukman et al. (2020) suggested an unbiased modified ridge-type (UMRT) estimator with prior information and discussed the performance of the new estimator over the OLS estimator, the RR estimator and the modified ridge-type estimator (MRT) using the matrix mean square error criterion.

1.3. Purpose Statement

The purpose of this study is to examine several different classes of biased estimators when the input matrix of the design is ill-conditioned. These estimators are the ridge regression estimator (1970), the contraction estimator (1973), the Liu estimator (1993), the two-parameter estimator (2007), and their restricted cases subject to the exact constraint $\boldsymbol{R\beta = r}$. Then with the aim of conducting a distinct comparison from comparisons found profusely in the literature, the concerned estimators have been compared after separating them into two key groups one under the classic linear regression model the other under the restricted linear regression model. The comparison was applied once theoretically by using the criterion of the MMSE, and once practically by conducting a simulation study and two different real-life applications based on theories that have not been previously applied in the literature, to the best of the author's knowledge.

1.4. Organization of the Study

In this study, the chapters are arranged as follows: in Chapter 2, some preliminaries that are needed in subsequent chapters were introduced, involving the definition of the linear regression model, linear estimators, and some common criteria for the evaluation of the goodness of estimators. In Chapter 3, a deep explanation of the problem of multicollinearity includes its definition, reasons of occurrence, effects on the linear

model, and some ways to deal with and overcome it were presented. While in Chapter 4 the ordinary least squares estimator, and some restricted and unrestricted biased estimators were presented and discussed, which are as follows; As unrestricted biased estimators: the ridge regression estimator, the contraction estimator, the Liu estimator, and the two-parameter estimator. As restricted biased estimators: the restricted least squares estimator, the restricted ridge regression estimator, the restricted contraction estimator, the modified restricted Liu estimator, and the restricted two-parameter estimator. In Chapter 5, some theories that summarize the comparisons between the restricted and the unrestricted estimators under the MMSE criterion were shown. In Chapter 6, two numerical examples based on theoretical comparisons provided in the previous chapter were undertaken to evaluate the performance of estimators in respect of both the scalar mean square error (SMSE) and the matrix mean square error (MMSE) criteria. In Chapter 7, a Monte-Carlo simulation was implemented by MATLAB. In Chapter 8, all obtained results have been summarized in the conclusion.

2. PRELIMINARIES

One of the primary aims of statistical research is to study causation and the impact of changes in estimators or independent variables on the dependent variable. Linear models, which are widely employed in describing the relationship between variables, inferring conclusions and making predictions about subjects, are utilized to accomplish this aim too. In this chapter, some preliminary information for the linear regression model that will be needed in subsequent sections are presented, some criteria used to determine the goodness of estimators are given too.

2.1. Linear Models

Assume that the outcome of any procedure is denoted by $\boldsymbol{y}$ variable, that is called the dependent variable, since it depends on p of explanatory variables denoted by $\boldsymbol{X}_i,\ i = 1, \dots, p$. Then the behavior of $\boldsymbol{y}$'s can be expressed by a relationship given by

$$\begin{aligned} \boldsymbol{y} &= f\left(\boldsymbol{X_1}, \dots, \boldsymbol{X_p}, \beta_1, \dots, \beta_p\right) + \boldsymbol{e} \\ &= \beta_1 \boldsymbol{X_1} + \cdots + \beta_p \boldsymbol{X_p} + \boldsymbol{e}. \end{aligned} \tag{2.1}$$

Here f denotes a well-defined function, β_i are the parameters that describe the function and contribution of the independent variables, while the stochastic nature of the relationship between $\boldsymbol{y}$ and the $\boldsymbol{X}$'s is reflected by the term $\boldsymbol{e}$. When $\boldsymbol{e}$ is equal to 0, the relationship (2.1) is referred to as the mathematical model; otherwise, it is referred to as the statistical model. Also, if the relationship (2.1) is linear in parameters, it is called linear; otherwise, it is called nonlinear (Rao et al., 2008).

Since $\boldsymbol{y}$ is the outcome of a predetermined variables $\boldsymbol{X}_1, \dots, \boldsymbol{X}_p$, hence both are known, as a result the description of a linear model relies on the knowledge of the parameters. To determine these parameters many statistical estimation approaches are found. For instance, method of maximum likelihood (ML), method of moments, least squares method etc. The expression linear model can be used in a number of science fields depending on the context. Statistically, some of the underlying models of this term are the linear regression model and the time series model, in this study we are mainly focusing on the linear regression model.

2.2. Linear Regression Model

The outcome $\boldsymbol{y}$ or the study variable can be influenced by one or more independent variable(s). Nonetheless, in real life experiments it is rare for a study variable to be

affected by only one predictor variable. The linear regression model consisting of more than independent variable is referred to as multiple. The standard multiple linear regression model is expressed as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{2.2}$$

where $\boldsymbol{y}$ is an $n \times 1$ observation vector of dependent variable, $\boldsymbol{X}$ is an $n \times p$ full column rank observation matrix of p independent variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of independent and identically distributed $(0, \sigma^2)$ random errors. Whereas n is the observations' number, p is the independent variables' number (Montgomery et al., 2021).

The main objective of multiple linear regression revolves around various points, some of them according Rencher and Schaalje (2008) are the prediction, the description or explanation of observed data, the estimation of the parameters, selection of the independent variables and controlling the output. In addition, regression analysis is widely used for purpose of forecasting and control, where its applications overlap considerably with machine learning. In some other cases it can be used to investigate causative relationships between the independent and dependent variables. Assumptions of the model (2.2) can be listed as follows, see (Rao et al., 2008):

- $\boldsymbol{X}$ is a non-stochastic matrix, while $\boldsymbol{X'X}$ is a non-singular matrix with $rank(\boldsymbol{X'X}) = p$.
- $rank(\boldsymbol{X}) = p$. This means the column vectors of the matrix $\boldsymbol{X}$ are linear independent. There are no linear or near-linear relationships between regressors. If this assumption is not met, the parameter estimates will be inconsistent, and the variances of the estimates may be large.
- $n \geq p$. In order for $\boldsymbol{X}$ matrix to have a full column rank, the number of observations must be greater than the number of regressor in the model.
- $E(\boldsymbol{e}) = 0$. In words, the mean of the error term is equal to zero. In cases where this assumption is not met, parameter estimates are larger or smaller than their actual values which means biased parameter estimates (Graybill, 1961).
- $Var(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$. The variance of the error term is constant and does not change depending on the changes in the independent variable, i.e. the errors are of constant variance (Myers, 1990). Also, $cov(e_i, e_j) = 0, i, j = 1,2, \dots, n$ that is, the

errors are unrelated to each other. If this assumption is not fulfilled, then the error terms include autocorrelation.

- $\lim_{n\to\infty}(\boldsymbol{X'X}/n) = \Delta$. An assumption of $\boldsymbol{X'X}/n$ matrix being convergent as the observations number approaches infinity. The limit of $\boldsymbol{X'X}/n$ matrix must be a non-singular and non-stochastic matrix.

When these assumptions are fulfilled over the regression model, the OLS method is often used to estimate the unknown regression parameters. In this manner, parameter estimates are found in a way that minimizes the sum of error squares as much as possible.

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} e_i^2 \\ &= \boldsymbol{e}'\boldsymbol{e} \\ &= (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}, \end{aligned} \tag{2.3}$$

if we take the partial derivatives of function (2.3) with respect to $\boldsymbol{\beta}$ and make it equal to zero

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{X\beta} - \boldsymbol{X}'\boldsymbol{y} = 0,$$

$$\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{X}'\boldsymbol{y}, \tag{2.4}$$

and thus, the OLS equation is obtained as

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{2.5}$$

where $\boldsymbol{X'X}$ is a $p \times p$ correlation matrix between the independent variables and $\boldsymbol{X}'\boldsymbol{y}$ is a $p \times 1$ correlation vector between the independent variables and the dependent variable. If $\boldsymbol{X}$ is not a linearly independent matrix, then $(\boldsymbol{X'X})^{-1}$may not be exist. In such a case, some different solutions of the normal equation (2.4) will be available.

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{G}\boldsymbol{X}'\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{G}\boldsymbol{X}'\boldsymbol{X})\boldsymbol{w}, \tag{2.6}$$

where $\boldsymbol{w}$ is an arbitrary vector, $\boldsymbol{G} = (\boldsymbol{X'X})^{-}$ is the generalized inverse of the $\boldsymbol{X'X}$ matrix, and since it is not a unique matrix the normal equation (2.4) has more than one solution (Rao et al., 2008). Therefore, $\widehat{\boldsymbol{\beta}}$ can no longer be considered as an estimator. Hence, it is preferable to use another estimation method for the regression coefficients.

The experimental predictor of $\boldsymbol{y}$ is

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}. \tag{2.7}$$

For all $\widehat{\boldsymbol{\beta}}$ solutions of $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{y}$, $\widehat{\boldsymbol{y}}$ has the same value. The estimator of the error term $\boldsymbol{e}$ is obtained as follows, where $\widehat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}$ parameter

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}. \tag{2.8}$$

When the estimated error sum of squares is divided by $n - p$, an unbiased and consistent estimator of σ_e^2 is obtained, as follows:

$$\hat{\sigma}_e^2 = \frac{\hat{\boldsymbol{e}}'\,\hat{\boldsymbol{e}}}{n-p}. \tag{2.9}$$

The properties of $\widehat{\boldsymbol{\beta}}$ estimator are as follows:

$$E(\widehat{\boldsymbol{\beta}}) = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}X'y] = (\boldsymbol{X}'\boldsymbol{X})^{-1}X'E(y) = \beta, \tag{2.10}$$

$$\begin{aligned} var(\widehat{\boldsymbol{\beta}}) &= var\,[(\boldsymbol{X}'\boldsymbol{X})^{-1}X'y] \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}X'\,\mathrm{var}(\boldsymbol{y})X(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}, \end{aligned} \tag{2.11}$$

where $E(\widehat{\boldsymbol{\beta}})$ and $var(\widehat{\boldsymbol{\beta}})$ are the expected value and the covariance matrix of $\widehat{\boldsymbol{\beta}}$ estimator, respectively. The OLS estimator $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator that has the smallest variance among all other linear unbiased estimators. And according to the Gauss-Markov theorem the OLS estimator is considered the best linear unbiased estimator (BLUE) for $\boldsymbol{\beta}$ parameter (Albert, 1973). Therefore, due to its suitable statistical properties it has been used for a long time. But, if the model's assumptions are not met, the parameters estimated by the OLS method may not be reliable. Although theoretically some results can be obtained under the assumption that the columns of the $\boldsymbol{X}$ matrix are linear independent, applications are generally encountered in cases where the columns of the design matrix are linearly dependent. In this case, the problem of multicollinearity arises, and the matrix $\boldsymbol{X}'\boldsymbol{X}$ becomes ill-conditioned. In datasets with multicollinearity, the OLS estimates of the parameters are known to be great in absolute values (Montgomery et al., 2012). In addition, the sample variances increase. As a result, the obtained confidence intervals of the parameters become wider, and significant regression coefficients can be removed. Due to such problems faced when the OLS method is used for estimation of regression parameters, different methods of estimation become needed.

2.3. Linear Estimators

Estimating or finding approximate values of the real yet undefined vector of regression parameters $\boldsymbol{\beta}$ in the linear regression model is now the mission to pay attention to, based on the assumptions stated in the previous section and $(\boldsymbol{y}, \boldsymbol{X})$ observations. This can be done by substituting the actual values with suitable estimators $\widetilde{\boldsymbol{\beta}}$. The choosing of linear estimators in $\boldsymbol{y}$ can be justified that linear estimators are more easily suited to statistical analysis, particularly the investigation of consistency, unbiasedness, the confidence intervals' construction, and so on.

Definition 2.3.1. *The linear estimators of the parameter vector* $\boldsymbol{\beta}: p \times 1$ *is defined as:*

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{a}, \tag{2.12}$$

here $\boldsymbol{A}: p \times n$ *and* $\boldsymbol{a}: p \times 1$ *are two non-stochastic matrices. if* $\boldsymbol{a} = 0$ *the linear estimator* $\widetilde{\boldsymbol{\beta}}$ *is called homogeneous linear estimator, else* $\widetilde{\boldsymbol{\beta}}$ *is called heterogeneous linear estimator of* β *(Rao et al., 2008).*

From this definition, it is seen that the OLS estimator of $\boldsymbol{\beta}$ is a homogeneous linear estimator with $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{a} = 0$.

2.4. Criteria Used to Determine the Goodness of Estimators

As the model's goodness-of-fit usually being measured by the sum of squared errors, the goodness of a linear estimator $\widetilde{\boldsymbol{\beta}}$ can be assessed by using the quadratic loss function. Here we present the definition of the loss function:

Definition 2.4.1. *Let* $\widehat{\boldsymbol{\theta}}$ *be an estimator of* $\boldsymbol{\theta}$ *parameter in the parameter space* Ω*. The function* $L(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ *which satisfies the following features is called the loss function:*

a) $\forall\, \boldsymbol{\theta} \in \Omega$ *and* $\forall\, \widehat{\boldsymbol{\theta}}, L(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) > 0$,

b) For $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, $L(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = 0$.

Bhattacharya (1966) defined the standard loss function as

$$L(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) = \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|^2. \tag{2.13}$$

Definition 2.4.2. *Let the loss function of* $\widetilde{\boldsymbol{\beta}}$ *be represented as* $L(\widetilde{\boldsymbol{\beta}})$*, where* $\widetilde{\boldsymbol{\beta}}$ *is an estimator of* $\boldsymbol{\beta}$*. Then the expected loss of* $\widetilde{\boldsymbol{\beta}}$ *estimator under the loss function is as follows:*

$$R(\tilde{\beta}) = E[L(\tilde{\beta})]. \tag{2.14}$$

The expected loss of an estimator is also known as the risk of this estimator. For more about loss function and balanced loss functions see (Dey et al., 1999) and for quadratic loss see (James and Stein, 1992).

In the subsequence sections, some of used criteria to evaluate the performance of the linear estimators that can be classified under the loss function.

2.4.1. Matrix loss function and mean square error

The matrix loss function of $\widetilde{\boldsymbol{\beta}}$ estimator of $\boldsymbol{\beta}$ parameter is defined as

$$\mathrm{L}(\widetilde{\boldsymbol{\beta}}) = (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})', \tag{2.15}$$

according to this loss function the risk or the matrix mean square error is:

$$MMSE(\widetilde{\boldsymbol{\beta}}) = E\left[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right]. \tag{2.16}$$

Since the bias of $\widetilde{\boldsymbol{\beta}}$ estimator i.e., the difference between $\boldsymbol{\beta}$ and $E(\widetilde{\boldsymbol{\beta}})$, is presented as $Bias(\widetilde{\boldsymbol{\beta}})$ then (2.16) equation can be expressed as follows:

$$MMSE(\widetilde{\boldsymbol{\beta}}) = var(\widetilde{\boldsymbol{\beta}}) + Bias\,(\widetilde{\boldsymbol{\beta}})Bias\,(\widetilde{\boldsymbol{\beta}})', \tag{2.17}$$

where $var(\widetilde{\boldsymbol{\beta}})$ is the dispersion matrix as

$$Var(\widetilde{\boldsymbol{\beta}}) = E\left[\left(\widetilde{\boldsymbol{\beta}} - E(\widetilde{\boldsymbol{\beta}})\right)\left(\widetilde{\boldsymbol{\beta}} - E(\widetilde{\boldsymbol{\beta}})\right)'\right]. \tag{2.18}$$

$$Bias(\widetilde{\boldsymbol{\beta}}) = E(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}. \tag{2.19}$$

See (Trenkler, 1983) (Toutenburg and Trenkler, 1990) and (Rao and Toutenburg, 1995).

2.4.2. Scalar loss function and mean square error

The scalar loss function of $\widetilde{\boldsymbol{\beta}}$ estimator of $\boldsymbol{\beta}$ parameter is defined as follows:

$$L(\widetilde{\boldsymbol{\beta}}) = (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{2.20}$$

the risk obtained according to this loss function is called scalar mean square error

$$SMSE(\widetilde{\boldsymbol{\beta}}) = E\left[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\right], \tag{2.21}$$

the $SMSE$ of an estimator can be expressed also as the trace of the MMSE. This can be written in the following formula

$$SMSE(\widetilde{\boldsymbol{\beta}}) = tr\,E\left[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] = tr\,MMSE(\widetilde{\boldsymbol{\beta}}). \tag{2.22}$$

The use of SMSE means that the non-diagonal elements of the MMSE are ignored and all their differences $(\widetilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta})$ are equally weighted. Therefore, MMSE is a more

comprehensive and more preferred criterion to statisticians than the SMSE. However, since it is more difficult to be calculated, SMSE is generally more used in practice.

2.4.3. Predictive loss function and mean square error

Consider a linear regression model $\boldsymbol{y}\sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ where $\boldsymbol{y}$ is $n \times 1$and $\boldsymbol{\beta}$ is $p \times 1$. Then the predictive loss function of $\widetilde{\boldsymbol{\beta}}$ estimator of $\boldsymbol{\beta}$ is defined as

$$L_p(\widetilde{\boldsymbol{\beta}}) = (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}), \tag{2.23}$$

see (Gelfand and Ghosh, 1998). Generally, the estimation of $\boldsymbol{y}$ for future $\boldsymbol{X}$ values is made using the $\boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ predictor. Now the risk according to (2.23) loss function is

$$PSMSE(\widetilde{\boldsymbol{\beta}}) = E\left[(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})\right], \tag{2.24}$$

and it is also called the predictive scalar mean square error (PSMSE). Meanwhile, the predictive matrix mean square error (PMMSE) is as follows

$$PMMSE(\widetilde{\boldsymbol{\beta}}) = E[(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})']. \tag{2.25}$$

2.4.4. Weighted quadratic loss function and its quadratic risk

Let $\boldsymbol{W}: p \times p$ be a positive defined symmetric matrix, then the weighted quadratic loss function of $\widetilde{\boldsymbol{\beta}}$ estimator of $\boldsymbol{\beta}$ is defined as

$$L_w(\widetilde{\boldsymbol{\beta}}) = (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{W}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{2.26}$$

see (Bhattacharya, 1966; Stein, 1956). If $\boldsymbol{W} = \boldsymbol{I}_p$, then the weighted quadratic loss function equals the scalar loss function ($L_w(\widetilde{\boldsymbol{\beta}}) = SMSE(\widetilde{\boldsymbol{\beta}})$). In prediction problems, it can sometimes be useful to use the $\boldsymbol{W} \neq \boldsymbol{I}_p$. For instance, if the diagonal elements of $\boldsymbol{W}$ are chosen to be different from each other, then the $(\widetilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^2$ distance squares will have a greater or lesser effect on the total loss. Corresponding to (2.26) loss function the quadratic risk or the weighted scalar mean square error (WSMSE) is

$$WSMSE(\widetilde{\boldsymbol{\beta}}) = E\left[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{W}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\right]. \tag{2.27}$$

The weighted matrix mean square error (WMMSE) is

$$WMMSE(\widetilde{\boldsymbol{\beta}}) = E[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{W}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})']. \tag{2.28}$$

If $\boldsymbol{W} = \boldsymbol{X}'\boldsymbol{X}$, then $WMMSE(\widetilde{\boldsymbol{\beta}})$ is referred to as predictive matrix mean square error (Rao et al., 2008).

2.5. Some Definitions and Theorems

The following definitions and theorems are considerably used in studies that aim to determine the goodness of estimators and will be needed in subsequent chapters.

Definition 2.5.1. *Assume that* $\boldsymbol{A}: m \times m$ *be a symmetric matrix and* $\boldsymbol{x}: m \times 1$ *be any vector. Then in* $\boldsymbol{x}$*, the quadratic form is defined as*

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i,j} a_{ij}x_i x_j. \tag{2.29}$$

Definition 2.5.2. *If the quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, $\forall x \neq 0$, *then* $\boldsymbol{A}$ *is called a positive definite (pd) matrix i.e.,* $\boldsymbol{A} > 0$ *(Rao et al., 2008).*

Definition 2.5.3. *If the quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ $\forall x \neq 0$, *then* $\boldsymbol{A}$ *is called a non-negative definite (nnd) matrix i.e.,* $\boldsymbol{A} \geq 0$ *(Rao et al., 2008).*

Definition 2.5.4. *A square matrix* $\boldsymbol{A}: m \times m$ *is considered to be orthogonal if* $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I} = \boldsymbol{A}'\boldsymbol{A}$ *(Rao et al., 2008).*

Definition 2.5.5. *Let* $\boldsymbol{A}: m \times m$ *be a square matrix, then* $q(\lambda) = |\boldsymbol{A} - \lambda\boldsymbol{I}|$ *is a* m^{th} *order polynomial in* λ*, the* m *roots* λ_1*, ...,* λ_m *of the characteristic equation* $q(\lambda)$ *which satisfies the equation* $|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$ *are called eigenvalues of* $\boldsymbol{A}$ *matrix (Rao et al., 2008).*

Definition 2.5.6. *Let* $\boldsymbol{A}: m \times m$ *and* $\boldsymbol{B}: m \times m$ *be any matrices. Then the roots* $\lambda_i = \lambda_i^B(\boldsymbol{A})$ *of the equation* $|\boldsymbol{A} - \lambda\boldsymbol{B}| = 0$ *are called the eigenvalues of* $\boldsymbol{A}$ *matrix in* $\boldsymbol{B}$ *(Rao et al., 2008).*

Definition 2.5.7. *Let* $\widetilde{\boldsymbol{\beta}}_1$ *and* $\widetilde{\boldsymbol{\beta}}_2$ *be two estimators of the parameter* $\boldsymbol{\beta}$*. If* $R(\widetilde{\boldsymbol{\beta}}_2) - R(\widetilde{\boldsymbol{\beta}}_1) \geq 0$ *then the* $\widetilde{\boldsymbol{\beta}}_1$ *estimator is superior to the* $\widetilde{\boldsymbol{\beta}}_2$ *estimator under the loss function* $L(\widetilde{\boldsymbol{\beta}})$ *(Rao et al., 2008).*

Theorem 2.5.1. *Let* $\boldsymbol{A}$ *be an* $m \times m$ *matrix and* $\boldsymbol{G}$ *be a* $n \times m$ *matrix of* $Rank(\boldsymbol{G}) = m \leq n$*. Then* $\boldsymbol{G}\boldsymbol{A}\boldsymbol{G}' \geq 0$ if and $only$ if $\boldsymbol{A} \geq 0$ *(Rao and Toutenburg, 1995).*

Theorem 2.5.2. *Let* $\boldsymbol{A}: m \times m$ *be a symmetric pd matrix,* $\boldsymbol{a}: m \times 1$ *be a vector and* $\alpha > 0$ *be an arbitrary scalar. Then* $\alpha\boldsymbol{A} - \boldsymbol{a}\boldsymbol{a}' \geq 0$ *if and only if* $\boldsymbol{a}'\boldsymbol{A}^{-1}\boldsymbol{a} \leq \alpha$ *(Farebrother, 1976).*

Theorem 2.5.3. *Let* $\boldsymbol{A}: n \times n$ *be symmetric matrix,* $\boldsymbol{a}: n \times 1$ *be a vector and* $\alpha > 0$ *be any scalar. Then the following statements are equivalent*

(i) $\alpha\boldsymbol{A} - \boldsymbol{a}\boldsymbol{a}' \geq 0$.

(ii) $\boldsymbol{A} \geq 0$, $A \in \mathcal{R}(\boldsymbol{A})$ *and* $\boldsymbol{a}'\boldsymbol{A}^{-}\boldsymbol{a} \leq \alpha$, *with* $\boldsymbol{A}^{-}$ *being any generalized inverse of* $\boldsymbol{A}$ *matrix (Rao and Toutenburg, 1995).*

Theorem 2.5.4. *Let* $\boldsymbol{A}: n \times n$ *and* $\boldsymbol{B}: n \times n$ *be any two matrices, when* $\boldsymbol{A} > 0$ *or (* $\boldsymbol{A} \geq 0$*) and* $\boldsymbol{B} > 0$*, then* $\boldsymbol{A} - \boldsymbol{B} \geq 0$ *if and only if* $\lambda_i^B(\boldsymbol{A}) \leq 1 \ \ (i = 1, \dots, n)$ *(Rao and Toutenburg, 1995).*

Theorem 2.5.5. *Let* $\boldsymbol{A}: n \times n$ *and* $\boldsymbol{B}: n \times n$ *be any two matrices. If* $\boldsymbol{A} \geq 0$ *and* $\boldsymbol{B} > 0$*, then* $\lambda_i^B(\boldsymbol{A}) \geq 0$ *(Rao et al., 2008).*

3. MULTICOLLINEARITY

In the multiple linear regression model, it is assumed that the explanatory variables which form the columns of the design matrix are independent. However, in practice, there can usually be a linear or near-linear relationship among the columns of the $\boldsymbol{X}$ matrix. This case of independence assumption's violation of the explanatory variables is referred to as the multicollinearity problem.

Let $\boldsymbol{X}_j$ be the j-th column vector of $\boldsymbol{X}$ matrix and suppose that there is a set of constants $\{t_1, t_2, \dots, t_p\}$. Such that the $\boldsymbol{X}$ matrix's vectors are put in such a form

$$\sum_{j=1}^{p} t_j \boldsymbol{X}_j = 0, \tag{3.1}$$

if at least one of the scalars t_j is not zero, then the $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_p\}$ vectors are linearly dependent. When Eq. (3.1) holds exactly for a subset of $\boldsymbol{X}$'s columns, the model is called a full multicollinear model. In such a case $Rank(\boldsymbol{X}'\boldsymbol{X}) < p$, therefore the $\boldsymbol{X}'\boldsymbol{X}$ matrix will be singular which means that $(\boldsymbol{X}'\boldsymbol{X})^{-1}$matrix cannot be found. On the other hand, if the equation (3.1) is approximately true for some subsets of the columns of the $\boldsymbol{X}$ matrix, then the $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_p\}$ vectors are approximately linearly dependent, and the model is called an approximate multicollinear model. If the model suffers from multicollinearity, then a negligible change in $\boldsymbol{X}'\boldsymbol{X}$ matrix leads to a significant change in $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ matrix. Likewise, when some diagonal elements of the matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ grow too large, some elements of the OLS estimator will have a considerable variance. The explanatory variables are considered orthogonal if there are no linear relationships between them and their correlation matrix is indicated by $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$ (Montgomery et al., 2012).

3.1. Reasons of the Multicollinearity

There are many reasons for multicollinearity. Some of these are as follow (Montgomery et al., 2012):

1. Data collection used technique: Multicollinearity problem may appear when the analyst samples only a subspace of the independent variables' region defined, even though it is not the case in the population.
2. Constraints on the model or in the population: Multicollinearity resulting from physical constraints on the model or in the population is performed by preserving the real relationship, that is, strong dependency that exists in the population as well in the sample.
3. Model specification: If the range of the explanatory variables in the $\boldsymbol{X}$ matrix is small, multicollinearity problem is said to exist when polynomial term is being added to the regression model.
4. An over-defined model: When the number of the variables is over the number of observations, the model becomes over-defined. These types of models are mostly encountered in fields such as behavioral sciences (Ethology) and medicine. In case the model is over-defined, in order of priorities some of the independent variables must be removed from the model.

3.2. Some Ways to Detect the Multicollinearity

There are many ways to detect multicollinearity. Some of those are as follows:

1. Correlation coefficients between independent variables: The non-diagonal elements of the $\boldsymbol{X'X}$ matrix are called correlation coefficients and are denoted by r_{ij}. Geometrical; r_{ij} is the cosine of the angle between vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ (Farrar and Glauber, 1967). If r_{ij} is close to one, then there is a high degree of multicollinearity between the vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$.
2. The determinant of the correlation matrix: The determinant value of $\boldsymbol{X'X}$ is in the range $[0,1]$. If $|\boldsymbol{X'X}| = 0$, the column vectors of the $\boldsymbol{X}$ matrix are said to be linearly dependent. When $|\boldsymbol{X'X}| = 1$, it is said that the columns of the $\boldsymbol{X}$ matrix are orthogonal, but nothing can be declaring about linear dependence of the independent variable. Briefly, as the $|\boldsymbol{X'X}|$ value approaches zero, the degree of multicollinearity increases (Farrar and Glauber, 1967).

3. Coefficient of determination: The coefficient of determination obtained from the regression of $\boldsymbol{X}_j$ variable on the other $p-1$ independent variable is denoted by R_j^2. Which is represented as

$$R_j^2 = 1 - C_{jj}^{-1}, \;\; j = 1,2,\ldots,p \tag{3.2}$$

where C_{jj}^{-1} is the diagonal elements of the inverse of the correlation matrix. If the coefficient of determination is close to one, then a linear dependence is said to exist between $\boldsymbol{X}_j$ and a subset of the remaining columns of the $\boldsymbol{X}$ matrix, and if the columns of $\boldsymbol{X}$ are orthogonal to each other, then the values of R_j^2 equal zero and this means that the multicollinearity problem does not exist.

4. Variance Inflation Factor (VIF): The j-diagonal element of the $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ matrix, which was proposed by (Farrar and Glauber, 1967) to determine the multicollinearity problems, was named VIF by (Marquaridt, 1970). If C_{jj} in (3.2) equation is taken as $C_{jj} = VIF_j$, then VIF_j can be written this way $VIF_j = \frac{1}{1-R_j^2}$. In linear models' studies, the number of independent variables the number of VIF values are. If there is no linear dependence between $\boldsymbol{X}_j$ and other independent variables, the R_j^2 value will be very small, then VIF_j value approaches one. On the other hand, if there is a linear dependence between $\boldsymbol{X}_j$ and other independent variables, then the value of R_j^2 will be close to one, then VIF_j value will be very large. If any *VIF* value is greater than 10, then it is said that multicollinearity exists (Rawlings et al., 2001).
5. Condition number: The condition number of a matrix can be used to define the internal relationship between regressors. If a very small change in the non-singular $\boldsymbol{A}$ matrix corresponds to a very large change in the $\boldsymbol{A}^{-1}$ matrix, then $\boldsymbol{A}$ matrix is called ill-conditioned. By taking this situation into account it can be said that the ill-condition of the model and its internal relationship are two interrelated things. Thus, the condition number also used to measure the degree of ill-condition. Let λ_i be the largest eigenvalue, and λ_j be the smallest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$ matrix, then:

$$Condition\ Number\ (CN) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{\lambda_i}{\lambda_j}. \tag{3.3}$$

The condition number measures the sensitivity of the regression against small changes in the data (Montgomery et al., 2012). If $\lambda_{min} = 0$ then the

condition number is infinite which means perfect multicollinearity between the regressors. If $\lambda_{max} = \lambda_{min}$ then the condition number equals 1 and the regressors can be stated to be orthogonal. Severe multicollinearity is demonstrated by the large values of condition number. Usually, a condition number smaller than 100 indicates no multicollinearity among the regressors, a condition number between 100 and 1000 indicates moderate degree of multicollinearity, and a condition number greater than 1000 shows high degree of multicollinearity. Worthy of mention, in some studies we have observed that some researchers use another relevant formula for the condition number along with another grading scales of multicollinearity, given as $\sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$, see (Kejian, 1993), (Rao et al., 2008) and (Kibria and Banik, 2016). This formula was suggested for the first time by (Vinod and Ullah, 1981). In this case, if the condition number is around 10, weak multicollinearity can be said to exist, if the condition number is between 30 and 100, then there are multiple relations from moderate to a strong degree, and if the condition number is over 100, there is a serious multiple correlation problem (Rawlings et al., 2001).

3.3. Effects of Multicollinearity

The presence of an internal linear relationship between the independent variables in the linear regression model causes a number of issues with OLS estimations and hypothesis testing. To explain the impact of multicollinearity on OLS estimations, consider a linear regression model with two explanatory variables $\boldsymbol{X_1}$ and $\boldsymbol{X}_2$,

$$\boldsymbol{y} = \beta_1 \boldsymbol{X}_1 + \beta_2 \boldsymbol{X}_2 + \boldsymbol{e}. \tag{3.4}$$

Then the obtaining of the OLS estimator of this model passes through some of the normal equations

$$\boldsymbol{X}'\boldsymbol{X}\,\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}.$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}. \tag{3.5}$$

here r_{12} is the correlation between $\boldsymbol{X}_1$and $\boldsymbol{X}_2$, and r_{jy} is the correlation between $\boldsymbol{X}_\text{j}$ and $\boldsymbol{y}$. The inverse of $\boldsymbol{X}'\boldsymbol{X}$ is

$$C = (\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix}. \tag{3.6}$$

By multiplying (3.5) with (3.6) the estimates of the regression coefficients are

$$\widehat{\boldsymbol{\beta}}_1 = \frac{r_{1y} - r_{12}\, r_{2y}}{1-r_{12}^2}, \qquad \widehat{\boldsymbol{\beta}}_2 = \frac{r_{2y} - r_{12}\, r_{1y}}{1-r_{12}^2}. \tag{3.7}$$

If there is a full or exact linear relation between $\boldsymbol{X}_1$and $\boldsymbol{X}_2$ i.e., $|r_{12}| = 1$, then the rank of the $X'X$ matrix decreases. Therefore, the OLS estimator given by (2.4) cannot be found. When there is a near-linear relation between $\boldsymbol{X}_1$and $\boldsymbol{X}_2$, i.e., $|r_{12}| \approx 1$. In this case, $cov(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \to \pm\infty$. This means that the approximate collinearity between $\boldsymbol{X}_1$and $\boldsymbol{X}_2$ vectors causes growth in the OLS estimates' variances of the regression coefficients (Montgomery et al., 2012). Multicollinearity causes the OLS estimates to be extremely large in absolute value. This can be seen by examining the squared distance between the OLS estimator $\widehat{\boldsymbol{\beta}}$ and the parameter vector $\boldsymbol{\beta}$

$$L^2 = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{3.8}$$

the expected squared distance is

$$\begin{aligned} E(L^2) &= E(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)^2 \\ &= \sum\nolimits_{j=1}^{p} var(\hat{\beta}_j) \\ &= \sum\nolimits_{j=1}^{p} Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}) = \sigma^2 tr(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2 \sum\nolimits_{j=1}^{p} \frac{1}{\lambda_j}. \end{aligned} \tag{3.9}$$

This means that if at least one eigenvalue j is too small, then the distance between the OLS estimator and $\boldsymbol{\beta}$ will be very large. That is, the OLS estimation vector is bigger than the parameter vector, and this shows that the OLS method produces great estimated regression coefficients in absolute values (Montgomery et al., 2012).

Approximate multicollinearity adversely affects the hypothesis testing of the parameters. Since at the presence of multicollinearity in the model the standard errors of the regression coefficients will be large, the variances of the OLS estimators will be large, as well as the value of test statistic t will decrease in absolute value. This shall cause

independent variables to be considered insignificant, even if it is truly important in explaining the dependent variable.

The following hypothesis was established to measure the statistical importance of each individual parameter

$$H_0: \boldsymbol{\beta}_j = 0$$

$$H_1: \boldsymbol{\beta}_j \neq 0.$$

The t statistic used to test the H_0 hypothesis against the H_1 hypothesis.

$$t_j = \frac{\hat{\beta}_j}{\sqrt{\sigma^2/(1 - {R_j}^2)}} = \hat{\beta}_j \sqrt{\frac{1 - {R_j}^2}{\sigma^2}}. \tag{3.10}$$

In the presence of multicollinearity, when $R_j^2 \to 1$, t_j value approaches zero. As a result, even if it is truly significant, it can be concluded from t statistic that the $\boldsymbol{\beta}_j$ parameter is not different from zero and that the independent variable $\boldsymbol{X_j}$ does not affect the dependent variable. On the other hand, even if it is concluded that the parameters are significant in the F test. when t test is performed for each parameter separately, it can be concluded that the parameters are insignificant. This is a problem caused by multicollinearity.

3.4. Methods for Overcoming Multicollinearity

Some of the methods that can be used to significantly mitigate the effects of collinearity are as follow:

1. Additional data collection: Farrar and Glauber (1967), suggested additional data collection to address multicollinearity. However, it is not always possible to use this method because of limitations on the model or in the population.
2. Removing variables that cause multicollinearity: Variables that cause multicollinearity are detected and discarded from the model by using precise procedures. Applying this method, makes variables approach orthogonality and variance of the OLS estimator decreases. However, in this case, one or more variables that influence the dependent variable can be omitted from the model.
3. Transforming variables: Applying an appropriate transformation to independent variables can solve multicollinearity problems. However, this method may cause a change in the relationship between the dependent variable and the independent variables.

These methods are intended to modify the data. Nonetheless, it may not always be possible to make changes to the data. These methods may not be applicable due to time, material inadequacies or lack of enough data sources. Even if applicable, it may not always give a certain solution.

If there is multicollinearity in the linear regression model, then the OLS estimator given in (2.5) is still the best unbiased estimator. However, its variance is very large. Therefore, the OLS estimator will be unstable. Hence, to circumvent this problem remedial actions become needed without changing the regressors or removing any of them. It has been concluded that using biased estimators that can reduce parameter variances is an appropriate approach to solve the problem.

4. SOME BIASED LINEAR ESTIMATORS

In statistics, an estimator's bias is the discrepancy between the estimated parameter's real value and its expected value. If an estimator is either underestimate or overestimate a population parameter, then it is considered to be biased. So, in this chapter, some biased estimators that are widely used in statistics and econometrics will be included and divided into two main groups. In addition, the statistical properties of these estimators will be presented.

4.1. Unrestricted Biased Estimators

A restricted model is one in which the regression coefficients β_i are subject to a series of restrictions. In contrast, when the classical model is not exposed to any kind of restrictions then it will be referred to as an unrestricted model. Based on this, the estimators of each model type take their names as restricted and unrestricted.

4.1.1. Ridge regression estimator

One of the best-known approaches to circumvent plenty of the obstacles associated with the OLS estimates is the RR estimator that was originally introduced by Hoerl and Kennard (1970) through the acceptance of some bias to reduce variance. As mentioned in earlier chapters, according to the Gauss-Markov theorem, the OLS estimator is an unbiased estimator of $\boldsymbol{\beta}$ and has minimum variance within the entire class of linear unbiased estimators. However, its variance may not always be insignificantly small. In order to overcome variance growth, Hoerl and Kennard (1970) with the risk of losing the unbiasedness property, proposed the RR estimation method. The used logic of deriving

the RR estimator is based on minimizing the square distance. So, regarding this technique, the sum of the error squares, where $\widetilde{\boldsymbol{\beta}}$ any estimator of unknown parameter $\boldsymbol{\beta}$, is as follows:

$$\begin{aligned} S(\widetilde{\boldsymbol{\beta}}) &= (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) \\ &= (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) \\ &= S_{min} + \phi(\widetilde{\boldsymbol{\beta}}). \end{aligned} \tag{4.1}$$

where $(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})$ is the square of bias that caused by using $\widehat{\boldsymbol{\beta}}$ instead of $\widetilde{\boldsymbol{\beta}}$. When $\boldsymbol{X}'\boldsymbol{X}$ is an ill-conditioned matrix, the distance between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ increases. Therefore, the RR estimation method intends to set the squared distance of the $\widetilde{\boldsymbol{\beta}}$ estimator to a minimum. There can be more than one estimator carrying the same sum of error squares. Among these, the estimator with the smallest distance should be selected. Let $\phi_0 > 0$ be a constant that stands for the error sum of squares. In this case, there is a $\{\widetilde{\boldsymbol{\beta}}\}$ set of estimators that satisfy the following equation

$$S(\widetilde{\boldsymbol{\beta}}) = S_{min} + \phi_0. \tag{4.2}$$

To find the estimator with the smallest distance (error) within this set, let $1/k$ be the Lagrange multiplier then the following equation is investigated

$$min_{\widetilde{\beta}} \left\{ \widetilde{\boldsymbol{\beta}}'\widetilde{\boldsymbol{\beta}} + \left(\frac{1}{k}\right)\left[(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) - \phi_0\right] \right\}. \tag{4.3}$$

When the partial derivatives of (4.3) with respect to $\widetilde{\boldsymbol{\beta}}$ and $1/k$ are equal to zero, the following equations are obtained

$$\widetilde{\boldsymbol{\beta}} + \frac{1}{k}(\boldsymbol{X}'\boldsymbol{X})(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}) = 0 \tag{4.4}$$

$$\phi_0 = (\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}). \tag{4.5}$$

The solution of (4.3) by using the last equations was found by Hoerl and Kennard (1970) and named the RR estimator of $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}}_k = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}\,, k \geq 0. \tag{4.6}$$

The scalar k has many names in the literature, such as "shrinkage", "Ridge" or "bias" parameter, k will be mentioned as a ridge parameter all over this study. The choice of k affects the performance of the ridge estimator. When $k = 0$ the RR estimator is equal to the OLS estimator. Let $\boldsymbol{S}_k = \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}$ then $\widehat{\boldsymbol{\beta}}_k$ can be written as

$$\widehat{\boldsymbol{\beta}}_k = \boldsymbol{S}_k^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{4.7}$$

The expected value of $\widehat{\boldsymbol{\beta}}_k$ is

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_k) &= E[(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}] \\ &= E[(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e})] \\ &= [(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'E(\boldsymbol{e})] \\ &= (\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{S}_k^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} - k\,\boldsymbol{S}_k^{-1}\boldsymbol{\beta}, \end{aligned} \tag{4.8}$$

the part $(-k\,\boldsymbol{S}_k^{-1}\boldsymbol{\beta})$ of this equality expresses bias. So,

$$\begin{aligned} Bias(\widehat{\boldsymbol{\beta}}_k) &= E(\widehat{\boldsymbol{\beta}}_k) - \boldsymbol{\beta} \\ &= \boldsymbol{\beta} - k\,\boldsymbol{S}_k^{-1}\boldsymbol{\beta} - \boldsymbol{\beta} \\ &= -k\,\boldsymbol{S}_k^{-1}\boldsymbol{\beta}. \end{aligned} \tag{4.9}$$

While the Variance-covariance matrix is as follows:

$$\begin{aligned} Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{k}}) &= Var[(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}] \\ &= (\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'Var(\boldsymbol{y})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1} \\ &= \sigma^2\boldsymbol{S}_k^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{S}_k^{-\mathbf{1}}. \end{aligned} \tag{4.10}$$

By using (4.9) and (4.10) the MMSE matrix of ridge regression estimator is

$$\begin{aligned} MMSE(\widehat{\boldsymbol{\beta}}_k) &= Var(\widehat{\boldsymbol{\beta}}_k) + Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{k}})Bias(\widehat{\boldsymbol{\beta}}_k)' \\ &= \boldsymbol{S}_k^{-1}(\sigma^2\,\boldsymbol{X}'\boldsymbol{X} + k^2\boldsymbol{\beta}\boldsymbol{\beta}')\boldsymbol{S}_k^{-1}. \end{aligned} \tag{4.11}$$

Now let $\lambda_1, \lambda_2, \ldots, \lambda_p$ to be the eigenvalues of the $\boldsymbol{X}'\boldsymbol{X}$ matrix, then the SMSE equation of the RR estimator is

$$\begin{aligned} SMSE(\widehat{\boldsymbol{\beta}}_k) &= tr\left\{Var(\widehat{\boldsymbol{\beta}}_k) + Bias(\widehat{\boldsymbol{\beta}}_k)Bias(\widehat{\boldsymbol{\beta}}_k)'\right\} \\ &= \sigma^2 tr\left(\boldsymbol{S}_k^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{S}_k^{-1}\right) + k^2\boldsymbol{\beta}'\boldsymbol{S}_k^{-\mathbf{2}}\boldsymbol{\beta} \\ &= \sigma^2\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j+k)^2} + k^2\sum_{j=1}^{p}\frac{\beta_j^2}{(\lambda_j+k)^2}. \end{aligned} \tag{4.12}$$

Or in another way,

$$SMSE(\widehat{\boldsymbol{\beta}}_k) = tr[MMSE(\widehat{\boldsymbol{\beta}}_k)] = \sum_{j=1}^{p} \frac{\sigma^2\lambda_j + k^2\beta_j^2}{(\lambda_j + k)^2}. \quad (4.13)$$

Since $k > 0$, when k increases, bias increases, and variance decreases. At Eq. (4.12) the first part before the plus mark indicates the total variance value which is a continuous and monotone decreasing function of the bias parameter k, and the second part after the plus mark indicates the square of the bias value which is a continuous and monotone increasing function of the k bias parameter (Hoerl and Kennard, 1970b). Thus, comparing to the OLS estimator, the RR estimator is a good estimator when the decrease in total variance is greater than the increase in the bias square. Since the MMSE value of $\widehat{\boldsymbol{\beta}}_k$ will be smaller than the variance of $\widehat{\boldsymbol{\beta}}$. So, when using the RR estimator, the k value should be selected considering that the decrease in variance is greater than the increase in the square of bias.

Selection of ridge regression parameter k

The reason for the debate over the RR estimator is related to the selection of bias parameter k. Several scholars have developed and proposed a variety of approaches for estimating k. Just to mention a few, (Hoerl and Kennard, 1970b), (Hoerl et al., 1975), (McDonald and Galarneau, 1975), (Lawless and Wang, 1976), (Dempster et al., 1977), (Gibbons, 1981), (Singh and Tracy, 1999), (Wencheko, 2000), (Kibria, 2003), (Khalaf and Shukur, 2005), (Alkhamisi and Shukur, 2008), (Muniz and Kibria, 2009), (M. H. J. Gruber, 2010), (Mansson et al., 2010), (Muniz et al., 2012), (Hefnawy and Farag, 2014), (Aslam, 2014), (Arashi and Valizadeh, 2015) and very recently (Kibria and Banik, 2016). The following are some different selection methods of ridge parameter k mentioned in the literature:

1. Ridge Trace method: For a suitable selection of k (Hoerl and Kennard, 1970) suggested the ridge trace. Ridge trace is the graph of the elements of $\widehat{\boldsymbol{\beta}}_k$ corresponding to k, for all values of k that are usually within the closed interval [0,1]. After a particular value of k, the estimates become stationary. The smallest k value that makes $\widehat{\boldsymbol{\beta}}_k$ stable is taken as an appropriate value for ridge estimation. Hoerl and Kennard (1970) recommended the range of k as $0 \leq k \leq 1$. However, this range may not be valid for all datasets (Vinod and Ullah, 1981). In addition, they suggested graphing the ridge trace after standardizing the data in order to compare the regression coefficients in different

units. If the multicollinearity is severe, the stability in the regression coefficients will be seen from the ridge trace. As k increases, some ridge estimates will be quite different from each other, and in some k values $\widehat{\boldsymbol{\beta}}_k$ values will be stable. In general, the smallest k value at which $\widehat{\boldsymbol{\beta}}_k$ estimates are stationary should be determined. This value is expected to give estimates with a smaller MMSE than the OLS.

2. Minimizing the value of the SMSE: This method is summed up by taking the derivative of the SMSE of the RR estimator with respect to the parameter k and then make it equal to zero. By using this technique (Hoerl and Kennard, 1970b) presented the following estimation of k

$$\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}^2_{max}}, \tag{4.14}$$

where $\hat{\sigma}^2$ is the unbiased estimator of σ^2 parameter and $\hat{\alpha}^2_{max}$ is the maximum squared value of the OLS estimators in the canonical form (Kibria, 2003).

3. If the values of α_i are too small, where α_i's are the regression coefficients in the canonical form, then the arithmetic mean is not a good choice because the bias of k is too large. Therefore, (Hoerl et al., 1975) by using the harmonic mean of k values proposed k to be

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}. \tag{4.15}$$

And by conducting a simulation study, they showed that the ridge estimator for $\hat{k}_{HKB}$ has better MMSE than the OLS.

Despite the efficiency and popularity of the RR estimator dealing with multicollinearity, it also has some drawbacks. Many of the RR estimator problems stem from its dependency on k. For instance, when $k \to \infty, \widehat{\boldsymbol{\beta}}(k) \to 0$ which is a stable biased estimator of $\boldsymbol{\beta}$, and when $k \to 0, \widehat{\boldsymbol{\beta}}(k) \to \widehat{\boldsymbol{\beta}}$ which is an unbiased but unstable estimator of $\boldsymbol{\beta}$. That is, $\widehat{\boldsymbol{\beta}}(k)$ traces are taking place from $\widehat{\boldsymbol{\beta}}$ to 0 of the parameter space. The estimated distance between $\widehat{\boldsymbol{\beta}}(k)$ and $\boldsymbol{\beta}$ tends to decrease with k increasing from zero. However, to generate such a point estimate, there is no clear-cut, a single value of k.

4.1.2. Modified ridge regression estimator

The obstacles found in the RR estimator led Swindel (1976) to suggest an alternative estimator called the modified ridge regression (MRR) as a solution to the problems of minimization.

$$\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b}) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{b}), \quad k \geq 0, \tag{4.16}$$

where $\boldsymbol{b}: p \times 1$ is a non-stochastic vector that carries prior information about $\boldsymbol{\beta}$. Therefore, $\boldsymbol{b}$ should be chosen in a way that reflects the prior information and hypotheses related to $\boldsymbol{\beta}$ properly. Like the RR estimator, the MRR estimator aims to reduce the effect of multicollinearity by adding the k constant to the diagonal elements of the $\boldsymbol{X}'\boldsymbol{X}$ matrix.

Some results related to the MRR estimator can be obtained after applying some matrix operations to (4.16)

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b}) &= (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{b}) \\
&= (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{b}) \\
&= (\boldsymbol{I} + k(\boldsymbol{X}'\boldsymbol{X})^{-1})^{-1}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{b} + \boldsymbol{b} + k(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{b}\right) \\
&= (\boldsymbol{I} + k(\boldsymbol{X}'\boldsymbol{X})^{-1})^{-1}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{b}\right) + \boldsymbol{b}.
\end{aligned} \tag{4.17}$$

Since it can be expressed in the previous form, the following results can be reached:

- If $\boldsymbol{b} = \widehat{\boldsymbol{\beta}}$, then $\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b}) = \widehat{\boldsymbol{\beta}}$. Using the OLS estimator as prior information makes the MR estimator equals the OLS estimator.
- When $k = 0$, then $\widehat{\boldsymbol{\beta}}(0,\boldsymbol{b}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \widehat{\boldsymbol{\beta}}$ is the OLS estimator.
- $\lim_{k\to\infty} \widehat{\boldsymbol{\beta}}(k,\boldsymbol{b}) = \boldsymbol{b}$. i.e., when k becomes close to infinity, $\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})$ approaches $\boldsymbol{b}$.
- When $= 0$, then $\widehat{\boldsymbol{\beta}}(k,0) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y} = \widehat{\boldsymbol{\beta}}_k$ is the RR estimator.

These indicate that the MRR technique yields an estimate with a shorter length than the OLS, i.e. $\left\|\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right\| \leq \left\|\widehat{\boldsymbol{\beta}}\right\|$. The expected value of the MRR estimator is

$$\begin{aligned}
E\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) &= E\left(\boldsymbol{S}_k^{-1}(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{b})\right) \\
&= \boldsymbol{S}_k^{-1}(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + k\boldsymbol{b}) \\
&= \boldsymbol{S}_k^{-1}\boldsymbol{S}\boldsymbol{\beta} + k\boldsymbol{S}_k^{-1}\boldsymbol{b}.
\end{aligned} \tag{4.18}$$

Since $E\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) \neq \boldsymbol{\beta}$, it can be seen that the MRR estimator is biased. And its bias is

$$
\begin{aligned}
Bias\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) &= E\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) - \boldsymbol{\beta} \\
&= \boldsymbol{S}_k{}^{-1}\boldsymbol{S}\boldsymbol{\beta} + k\boldsymbol{S}_k{}^{-1}\boldsymbol{b} - \boldsymbol{\beta} \\
&= \left(\boldsymbol{S}_k{}^{-1}\boldsymbol{S} - \boldsymbol{I}_p\right)\boldsymbol{\beta} + k\boldsymbol{S}_k{}^{-1}\boldsymbol{b} \\
&= \left(\boldsymbol{S}_k{}^{-1}\left(\boldsymbol{S} + k\boldsymbol{I}_p - k\boldsymbol{I}_p\right) - \boldsymbol{I}_p\right)\boldsymbol{\beta} + k\boldsymbol{S}_k{}^{-1}\boldsymbol{b} \\
&= \left(\boldsymbol{S}_k{}^{-1}\boldsymbol{S}_k - k\boldsymbol{S}_k{}^{-1} - \boldsymbol{I}_p\right)\boldsymbol{\beta} + k\boldsymbol{S}_k{}^{-1}\boldsymbol{b} \\
&= -k\boldsymbol{S}_k{}^{-1}\boldsymbol{\beta} + k\boldsymbol{S}_k{}^{-1}\boldsymbol{b} \\
&= -k\boldsymbol{S}_k{}^{-1}(\boldsymbol{\beta} - \boldsymbol{b}). \qquad (4.19)
\end{aligned}
$$

As the prior information approaches the real unknown parameter vector, i.e., $\boldsymbol{b} \rightarrow \boldsymbol{\beta}$, the bias of the MRR estimator decreases, but it is obvious that if $\boldsymbol{b} = \boldsymbol{\beta}$ is supposed. In other words, if the prior information is correct, the MRR estimator will be unbiased regardless of the value of k.

MRR estimator's variance-covariance matrix is

$$
\begin{aligned}
Var\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) &= Var\left(\boldsymbol{S}_k{}^{-1}(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{b})\right) \\
&= \boldsymbol{S}_k{}^{-1}Var(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{b})\boldsymbol{S}_k{}^{-1} \\
&= \sigma^2\boldsymbol{S}_k{}^{-1}\boldsymbol{S}\boldsymbol{S}_k{}^{-1}. \qquad (4.20)
\end{aligned}
$$

Note that the variance-covariance matrix of the MRR estimator is the same as the variance-covariance matrix of the RR estimator. By using the results given in (4.19) and (4.20) the MMSE of the MRR estimator is as

$$
\begin{aligned}
MMSE\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) &= Var\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right) + Bias\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right)Bias\left(\widehat{\boldsymbol{\beta}}(k,\boldsymbol{b})\right)' \\
&= \sigma^2\boldsymbol{S}_k{}^{-1}\boldsymbol{S}\boldsymbol{S}_k{}^{-1} + k^2\boldsymbol{S}_k{}^{-1}(\boldsymbol{\beta} - \boldsymbol{b})(\boldsymbol{\beta} - \boldsymbol{b})'\boldsymbol{S}_k{}^{-1}. \qquad (4.21)
\end{aligned}
$$

Swindel (1976) compared the OLS estimator with the MRR estimator according to the MMSE criterion and showed that the MRR estimator is superior to the OLS. Also, Pliskin (1987) compared the MRR estimator and the RR estimator by using the same k value and gave the necessary and sufficient condition for the MRR estimator to be superior to the RR estimator.

4.1.3. Contraction estimator

Mayer, and Willke (1973) presented a class of shrunken estimators

$$
\widehat{\boldsymbol{\beta}}(\rho) = (1+\rho)^{-1}\widehat{\boldsymbol{\beta}}\,, \quad \rho > 0, \qquad (4.22)
$$

where it shrinks every element of the OLS estimator by $(1+\rho)^{-1}$. These type of estimators have passed in the literature under various names e.g. the shrinkage estimator, the shrunken estimator and the contraction estimator, the last appellation was given by M. Gruber and Rao (1982) also demonstrate that ridge and contraction type of estimators as special cases of the Bayes estimator and show the superiority condition of the contraction estimator over the ridge estimator. Since $\lim_{\boldsymbol{\rho \to 0}} (1+\rho)^{-1} = 1$, and $\lim_{\boldsymbol{\rho \to \infty}} (1+\rho)^{-1} = 0$, $\widehat{\boldsymbol{\beta}}(\rho)$ can be put in a different form

$$\widehat{\boldsymbol{\beta}}(d) = d\widehat{\boldsymbol{\beta}}\,, \tag{4.23}$$

where $d = (1+\rho)^{-1}$. $\widehat{\boldsymbol{\beta}}(d)$ is the linear d function which makes calculations much easier. However, the downside of $\widehat{\boldsymbol{\beta}}(d)$ is that it shrinks all elements of $\widehat{\boldsymbol{\beta}}$ with the same factor. The bias and the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}(d)$ are respectively

$$\begin{aligned} Bias\left(\widehat{\boldsymbol{\beta}}(d)\right) &= E\left(\widehat{\boldsymbol{\beta}}(d)\right) - \boldsymbol{\beta} \\ &= d\boldsymbol{\beta} - \boldsymbol{\beta} = (d-1)\boldsymbol{\beta}, \end{aligned} \tag{4.24}$$

$$Var\left(\widehat{\boldsymbol{\beta}}(d)\right) = \sigma^2 d^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{4.25}$$

From previous equations the MMSE of the contraction estimator is as follows:

$$MMSE\left(\widehat{\boldsymbol{\beta}}(d)\right) = \sigma^2 d^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} + (d-1)^2 \boldsymbol{\beta}\boldsymbol{\beta}'. \tag{4.26}$$

4.1.4. Liu estimator

In spite of ridge estimator efficiency in practice, k is a complex function. The bias parameter k has numerous selection methods and its preference depends on the analyzer. Since there is no consensus on the manner of choosing k, new estimation methods have been investigated in the case of linear model morbidity. One of those methods is the Liu estimator, by combining the ridge and the contraction estimators' advantages Kejian (1993) introduced a new estimator

$$\widehat{\boldsymbol{\beta}}_d = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{y} + d\widehat{\boldsymbol{\beta}})\,, \qquad 0 < d < 1. \tag{4.27}$$

This estimator was later named Liu estimator by Akdeniz and Kaçiranlar (1995) and M. H. J. Gruber (1998). The estimator of Liu was obtained by using the OLS estimate on the model caused by augmenting $d\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{\varepsilon}'$ to (2.2). The advantage of $\widehat{\boldsymbol{\beta}}_d$ over $\widehat{\boldsymbol{\beta}}_k$ is because d is a linear function, so the selection of d is easier than the selection of k. $\widehat{\boldsymbol{\beta}}_d$ also can be written as a linear transformation of the OLS estimator in the following form:

$$\widehat{\boldsymbol{\beta}}_d = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})\,\widehat{\boldsymbol{\beta}}. \tag{4.28}$$

From this, we can see that when $d = 1$, $\widehat{\boldsymbol{\beta}}_d = \widehat{\boldsymbol{\beta}}$ and $\|\widehat{\boldsymbol{\beta}}_d\| < \|\widehat{\boldsymbol{\beta}}\|$. Also, the expected value of $\widehat{\boldsymbol{\beta}}_d$

$$E(\widehat{\boldsymbol{\beta}}_d) = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})\boldsymbol{\beta}\,, \tag{4.29}$$

demonstrates that $\widehat{\boldsymbol{\beta}}_d$ is a biased estimator. Furthermore, with supposing that $\boldsymbol{A}_d = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})$, the bias and the variance-covariance matrix of the Liu estimator are respectively

$$Bias(\widehat{\boldsymbol{\beta}}_d) = (\boldsymbol{A}_d - 1)\boldsymbol{\beta}, \tag{4.30}$$

$$Var(\widehat{\boldsymbol{\beta}}_d) = \sigma^2\boldsymbol{A}_d(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}_d'. \tag{4.31}$$

By using the equations (4.30) and (4.31), Liu estimator's MMSE matrix is as follows:

$$MMSE(\widehat{\boldsymbol{\beta}}_d) = \sigma^2\boldsymbol{A}_d(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}_d' + (\boldsymbol{A}_d - 1)\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{A}_d - 1). \tag{4.32}$$

As well as the SMSE value is

$$SMSE(\widehat{\boldsymbol{\beta}}_d) = \sigma^2\sum_{i=1}^{p}\frac{(\lambda_i + d)^2}{\lambda_i(\lambda_i + 1)} + (d-1)^2\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i + 1)^2}\,, \tag{4.33}$$

where α_i is the i-th element of $\boldsymbol{\alpha} = \boldsymbol{T}'\boldsymbol{\beta}$ vector used to write the model (2.2) in the canonical form, given the orthogonal matrix $\boldsymbol{T}$ of the orthonormal eigenvectors of the $\boldsymbol{X}'\boldsymbol{X}$ matrix. The equation of the bias parameter that makes the $SMSE(\widehat{\boldsymbol{\beta}}_d)$ minimum is given by Liu (1993) as follows:

$$\begin{aligned} d_{opt} &= \sum_{i=1}^{p}\frac{\alpha_i^2 - \sigma^2}{(\lambda_i + 1)^2}\Big/\sum_{i=1}^{p}\frac{\sigma^2 + \lambda_i\alpha_i^2}{\lambda_i(\lambda_i + 1)^2} \\ &= 1 - \frac{\sigma^2\left(\sum_{i=1}^{p}\frac{1}{\lambda_i(\lambda_i + 1)}\right)}{\sum_{i=1}^{p}\frac{\sigma^2 + \lambda_i\alpha_i^2}{\lambda_i(\lambda_i + 1)^2}}. \end{aligned} \tag{4.34}$$

However, d_{opt} is not very handy in practice since it depends on the unknown parameters σ^2 and α_i^2. If the unbiased estimators $\hat{\alpha}_i^2 - \hat{\sigma}^2/\lambda_i$ and $\hat{\sigma}^2$ are used instead, then the following $\hat{d}_{opt}$ will be obtained as follows:

$$\hat{d}_{opt} = 1 - \hat{\sigma}^2\left(\sum_{i=1}^{p}\frac{1}{\lambda_i(\lambda_i + 1)}\Big/\sum_{i=1}^{p}\frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2}\right). \tag{4.35}$$

Liu (1993) named (4.35) the minimum MMSE estimate and gave a generalized form of it. In addition, different selection methods of d were conducted.

4.1.5. The wo-parameter estimator

With the aim of reaching new estimators that are more balanced and less biased, statisticians have started a new trend towards finding new estimation methods with more than one parameter to find new estimators encountering multicollinearity whose length is closer to β than $\widehat{\boldsymbol{\beta}}$. For instance, Özkale and Kaciranlar (2007) introduced the following two-parameter estimator by grafting the contraction estimator into the MRR estimator:

$$\widehat{\boldsymbol{\beta}}(k,d) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\left(\boldsymbol{X}'\boldsymbol{y} + kd\widehat{\boldsymbol{\beta}}\right), \quad k > 0\,, 0 < d < 1. \tag{4.36}$$

Another two parameter estimator was obtained by Yang and Chang (2010)

$$\widehat{\boldsymbol{\beta}}^{*}(k,d) = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}\,, \quad k > 0\,, 0 < d < 1. \tag{4.37}$$

Many other two type parameter estimators were suggested separately in the literature, see (Sakallıoğlu and Kaçıranlar, 2008) and (Dorugade, 2014). Both two-parameter (TP) estimators in (4.36) and (4.37) can be obtained as a solution of the minimization problems, or similar to that of the Liu estimator it can be obtained by augmenting the linear stochastic constraints $kd\widehat{\boldsymbol{\beta}} = k\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $(d-k)\widehat{\boldsymbol{\beta}}(k) = \boldsymbol{\beta} + \boldsymbol{\varepsilon}'$ to (2.2) model one for (4.36) and the other for (4.37) respectively, and then using the least squares estimator. Here $\boldsymbol{\varepsilon}: p \times 1$ is a random vector with a mean of zero, variance-covariance matrix of $\sigma^2\boldsymbol{I}$ and both random errors $\boldsymbol{e}$ and $\boldsymbol{\varepsilon}$ are uncorrelated i.e., $E(\boldsymbol{e}\boldsymbol{\varepsilon}') = 0$.

From this section on, the name two parameter estimator will refer to the two-parameter estimator $\widehat{\boldsymbol{\beta}}(k,d)$ of Özkale and Kaciranlar in (4.36) that will be discussed throughout the study. The TP estimator unifies the advantages of ridge and contraction estimators, also considers a general estimator that contains the ordinary least squares, the ridge regression, the Liu, and the contraction estimators as special cases. This feature can be shown by choosing different values for k and d in $\widehat{\boldsymbol{\beta}}(k,d)$, as in the following equations:

- $\lim_{d\to 1}\widehat{\boldsymbol{\beta}}(k,d) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ which is the OLS estimator.
- $\lim_{k\to 0}\widehat{\boldsymbol{\beta}}(k,d) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ which is the OLS estimator.
- $\lim_{d\to 0}\widehat{\boldsymbol{\beta}}(k,d) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is the RR estimator.
- $\widehat{\boldsymbol{\beta}}(1,d) = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{y} + d\widehat{\boldsymbol{\beta}})$ is the Liu estimator.
-

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(k,d) &= (\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\left(\boldsymbol{X}'\boldsymbol{y}+kd\widehat{\boldsymbol{\beta}}\right)\\ &= [\boldsymbol{I}+k(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\boldsymbol{X}'\boldsymbol{y}+kd\widehat{\boldsymbol{\beta}}\right)\\ &= [\boldsymbol{I}+k(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1}\left(\widehat{\boldsymbol{\beta}}+kd(\boldsymbol{X}'\boldsymbol{X})^{-1}\widehat{\boldsymbol{\beta}}\right)\\ &= [\boldsymbol{I}+k(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1}\left(\widehat{\boldsymbol{\beta}}-d\widehat{\boldsymbol{\beta}}\right)+d\widehat{\boldsymbol{\beta}}.\end{aligned}$$

After putting the $\widehat{\boldsymbol{\beta}}(k,d)$ on the previous form it can be said that $\lim_{k\to\infty}\widehat{\boldsymbol{\beta}}(k,d)=d\widehat{\boldsymbol{\beta}}$ is the contraction estimator

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(k,d) &= (\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}\left(\boldsymbol{X}'\boldsymbol{y}+kd\widehat{\boldsymbol{\beta}}\right)\\ &= (\boldsymbol{X}'\boldsymbol{X}+k\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{y}+kd\boldsymbol{I})\widehat{\boldsymbol{\beta}}.\end{aligned}$$

We see that the parameters of shrinkage related to each $\widehat{\boldsymbol{\beta}}$ differ. Therefore, the contraction estimator's drawback was overcome.

For the representation, $\widehat{\boldsymbol{\beta}}(k,d)=\boldsymbol{S}_k^{-1}\boldsymbol{S}_{kd}^{-1}\widehat{\boldsymbol{\beta}}$, where $\boldsymbol{S}_k=\boldsymbol{S}+k\boldsymbol{I}$, $\boldsymbol{S}_{kd}=\boldsymbol{S}+kd\boldsymbol{I}$ and $\boldsymbol{S}=\boldsymbol{X}'\boldsymbol{X}$ the biased and dispersion matrix respectively are as follows:

$$Bais\left(\widehat{\boldsymbol{\beta}}(k,d)\right)=(\boldsymbol{S}_k^{-1}\boldsymbol{S}_{kd}-\boldsymbol{I})\boldsymbol{\beta}=k(d-1)\boldsymbol{S}_k^{-1}\boldsymbol{\beta}. \tag{4.38}$$

$$Var\left(\widehat{\boldsymbol{\beta}}(k,d)\right)=\sigma^2\boldsymbol{S}_k^{-1}\boldsymbol{S}_{kd}\boldsymbol{S}^{-1}\boldsymbol{S}_{kd}\boldsymbol{S}_k^{-1}. \tag{4.39}$$

By using (4.38) and (4.39) equations the MMSE for the TP estimator is obtained

$$MMSE\left(\widehat{\boldsymbol{\beta}}(k,d)\right)=\sigma^2\boldsymbol{S}_k^{-1}\boldsymbol{S}_{kd}\boldsymbol{S}^{-1}\boldsymbol{S}_{kd}\boldsymbol{S}_k^{-1}+k^2(d-1)^2\boldsymbol{S}_k^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{S}_k^{-1}. \tag{4.40}$$

Özkale and Kaciranlar (2007) presented the necessary and sufficient conditions for the two-parameter estimator to be superior to the OLS estimator according to the MMSE criterion. They also introduced a selection method for k and d bias parameters. Besides M. H. J. Gruber (2010) noticed that the two-parameter estimator $\hat{\beta}(k,d)$ is a special case of the Bayes, the minimax, generalized ridge regression estimators, and the mixed estimators.

4.2. Restricted Biased Estimators

In some cases, the classical linear regression model given in (2.2) is estimated subject to certain prior limitations on the model's unknown parameter vector β in the hope to provide better estimators than the OLS estimator. These prior limitations may be stated in the form of the linear equality restrictions

$$\boldsymbol{R}\boldsymbol{\beta}=\boldsymbol{r}, \tag{4.41}$$

where $\boldsymbol{R}$ is an $m \times p$ matrix of known prior information, its $rank(\boldsymbol{R}) = m < p$, $\boldsymbol{r}$ is an $m \times 1$ vector, and the linear constraints "m" involved in (2.2) model are independent of each other. In this research, the restricted linear model (RLM) denotes incorporation between (2.2) and (4.41).

4.2.1. Restricted least squares estimator

In order to use the sample information and the prior information simultaneously, the sum of the error squares $(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$ should be minimized subject to the constraint $\boldsymbol{R\beta} = \boldsymbol{r}$. The Lagrange function for this is

$$S(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) - 2\boldsymbol{\lambda}'(\boldsymbol{R\beta} - \mathbf{r}), \tag{4.42}$$

where $\boldsymbol{\lambda}$: $m \times 1$ is the Lagrange multiplier vector. When the partial derivatives of the function $S(\boldsymbol{\beta}, \boldsymbol{\lambda})$ are taken with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ respectively and make them equal zero,

$$\frac{1}{2}\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = -\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{X\beta} - \boldsymbol{R}'\boldsymbol{\lambda} = 0$$

$$\frac{1}{2}\frac{\partial S(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = -\boldsymbol{R\beta} + \boldsymbol{r} = 0, \tag{4.43}$$

are obtained. By solving the normal equations given in (4.43) the restricted least square (RLS) estimator is obtained

$$\widehat{\boldsymbol{\beta}}_r = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\left(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}\right). \tag{4.44}$$

This approach generates an unbiased estimator when the constraints are correct, and contributes to a smaller variance in sampling than the OLS estimator. If the constraints are incorrect, then the sampling variance remains reduced, but the estimator is a biased one. The drawback of using exact restrictions is that the estimator will have more risk than the OLS estimator because of this potential bias, see (Güler and Kaçıranlar, 2009).

If you pay attention to the RLS equation, you can see that if a transformation can be done, then it is obvious that $\widehat{\boldsymbol{\beta}}_r$ fulfills the (4.41) restriction

$$\boldsymbol{R}\widehat{\boldsymbol{\beta}}_r = \boldsymbol{R}\widehat{\boldsymbol{\beta}} + \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\left(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{r}. \tag{4.45}$$

The expected value of the RLS

$$E\left(\widehat{\boldsymbol{\beta}}_r\right) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R\beta}), \tag{4.46}$$

this demonstrates that if the constraint is true then $\widehat{\boldsymbol{\beta}}_r$ is unbiased, otherwise it is biased. Also, the variance-covariance matrix is

$$Var\left(\widehat{\boldsymbol{\beta}}_r\right) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} - \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{4.47}$$

As previously shown, the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_r$ depends on $\boldsymbol{R}$. And the total variance of $\widehat{\boldsymbol{\beta}}_r$ will always be smaller compared to the total variance of $\widehat{\boldsymbol{\beta}}$. That is,

$$Var(\widehat{\boldsymbol{\beta}}) - Var\,(\widehat{\boldsymbol{\beta}}_r) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\,\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1} \geq 0. \quad (4.48)$$

The expression at the right of the equation (4.48) is a nnd matrix. Therefore, when the constraint $\boldsymbol{R\beta} = \boldsymbol{r}$ is true, it is effective to apply linear restrictions on the model (2.2).

The MMSE of the restricted least squares estimator is

$$\begin{aligned} MMSE\big(\widehat{\boldsymbol{\beta}}_r\big) &= \sigma^2\boldsymbol{M}_0\,\boldsymbol{X}'\boldsymbol{X}\boldsymbol{M}_0 \\ &= \sigma^2\boldsymbol{S}^{-1} - \sigma^2\boldsymbol{S}^{-1}\,\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}^{-1}\,\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{S}^{-1}, \end{aligned} \quad (4.49)$$

where $\boldsymbol{M}_0 = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\,\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}^{-1}\,\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{S}^{-1}$, $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}$. The RLS is frequently applied as an unbiased estimation approach. However, at the existence of the multicollinearity problem in practical regression analysis the unknown coefficients estimated by the RLS estimator are usually unstable and give misleading information. Therefore, the need for alternative, more stable estimators is still felt.

4.2.2. Restricted ridge regression estimator

When it comes to the RRR estimator, there is no one version of it. For a while of time the concept of restricted ridge estimator only has been indicating to the one that was proposed by Sarkar (1992) which is as follows:

$$\boldsymbol{\beta}^*(k) = \boldsymbol{W}\widehat{\boldsymbol{\beta}}_r\,, \boldsymbol{W} = (\boldsymbol{I}_p + k(\boldsymbol{X}'\boldsymbol{X})^{-1})^{-1}, k \geq 0, \quad (4.50)$$

where $\widehat{\boldsymbol{\beta}}_r$ is the RLS estimator given in (4.44). $\boldsymbol{\beta}^*(k)$ drops the components of $\widehat{\boldsymbol{\beta}}_r$ towards zero, that is $\boldsymbol{\beta}^*(0) = \widehat{\boldsymbol{\beta}}_r$ and $\lim_{k\to\infty} \boldsymbol{\beta}^*(k) = 0$. Also, this approach reduces the regression parameter variance and renders the estimation stable by adding $\boldsymbol{W}$ to $\widehat{\boldsymbol{\beta}}_r$ which generally is applied to the OLS estimator to obtain the ridge regression estimator:

$$\widehat{\boldsymbol{\beta}}(k\,) = \boldsymbol{W}\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p)^{-1}\boldsymbol{X}'\boldsymbol{y}. \quad (4.51)$$

Nonetheless, it is easy to see that this procedure does not satisfy linear restriction (4.41) for every outcome of $\boldsymbol{y}$. Because Sarkar does not presume that the restriction (4.41) must always be valid when $\boldsymbol{\beta}^*(k)$ is used. Therefore, from this point of view and by affirming these concepts Groß (2003) introduced a new RRR estimator for $\boldsymbol{\beta}$ with respect to $\boldsymbol{R\beta} = \boldsymbol{r}$ restrictions

$$\widehat{\boldsymbol{\beta}}_r(k) = \widehat{\boldsymbol{\beta}}(k\,, \boldsymbol{\beta}_0) - \boldsymbol{S}_k^{-1}\,\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_k^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}}(k\,, \boldsymbol{\beta}_0) - \boldsymbol{r}),\ k \geq 0, \quad (4.52)$$

where $\boldsymbol{S}_k = \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p$, $\boldsymbol{\beta}_0 = \boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{r}$, and by letting $\boldsymbol{b} = \boldsymbol{\beta}_0$ in the MRR given at (4.16) we obtain $\widehat{\boldsymbol{\beta}}(k\,, \boldsymbol{\beta}_0) = \boldsymbol{S}_k^{-1}(\boldsymbol{X}'\boldsymbol{y} + k\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{R}')^{-1}\boldsymbol{r})$. Zhong and Yang (2007) showed that the restricted ridge regression estimator can be obtained by minimizing the sum of squared residuals with a spherical restriction and a linear restriction (4.41). Thus, the restricted linear regression is transformed into an optimization problem with two restrictions

$$\begin{aligned} & min_{\beta}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\,(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ & s.t.\ \ \boldsymbol{\beta}'\boldsymbol{\beta} \le \phi^2 \\ & \qquad \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}. \end{aligned}$$

Another presentation of $\widehat{\boldsymbol{\beta}}_r(k)$ may be introduced as follows:

$$\widehat{\boldsymbol{\beta}}_r(k) = \widehat{\boldsymbol{\beta}}(k) + \boldsymbol{S}_k^{-1}\,\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_k^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}(k\,)),\ \ k \ge 0, \tag{4.53}$$

where $\widehat{\boldsymbol{\beta}}(k)$ is the RR estimator of $\boldsymbol{\beta}$. We refer to (4.52) as the RRR estimator. If linear restrictions $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ are assumed to hold, then it is more appropriate to shrink towards $\boldsymbol{\beta}_0$ the shortest vector which satisfies the restrictions, rather than towards the zero vector. Groß (2003) proved that $\widehat{\boldsymbol{\beta}}_r(0) = \widehat{\boldsymbol{\beta}}_r$ and $\lim_{k\to\infty} \widehat{\boldsymbol{\beta}}_r(k) = \boldsymbol{\beta}_0$. The expected value of $\widehat{\boldsymbol{\beta}}_r(k)$ is

$$E(\widehat{\boldsymbol{\beta}}_r(k)) = \boldsymbol{\beta} - k\boldsymbol{M}_k\boldsymbol{\beta}. \tag{4.54}$$

Also, the bias of $\widehat{\boldsymbol{\beta}}_r(k)$

$$Bias(\widehat{\boldsymbol{\beta}}_r(k)) = E(\widehat{\boldsymbol{\beta}}_r(k)) - \boldsymbol{\beta} = -k\boldsymbol{M}_k\boldsymbol{\beta}, \tag{4.55}$$

where $\boldsymbol{M}_k$ is a non-zero symmetric matrix, equals to $\boldsymbol{S}_k^{-1} - \boldsymbol{S}_k^{-1}\,\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_k^{-1}\,\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{S}_k^{-1}$, and the variance-covariance matrix is

$$\begin{aligned} Var\big(\widehat{\boldsymbol{\beta}}_r(k)\big) &= E\big[\big(\widehat{\boldsymbol{\beta}}_r(k) - E(\widehat{\boldsymbol{\beta}}_r(k))\big)\big(\widehat{\boldsymbol{\beta}}_r(k) - E(\widehat{\boldsymbol{\beta}}_r(k))\big)'\big] \\ &= \sigma^2\boldsymbol{M}_k\,\boldsymbol{X}'\boldsymbol{X}\boldsymbol{M}_k. \end{aligned} \tag{4.56}$$

For the proofs of (4.54), (4.55) and (4.56) see (Zhong and Yang, 2007). By making use of the two previous equations the MMSE of $\widehat{\boldsymbol{\beta}}_r(k)$ is obtained as

$$MMSE\left(\widehat{\boldsymbol{\beta}}_r(k)\right) = \sigma^2\boldsymbol{M}_k\,\boldsymbol{X}'\boldsymbol{X}\boldsymbol{M}_k + k^2\boldsymbol{M}_k\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{M}_k, \tag{4.57}$$

where $\boldsymbol{M}_k = \boldsymbol{S}_k^{-1} - \boldsymbol{S}_k^{-1}\,\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_k^{-1}\,\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{S}_k^{-1}$. The mean square error of $\widehat{\boldsymbol{\beta}}_r(k)$ is

$$SMSE\left(\widehat{\boldsymbol{\beta}}_r(k)\right) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i(\lambda_i + k - r_{ii}^*)^2}{(\lambda_i + k)^4} + k^2 \left[\sum_{i=1}^{p} \frac{\alpha_i(\lambda_i + k - r_{ii}^*)}{(\lambda_i + k)^2}\right]^2. \quad (4.58)$$

Although this estimator has fewer MMSE compared to the RLS estimator, it is very complex in practical implementation and a compacter estimator is supposed to be obtained. Therefore, there was remained a need to find a better estimator to overcome the multicollinearity of the RLM.

4.2.3. Restricted contraction estimator

The restricted contraction estimator is as follows:

$$\widehat{\boldsymbol{\beta}}_R(d) = d\widehat{\boldsymbol{\beta}} + \boldsymbol{R}'[\boldsymbol{R}\boldsymbol{R}']^{-1}\left(r - \boldsymbol{R}d\widehat{\boldsymbol{\beta}}\right), \quad (4.59)$$

see (Özkale and Kaciranlar, 2007).

4.2.4. Restricted Liu estimator

As the RRR estimator, two restricted Liu estimators are available in the literature for the multiple linear regression parameters. The first estimator was proposed by Kaçiranlar et al. (1999) by following the method Sarkar (1992) has used while obtaining the restricted ridge estimator. That is Sarkar replaced the OLS estimator with the RLS estimator in the RR estimator, since the ridge and the Liu estimators have both been shown as linear transformations of the OLS estimator in the sense that $\widehat{\boldsymbol{\beta}}_k = \mathbf{W}\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_d = \boldsymbol{A}_d\widehat{\boldsymbol{\beta}}$ where $\boldsymbol{W} = (\boldsymbol{I}_p + k(\boldsymbol{X}'\boldsymbol{X})^{-1})^{-1}$ and $\boldsymbol{A}_d = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})$. Kaçiranlar et al. (1999) as well have replaced the OLS estimator by the RLS estimator in Liu estimator, see (Akdeniz and Kaçiranlar, 2001), hence the restricted estimator resulting from this action is

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{rd} &= \boldsymbol{A}_d\widehat{\boldsymbol{\beta}}_r \\ &= (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})\widehat{\boldsymbol{\beta}}_r. \end{aligned} \quad (4.60)$$

Yet, $\widehat{\boldsymbol{\beta}}_{rd}$ does not satisfy the linear restrictions (4.41) and as a result it cannot be considered a restricted estimator with respect to restrictions (4.41). Precisely from this vantage point, the second RL estimator has come which closely resembles the RRR estimator proposed by Groß (2003). When the RRR estimator combines the two estimators MR proposed by Swindel (1976) and the RLS estimator; the second RL estimator combines the modified Liu estimator with the RLS estimator

$$\widehat{\boldsymbol{\beta}}_r(d) = \widehat{\boldsymbol{\beta}}_d + (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}_d), \quad (4.61)$$

where $\widehat{\boldsymbol{\beta}}_d$ is the Liu estimator. In order to avoid overlapping of names of the two restricted Liu estimators, Özkale and Kaciranlar suggested calling $\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(d)$ as modified restricted Liu (MRL) estimator, see (Özkale and Kaciranlar, 2007). If the bias of $\widehat{\boldsymbol{\beta}}_r(d)$ is given as follows:

$$Bias(\widehat{\boldsymbol{\beta}}_r(d)) = (d-1)\boldsymbol{M}_1\boldsymbol{\beta}, \tag{4.62}$$

where $\boldsymbol{M}_1 = \boldsymbol{S}_1^{-1} - \boldsymbol{S}_1^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_1^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{S}_1^{-1}$ and $\boldsymbol{S}_1 = \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I}$. The variance-covariance matrix is

$$Var\big(\widehat{\boldsymbol{\beta}}_r(d)\big) = \sigma^2\boldsymbol{M}_1\boldsymbol{S}_d\boldsymbol{S}^{-1}\boldsymbol{S}_d\boldsymbol{M}_1. \tag{4.63}$$

Then the matrix mean square error of the modified restricted Liu is obtained as

$$MMSE\big(\widehat{\boldsymbol{\beta}}_r(d)\big) = \sigma^2\boldsymbol{M}_1\boldsymbol{S}_d\boldsymbol{S}^{-1}\boldsymbol{S}_d\boldsymbol{M}_1 + (d-1)^2\boldsymbol{M}_1\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{M}_1. \tag{4.64}$$

4.2.5. Restricted two-parameter estimator

Under the same logic they had used obtaining the two-parameter (TP) estimator before Özkale and Kaçiranlar (2007) derived the restricted two-parameter (RTP) estimator in order to find an improved estimator that minimizes the squared distance of $\widehat{\boldsymbol{\beta}}$ toward the regression parameter. The new estimator was derived as a solution to minimization problems with the addition of additional restrictions on the parameter, thus the resulting estimator is as follows:

$$\widehat{\boldsymbol{\beta}}_r(k,d) = \widehat{\boldsymbol{\beta}}(k,d) + \boldsymbol{S}_k^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_k^{-1}\boldsymbol{R}']^{-1}\left(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}(k,d)\right), k > 0, 0 < d < 1. \tag{4.65}$$

As the TP estimator considers a general case of the OLS, the RR, the Liu, and the contraction estimator, the restricted version of it, that is the RTP estimator considers a general case of the RLS, the RRR introduced by GroB (2003), the modified restricted Liu (MRL) mentioned by Özkale and Kaciranlar (2007), and the restricted contraction (RC) estimator. The following equations illustrate some relationships expressed as

- $\lim_{k\to 0}\widehat{\boldsymbol{\beta}}_r(k,d) = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}})$ is the RLS estimator.
- $\lim_{d\to 0}\widehat{\boldsymbol{\beta}}_r(k,d) = \widehat{\boldsymbol{\beta}}(k) + \boldsymbol{S}_k^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}_k^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}(k))$ is the RRR estimator.
- $\lim_{k\to\infty}\widehat{\boldsymbol{\beta}}_r(k,d) = \widehat{\boldsymbol{\beta}}_R(d) = d\widehat{\boldsymbol{\beta}} + \boldsymbol{R}'[\boldsymbol{R}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R}d\widehat{\boldsymbol{\beta}})$ is the RC estimator.
- For $k = 1$,

$\widehat{\boldsymbol{\beta}}_r(1,d) = \widehat{\boldsymbol{\beta}}_r(d) = \widehat{\boldsymbol{\beta}}_d + (\boldsymbol{X'X} + \boldsymbol{I})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X} + \boldsymbol{I})^{-1}\boldsymbol{R'}]^{-1}(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}_{\boldsymbol{d}})$ is the MRL estimator.

By putting the RTP estimator given in (4.65) on another form

$$\widehat{\boldsymbol{\beta}}_r(k,d) = \boldsymbol{M}(k)\boldsymbol{S}_{kd}\widehat{\boldsymbol{\beta}} + \boldsymbol{S}_k^{-1}\boldsymbol{R'}[\boldsymbol{R}\boldsymbol{S}_k^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{r}, \tag{4.66}$$

where $\boldsymbol{M}_k = \boldsymbol{S}_k^{-1} - \boldsymbol{S}_k^{-1}\boldsymbol{R'}[\boldsymbol{R}\boldsymbol{S}_k^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{R}\boldsymbol{S}_k^{-1}$, then the variance of $\hat{\beta}_r(k,d)$ is

$$Var\left(\widehat{\boldsymbol{\beta}}_r(k,d)\right) = \sigma^2\boldsymbol{M}_k\boldsymbol{S}_{kd}\boldsymbol{S}^{-1}\boldsymbol{S}_{kd}\boldsymbol{M}_k. \tag{4.67}$$

And the bias is

$$Bias\left(\widehat{\boldsymbol{\beta}}_r(k,d)\right) = \boldsymbol{M}_k[\boldsymbol{S}_{kd} - \boldsymbol{S}_k]\boldsymbol{\beta} = k(d-1)\boldsymbol{M}_k\boldsymbol{\beta}. \tag{4.68}$$

Then the MMSE of $\widehat{\boldsymbol{\beta}}_r(k,d)$, when the condition $\boldsymbol{R\beta} = \boldsymbol{r}$ is presumed to hold true, is

$$MMSE\left(\widehat{\boldsymbol{\beta}}_r(k,d)\right) = \sigma^2\boldsymbol{M}_k\boldsymbol{S}_{kd}\boldsymbol{S}^{-1}\boldsymbol{S}_{kd}\boldsymbol{M}_k + k^2(d-1)^2\boldsymbol{M}_k\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{M}_k. \tag{4.69}$$

5. COMPARASIONS ACCORDING TO MMSE CRITERION

There are many criteria to compare two estimators of an unknown parameter. Some of these criteria have been mentioned previously in chapter two. In this research, the author believes that the most appropriate criterion for use is the MMSE, which is a commonly accepted criterion for gauging an estimator's performance of $\boldsymbol{\beta}$ because it contains all relevant estimator quality information, also it includes the comparison in terms of bias and in terms of dispersion matrix. The MMSE of an estimator $\widetilde{\boldsymbol{\beta}}$ is defined as

$$\begin{aligned} MMSE(\widetilde{\boldsymbol{\beta}}) &= E\left[(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] \\ &= Var(\widetilde{\boldsymbol{\beta}}) + \left[Bias(\widetilde{\boldsymbol{\beta}})\right]\left[Bias(\widetilde{\boldsymbol{\beta}})\right]'. \end{aligned} \tag{5.1}$$

Let $\widetilde{\boldsymbol{\beta}}_1$and $\widetilde{\boldsymbol{\beta}}_2$ be two estimators of $\boldsymbol{\beta}$ parameter, in the sense of MMSE criterion the estimator $\widetilde{\boldsymbol{\beta}}_2$ is said to be superior to $\widetilde{\boldsymbol{\beta}}_1$ if and only if $MMSE(\widetilde{\boldsymbol{\beta}}_1) - MMSE(\widetilde{\boldsymbol{\beta}}_2) \geq 0$. Since the SMSE is the trace of MMSE

$$SMSE(\widetilde{\boldsymbol{\beta}}) = tr\left[MMSE(\widetilde{\boldsymbol{\beta}})\right] = tr\left(Var(\widetilde{\boldsymbol{\beta}})\right) + \left[Bias(\widetilde{\boldsymbol{\beta}})\right]'\left[Bias(\widetilde{\boldsymbol{\beta}})\right]. \tag{5.2}$$

Then if the estimator $\widetilde{\boldsymbol{\beta}}_2$ dominates the estimator $\widetilde{\boldsymbol{\beta}}_1$ in the sense of the MMSE criterion, $\widetilde{\boldsymbol{\beta}}_2$ dominates $\widetilde{\boldsymbol{\beta}}_1$ in the sense of the SMSE, i.e. $SMSE(\widetilde{\boldsymbol{\beta}}_1) \geq SMSE(\widetilde{\boldsymbol{\beta}}_2)$. The reverse conclusion does not necessarily hold true. Hence, the MMSE is regarded as a stronger

criterion than the SMSE, see (Rao and Toutenburg, 1995). Using the SMSE as a search basic criterion implies that we disregard MMSE's off-diagonal components. Therefore, it may be more rational to compare two estimators in the sense of the MMSE criterion.

5.1. Comparison of $\widehat{\boldsymbol{\beta}}_r(k,d)$ to $\widehat{\boldsymbol{\beta}}(k,d)$ by the MMSE Criterion

The expected loss or the $MMSE$ of the two-parameter estimator and the restricted two parameter estimator are given respectively by (4.40) and (4.69). the difference between the two estimator's MMSEs is

$$\Delta = MMSE\left(\widehat{\boldsymbol{\beta}}(k,d)\right) - MMSE\left(\widehat{\boldsymbol{\beta}}_r(k,d)\right) = \boldsymbol{B} - \boldsymbol{M}_k\boldsymbol{S}_k\boldsymbol{B}\boldsymbol{S}_k\boldsymbol{M}_k\,, \tag{5.3}$$

where $\boldsymbol{B} = MMSE(\widehat{\boldsymbol{\beta}}(k,d))$. Özkale and Kaciranlar (2007) derived the necessary and sufficient condition for $\boldsymbol{\Delta}$ to be nnd matrix by applying a theorem from (Graybill, 1976) (Graybill, 1983) and gave the following theorem:

Theorem 5.1.1. *The $\widehat{\boldsymbol{\beta}}_r(k,d)$ estimator of $\boldsymbol{\beta}$ is superior to the $\widehat{\boldsymbol{\beta}}(k,d)$ estimator by the criterion of MMSE if and only if $\lambda_{max}(\boldsymbol{B}^{-1}\boldsymbol{M}_k\boldsymbol{S}_k\boldsymbol{B}\boldsymbol{S}_k\boldsymbol{M}_k) \leq 1$, where λ_{max} is the maximum eigenvalue of $\boldsymbol{B}^{-1}\boldsymbol{M}_k\boldsymbol{S}_k\boldsymbol{B}\boldsymbol{S}_k\boldsymbol{M}_k$.*

The MMSE criterion comparison between $\widehat{\boldsymbol{\beta}}(k,d)$ and $\widehat{\boldsymbol{\beta}}_r(k,d)$ includes the MMSE comparisons between the OLS and the RLS as k approaches zero, the ridge regression and the restricted ridge regression as d approaches zero, the Liu and the MRL when $k = 1$. Hence, we have the following theorems on the comparability of these estimators using the MMSE criterion:

Theorem 5.1.2. *A necessary and sufficient condition for the RLS estimator of $\boldsymbol{\beta}$ to be superior to the OLS estimator by the MMSE criterion is $\lambda_{max}(1 - \boldsymbol{R}'[\boldsymbol{R}\boldsymbol{S}^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{S}^{-1}) \leq 1$.*

Theorem 5.1.3. *The restricted ridge estimator of $\boldsymbol{\beta}$ dominates the ridge regression estimator by the criterion of MMSE if and only if $\lambda_{max}(\boldsymbol{A}^{-1}\boldsymbol{M}_k\boldsymbol{S}_k\boldsymbol{A}\boldsymbol{S}_k\boldsymbol{M}_k) \leq 1$, where $A = \sigma^2\boldsymbol{S}_k^{-1}\boldsymbol{S}\boldsymbol{S}_k^{-1} + k^2\boldsymbol{S}_k^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{S}_k^{-1}$.*

Theorem 5.1.4. *The necessary and sufficient condition for the MRL estimator of $\boldsymbol{\beta}$ to dominate the Liu estimator by the MMSE criterion is given by $\lambda_{max}(\boldsymbol{C}^{-1}\boldsymbol{M}_1\boldsymbol{S}_1\boldsymbol{C}\boldsymbol{S}_1\boldsymbol{M}_1) \leq 1$, where $\boldsymbol{C} = \sigma^2\boldsymbol{S}_1^{-1}\boldsymbol{S}_d\boldsymbol{S}^{-1}\boldsymbol{S}_d\boldsymbol{S}_1^{-1} + (d-1)^2\boldsymbol{S}_1^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{S}_1^{-1}$.*

Theorem 5.1.2 was introduced by Toro-Vizcarrondo and Wallace (1968) when it is assumed that the restrictions in (4.41) are not true, while Theorems 5.1.3 and 5.1.4 were

introduced for the first time by Özkale and Kaciranlar (2007). Notice that the results of the comparison depend on the unknown parameters $\boldsymbol{\beta}$ and σ^2, besides the d and k choices. We cannot exclude, because of such unknown parameters that the conditions obtained in the theorems will hold. Therefore, for certain values of $\boldsymbol{\beta}$, σ^2, d, and k the restricted estimators will perform better than the other estimators. Consequently, using a suitable estimate of d and k or utilizing prior information about these parameters, and replacing $\boldsymbol{\beta}$ and σ^2 with their unbiased estimators leads to feasible results.

5.1.1. Selection of the parameters k and d

Trying to find appropriate bias parameters is an integral part of any study of biased estimators in the linear model. One of the most common methods used to estimate bias parameters is to propose the estimators of the biasing parameters that yield minimum MMSE. Nevertheless, due to the inability to diagonalize $\boldsymbol{M}_k$ and $\boldsymbol{S}_k$ by the same orthogonal matrix, no particular rule for selecting k and d may be suggested in the sense of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}(k,d)$ that is guaranteed to generate minimum MMSE, see (Özkale, 2014).

It is well known that orthogonal transformation can transform a linear regression model into a canonical form. Let $\boldsymbol{Z}=\boldsymbol{XQ}$, $\boldsymbol{\alpha}=\boldsymbol{Q'\beta}$ and Q is $p\times p$ orthogonal matrix such that $\boldsymbol{Z'Z}=\boldsymbol{Q'X'XQ}=\boldsymbol{Q'SQ}=\boldsymbol{\Lambda}=diag(\lambda_1,\dots,\lambda_p)$, where $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_p\geq 0$ are the order eigenvalues of $\boldsymbol{X'X}$, then the canonical form of model (2.2) can be obtained as

$$\boldsymbol{y}=\boldsymbol{Z\alpha}+\boldsymbol{e}, \tag{5.4}$$

since $\widehat{\boldsymbol{\alpha}}(k,d)=\boldsymbol{Q'}\widehat{\boldsymbol{\beta}}(k,d)$ and $MMSE(\widehat{\boldsymbol{\alpha}}(k,d))=\boldsymbol{Q'}MMSE\left(\widehat{\boldsymbol{\beta}}(k,d)\right)\boldsymbol{Q}$, based on (4.40) the MMSE of $\widehat{\boldsymbol{\alpha}}(k,d)$ can be written as

$$MMSE\left(\widehat{\boldsymbol{\alpha}}(k,d)\right)=\sigma^2(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}+kd\boldsymbol{I})\boldsymbol{\Lambda}^{-1}(\boldsymbol{\Lambda}+kd\boldsymbol{I})(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}+k^2(d-1)^2(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}\boldsymbol{\alpha\alpha'}(\boldsymbol{\Lambda}+k\boldsymbol{I})^{-1}. \tag{5.5}$$

Optimal values for d and k can be derived by minimizing the quadratic function of $\boldsymbol{d}$, which can be written as

$$f(k,d)=tr\left[MMSE\left(\widehat{\boldsymbol{\alpha}}(k,d)\right)\right]=\sum_{i=1}^{p}\frac{\sigma^2(\lambda_i+kd)^2+k^2(\mathrm{d}-1)^2\alpha_i^2\lambda_i}{\lambda_i(\lambda_i+k)^2}. \tag{5.6}$$

The d-value that minimizes the function $f(k,d)$ can be found by differentiating $f(k,d)$ with respect to d when k is fixed.

$$\frac{\partial f(k,d)}{\partial d} = \sum_{i=1}^{p} \frac{2\sigma^2 k(\lambda_i + kd) + 2k^2(d-1)\alpha_i^2\lambda_i}{\lambda_i(\lambda_i + k)^2}. \tag{5.7}$$

By equating it to zero, and after replacing the unknown parameters σ^2 and α_i^2 with their unbiased estimators, we get the optimal estimator of d for the fixed value k

$$\hat{d}_{opt} = \frac{\sum_i^p \frac{(k\hat{\alpha}_i^2 - \hat{\sigma}^2)}{(\lambda_i + k)^2}}{\sum_i^p \frac{k(\hat{\sigma}^2 + \hat{\alpha}_i^2\lambda_i)}{\lambda_i(\lambda_i + k)^2}}. \tag{5.8}$$

As was mentioned before in Section 4.1.5, the TP estimator leads to the Liu estimator when $k = 1$. Therefore, when $k = 1$ the value of $\hat{d}_{opt}$ in (5.8) decreases to be equal to

$$\hat{d}_{opt} = \frac{\sum_i^p \frac{(\hat{\alpha}_i^2 - \hat{\sigma}^2)}{(\lambda_i + 1)^2}}{\sum_i^p \frac{(\hat{\sigma}^2 + \hat{\alpha}_i^2\lambda_i)}{\lambda_i(\lambda_i + 1)^2}}. \tag{5.9}$$

This estimate of d was given by Liu (1993). Similarly, the k value that minimizes the function $f(k,d)$ for fixed d-value can be derived by differentiating $f(k,d)$ with respect to d

$$\frac{\partial f(k,d)}{\partial k} = \sum_{i=1}^{p} \frac{2\sigma^2(\lambda_i + kd)(\mathrm{d}-1) + 2k(d-1)^2\alpha_i^2\lambda_i}{(\lambda_i + k)^3}, \tag{5.10}$$

and then equating the numerator to zero. The orientation toward equating only the numerator to zero not all the fraction came from what Hoerl and Kennard (1970) have done during the procedures of finding the optimal k value as $= \frac{\sigma^2}{\alpha_i^2}$. This value of k was obtained by minimizing or partial deriving $tr[MMSE(\hat{\boldsymbol{\alpha}}(k))]$ in respect to k, where $\hat{\boldsymbol{\alpha}}(k)$ is the ridge regression estimator in canonical form, then equating the numerator to zero. Similarly, Özkale and Kaciranlar (2007) derived the k-value by equating the numerator of $\frac{\partial f(k,d)}{\partial k}$ to zero which can be concluded as

$$k = \frac{\sigma^2}{\alpha_i^2 - d(\frac{\sigma^2}{\lambda_i} + \alpha_i^2)}. \tag{5.11}$$

Due to the full dependency of the optimal value of k on the unknown parameters σ^2 and α_i^2, when d is fixed. They must be estimated from the observed data and be

replaced by their unbiased estimators as suggested by Hoerl and Kennard (1970) and Kibria (2003), so the estimated optimal k value is

$$\hat{k} = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2 - d(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2)}. \tag{5.12}$$

Another estimator of k was proposed by Hoerl et al. (1975) by taking the harmonic mean of k values that had been found by Hoerl and Kennard (1970b), whereas Kibria (2003) proposed estimators of k by using the arithmetic and geometric means of k values found by Hoerl and Kennard (1970) too. With the same idea, Özkale and Kaciranlar (2007) proposed estimators of k which minimize $f(k, d)$. And they are as follows:

$$\hat{k}_{HM} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\left[\hat{\alpha}_i^2 - d(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2)\right]} \tag{5.13}$$

$$\hat{k}_{AM} = \frac{1}{p}\sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2 - d(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2)}, \tag{5.14}$$

$$\hat{k}_{GM} = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^{p}\left[\hat{\alpha}_i^2 - d(\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2)\right]\right)^{1/p}}, \tag{5.15}$$

which respectively are the harmonic mean, the arithmetic mean and the geometric mean of $\hat{k}$ values in (5.12). Lukman et al. (2020) developed a new estimator for the bias estimator k, which we will refer to as the Lukman bias estimator from now on

$$\hat{k}_{LM} = \frac{p\hat{\sigma}^2}{(1+d)[(\hat{\beta} - J)(\hat{\beta} - J)' - \hat{\sigma}^2 tr(\Lambda^{-1})]}, \tag{5.16}$$

where $J = \frac{\sum_{i=1}^{p}\hat{\beta}_i}{p}$. When k is negative, they suggested using:

$$\hat{k}_{rep} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\alpha_i^2}. \tag{5.17}$$

As was shown before when d approaches zero, $\hat{\beta}(k, d)$ approaches the RR estimator. Thus, when d approaches zero, the k estimators in (5.13), (5.14), and (5.15) decrease to k estimators given by Hoerl et al. (1975) and Kibria (2003). Other values of k may be found in (Muniz and Kibria, 2009) (Aslam, 2014) and (Kibria and Banik, 2016).

Because k must always be positive, the following theorem shows the situation of the positivity of the estimators in Eqs. (5.13) - (5.15). Where if k values in (5.11) are restricted to be positive, the positivity of the estimators may be achieved.

Theorem 5.1.5. If

$$\hat{d} < min\left\{\frac{\hat{\alpha}_i^2}{\frac{\hat{\sigma}^2}{\lambda_i} + \hat{\alpha}_i^2}\right\}, \tag{5.18}$$

for all i, then the $\hat{k}_{HM}$, $\hat{k}_{AM}$ and $\hat{k}_{GM}$ are always positive. Check the proof in (Özkale and Kaciranlar, 2007).

It is apparent that $\hat{d}_{opt}$ in (5.8) is dependent on the value of k and the estimators of k are dependent on the value of d. One of the used procedures to avoid looping in choosing and estimating k and d parameters is the iterative method.

(i) Calculate $\hat{d}$ from (5.18).

(ii) By using $\hat{d}$ in (i), estimate $\hat{k}_{HM}$, $\hat{k}_{AM}$, $\hat{k}_{GM}$ or $\hat{k}_{LM}$.

(iii) Find $\hat{d}_{opt}$ in (5.8) by using obtained $\hat{k}_{HM}$, $\hat{k}_{AM}$, $\hat{k}_{GM}$ or $\hat{k}_{LM}$.

(iv) If $\hat{d}_{opt}$ is negative replace $\hat{d}_{opt} = \hat{d}$. It should be noted that $\hat{d}_{opt}$ is always less than one, although it may be less than zero. In addition, $\hat{d}$ is always less than one and greater than zero.

(v) If any value of $\hat{k}_{HM}$, $\hat{k}_{AM}$, $\hat{k}_{GM}$ or $\hat{k}_{LM}$ is negative replace it with $\hat{k}_{rep}$ from (5.17).

6. REAL-LIFE APPLICATIONS

In all of this chapter's sections, our analysis and computations are performed by using (R2018b) version of the MATLAB program. The program was implemented on a group of biased estimators that were discussed in Chapter 4 with the aim of comparing their performance. The comparisons are conducted in the respect of MMSE given the theories presented in Chapter 5. Also, we accept the trace of MMSE or what being expressed the SMSE as a measurement to do the scalar comparisons for the real data examples following many of authors among them (Hoerl and Kennard, 1970a) and (Akdeniz and Erol, 2003). The MMSE formulas of the restricted and unrestricted estimators were given in Chapter 4. As these estimators are fundamentally separate into two subgroups, the comparison primarily occurs between these two subgroups: the unrestricted estimators OLS, RR, Liu, and TP, and the restricted estimators RLS, RRR, MRL, RTP.

6.1. Numerical Example 1: Portland Cement Data

To illustrate the Theories shown in the previous chapter, we consider the data set on Portland cement originally due to (Woods et al., 1932). This data set is often referred to as Hald data since it was analyzed by (Hald, 1952). And also, then it has been widely analyzed by many authors, among them: (Hamaker, 1962) (Kaçiranlar et al., 1999), (Sakallioğlu et al., 2001), (Liu, 2003), (Yang and Chang, 2010), (Li and Yang, 2011), (Özkale, 2012), (Li and Yang, 2016) and most recently by (Ahmad and Aslam, 2020). These data came from an experimental investigation of the heat generated during the setting and hardening of variously composed Portland cement and discussed the dependence of the heat on the percentages of four compounds in the clinkers from which the cement was manufactured. This data set consists of the four compounds: Tricalcium aluminate, tricalcium silicate, tetracalcium aluminoferrite, and dicalcium silicate. As noticed by Woods, the great thicknesses in some constructions of large works such as dams severely impede the outflow of heat. This property is of interest to study. When the subsequent cooling to the surrounding temperature takes place, the consequent increase in temperature while the cement is hardening leads to contractions and cracking.

Before fitting a linear regression model, the dependent and independent variables were taken as follows:

$\boldsymbol{y}$: heat evolved in calories/gram of cement, after 180 days of curing.

$\boldsymbol{X}_1$: percentage of tricalcium aluminate ($3CaOAl_2O_3$) in the mixture.

$\boldsymbol{X}_2$: percentage of tricalcium silicate ($3CaOSiO_2$) in the mixture.

$\boldsymbol{X}_3$: percentage of tetracalcium aluminoferita ($4CaOAl_2O_3Fe_2O_3$) in the mixture.

$\boldsymbol{X}_4$: percentage of β-dicalcium silicate ($2CaOSiO_2$) in the mixture.

Subsequently, our data is assembled as follows:

$$\boldsymbol{X} = \begin{pmatrix} 7 & 26 & 6 & 60 \\ 1 & 29 & 15 & 52 \\ 11 & 56 & 8 & 20 \\ 11 & 31 & 8 & 47 \\ 7 & 52 & 6 & 33 \\ 11 & 55 & 9 & 22 \\ 3 & 71 & 17 & 6 \\ 1 & 31 & 22 & 44 \\ 2 & 54 & 18 & 22 \\ 21 & 47 & 4 & 26 \\ 1 & 40 & 23 & 34 \\ 11 & 66 & 9 & 12 \\ 10 & 68 & 8 & 12 \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} 78.5 \\ 74.3 \\ 104.3 \\ 87.6 \\ 95.9 \\ 109.2 \\ 102.7 \\ 72.5 \\ 93.1 \\ 115.9 \\ 83.8 \\ 113.3 \\ 109.4 \end{pmatrix}.$$

In this section we will analyze and interpret these data by following the original analysis of (Woods et al., 1932), and fit a linear model without intercept (homogeneous model). In this data set we have $n = 13$ experimental units and $p = 4$ unknown regression coefficients. the eigenvalues of $\boldsymbol{X'X}$ were obtained as

$$\lambda_1 = 44663.303\,, \qquad \lambda_2 = 5965.339\,, \qquad \lambda_3 = 809.952\,, \qquad \lambda_4 = 105.406.$$

The Condition number (CN) $= \frac{\lambda_{max}}{\lambda_{min}} = 423.7271$ indicates a moderate level of multicollinearity. It is worth noting that if $\boldsymbol{X}$ is standardized (normalized) such that $\boldsymbol{X'X}$ is a matrix of correlation, then the condition number almost doubles to 1376.8806 and this refers to a serious level of the problem. The researcher does not prefer data standardizing in our case because during the program's running it was noticed that standardizing causes problems with the selected values of k and d.

Since the CN is an overall multicollinearity diagnostic measure it shows the degree of collinearity in the whole model and does not specify which regressors may be the reason for collinearity among regressors. So, there is a need for an individual multicollinearity diagnostic measure, which is the variance inflation factors (VIF). The VIF computed from the correlation matrix of the independent variables for the example are as follows

$$VIF_1 = 38.496, \qquad VIF_2 = 254.423, \qquad VIF_3 = 46.868\,, \qquad VIF_4 = 282.513$$

And these factors indicate that the estimates of β_1, β_2, β_3 and β_4 would be seriously affected by the very near-singularity in $\boldsymbol{X}$ matrix, since the guideline of Marquardt (1970) for severe collinearity is $VIF > 10$.

k and d values were obtained by following the selection instructions given in the previous chapter along with setting an initial value for d equals 0.5 in Table 6.1.

Table 6.1. *The selected k and d values of Hald data*

	k values	d optimum
Harmonic mean	6.7895	0.7302
Arithmetic mean	73.4301	0.9601
Geometric mean	15.5485	0.8729

For the linear restriction, we follow (Najarian, Arashi anad Kibria, 2013) taking $\boldsymbol{R}$ and $\boldsymbol{r}$ as follows

$$\boldsymbol{R} = [\,1 \quad -1 \quad 1 \quad 0\,], \qquad \boldsymbol{r} = [0].$$

The values of all our biased estimators are obtained, and their respective SMSE values are computed then summarized in Table 6.2.

Table 6.2. *The coefficient estimates and the SMSE values of OLS, RR, Liu, TP, RLS, RRR, MRL, and RTP estimators for Hald data*

	k, d	β_1	β_2	β_3	β_4	SMSE
OLS	(0, 0)	2.1930	1.1533	0.7585	0.4863	**0.0638**
RR	(6.7895, 0)	2.1044	1.1737	0.6952	0.4996	**0.0572**
	(73.4301,0)	1.5756	1.2901	0.3472	0.5744	**0.0264**
	(15.5485, 0)	2.0040	1.1966	0.6248	0.5145	**0.0502**
Liu	(0, 0.7302)	2.1893	1.1542	0.7558	0.4869	**0.4153**
	(0, 0.9601)	2.1925	1.1535	0.7581	0.4864	**0.4158**
	(0, 0.8729)	2.1913	1.1537	0.7573	0.4866	**0.4156**
TP	(6.7895, 0.7302)	2.1691	1.1588	0.7414	0.4899	**0.0620**
	(73.4301, 0.9601)	2.1684	1.1588	0.7421	0.4898	**0.0619**
	(15.5485, 0.8729)	2.1690	1.1588	0.7415	0.4899	**0.0620**
RLS	(0, 0)	1.3374	1.3735	0.0361	0.6271	**0.0085**
RRR	(6.7895, 0)	1.3313	1.3730	0.0417	0.6266	**0.0084**
	(73.4301,0)	1.2768	1.3685	0.0917	0.6228	**0.0073**
	(15.5485, 0)	1.3237	1.3724	0.0487	0.6261	**0.0082**
MRL	(0, 0.7302)	1.3372	1.3735	0.0363	0.6270	**0.0085**
	(0, 0.9601)	1.3374	1.3735	0.0361	0.6270	**0.0085**
	(0, 0.8729)	1.3374	1.3735	0.0361	0.6270	**0.0085**
RTP	(6.7895, 0.7302)	1.3362	1.3735	0.0373	0.6267	**0.0085**
	(73.4301, 0.9601)	1.3412	1.3758	0.0346	0.6237	**0.0085**
	(15.5485, 0.8729)	1.3369	1.3738	0.0369	0.6263	**0.0085**

From Table 6.2, we can easily see that the restricted estimators are performing better than the unrestricted estimators as they have smaller SMSE values comparing to their counterparts. On the unrestricted estimation side, the worst estimate occurs by the Liu estimator as it has the biggest SMSE value. On the other hand, all the restricted estimators approximately operate at the same level of goodness with only the arithmetic mean restricted ridge regression estimator being the exception. Generally, the RR estimator is superior to the unrestricted estimator set, and the RRR estimator is superior to the restricted estimator set. Furthermore, the RRR estimator outperforms all other biased estimators by the SMSE criterion. It should be noted that different selection of k and d undoubtedly leads to differences in estimation and the SMSE values. Also, the table shows no significant differences in the regression coefficients' estimates and the SMSE values of the MRL estimator. The researcher attributes this to the chosen d values being close to each other.

Table 6.3. *The superiority according to MMSE*

Superiority of	**Condition of superiority**	**Programme's results**
RLS to OLS by MMSE criterion	$\lambda_{max}(1 - \boldsymbol{R}'[\boldsymbol{RS^{-1}R}']^{-1}\boldsymbol{RS^{-1}}) \leq 1$	**1**
RRR to RR by MMSE criterion	$\lambda_{max}(\boldsymbol{A^{-1}M_kS_kAS_kM_k}) \leq 1$	**1**
		1.0014
		1.001
MRL to LE by MMSE criterion	$\lambda_{max}(\boldsymbol{C^{-1}M_1S_1CS_1M_1}) \leq 1$	**1**
		1
		1
RTP to TP by MMSE criterion	$\lambda_{max}(\boldsymbol{B^{-1}M_kS_kBS_kM_k}) \leq 1$	**1**
		1.0012
		1

Similarly, the estimators were compared concerning the MMSE criterion in Table 6.3. We notice that the RLS estimator is superior to the OLS estimator, the RRR estimators are superior to the RR estimators, the MRL estimators are superior to the Liu estimators, and the RTP estimators are superior to the TP estimators since the data fulfills the necessary and sufficient conditions mentioned in the Theorems (5.1.1), (5.1.2), (5.1.3) and (5.1.4), respectively.

6.2. Numerical Example 2: Total National Research and Development Expenditures

To make the previous theoretical parts more clear another application is implemented, this time on the Total National Research and Development Expenditures'(TNRDE) data as a Percent of Gross National Product (GNP) for some countries in the time period between 1972 and 1986, were originally given by (Gruber, 1998) and also cited by (Akdeniz and Erol, 2003), (Zhong and Yang, 2007), (Yang and Cui, 2011), (Najarian et al., 2013) and (Şiray, 2014) to compare some biased estimators. The regression on this data set represents the relationship between the dependent variable $\boldsymbol{y}$ which stands for the percentage spent by the United States and another four independent variables $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_3}$ and $\boldsymbol{X_4}$. The variable $\boldsymbol{X_1}$ represents the percent spent by France, $\boldsymbol{X_2}$ represents the percent spent by West Germany, $\boldsymbol{X_3}$ represents the percent spent by Japan, and $\boldsymbol{X_4}$ represents the percent spent by the former Soviet Union. The data set consists of 10 observations shown in Table 6.4.

Table 6.4. *Total national research and development expenditures as a present of GNP by country: 1972-1986*

Year	$\boldsymbol{y}$	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$
1972	2.3	1.9	2.2	1.9	3.7
1975	2.2	1.8	2.2	2.0	3.8
1979	2.2	1.8	2.4	2.1	3.6
1980	2.3	1.8	2.4	2.2	3.8
1981	2.4	2.0	2.5	2.3	3.8
1982	2.5	2.1	2.6	2.4	3.7
1983	2.6	2.1	2.6	2.6	3.8
1984	2.6	2.2	2.6	2.6	4.0
1985	2.7	2.3	2.8	2.8	3.7
1986	2.7	2.3	2.7	2.8	3.8

If we fit a linear model without intercept (homogeneous model) to the data, then the $CN = 8776.382$ which means that the $\boldsymbol{X}$ matrix is ill-conditioned and the regressors suffer a serious level of multicollinearity. Under this model the eigenvalues of $\boldsymbol{X'X}$ are

$$\lambda_1 = 302.9626, \qquad \lambda_2 = 0.7283\,, \qquad \lambda_3 = 0.0446\,, \qquad \lambda_4 = 0.0345.$$

The variance inflation factors computed from the correlation matrix of the independent variables are

$$VIF_1 = 6.910, \qquad VIF_2 = 21.581, \qquad VIF_3 = 29.756\,, \qquad VIF_4 = 1.795,$$

and these factors indicate that the estimates of $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ would be affected by the very near-singularity in $\boldsymbol{X}$ matrix. In this case, the near-singularity is known to be due the near-redundancy between $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$. When we fit a linear model with intercept (inhomogeneous model) by adding an $n \times 1$ vector all of its elements equal one to the design matrix, the size of design matrix becomes 10×5. Here we still have $n = 10$ observations, but there are now $p = 5$ unknown regression coefficients. Under this case the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ are

$\lambda_1 = 312.932, \quad \lambda_2 = 0.7536, \quad \lambda_3 = 0.0453, \quad \lambda_4 = 0.0372\,, \quad \lambda_5 = 0.0019.$

And the condition number is 168129.285 this indicates the existence of a severe degree of multicollinearity among the regressors. In this section the analysis and results are given only for the homogeneous model. By setting the initial value of d equals to 0.3, the selected k and d values are as follows:

Table 6.5. *The selected k and d values for Total national research and development expenditures data*

	k **values**	d **optimum**
Harmonic mean	0.0202	0.1940
Arithmetic mean	0.0988	0.6378
Geometric mean	0.0406	0.4742

For the linear restriction $\boldsymbol{R\beta} = \boldsymbol{r}$, we use the $\boldsymbol{R}$ vector suggested by (Yang and Cui, 2011) and extract the value of $\boldsymbol{r}$

$$\boldsymbol{R} = [\,1 \quad -2 \quad -2 \quad -2\,], \qquad \boldsymbol{r} = [-2.775].$$

The values of all our biased estimators are obtained and their respective MMSE values are computed then summarized in Table 6.6.

Table 6.6. *The coefficient estimates and the SMSE values of OLS, RR, Liu, TP, RLS, RRR, MRL, and RTP estimators for TNRDE data*

	$(k\,, d)$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_3$	$\boldsymbol{\beta}_4$	**SMSE**
OLS	(0, 0)	0.6455	0.0896	0.1436	0.1526	**0.0808**
RR	(0.0202, 0)	0.5053	0.1293	0.1965	0.1685	**0.0360**

	(0.0988, 0)	0.3370	0.1832	0.2464	0.1918	**0.0079**
	(0.0406, 0)	0.4319	0.1525	0.2208	0.1773	**0.0207**
Liu	(0, 0.1940)	0.2877	0.1871	0.2161	0.2335	**9.0235**
	(0, 0.6378)	0.4846	0.1334	0.1762	0.1890	**9.0424**
	(0, 0.4742)	0.4120	0.1532	0.1909	0.2054	**9.0321**
TP	(0.0202, 0.1940)	0.5325	0.1216	0.1862	0.1654	**0.0432**
	(0.0988, 0.6378)	0.5337	0.1235	0.1808	0.1668	**0.0450**
	(0.0406, 0.4742)	0.5332	0.1227	0.1842	0.1656	**0.0440**
RLS	(0,0)	-0.3848	0.3699	0.5741	0.2511	**0.0451**
RRR	(0.0202, 0)	-0.3836	0.3973	0.5525	0.2459	**0.0195**
	(0.0988, 0)	-0.3851	0.4235	0.5240	0.2475	**0.0047**
	(0.0406, 0)	-0.3836	0.4098	0.5411	0.2449	**0.0113**
MRL	(0, 0.1940)	-0.3953	0.4321	0.4806	0.2772	**0.0028**
	(0, 0.6378)	-0.3790	0.4433	0.5106	0.2442	**0.0196**
	(0, 0.4742)	-0.3850	0.4392	0.4995	0.2564	**0.0115**
RTP	(0.0202, 0.1940)	-0.3833	0.3977	0.5530	0.2452	**0.0236**
	(0.0988, 0.6378)	-0.3805	0.4277	0.5323	0.2373	**0.0252**
	(0.0406, 0.4742)	-0.3821	0.4115	0.5436	0.2414	**0.0243**

Corresponding to the results obtained in the preceding example, from Table 6.6 we can observe that the restricted estimators are performing better than the unrestricted estimators in the sense of smaller SMSE. Puzzlingly, the Liu estimator provides estimates with SMSE values that are remarkably far away from the other unrestricted estimators. The RR estimator is superior to the unrestricted estimator set, and the RRR estimator is superior to all the restricted and unrestricted estimators in the sense of SMSE criterion. Also, it was noted that the RR and RRR estimators of $\hat{k}_{AM}$ perform better compering to the RR and RRR estimators of $\hat{k}_{HM}$ and $\hat{k}_{GM}$ as well as outperform all other biased estimators.

Table 6.7. *The superiority according to MMSE*

Superiority of by MMSE criterion	**Condition of superiority**	**Programme's results**
RLS to OLS by MMSE criterion	$\lambda_{max}(1 - \boldsymbol{R}'[\boldsymbol{RS^{-1}R}']^{-1}\boldsymbol{RS^{-1}}) \leq 1$	**1**
RRR to RR by MMSE criterion	$\lambda_{max}(\boldsymbol{A^{-1}M}_k\boldsymbol{S}_k\boldsymbol{AS}_k\boldsymbol{M}_k) \leq 1$	**1.0009**
		1.0047

		1.0020
MRL to LE by MMSE criterion	$\lambda_{max}(\boldsymbol{C^{-1}M_1S_1CS_1M_1}) \leq 1$	**1.0036**
		1.0314
		1.0169
RTP to TP by MMSE criterion	$\lambda_{max}(\boldsymbol{B^{-1}M_kS_kBS_kM_k}) \leq 1$	**1.0002**
		1.0021
		1.0002

When the estimators are compared according to MMSE criterion, the RLS estimator is superior to the OLS estimator, the RRR estimators is superior to the RR estimators, the MRL estimators is superior to the Liu estimators, and the RTP estimators is superior to the TP estimators, since the data fulfills the necessary and sufficient conditions mentioned in the Theorems (5.1.1), (5.1.2), (5.1.3) and (5.1.4).

7. MONTE-CARLO SIMULATION STUDY

For the sake of drawing an extensive and generalizable conclusion of relative characteristics of our pre-discussed estimators further than the results of the investigation that had been on the real data, a simulation study has been conducted.

7.1. Simulation Essence

In this section, we perform a Monte-Carlo simulation study by using MATLAB program. By following McDonald and Galarneau (1975) as many other researchers like Kejian (1993) the generation of independent variables was based on subsequent equation

$$X_{ij} = (1-\gamma^2)^{1/2} Z_{ij} + \gamma Z_{ip}, i = 1,2,\dots,n, j = 1,2,\dots,p, \qquad (7.1)$$

where Z_{ij} are independent standard normal pseudo-random numbers and γ is defined in such a way that the correlation between any two explanatory variables is provided by γ^2. The independent variables were standardized resulting in $\boldsymbol{X'X}$ being in correlation form. The observations on the dependent variables were generated by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i, i = 1,2,\dots,n, \qquad (7.2)$$

where e_i are independent normal $(0, \sigma^2)$ pseudo-random numbers. Also, $\boldsymbol{y}$ is standardized such that $\boldsymbol{X'y}$ represents the vector of the dependent variable's correlations with each explanatory variable. In this study a comprehensive simulation is intended; therefore, three different sample sizes $n = 25, 50, 100$ for two different independent variables numbers $p = 4, 7$ are adopted, and four distinct values of correlations $\gamma = 0.80, 0.90, 0.95, 0.99$ are considered to demonstrate a weak, strong and severe level of

multicollinearity among the regressors. Also, for the error standard deviation two different values $\sigma = 1, 5$ are investigated, and the intercept β_0 is presumed to be identically zero.

According to Newhouse and Oman (1971) if the mean square error is a function of $\boldsymbol{\beta}$, σ^2 and k and if the independent variables are fixed, then the MMSE is minimized when $\boldsymbol{\beta}$ is the normalized eigenvector corresponding to the largest eigenvalue of $\boldsymbol{X'X}$ matrix subject to constraint that $\boldsymbol{\beta'\beta} = 1$. Consequently, we selected the eigenvector that makes $\boldsymbol{\beta'\beta} = 1$ as a parameter vector. Now coming to the exact linear restrictions, when $p = 4$ the $\boldsymbol{R\beta} = \boldsymbol{r}$ was specified supposing $\boldsymbol{r} = [0]$ then $\boldsymbol{R} = [1 \quad -1 \quad 1 \quad 0]$, and when $p = 7$ the linear restriction was specified supposing that $\boldsymbol{R} = [1 \quad 0 \quad 1 \quad -1 \quad 1 \quad 1 \quad 0]$ as a result $\boldsymbol{r} = [1.0640]$. Notice that to ensure $rank(\boldsymbol{R}) = m < p$ we designed R matrix to be $1 \times p$.

In simulation study, for each σ, γ, p and n the experiment is performed 3000 times. The estimated mean square error (EMSE) is computed as

$$EMSE(\widehat{\boldsymbol{\alpha}}) = \frac{1}{RN}\sum_{i=1}^{RN}(\widehat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha})'(\widehat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}), \tag{7.3}$$

where $\boldsymbol{\alpha}$ is the parameter vector in the canonical form, $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{i}}$ is one of α estimators in the ith replication, and RN is the replications number of the experiment which in our study end up to 3000.

7.2. Simulation Results

The values of the parameter coefficients and the eigenvalues of $\boldsymbol{X'X}$ for the different values of p and n resulting from the simulation are assembled in Tables 7.1-7.6.

Table 7.1. *Values of $\boldsymbol{\beta}$ and λ for all γ used in simulation when p=4 and n=25*

$\boldsymbol{\gamma}$	**0.80**	**0.90**	**0.95**	**0.99**
$\boldsymbol{\lambda_1}$	3.3251	1.5832	0.7666	0.1478
$\boldsymbol{\lambda_2}$	5.0517	2.7095	1.4265	0.306
$\boldsymbol{\lambda_3}$	11.5818	6.187	3.1654	0.634
$\boldsymbol{\lambda_4}$	76.0413	85.5203	90.6414	94.9122
$\boldsymbol{\beta_1}$	0.4833	0.4933	0.4972	0.4995
$\boldsymbol{\beta_2}$	0.4605	0.4815	0.4912	0.4984
$\boldsymbol{\beta_3}$	0.5203	0.508	0.5033	0.5005
$\boldsymbol{\beta_4}$	0.5326	0.5164	0.5082	0.5016

Table 7.2. *Values of $\boldsymbol{\beta}$ and λ for all γ used in simulation when p=4 and n=50*

$\boldsymbol{\gamma}$	**0.80**	**0.90**	**0.95**	**0.99**
$\boldsymbol{\lambda_1}$	8.2014	3.8871	1.8814	0.3631
$\boldsymbol{\lambda_2}$	17.4692	9.4835	4.9086	0.9977
$\boldsymbol{\lambda_3}$	20.8821	11.2794	5.8173	1.177
$\boldsymbol{\lambda_4}$	149.4473	171.3499	183.3927	193.4622
$\boldsymbol{\beta_1}$	0.4796	0.906	0.4955	0.4992
$\boldsymbol{\beta_2}$	0.4955	0.4974	0.4985	0.4997
$\boldsymbol{\beta_3}$	0.4873	0.4933	0.4967	0.4994
$\boldsymbol{\beta_4}$	0.5357	0.5183	0.5092	0.5018

Table 7.3. *Values of $\boldsymbol{\beta}$ and λ for all γ used in simulation when p=4 and n=100*

$\boldsymbol{\gamma}$	**0.80**	**0.90**	**0.95**	**0.99**
$\boldsymbol{\lambda_1}$	8.1721	3.8203	1.8575	0.3664
$\boldsymbol{\lambda_2}$	23.1251	11.7848	5.9776	1.2185
$\boldsymbol{\lambda_3}$	27.9873	13.9167	6.9404	1.394
$\boldsymbol{\lambda_4}$	336.7155	366.4782	381.2245	393.0211
$\boldsymbol{\beta_1}$	0.4959	0.4977	0.4988	0.4997
$\boldsymbol{\beta_2}$	0.484	0.4935	0.4971	0.4995
$\boldsymbol{\beta_3}$	0.4938	0.4964	0.498	0.4996
$\boldsymbol{\beta_4}$	0.5254	0.5122	0.506	0.5012

Table 7.4. *Values of $\boldsymbol{\beta}$ and λ for all γ used in simulation when p=7 and n=25*

$\boldsymbol{\gamma}$	**0.80**	**0.90**	**0.95**	**0.99**
$\boldsymbol{\lambda_1}$	2.2178	1.0438	0.506	0.0986
$\boldsymbol{\lambda_2}$	4.0255	2.3015	1.2472	0.2700
$\boldsymbol{\lambda_3}$	7.6492	4.4273	2.3809	0.5023
$\boldsymbol{\lambda_4}$	10.6060	6.0798	3.2960	0.7142
$\boldsymbol{\lambda_5}$	13.9984	8.3502	4.6085	1.0023
$\boldsymbol{\lambda_6}$	22.1580	13.0122	7.0590	1.4948
$\boldsymbol{\lambda_7}$	107.3451	132.7852	148.9024	163.9178
$\boldsymbol{\beta_1}$	0.388	0.3842	0.3814	0.3786
$\boldsymbol{\beta_2}$	0.345	0.3613	0.3695	0.3763
$\boldsymbol{\beta_3}$	0.407	0.3909	0.3837	0.3789
$\boldsymbol{\beta_4}$	0.383	0.3813	0.3797	0.3783

$\boldsymbol{\beta_5}$	0.325	0.3533	0.3667	0.3761
$\boldsymbol{\beta_6}$	0.327	0.3541	0.3663	0.3756
$\boldsymbol{\beta_7}$	0.453	0.4164	0.3975	0.3819

Table 7.5. *Values of $\boldsymbol{\beta}$ and λ for all γ used in simulation when p=7 and n=50*

$\boldsymbol{\gamma}$	**0.80**	**0.90**	**0.95**	**0.99**
$\boldsymbol{\lambda_1}$	2.4025	1.0645	0.5032	0.0965
$\boldsymbol{\lambda_2}$	8.3538	4.0858	0.2161	0.3978
$\boldsymbol{\lambda_3}$	9.0849	4.4600	2.2062	0.4382
$\boldsymbol{\lambda_4}$	10.0233	4.9840	2.4925	0.5023
$\boldsymbol{\lambda_5}$	13.1696	6.4999	2.2227	0.6399
$\boldsymbol{\lambda_6}$	17.8777	8.7072	4.2872	0.8459
$\boldsymbol{\lambda_7}$	282.0882	313.1985	328.2722	340.0793
$\boldsymbol{\beta_1}$	0.3677	0.3734	0.3758	0.3776
$\boldsymbol{\beta_2}$	0.3820	0.3794	0.3785	0.3780
$\boldsymbol{\beta_3}$	0.3719	0.3753	0.3768	0.3778
$\boldsymbol{\beta_4}$	0.3685	0.3742	0.3763	0.3777
$\boldsymbol{\beta_5}$	0.3715	0.3764	0.3762	0.3776
$\boldsymbol{\beta_6}$	0.3760	0.3773	0.3777	0.3779
$\boldsymbol{\beta_7}$	0.4067	0.3913	0.3844	0.3792

Table 7.76. *Values of $\boldsymbol{\beta}$ and λ for all γ used in simulation when p=7 and n=100*

$\boldsymbol{\gamma}$	**0.80**	**0.90**	**0.95**	**0.99**
$\boldsymbol{\lambda_1}$	6.5180	2.9371	1.3974	0.2689
$\boldsymbol{\lambda_2}$	22.0968	11.4587	5.8252	1.1803
$\boldsymbol{\lambda_3}$	24.9985	12.9131	6.5884	1.3505
$\boldsymbol{\lambda_4}$	31.0196	16.2672	8.3721	1.7292
$\boldsymbol{\lambda_5}$	42.1492	21.6294	10.8464	2.1422
$\boldsymbol{\lambda_6}$	46.1978	23.9487	12.0142	2.3668
$\boldsymbol{\lambda_7}$	520.0202	603.8457	647.9564	683.9622
$\boldsymbol{\beta_1}$	0.3844	0.3806	0.3791	0.3781
$\boldsymbol{\beta_2}$	0.3711	0.3747	0.3763	0.3776
$\boldsymbol{\beta_3}$	0.3880	0.3822	0.3798	0.3783
$\boldsymbol{\beta_4}$	0.3499	0.3662	0.3728	0.3771
$\boldsymbol{\beta_5}$	0.3733	0.3721	0.3763	0.3776
$\boldsymbol{\beta_6}$	0.3511	0.3666	0.3729	0.3771

$\boldsymbol{\beta_7}$	0.4230	0.3993	0.3883	0.3800

Tables 7.7-7.12 describe the simulation findings of 3000 replications. On the tables, the condition number (CN) of each n and p altered experiment is also given. Deserves to be mentioned that the k's used in the simulation are the harmonic ($\hat{k}_{HM}$), the arithmetic ($\hat{k}_{AM}$) means of $\hat{k}$ and the Lukman bias estimator ($\hat{k}_{LM}$), respectively.

Table 7.7. *Estimated k and d and EMSE for the restricted and unrestricted biased estimators when n=25, p=4*

			$\gamma = 0.80$				$CN = 22.869$			
			EMSE of the unrestricted biased estimators				**EMSE of the restricted biased estimators**			
σ	k	d	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.793**	**0.019**	0.598	0.293	0.375	0.254	0.576	0.330	0.402	0.298
	511.083	**0.377**	0.598	0.291	0.490	0.301	0.576	0.338	0.487	0.333
	34.831	**0.191**	0.598	0.177	0.456	0.257	0.576	0.237	0.461	0.294
5	**0.705**	**0.285**	373.965	338.165	293.513	348.819	270.899	244.860	213.885	251.926
	249.299	**0.801**	373.965	128.566	360.486	348.161	270.899	94.396	261.938	262.714
	4.063	**0.330**	373.965	335.978	296.369	348.759	270.899	243.165	215.954	251.774
			$\gamma = 0.90$				CN =54.017			
σ	k	d	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.235**	**0.048**	1.171	0.462	0.597	0.485	0.996	0.440	0.554	0.458
	534.776	**0.320**	1.171	0.279	0.819	0.512	0.996	0.308	0.721	0.494
	15.147	**0.204**	1.171	0.239	0.763	0.470	0.996	0.267	0.678	0.447
5	**0.388**	**0.299**	731.801	658.704	484.220	682.113	538.550	483.644	359.029	500.031
	338.892	**0.830**	731.801	185.110	698.035	680.775	538.550	134.155	517.653	524.578
	3.780	**0.344**	731.801	654.337	492.289	681.983	538.550	480.206	364.974	499.712
			$\gamma = 0.95$				CN =118.234			
σ	k	d	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**1.663**	**0.078**	2.321	0.780	0.883	0.949	1.860	0.662	0.757	0.797
	3.717e+3	**0.296**	2.321	0.324	1.324	0.936	1.860	0.323	1.099	0.830
	7.715	**0.250**	2.321	0.379	1.229	0.908	1.860	0.361	1.021	0.777
5	**0.206**	**0.308**	1.451e+3	1.302e+3	759.952	1.352e+3	1.081e+3	967.253	575.340	1.003e+3

	1.021e+3	**0.856**	1.451e+3	251.251	1.370e+3	1.349e+3	1.081e+3	179.788	1.039e+3	1.058e+3
	1.582	**0.352**	1.451e+3	1.294e+3	780.193	1.351e+3	1.081e+3	960.315	590.385	1.002e+3
					$\boldsymbol{\gamma = 0.99}$			$\boldsymbol{CN = 642.202}$		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**0.640**	**0.129**	11.531	3.261	1.959	4.661	8.876	2.445	1.633	3.579
	34.767	**0.319**	11.531	0.705	4.385	4.396	8.876	0.533	3.604	3.629
	3.622	**0.331**	11.531	1.563	3.673	4.436	8.876	1.238	2.961	3.519
5	**0.045**	**0.320**	7.207e+3	6.454e+3	2.2822e+3	6.711e+3	5.469e+3	4.872e+3	1.855e+3	5.068e+3
	600.704	**0.890**	7.207e+3	349.139	6.712e+3	6.696e+3	5.469e+3	228.653	5.359e+3	5.405e+3
	1.470	**0.362**	7.207e+3	6.410e+3	2.369e+3	6.710e+3	5.469e+3	4.837e+3	1.926e+3	5.065e+3

Table 7.8. *Estimated k and d and EMSE for the restricted and unrestricted biased estimators when n=50, p=4*

					$\boldsymbol{\gamma = 0.80}$			$\boldsymbol{CN = 18.222}$		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**3.391**	**0.005**	0.234	0.154	0.175	0.108	0.260	0.191	0.211	0.156
	1.420e+3	**0.468**	0.234	0.277	0.216	0.141	0.260	0.313	0.245	0.182
	248.282	**0.193**	0.234	0.116	0.201	0.113	0.260	0.163	0.232	0.160
5	**1.688**	**0.263**	146.379	132.884	131.807	136.289	113.348	102.422	102.079	105.894
	757.506	**0.776**	146.379	57.127	143.658	136.141	113.348	41.939	111.272	106.084
	4.992	**0.301**	146.379	132.004	132.415	136.272	113.348	101.756	102.520	105.880
					$\boldsymbol{\gamma = 0.90}$			$\boldsymbol{CN = 44.081}$		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.971**	**0.023**	0.457	0.240	0.308	0.202	0.443	0.260	0.320	0.238
	1.44e+3	**0.359**	0.457	0.229	0.388	0.227	0.443	0.267	0.385	0.258

	29.133	**0.176**	0.457	0.144	0.360	0.198	0.443	0.186	0.362	0.235
5	**0.947**	**0.295**	285.580	257.636	235.923	265.795	223.284	200.214	184.526	208.516
	1.07e+3	**0.809**	285.580	83.568	278.122	265.481	223.284	59.806	217.596	209.152
	2.492	**0.332**	285.580	255.912	237.800	265.753	223.284	198.904	185.890	208.488
					$\gamma = 0.95$			$CN = 97.476$		
σ	k	d	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.440**	**0.050**	0.911	0.385	0.517	0.392	0.804	0.369	0.481	0.393
	660.995	**0.302**	0.911	0.215	0.674	0.398	0.804	0.245	0.610	0.401
	160.254	**0.186**	0.911	0.206	0.630	0.374	0.804	0.236	0.572	0.383
5	**0.506**	**0.309**	569.090	511.692	406.871	529.574	446.939	399.147	321.003	417.243
	852.577	**0.838**	569.090	120.004	548.970	528.806	446.939	83.164	431.763	418.920
	4.337	**0.347**	569.090	508.260	412.491	529.482	446.939	396.537	325.054	417.186
					$\gamma = 0.99$			$CN = 532.792$		
σ	k	d	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**1.152**	**0.108**	4.592	1.411	1.285	1.905	3.707	1.132	1.093	1.600
	193.966	**0.292**	4.592	0.429	2.158	1.797	3.707	0.367	1.800	1.546
	6.382	**0.292**	4.592	0.685	1.964	1.805	3.707	0.622	1.605	1.555
5	**0.110**	**0.321**	2.870e+3	2.574e+3	1.206e+3	2.670e+3	2.260e+3	2.011e+3	1.004e+3	2.109e+3
	3.921e+3	**0.885**	2.870e+3	196.285	2.708e+3	2.666e+3	2.260e+3	125.942	2.146e+3	2.123e+3
	1.621	**0.359**	2.870e+3	2.556e+3	1.246e+3	2.669e+3	2.260e+3	1.998e+3	1.033e+3	2.108e+3

Table 7.9. *Estimated k and d and EMSE for the restricted and unrestricted biased estimators when n=100, p=4*

					$\gamma = 0.80$			$CN = 41.203$		
σ	k	d	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**3.423**	**0.006**	0.206	0.134	0.157	0.097	0.268	0.202	0.226	0.172
	1.43e+3	**0.430**	0.206	0.199	0.190	0.114	0.268	0.253	0.252	0.173

	70.316	**0.200**	0.206	0.088	0.178	0.096	0.268	0.161	0.243	0.166
5	**2.117**	**0.327**	128.571	115.592	117.419	119.910	94.347	84.570	86.266	87.569
	1.04e+3	**0.794**	128.571	42.266	126.721	119.701	94.377	30.901	93.118	91.482
	7.546	**0.364**	128.571	114.769	117.878	119.881	94.347	83.922	86.602	87.504
				$\gamma = 0.90$				**$CN = 95.929$**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**3.001**	**0.029**	0.423	0.218	0.291	0.192	0.430	0.257	0.322	0.238
	491.882	**0.341**	0.423	0.163	0.359	0.198	0.430	0.215	0.375	0.239
	31.587	**0.184**	0.423	0.118	0.337	0.182	0.430	0.176	0.357	0.229
5	**1.107**	**0.343**	264.429	236.678	222.487	246.565	196.469	175.072	165.650	182.368
	3.70e+3	**0.825**	264.429	64.720	258.783	246.040	196.469	46.594	192.962	191.132
	3.375	**0.380**	264.429	235.037	224.042	246.503	196.469	173.777	166.800	182.246
				$\gamma = 0.95$				**$CN = 205.239$**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.476**	**0.062**	0.855	0.353	0.498	0.376	0.756	0.349	0.472	0.373
	560.278	**0.304**	0.855	0.171	0.636	0.363	0.756	0.212	0.580	0.373
	46.987	**0.197**	0.855	0.180	0.598	0.353	0.756	0.216	0.550	0.359
5	**0.577**	**0.351**	534.328	477.241	393.739	498.150	400.932	356.203	297.104	372.319
	5.48e+3	**0.851**	534.328	90.640	518.368	496.873	400.932	63.567	392.257	390.818
	2.311	**0.388**	534.328	473.986	398.530	498.023	400.932	353.635	300.638	372.092
				$\gamma = 0.99$				**$CN = 1072.73$**		
	k	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**1.205**	**0.124**	4.276	1.274	1.297	1.798	3.366	1.127	1.078	1.455
	157.228	**0.296**	4.276	0.401	2.050	1.672	3.366	0.372	1.684	1.431
	4.943	**0.293**	4.276	0.593	1.893	1.682	3.366	0.522	1.536	1.404
5	**0.128**	**0.357**	2.672e+3	2.383e+3	1.229e+3	2.491e+3	2.032e+3	1.801e+3	984.607	1.889e+3
	1.254e+4	**0.889**	2.672e+3	126.367	2.536e+3	2.484e+3	2.032e+3	81.385	1.989e+3	1.992e+3

	1.840	**0.394**	2.672e+3	2.367e+3	1.263e+3	2.490e+3	2.032e+3	1.788e+3	1.010e+3	1.888e+3

Table 7.10. *Estimated k and d and EMSE for the restricted and unrestricted biased estimators when n=25, p=7*

					$\gamma = 0.80$			$CN =$ **48.402**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**3.897**	**0.073**	1.040	0.392	0.705	0.440	0.737	0.312	0.527	0.334
	635.578	**0.435**	1.040	0.319	0.859	0.467	0.737	0.279	0.626	0.377
	35.491	**0.222**	1.040	0.209	0.796	0.412	0.737	0.178	0.585	0.314
5	**0.474**	**0.376**	650.111	576.111	525.541	607.231	460.729	413.780	379.841	430.028
	357.531	**0.870**	650.111	150.168	632.075	603.625	460.729	121.305	449.895	453.801
	0.732	**0.403**	650.111	572.566	529.967	607.124	460.729	411.209	382.942	429.779
					$\gamma = 0.90$			$CN = 127.210$		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.920**	**0.119**	1.963	0.605	1.116	1.028	1.388	0.478	0.845	0.617
	1.023e+3	**0.401**	1.963	0.254	1.417	0.972	1.388	0.222	1.044	0.623
	17.216	**0.267**	1.963	0.279	1.303	0.750	1.388	0.228	0.969	0.563
5	**0.266**	**0.402**	1.227e+3	1.077e+3	884.755	1.146e+3	867.562	770.563	646.895	808.315
	174.219	**0.881**	1.227e+3	212.230	1.180e+3	1.139e+3	867.562	171.363	845.920	866.450
	0.504	**0.427**	1.227e+3	1.070e+3	896.137	1.146e+3	867.562	765.404	654.890	807.696
					$\gamma = 0.95$			$CN =$**294.289**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.002**	**0.153**	3.816	1.026	1.727	1.595	2.699	0.806	1.332	1.185
	408.994	**0.386**	3.816	0.227	2.337	1.401	2.699	0.200	1.754	1.133
	10.928	**0.314**	3.816	0.427	2.128	1.431	2.699	0.338	1.610	1.068
5	**0.141**	**0.416**	2.385e+3	2.081e+3	1.493e+3	2.229e+3	1.686e+3	1.488e+3	1.117e+3	1.569e+3
	536.684	**0.892**	2.385e+3	297.815	2.268e+3	2.212e+3	1.686e+3	241.884	1.657e+3	1.709e+3

	0.471	**0.441**	2.385e+3	2.067e+3	1.520e+3	2.228e+3	1.686e+3	1.477e+3	1.137e+3	1.568e+3
				$\gamma = 0.99$					**CN =1662.300**	
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**0.613**	**0.190**	18.647	4.358	4.615	7.722	13.198	3.401	3.742	5.718
	52.212	**0.385**	18.647	0.427	7.810	6.580	13.198	0.373	6.174	5.364
	6.046	**0.376**	18.647	1.671	6.865	6.893	13.198	1.277	5.453	5.138
5	**0.030**	**0.429**	1.165e+4	1.011e+4	5.450e+3	1.089e+4	8.249e+3	7.233e+3	4.420e+3	7.666e+3
	391.269	**0.913**	1.165e+4	531.585	1.087e+4	1.080e+4	8.249e+3	0.445e+3	8.459e+3	8.614e+3
	0.429	**0.453**	1.165e+4	1.005e+4	5.615e+3	1.089e+4	8.249e+3	7.181e+3	4.511e+3	7.658e+3

Table 7.11. *Estimated k and d and EMSE for the restricted and unrestricted biased estimators when n=50, p=7*

				$\gamma = 0.80$					**$CN = 117.413$**	
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**4.097**	**0.081**	0.888	0.358	0.635	0.394	0.783	0.310	0.559	0.346
	790.860	**0.408**	0.888	0.221	0.747	0.377	0.783	0.189	0.658	0.342
	41.622	**0.211**	0.888	0.185	0.702	0.355	0.783	0.161	0.618	0.316
5	**0.537**	**0.381**	554.793	492.055	463.605	518.334	488.500	432.061	408.541	455.440
	878.785	**0.872**	554.793	119.744	542.017	514.952	488.500	102.961	477.652	461.550
	0.651	**0.401**	554.793	489.118	466.951	518.225	488.500	429.385	411.476	455.265
				$\gamma = 0.90$					**$CN = 294.220$**	
	k	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**2.937**	**0.120**	1.893	0.615	1.122	0.819	1.668	0.530	0.984	0.720
	487.588	**0.383**	1.893	0.190	1.385	0.718	1.668	0.162	1.220	0.653
	15.251	**0.255**	1.893	0.274	1.290	0.727	1.668	0.237	1.134	0.649
5	**0.260**	**0.398**	1.183e+3	1.043e+3	869.456	1.106e+3	1.042e+3	915.676	768.721	971.182
	594.012	**0.888**	1.183e+3	184.612	1.142e+3	1.098e+3	1.042e+3	156.894	1.010e+3	986.973

	0.315	**0.417**	1.183e+3	1.037e+3	880.062	1.106e+3	1.042e+3	909.915	778.004	970.745
				$\gamma = 0.95$					**CN = 652.400**	
	k	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**1.929**	**0.144**	3.907	1.109	1.809	1.659	3.441	0.952	1.589	1.460
	236.112	**0.374**	3.907	0.206	2.407	1.411	3.441	0.171	2.124	1.286
	10.185	**0.312**	3.907	0.444	2.225	1.468	3.441	0.385	1.957	1.314
5	**0.128**	**0.405**	2.442e+3	2.148e+3	1.529e+3	2.284e+3	2.150e+3	1.884e+3	1.363e+3	2.004e+3
	1.351e+3	**0.901**	2.442e+3	262.643	2.330e+3	2.264e+3	2.150e+3	0.222e+3	2.072e+3	2.042e+3
	0.160	**0.425**	2.442e+3	2.135e+3	1.556e+3	2.283e+3	2.150e+3	1.872e+3	1.387e+3	2.003e+3
				$\gamma = 0.99$					**CN =3523.237**	
	k	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**0.547**	**0.172**	19.995	4.981	4.701	8.329	17.597	4.264	4.252	7.335
	85.680	**0.382**	19.995	0.567	8.167	7.038	17.597	0.457	7.367	6.433
	11.122	**0.382**	19.995	1.893	7.302	7.417	17.597	1.655	6.536	6.661
5	**0.025**	**0.411**	1.250e+4	1.097e+4	5.970e+3	1.169e+4	1.100e+4	9.620e+3	5.438e+3	1.025e+4
	2.625e+4	**0.923**	1.250e+4	388.799	1.169e+4	1.159e+4	1.100e+4	319.652	1.054e+4	1.050e+4
	0.067	**0.431**	1.250e+4	1.091e+4	6.042e+3	1.169e+4	1.100e+4	9.559e+3	5.507e+3	1.025e+4

Table 7.12. *Estimated k and d and EMSE for the restricted and unrestricted biased estimators when n=100, p=7*

				$\gamma = 0.80$					**CN =79.782**	
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**5.490**	**0.039**	0.317	0.174	0.263	0.155	0.245	0.147	0.210	0.127
	4614.253	**0.483**	0.317	0.241	0.296	0.167	0.245	0.203	0.232	0.144
	102.477	**0.200**	0.317	0.111	0.282	0.144	0.245	0.098	0.222	0.120
5	**1.553**	**0.377**	198.163	175.597	184.073	185.082	150.989	135.542	141.164	140.212
	2167.309	**0.853**	198.163	51.744	195.962	183.987	150.989	43.690	149.572	149.713

	1.966	**0.401**	198.163	174.457	184.648	185.034	150.989	134.613	141.598	140.072
				$\gamma = 0.90$				**$CN = 205.589$**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**4.577**	**0.081**	0.653	0.278	0.494	0.309	0.498	0.233	0.394	0.246
	1702.663	**0.404**	0.653	0.171	0.565	0.282	0.498	0.147	0.442	0.244
	27.895	**0.197**	0.653	0.150	0.534	0.276	0.498	0.130	0.421	0.225
5	**0.787**	**0.405**	408.011	358.634	356.263	381.420	309.070	275.237	273.318	286.680
	702.633	**0.869**	408.011	82.828	400.538	378.630	309.070	70.223	305.145	311.437
	1.038	**0.426**	408.011	356.316	358.202	381.298	309.070	273.319	274.803	286.347
				$\gamma = 0.95$				**$CN = 463.703$**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**3.520**	**0.117**	1.327	0.455	0.868	0.602	1.010	0.380	0.699	0.476
	976.972	**0.373**	1.327	0.139	1.028	0.512	1.010	0.120	0.815	0.443
	18.627	**0.227**	1.327	0.221	0.964	0.533	1.010	0.187	0.769	0.429
5	**0.396**	**0.416**	829.409	726.137	657.913	775.736	628.939	557.535	510.236	583.092
	791.103	**0.884**	829.409	126.481	806.416	769.208	628.939	107.405	621.393	640.870
	0.567	**0.437**	829.409	721.469	663.919	775.468	628.939	553.625	514.961	582.375
				$\gamma = 0.99$				**$CN = 2543.556$**		
σ	***k***	***d***	**OLS**	**RR**	**LE**	**TP**	**RLS**	**RRR**	**MRL**	**RTP**
1	**1.360**	**0.163**	6.725	1.746	2.571	2.886	5.140	1.444	2.134	2.278
	185.821	**0.370**	6.725	0.238	3.594	2.408	5.140	0.200	2.973	2.102
	9.594	**0.342**	6.725	0.668	3.314	2.540	5.140	0.553	2.742	2.043
5	**0.079**	**0.424**	4203.171	3668.386	2407.509	3932.684	3209.213	2832.084	1974.738	2974.577
	751.011	**0.911**	4203.171	246.996	3987.508	3894.249	3209.213	207.005	3267.161	3335.638
	0.272	**0.444**	4203.171	3644.953	2454.065	3931.223	3209.213	2812.175	2013.169	2970.834

The results in Tables 7.7-7.12 summarize the average of the mean square errors values resulting from the simulation for each estimator, and show the averaged values for simulated harmonic, arithmetic means, and Lukman's bias of the parameter k and their corresponding average of simulated d-optimal for each k estimate method.

Subsequent sections include a detailed commentary on the results of the Tables 7.7-7.12 in respect of the restrictions and various variables γ, n, $p, k, d,$ and σ, which all or some were considered as comparison criteria previously in the literature, see (Lukman et al., 2020; Lukman et al., 2019; Najarian et al., 2013; Özkale, 2014), and the researcher believes of their impact on the simulation results. Two fundamental questions will guide commentary and observations. First, does the standard have an effect on the EMSE values in general? With explaining whether the effect is favorable or bad. Second, does the criterion influence the degree of superiority between the restricted and unrestricted estimators?

7.3. Performance as a Function of the Restrictions

Tables 7.7-7.12 show that the performance of the estimators varies depending on the linear constraints. The columns 8-11 of Tables clearly indicate that the restricted estimators show better performance than the unrestricted estimators by having smaller averaged mean square error values compared to their unconstrained counterparts. This efficiency in performance in detail applies for all the study's estimators regardless of n, p, k, d and for fixed σ and γ values. i.e, the $\widehat{\boldsymbol{\beta}}_r$ estimator is superior to the $\widehat{\boldsymbol{\beta}}$, the $\widehat{\boldsymbol{\beta}}_r(k)$ estimator is superior to the $\widehat{\boldsymbol{\beta}}_k$ estimator, the $\widehat{\boldsymbol{\beta}}_r(d)$ estimator is superior to the $\widehat{\boldsymbol{\beta}}_d$ estimator, and the $\widehat{\boldsymbol{\beta}}_r(k,d)$ estimator is superior to the $\widehat{\boldsymbol{\beta}}(k,d)$ estimator. This demonstrates what was presented in the fifth chapter's theories. Nonetheless, some exceptions were noticed in Tables 7.7, 7.8 and 7.9 when $\gamma = 0.80, 0.90$ and $\sigma = 1$, in these specific parts of Tables the unrestricted biased estimators show better performance than the restricted estimators by having fewer EMSE values. This might be explained by the fact that these cases create null to low degrees of multicollinearity, resulting in modest condition numbers. As a result, we conclude that in the absence of multicollinearity, unrestricted estimators outperform restricted estimators. It is worth noting that this case of role inversion between the two groups does not occur when $\sigma = 5$. This leads to another conclusion: that the restricted estimators do perform better when the error has a considerable standard deviation even in the absence of the multicollinearity. In other

words, as σ gets larger not only the EMSE values grows, but also does the degree of superiority of the restricted estimators over the unrestricted estimators.

7.4. Performance as a Function of γ

The previously stated fact, that γ^2 represents the correlation between any two explanatory variables offers reasonable justification for the gradual increase in the degree of multicollinearity by increasing the value of γ. The escalation of the condition numbers' values indicated an increase in the amount of multiple linearity from minor to moderate to severe. It is worth noting that when $p = 4$, and $n = 25, 50, 100$ the ill-conditioning of X ranges from null to a moderate level of multicollinearity without reaching the severe level, unlike for $p = 7$, and $n = 25, 50, 100$ which ranges from medium to severe level of multicollinearity with an unique expectation when $n = 25, \gamma = 0.80$ where $CN = 48.402$ indicates a negligible degree of internal correlation. The scaling of multicollinearity degree based on the CN was discussed in Chapter 3.

Generally, for fixed values of σ, n and p, Tables 7.7-7.12 show that every increase in the correlation value represented in γ resulted in an increase in the amount of averaged mean square error for both the restricted and unrestricted biased estimators. Moreover, as γ increases the performance gap between the OLS estimator and the other unrestricted estimators becomes wider, as well as the performance gap between the RLS estimator and the other restricted estimators. Since by γ growing the EMSE values of the OLS and the RLS estimators increase substantially unlike the other estimators. This may explain the genuine influence of the multicollinearity on the OLS and RLS estimation by making them unstable, have large variance, and have longer vector than $\boldsymbol{\beta}$'s vector $\|\boldsymbol{\beta}\| < \|\widehat{\boldsymbol{\beta}}\|$ and $\|\boldsymbol{\beta}\| < \|\widehat{\boldsymbol{\beta}}_{\boldsymbol{r}}\|$.

7.5. Performance as a Function of k and d

Firstly, it is good to point out that the k estimates used in the simulation are respectively the harmonic ($\hat{k}_{HM}$), the arithmetic ($\hat{k}_{AM}$) means of $\hat{k}$ and the Lukman bias estimator ($\hat{k}_{LM}$). The variation in the three values of k from each other every time for the identical simulation inputs is due to the varied nature of each k's equation. As is observed in tables the $\hat{k}_{AM}$ gives the largest estimate for k, while both $\hat{k}_{HM}$ and $\hat{k}_{LM}$ generate relatively close estimates, yet in some cases, there are no significant differences between them, in general $\hat{k}_{HM}$ is always smaller. The same conclusion applies to the three values of d-optimal, due to d-optimal being reliant on k's estimations. To avoid the negativity

in $k's$ estimates a small initial d value was needed, we used $\hat{d} = 0.0005$ trying to satisfy (5.18) inequality.

Principally, the EMSE for each level of correlations decreases down as k grows. Which justifies the RR of $\hat{k}_{AM}$ with the smallest EMSE value among the RR estimators and all other unrestricted biased estimators. Similarly, having the RRR of $\hat{k}_{AM}$ the smallest EMSE value among the other RRR estimators and outperforming all restricted biased estimators become more understandable. With regard to d what happened is the exact opposite, the largest estimate of d corresponds to the largest EMSE values of the Liu and MRL estimators. Although, a negative link between k and the EMSE is concluded, on the contrary to Lukman et al. (2019) a positive link between d and EMSE is concluded.

7.6. Performance as a Function of σ

A positive direct relationship between the σ and the EMSE have been determined based on the results of the Tables. Regardless of γ value, dramatic changes have been noticed occurring in the EMSE values after a small change in σ i.e., when σ increases the error term increase accordingly. Furthermore, the superiorities of the restricted biased estimators over the unrestricted biased estimators are increasingly effective.

In the unrestricted biased estimator group, when $\sigma = 1$, the RR estimators outperform the other unrestricted estimators in virtually all instances, Specifically, the RR of $\hat{k}_{AM}$. The second acceptable performance is shown by the TP estimator. Exceptions are unavoidable; for example, in Table 7.7, when $\gamma = 0.99$, some of Liu's estimations take center stage. However, when $\sigma = 5$ the best performance has occurred by $\hat{k}_{AM}$ of RR, then by the first and third estimated values of d at Liu, followed by the other ridge and the other Liu estimations, indicating an overlapped level of efficiency between Liu and ridge except for the second estimated d due of being larger than the other values. The same conclusion applies for the restricted biased estimators' counterparts, as each restricted estimator operates with a performance grade equivalent to its unconstrained version.

7.7. Performance as a Function of n

For fixed values of σ, γ, and p Tables 7.7-7.12 display, that the rise in the number of observations leads to a decrease in the amount of averaged mean square error for both

the restricted and unrestricted biased estimators. Regarding the differences between restricted and unrestricted estimators, no correlation can be established between the number of observations or sample size and the degree of the unrestricted estimators' superiority over the restricted estimators. It means that the unrestricted estimators outperform the restricted biased estimators regardless of n.

7.8. Performance as a Function of p

According to Tables 7.7-7.12, the mean square error of both groups, the restricted and unrestricted biased estimators values grow as the number of regressors p increases for any given n, γ, and σ. Furthermore, we can notice from Tables 7.1-7.6 that the increase in the number of regressors p leads to a decrease in the regression coefficients values.

8. CONCLUSION

In order to identify and compare the performance of linear estimators when the regression model suffers from the multicollinearity problem, this applied study have been conducted. And since many biased alternatives to the ordinary least squares have been suggested in literature with the aim of obtaining a substantial reduction in variance, many comparison studies emerged as a result. With the aim of making a different comparison we have decided to separate some biased estimators into two main groups, the act that transfers the comparison to a radically different level of comparing between two linear regression models instead of only comparing the performance of a random group of estimators. Also, the comparison was applied once theoretically by using the criterion of the MMSE. These two models are the restricted linear regression model and the classical linear regression model. So, what differentiates this study is the comparison core itself, as well as the simulation study that was applied to it.

The OLS, the RR, the contraction, the Liu, the TP, the RRR, the RC, the MRL and the RTP estimators were broadly presented, and their properties as well were discussed. Also, the issue of selecting the biasing parameters k and d have been considered. Based on the MMSE criterion, some theories of comparing the restricted and the unrestricted estimators have been represented and discussed. Then the theoretical findings have been illustrated in terms of both the SMSE and the MMSE criteria, by using two different real-life data sets followed by a comprehensive simulation study controlled by several dimensions: σ, γ, p, n, k and d. The commentary was considered separately under the previous core areas. From the real applications we may highlight the following results:

- Rather than the unrestricted estimators, the restricted estimators have been demonstrated as a noble alternative to the OLS in estimating when the problem of Multicollinearity is existent in the linear regression model.
- In the sense of SMSE, the RRR estimator have shown outperform all the restricted and unrestricted estimators, whereas the RR is superior to the unrestricted estimators set.
- The RR and RRR estimators of $\hat{k}_{AM}$ perform better compering to the RR and RRR estimators of $\hat{k}_{HM}$ and $\hat{k}_{GM}$.

From the simulation study we may highlight the following results:

- The restricted biased estimators outperform the unrestricted biased estimators when the $\boldsymbol{X'X}$ is ill-conditioned in the linear regression model. On the other hand,

In the absence of multicollinearity, the unrestricted estimators outperform the restricted estimators.

- Almost everywhere, the ridge regression estimator performs best in the unrestricted estimators' group and as the restricted ridge regression does in the restricted estimators' group.
- An increase in the value of the correlation coefficient results to an increase in the EMSE for all the estimators.
- As the standard deviation of the error grows, so does the EMSE for each estimator.
- As k increases, the EMSE for each level of correlations decreases.
- As d increases, the EMSE for each level of correlations increases.
- An increase in the sample size n leads to a decrease in the EMSE for all the estimators.
- For every given $n,\ \gamma$ and σ, the values of the restricted and unrestricted biased estimators grow as the number of regressors p increases.

Given that the results in both practical examples and the Monte-Carlo simulation showed the superiority of the restricted estimators to the unrestricted estimators, the researcher believes that restricted estimation as a contemporary science deserves more attention from statisticians. This research can be further developed by including more estimators in the comparison or by including the stochastic restricted estimators beside or instead of the exact restricted estimators. In the present study, the researcher focused on the exact restricted estimators, however in practical applications, we do not have exact orthogonal prior information like $\boldsymbol{R\beta = r}$, for instance, studies involve industrial structures, production planning, etc. So, the uncertainty of the stochastic estimation is more appropriate for economic and financial studies. The researcher had a hope that the comparison would also include the modified ridge estimator, the restricted modified ridge estimator (RMR), the contraction and the restricted contraction (RC) estimators, but to the best of the author's knowledge the MMSE and SMSE equations of the RMR estimator and the RC estimator not been presented in literature yet. We eagerly await studies that will address this deficit.

REFERENCES

Ahmad, S., and Aslam, M. (2020). Another proposal about the new two-parameter estimator for linear regression model with correlated regressors. *Communications in Statistics-Simulation and Computation*, 1-19.

Akdeniz, F., and Erol, H. (2003). Mean squared error matrix comparisons of some biased estimators in linear regression. *Communications in statistics-theory and methods, 32*(12), 2389-2413.

Akdeniz, F., and Kaçiranlar, S. (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. doi:10.1080/03610929508831585

Akdeniz, F., and Kaçiranlar, S. (2001). More on the new biased estimator in linear regression. *Sankhyā: The Indian Journal of Statistics, Series B*, 321-325.

Albert, A. (1973). The Gauss–Markov theorem for regression models with possibly singular covariances. *SIAM Journal on Applied Mathematics, 24*(2), 182-187.

Alkhamisi, M. A., and Shukur, G. (2008). Developing Ridge Parameters for SUR Model. *Communications in Statistics - Theory and Methods, 37*(4), 544-564. doi:10.1080/03610920701469152

Arashi, M., and Valizadeh, T. (2015). Performance of Kibria's methods in partial linear ridge regression model. *Statistical Papers, 56*(1), 231-246. doi:10.1007/s00362-014-0578-6

Aslam, M. (2014). Performance of Kibria's Method for the Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence. *Communications in Statistics - Simulation and Computation, 43*(4), 673-686. doi:10.1080/03610918.2012.712185

Baye, M. R.,and Parker, D. F. (1984). Combining ridge and principal component regression:a money demand illustration. *Communications in Statistics - Theory and Methods, 13*(2), 197-205. doi:10.1080/03610928408828675

Bhattacharya, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. *Annals of Mathematical Statistics, 37*(6), 1819-1824.

Chatterjee, S., and Hadi, A. S. (2009). *Sensitivity analysis in linear regression* (Vol. 327): John Wiley & Sons.

Dempster, A. P., Schatzoff, M., and Wermuth, N. (1977). A Simulation Study of Alternatives to Ordinary Least Squares. *72*(357), 77-91. doi:10.1080/01621459.1977.10479910

Dey, D. K., Ghosh, M., and Strawderman, W. E. (1999). On estimation with balanced loss functions. *Statistics & probability letters, 45*(2), 97-101.

Dorugade, A. V. (2014). A modified two-parameter estimator in linear regression. *Statistics in Transition. New Series, 15*(1), 23-36.

Farebrother, R. W. (1976). Further results on the mean square error of ridge regression. *Journal of the Royal Statistical Society. Series B (Methodological), 38*(3), 248-250.

Farrar, D. E.,and Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107.

Gelfand, A. E., and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika, 85*(1), 1-11.

Gibbons, D. G. (1981). A Simulation Study of Some Ridge Estimators. *76*(373), 131-139. doi:10.1080/01621459.1981.10477619

Graybill, F. A. (1961). *An introduction to linear statistical models*. Retrieved from

Graybill, F. A. (1976). *Theory and application of the linear model* (Vol. 183): Duxbury press North Scituate, MA.

Graybill, F. A. (1983). Matrices with Applications in Statistics (Belmont, CA: Wadsworth). *Graybill2Matrices With Applications in Statistics1983*.

Groß, J. (2003). Restricted ridge estimation. *Statistics & probability letters, 65*(1), 57-64.

Gruber, M. H. J. (1998). Improving efficiency by shrinkage, Statistics: Textbooks and Monographs, vol. 156. In: Marcel Dekker Inc., New York.

Gruber, M. H. J. (2010). *Regression estimators: A comparative study*: JHU Press.

Gruber, M. I. I. J., and Rao, P. S. R. S. (1982). Bayes estimators for linear models with less than full rank. *Communications in statistics-theory and methods, 11*(1), 59-69.

Güler, H., and Kaçıranlar, S. (2009). A comparison of mixed and ridge estimators of linear models. *Communications in Statistics—Simulation and Computation®, 38*(2), 368-401.

Hald, A. (1952). *Statistical theory with engineering applications*. Retrieved from

Hamaker, H. C. (1962). On multiple regression analysis. *Statistica Neerlandica, 16*(1), 31-56.

Hefnawy, A. E.,and Farag, A. (2014). A Combined Nonlinear Programming Model and Kibria Method for Choosing Ridge Parameter Regression. *43*(6), 1442-1470. doi:10.1080/03610918.2012.735317

Hoerl, A. E., Kannard, R. W., and Baldwin, K. F. (1975). Ridge regression:some simulations. *Communications in Statistics, 4*(2), 105-123. doi:10.1080/03610927508827232

Hoerl, A. E.,and Kennard, R. W. (1970a). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics, 12*(1), 69-82. doi:10.1080/00401706.1970.10488635

Hoerl, A. E.,and Kennard, R. W. (1970b). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics, 12*(1), 55-67. doi:10.1080/00401706.1970.10488634

Lawless, J. F, and Wang P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics - Theory and Methods, 5*(4), 307-323. doi:10.1080/03610927608827353

James, W., and Stein, C. (1992). Estimation with quadratic loss. In *Breakthroughs in statistics* (pp. 443-460): Springer.

Kaçıranlar, S., and Sakallıoğlu, S. (2001). Combining the Liu estimator and the principal component regression estimator.

Kaciranlar, S., Sakallioglu, S., and Akdeniz, F. (1998). Mean squared error comparisons of the modified ridge regression estimator and iiie restricted ridge regression estimator. *Communications in statistics-theory and methods, 27*(1), 131-138.

Kaçiranlar, S., Sakallioğlu, S., Akdeniz, F., Styan, G. P. H., and Werner, H. J. (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.

Kaçıranlar, S., Sakallıoğlu, S., Özkale, M. R., and Güler, H. (2011). More on the restricted ridge regression estimation. *Journal of Statistical Computation and Simulation, 81*(11), 1433-1448.

Kejian, L. (1993). A new class of blased estimate in linear regression. *Communications in Statistics - Theory and Methods, 22*(2), 393-402. doi:10.1080/03610929308831027

Khalaf, G., and Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics - Theory and Methods, 34*(5), 1177-1182. doi:10.1081/sta-200056836

Kibria, B. M. G. (2003). Performance of Some New Ridge Regression Estimators. *32*(2), 419-435. doi:10.1081/sac-120017499

Kibria, B. M. G., and Banik, S. (2016). Some ridge regression estimators and their performances. *15*, 206-238.

Li, Y., and Yang, H. (2011). Two kinds of restricted modified estimators in linear regression model. *Journal of Applied Statistics, 38*(7), 1447-1454. doi:10.1080/02664763.2010.505951

Li, Y., and Yang, H. (2012). A new Liu-type estimator in linear regression model. *53*(2), 427-437. doi:10.1007/s00362-010-0349-y

Li, Y., and Yang, H. (2016). More on the two-parameter estimation in the restricted regression. *Communications in statistics-theory and methods, 45*(24), 7184-7196.

Liu, K. (2003). Using Liu-Type Estimator to Combat Collinearity. *32*(5), 1009-1020. doi:10.1081/sta-120019959

Lukman, A. F., Ayinde, K., Aladeitan, B., and Bamidele, R. (2020). An unbiased estimator with prior information. *Arab Journal of Basic and Applied Sciences, 27*(1), 45-55.

Lukman, A. F., Ayinde, K., Binuomote, S., and Clement, O. A. (2019). Modified ridge-type estimator to combat multicollinearity: Application to chemical data. *Journal of Chemometrics, 33*(5), e3125.

Mansson, K., Shukur, G., and Kibria, B. G. (2010). On some ridge regression estimators: A Monte Carlo simulation study under different error variances. *Journal of Statistics, 17*(1), 1-22.

Marquaridt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics, 12*(3), 591-612.

McDonald, G. C., and Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association, 70*(350), 407-416. doi:10.1080/01621459.1975.10479882

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821): John Wiley & Sons.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis*: John Wiley & Sons.

Muniz, G., Kibria, B. M., and Shukur, G. (2012). On developing ridge regression parameters: a graphical investigation.

Muniz, G., and Kibria, B. M. G. (2009). On Some Ridge Regression Estimators: An Empirical Comparisons. *Communications in Statistics - Simulation and Computation, 38*(3), 621-630. doi:10.1080/03610910802592838

Myers, R. H. (1990). Classical and modern regression with applications, Raymond H. Myers. *Duxbury Classic Series.*

Najarian, S., Arashi, M., and Kibria, B. M. G. (2013). A Simulation Study on Some Restricted Ridge Regression Estimators. *Communications in Statistics - Simulation and Computation, 42*(4), 871-890. doi:10.1080/03610918.2012.659953

Newhouse, J. P., and Oman, S. D. (1971). An evaluation of ridge estimators.

Özkale, M. R. (2012). Combining the unrestricted estimators into a single estimator and a simulation study on the unrestricted estimators. *Journal of Statistical Computation and Simulation, 82*(5), 653-688.

Özkale, M. R. (2014). The relative efficiency of the restricted estimators in linear regression models. *41*(5), 998-1027. doi:10.1080/02664763.2013.859234

Özkale, M. R., and Kaciranlar, S. (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics—Theory and Methods, 36*(15), 2707-2725.

Radhakrishna Rao, C., Toutenburg, H., and Heumann, C. (2008). Linear models and generalizations: least squares and alternatives. In: Berlin, Germany: Springer.

Rao, C. R., and Toutenburg, H. (1995). Linear models. In *Linear models* (pp. 3-18): Springer.

Rawlings, J. O., Pantula, S. G., and Dickey, D. A. Applied regression analysis: a research tool. 1998. *Brooks/Cole Publishing Company, America*, 1-7.

Rawlings, J. O., Pantula, S. G., and Dickey, D. A. (2001). *Applied regression analysis: a research tool*: Springer Science & Business Media.

Rencher, A. C., and Schaalje, G. B. (2008). *Linear models in statistics*: John Wiley & Sons.

Sakallıoğlu, S., and Kaçıranlar, S. (2008). A new biased estimator based on ridge estimation. *Statistical Papers, 49*(4), 669-689.

Sakallioğlu, S., Kaçiranlar, S., and Akdeniz, F. (2001). Mean squared error comparisons of some biased regression estimators. *Communications in statistics-theory and methods, 30*(2), 347-361.

Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares methods of estimation. *Communications in statistics-theory and methods, 21*(7), 1987-2000.

Singh, S., and Tracy, D. S. (1999). Ridge-regression using scrambled responses. *Metrika, 41*(2), 147-157.

Şiray, G. Ü. (2014). Modified and Restricted r-k Class Estimators. *43*(24), 5130-5155. doi:10.1080/03610926.2012.744050

Stein, C. (1956). *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*. Retrieved from

Swindel, B. F. (1976). Good ridge estimators based on prior information. *Communications in Statistics - Theory and Methods, 5*(11), 1065-1075. doi:10.1080/03610927608827423

Toro-Vizcarrondo, C., and Wallace, T. D. (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association, 63*(322), 558-572.

Toutenburg, H., and Trenkler, G. (1990). Mean square error matrix comparisons of optimal and classical predictors and estimators in linear regression. *Computational Statistics and Data Analysis, 10*(3), 297-305.

Trenkler, G. (1983). A note on superiority comparisons of homogeneous linear estimators. *Communications in statistics-theory and methods, 12*(7), 799-808.

Vinod, H. D., and Ullah, A. (1981). *Recent advances in regression methods* (Vol. 41): Marcel Dekker Incorporated.

Wencheko, E. (2000). Estimation of the signal-to-noise in the linear regression model. *Statistical Papers, 41*(3), 327.

Woods, H., Steinour, H. H., and Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. *Industrial & Engineering Chemistry, 24*(11), 1207-1214.

Yang, H., and Chang, X. (2010). A new two-parameter estimator in linear regression. *Communications in Statistics—Theory and Methods, 39*(6), 923-934.

Yang, H., and Cui, J. (2011). A stochastic restricted two-parameter estimator in linear regression model. *Communications in statistics-theory and methods, 40*(13), 2318-2325.

Zhong, Z., and Yang, H. (2007). Ridge estimation to the restricted linear model. *Communications in Statistics—Theory and Methods, 36*(11), 2099-2115.

CURRICULUM VITAE

My name is Israa Shaltoot. I graduated from Shuhada Al-Maghazi High School in 2012. Then, I graduated from the Department of Mathematics, Faculty of Science at Al-Aqsa University in 2016. After that, I have got a scholarship to do my master studies in Applied Statistics at Eskişehir Technical University, Eskisehir.