A Class of Generalized Shrunken Least Squares Estimators in Linear Model

by

Xiaoming Liu

A Thesis submitted to
the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

Abstract

Modern data analysis often involves a large number of variables, which gives rise to the problem of multicollinearity in regression models. It is well-known that in a linear model, when the design matrix X is nearly singular, then the ordinary least squares (OLS) estimator may perform poorly because of its numerical instability and large variance. To overcome this problem, many linear or nonlinear biased estimators are studied. In this work we consider a class of generalized shrunken least squares (GSLS) estimators that include many well-known linear biased estimators proposed in the literature. We compare these estimators under the mean square error and matrix mean square error criteria. Moreover, a simulation study and two numerical examples are used to illustrate some of the theoretical results.

Acknowledgments

First of all, I would like to express my deepest sense of gratitude to my supervisor Dr. Liqun Wang. I thank him for his patient guidance, excellent advice and encouragement during my research. Without his help, this work would not be possible.

I would also like to thank my committee members Dr. Saumen Mandal and Dr. Yang Zhang for their helpful suggestions and generous assistance.

To all my friends, thank you for your support and encouragement during these two years of research work. I cannot list all the names here, but you are always on my mind.

Special thanks to the Faculty of Graduate Studies for the Manitoba Graduate Scholarship (MGS) and my supervisor who supported me by his Natural Sciences and Engineering Research Council (NSERC) research grant.

I would like to express my sincerest gratitude to my parents: Zhian Liu and Yingxian Wang, for their love, endless support and encouragement throughout my life.

Last but not the least, I extend thanks and appreciation to everyone who helped me directly or indirectly in completing this thesis.

Dedication

This thesis is dedicated to my parents
for their love, endless support
and encouragement.

Contents

List of Tables

Chapter 1

Introduction

1.1 Motivation

1.1.1 Linear regression model

Regression analysis is one of the most popular statistical methodologies for investigating and modeling the relationship between variables. This methodology is often used for prediction of the response variable. Applications of regression analysis are widely used in economics, biological sciences, social sciences, and many other fields.

The topic of regression analysis has a long history. Galileo (1632) came very close to proposing a theory related to the least squares method. The earliest form of regression was the theory of least squares by Gauss (1809). Galton (1886) invented the use of the regression line and explained the common phenomenon of regression toward the mean in biological context. Yule (1897) and Pearson (1903) extended his work to a more general statistical framework. Fisher (1922) introduced the modern regression model, combining the regression theory and the theory of least squares.

Since then, regression analysis has become one of the most fundamental and the most widely used techniques in applied statistics.

The most common form of regression analysis is linear regression. Consider the standard linear regression model

$$Y = X\beta + \varepsilon, \tag{1.1.1}$$

where Y is a $n \times 1$ vector of observations of the response variable, X is a $n \times p$ design matrix of the levels of the regressor variables, β is a $p \times 1$ vector of unknown regression coefficients, and ε is a $n \times 1$ vector of random errors.

Throughout this work, we make the following assumptions:

ASSUMPTION 1 *X is a matrix of fixed constants that has full column rank $2 \leq p \leq n$.*

ASSUMPTION 2 *The random error ε has zero mean and variance-covariance matrix $\sigma^2 I_n$*

Under these assumptions, the ordinary least squares (OLS) estimator $\hat{\beta}_{OLS}$ of β is a function of Y which minimizes the sum of squared errors (SSE)

$$S(\beta) = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta). \tag{1.1.2}$$

By taking the derivative to $S(\beta)$ with respect to β and equating it to zero, we have

$$\left.\frac{\partial S}{\partial \beta}\right|_{\hat{\beta}} = -2X'(Y - X\hat{\beta}) = 0, \tag{1.1.3}$$

and therefore

$$X'X\hat{\beta}_{OLS} = X'Y. \tag{1.1.4}$$

The above equation is called the least squares normal equation and it shows that the nature of $X'X$ plays a very important role in the behavior of $\hat{\beta}_{OLS}$. Now, if the X matrix has full column rank p, then $(X'X)^{-1}$ matrix exists, and thus, the OLS estimator of β is given by

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y. \tag{1.1.5}$$

In this case, by the Gauss-Markov theorem, the OLS estimator $\hat{\beta}_{OLS}$ is the best linear unbiased estimator (BLUE) for β. This value of β corresponds to a minimum of SSE because the second order partial derivative of SSE with respect to β is a positive definite matrix $(2X'X)$.

1.1.2 Multicollinearity

Modern data analysis often involves a large number of variables, which gives rise to the problem of multicollinearity in regression models. It describes the situation where the columns of the X matrix are nearly linearly dependent. This implies that the matrix $X'X$ is nearly singular. Consequently, the numerical computation of (1.1.5) will be unstable. Moreover, the variance of the OLS estimator $\hat{\beta}_{OLS}$, $\sigma^2(X'X)^{-1}$, will be very large. To be more precise, let the eigenvalues of $X'X$ be denoted by

$$\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p = \lambda_{min} > 0.$$

Then the MSE of $\hat{\beta}_{OLS}$ is

$$\begin{aligned} MSE(\hat{\beta}_{OLS}) &= E((\hat{\beta}_{OLS} - \beta)'(\hat{\beta}_{OLS} - \beta)) \\ &= \sigma^2 trace(X'X)^{-1} \\ &= \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}. \end{aligned} \tag{1.1.6}$$

Multicollinearity implies that some of the eigenvalues will be close to zero. Equation (1.1.6) implies that the squared distance from $\hat{\beta}_{OLS}$ to β tends to be very large. And also, the variance of $\hat{\beta}_{OLS}$ is large, implying that confidence intervals on β will be wide and $\hat{\beta}_{OLS}$ is quite unstable.

Multicollinearity can be measured the so-called condition number, which is defined as the square root of the largest eigenvalue (λ_{max}) over the smallest eigenvalue (λ_{min}) of $X'X$,

$$K = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}. \tag{1.1.7}$$

A large value of K indicates strong multicollinearity. Belsley et al. (1980) suggested that a condition number $K < 10$ indicates very weak dependence; a condition number $10 < K \leq 30$ indicates weak dependence that may be starting to affect the regression estimate; a condition number $30 < K \leq 100$ indicates moderate to strong dependencies; and a condition number $100 < K$ indicates a serious collinearity problem.

1.1.3 Examples

Example 1. Portland Cement

First, we use an example which is commonly used in literature to illustrate problem of multicollinearity. The data set on Portland Cement was first used by Woods, Steinour and Starke (1932), and since, the data set has been widely analysed by other researchers, e.g., Hald (1952), Hamaker (1962), Gorman and Toman (1966), Daniel and Wood (1980), Nomura (1988), Kaçiranlar et al. (1999), Liu (2003), Yang and Xu (2007) and Sakallioğlu and Kaçiranlar (2008).

This data set came from an experimental investigation of the heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced. The four compounds considered by Woods, Steinour and Starke (1932) are tricalium aluminate: $3CaO \cdot Al_2O_3$, tricalcium silicate: $3CaO \cdot SiO_2$, tetracalcium aluminaferrite: $4CaO \cdot Al_2O_3 \cdot Fe_2O_3$, and β-dicalcium silicate: $2CaO \cdot SiO_2$, which we will denote by X_1, X_2, X_3 and X_4, respectively. The dependent variable Y is the heat evolved in calories per gram of cement after 180 days of curing.

The data are given as follows:

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

Woods et al. (1932) fitted the data to a linear model without intercept. Later, other researchers, such as Hald (1952), Gorman and Toman (1966) and Daniel and Wood (1980), fit a linear model with intercept (inhomogeneous model) to the data. Under this model, there are $n = 13$ observations but $p = 5$ unknown regression coefficients. The eigenvalues of $X'X$ are $(44676.2059, 5965.4221, 809.9521, 105.4187, 0.0012)$. The condition number of $X'X$ is as high as 6056.3443, so $X'X$ may be considered as ill-conditioned. Since there exists very strong multicollinearity, the OLS estimator may not be reliable.

Example 2. Air pollution in U.S. cities

In a climatology study, Sokal and Rohlf (1981) collected data on air pollution in 41 American cities. These data are also analysed by Hand et al (1994) and Rabe-Hesketh et al (2004). The data consist of the annual mean concentration of sulphur

dioxide in micrograms per cubic meter and measures of six explanatory variables and averaged over three years $1969 - 1971$ for each city. The data set contains the following variables:

Dependent variable:

Y	Average concentration of sulfur dioxide in micrograms per cubic meter

Independent variables:

X_1	Average annual temperature in degrees Fahrenheit
X_2	Number of manufacturing enterprises employing 20 or more workers
X_3	Population size (1970 census) in thousands
X_4	Average annual wind speed in miles per hour
X_5	Average annual precipitation in inches
X_6	Average number of days with precipitation per year

There is a single dependent variable (SO_2). There are six independent variables, two of which concern human ecology (X_2, X_3) and four climatic averages for weather stations at these cities (X_1, X_4, X_5, X_6). The data are given in Table (1.1). Cities are in alphabetical order by state.

In linear model with intercept (inhomogeneous model), there are $n = 41$ observations and $p = 7$ unknown regression coefficients. We find eigenvalues of $X'X$ to be $(49691889.5554, 866994.7600, 279362.0648, 7668.3100, 2807.9339, 103.2730, 0.0956)$. The condition number of $X'X$ is 22798.5515, so $X'X$ may be considered as quite ill-conditioned.

Table 1.1: Air pollution in 41 U.S. cities

Cities	Y	X_1	X_2	X_3	X_4	X_5	X_6
Phoenix	10	70.3	213	582	6	7.05	36
Little Rock	13	61	91	132	8.2	48.52	100
San Francisco	12	56.7	453	716	8.7	20.66	67
Denver	17	51.9	454	515	9	12.95	86
Hartford	56	49.1	412	158	9	43.37	127
Wilmington	36	54	80	80	9	40.25	114
Washington	29	57.3	434	757	9.3	38.89	111
Jacksonville	14	68.4	136	529	8.8	54.47	116
Miami	10	75.5	207	335	9	59.8	128
Atlanta	24	61.5	368	497	9.1	48.34	115
Chicago	110	50.6	3344	3369	10.4	34.44	122
Indianapolis	28	52.3	361	746	9.7	38.74	121
Des Moines	17	49	104	201	11.2	30.85	103
Wichita	8	56.6	125	277	12.7	30.58	82
Louisville	30	55.6	291	593	8.3	43.11	123
New Orleans	9	68.3	204	361	8.4	56.77	113
Baltimore	47	55	625	905	9.6	41.31	111
Detroit	35	49.9	1064	1513	10.1	30.96	129
Minneapolis	29	43.5	699	744	10.6	25.94	137
Kansas	14	54.5	381	507	10	37	99
St. Louis	56	55.9	775	622	9.5	35.89	105
Omaha	14	51.5	181	347	10.9	30.18	98
Albuquerque	11	56.8	46	244	8.9	7.77	58
Albany	46	47.6	44	116	8.8	33.36	135
Buffalo	11	47.1	391	463	12.4	36.11	166
Cincinnati	23	54	462	453	7.1	39.04	132
Cleveland	65	49.7	1007	751	10.9	34.99	155
Columbia	26	51.5	266	540	8.6	37.01	134
Philadelphia	69	54.6	1692	1950	9.6	39.93	115
Pittsburgh	61	50.4	347	520	9.4	36.22	147
Providence	94	50	343	179	10.6	42.75	125
Memphis	10	61.6	337	624	9.2	49.1	105
Nashville	18	59.4	275	448	7.9	46	119
Dallas	9	66.2	641	844	10.9	35.94	78
Houston	10	68.9	721	1233	10.8	48.19	103
Salt Lake City	28	51	137	176	8.7	15.17	89
Norfolk	31	59.3	96	308	10.6	44.68	116
Richmond	26	57.8	197	299	7.6	42.59	115
Seattle	29	51.1	379	531	9.4	38.79	164
Charleston	31	55.2	35	71	6.5	40.75	148
Milwaukee	16	45.7	569	717	11.8	29.07	123

1.1.4 Linear biased estimators

To overcome problem of multicollinearity, many well-known linear biased estimators are used to reduce the multicollinearity in the literature. The two major types of linear biased estimators are the Stein estimator (1956, 1960) and the ordinary ridge regression (ORR) estimator proposed by Hoerl and Kennard (1970). Gui (1994) proposed a class of principal components estimator. Liu estimator (1993) combined the advantages of the Stein estimator (1956) and the ORR estimator. The properties of the Liu estimator were studied by Akdeniz and Kaçiranlar (1995, 2001), Arslan and Billor (2000), Kaçiranlar and Sakallioğlu (1999, 2001). Liu (2003) proposed a new Liu-type two-parameter estimator. Sakallioğlu and Kaçiranlar (2008) introduced the $k-d$ class estimator, by combining the OLS estimator, the ORR estimator and the Liu estimator.

The OLS estimator has minimum variance in the class of unbiased linear estimators, but there is no guarantee that this variance will be small. If we consider both biased and unbiased estimators, the natural measure of goodness is the mean square error. For any estimator $\hat{\beta}$ of β, its mean square error is defined as

$$\begin{aligned} MSE(\hat{\beta}) &= E[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)] \\ &= Var(\hat{\beta}) + [E(\hat{\beta})-\beta]'[E(\hat{\beta})-\beta], \end{aligned} \tag{1.1.8}$$

or symbolically

$$MSE = Var + (Bias)^2. \tag{1.1.9}$$

So there is a trade-off between variance and bias. Therefore it is possible to significantly reduce the MSE by introducing a small amount of bias but decrease

the variance by a large amount. Moreover, the confidence intervals on β will be much narrower by using such biased estimator. And also, the biased estimator $\hat{\beta}_{BIAS}$ will be more stable than the unbiased OLS estimator $\hat{\beta}_{OLS}$. The comparison between $\hat{\beta}_{OLS}$ and $\hat{\beta}_{BIAS}$ is illustrated in Table (1.2).

Table 1.2: OLS estimators and biased estimators of β

Estimator	Expected value	Variance
$\hat{\beta}_{OLS}$	$E(\hat{\beta}_{OLS}) = \beta$	large
$\hat{\beta}_{BIAS}$	$E(\hat{\beta}_{BIAS}) \neq \beta$	small

In the next section, we provide an outline of the thesis.

1.2 Thesis organization

In Chapter 2, we consider a class of generalized shrunken least squares (GSLS) estimators of Wang (1990) for dealing with multicollinearity. The GSLS estimators include many other linear biased estimators proposed in the literature. We can write these linear biased estimators in terms of the GSLS estimators.

The primary aims of Chapter 3 are to compare any two GSLS estimators under the mean square error (MSE) and matrix mean square error ($MMSE$) criteria. Based on the MSE criterion, we can find the locally optimal estimator among the GSLS estimators.

In Chapter 4, we discuss the admissibility of the special subclasses of the GSLS estimators. We prove that the proposed GSLS estimators have superior properties over Liu estimators, Liu-type ridge estimators and $k-d$ class estimators under the $MMSE$ and MSE criteria.

In Chapter 5, we evaluate the performance of thirteen biased estimators with a simulation study by considering different levels of multicollinearity and different levels of σ. The estimated mean squared error ($EMSE$) is computed for each of the biased estimators.

In Chapter 6, we use two real data examples which are commonly used in literature to compare our proposed estimator with ordinary ridge regression (ORR) estimators, Liu estimators, Liu-type ridge estimators and $k - d$ class estimators.

Chapter 2

Generalized Shrunken Least Squares Estimators

2.1 Introduction

Many procedures have been developed for obtaining biased estimators of regression coefficients. Wang (1990) introduces a class of linear estimators, which are called the generalized shrunken least squares (GSLS) estimators:

$$\hat{\beta}_{GS}(A) = PAP'\hat{\beta}_{OLS}, \tag{2.1.1}$$

where $A = \text{diag}(a_1, a_2, \cdots, a_p)$, $0 \leq a_i \leq 1$, $i = 1, 2, \cdots, p$. Here P is the orthogonal matrix such that

$$P'X'XP = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_p) = \Lambda$$

and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$.

Obviously, this class $\hat{\beta}_{GS}(A)$ includes OLS estimator if $A = I$, the identity matrix. This is the only linear unbiased estimator in the class of GSLS estimators.

We can write equation (2.1.1) in the following form using $X'X = P\Lambda P'$,

$$\begin{aligned} \hat{\beta}_{GS}(A) &= PAP'(X'X)^{-1}X'Y \\ &= PA\Lambda^{-1}P'X'Y \end{aligned} \tag{2.1.2}$$

In the next section, we can see that this class includes many other linear biased estimators.

2.2 Biased estimators

In this section, we introduce various well-known biased estimators. A summary of these estimators is given in Table 2.1.

2.2.1 Stein estimator

Stein estimator (1956):

$$\hat{\beta}_s = c\hat{\beta}_{OLS}, \tag{2.2.1}$$

where $0 < c < 1$.

We can write the Stein estimator in the form of GSLS estimator as

$$\hat{\beta}_s = PTP'\hat{\beta}_{OLS}, \tag{2.2.2}$$

where $T = cI$ and $0 < c < 1$.

The Stein estimator has a simple form of biased estimators. The Stein estimator shrinks the OLS estimator of β toward the origin by a factor c, $0 < c < 1$. But the shrinkage factor for each component of $\hat{\beta}_{OLS}$ is the same, so $\hat{\beta}_s$ may still be unstable.

2.2.2 Ridge regression estimator

The ordinary ridge regression (ORR) estimator of Hoerl and Kennard (1970):

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y, \tag{2.2.3}$$

where $k > 0$.

Let $(X'X + kI)^{-1} = PA\Lambda^{-1}P'$ using equation (2.1.2), then

$$\begin{aligned} R &= \Lambda P'(X'X + kI)^{-1}P \\ &= \Lambda(\Lambda + kI)^{-1}. \end{aligned} \tag{2.2.4}$$

We can write the ORR estimator in the form of GSLS estimator as

$$\hat{\beta}(k) = PRP'\hat{\beta}_{OLS}. \tag{2.2.5}$$

The ORR estimator is one of the most widely used biased estimators in practice. The procedure is based on adding a number k to the diagonal of $X'X$, so the choice of k will affects the performance of the ORR estimator. In practice, we often use a small k. If $X'X$ is very ill conditioned, a small k may not be large enough to reduce the condition number of $X'X + kI$ to a small number.

2.2.3 Principal components estimator

Gui (1994) introduces a class of principal components (PC) estimator:

$$\hat{\beta}_{PC}(W) = \sum_{i=1}^{p} w_i \tilde{\beta}^{(i)}, \tag{2.2.6}$$

where $W = \text{diag}(w_1, w_2, \cdots, w_p)$, $0 \leq w_i \leq \lambda_i^{-1}$, $\tilde{\beta}^{(i)} = p_i p_i' X'Y$, $i = 1, 2, \cdots, p$, $P = (p_1, p_2, \cdots, p_p)$ is the orthogonal matrix such that $X'X = P\Lambda P'$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$.

Since $\hat{\beta}_{PC}(W) = PWP'X'Y$, we can write the PC estimator in the form of GSLS estimator as

$$\hat{\beta}_{PC}(W) = PGP'\hat{\beta}_{OLS}, \tag{2.2.7}$$

where $G = \Lambda W$.

2.2.4 Liu estimator

To overcome the disadvantages of ORR estimator, Liu (1993) combine the advantages of Stein estimator with ORR estimator to combat multicollinearity. Liu estimator is defined by

$$\hat{\beta}(d) = (X'X + I)^{-1}(X'Y + d\hat{\beta}_{OLS}), \tag{2.2.8}$$

where $0 < d < 1$.

Liu estimator can be simplified to the following form,

$$\begin{aligned} \hat{\beta}(d) &= (X'X + I)^{-1}(I + d(X'X)^{-1})X'Y \\ &= P(\Lambda + I)^{-1}(I + d\Lambda^{-1})P'X'Y. \end{aligned} \tag{2.2.9}$$

Using equation (2.1.2), then

$$L = \Lambda(\Lambda + I)^{-1}(I + d\Lambda^{-1}). \tag{2.2.10}$$

We can write the Liu estimator in the form of GSLS estimator as

$$\hat{\beta}(d) = PLP'\hat{\beta}_{OLS}. \tag{2.2.11}$$

The Liu estimator $\hat{\beta}(d)$ is based on $\hat{\beta}_{OLS}$, which performs poorly and sometimes gives misleading information.

To avoid this problem, Liu (2003) introduces a new Liu-type estimator:

$$\hat{\beta}_{k,d} = (X'X + kI)^{-1}(X'Y - d\hat{\beta}^*), \tag{2.2.12}$$

where $k > 0, -\infty < d < \infty$ and $\hat{\beta}^*$ can be any estimator of β.

1. When $\hat{\beta}^* = \hat{\beta}_{OLS}$, we can write the Liu-type estimator in the form of GSLS estimator as

   $$\hat{\beta}_{k,d} = PL_{OLS}P'\hat{\beta}_{OLS}, \tag{2.2.13}$$

   where $L_{OLS} = \Lambda(\Lambda + kI)^{-1}(I - d\Lambda^{-1})$.

2. When $\hat{\beta}^* = \hat{\beta}(k)$, we can write the Liu-type estimator in the form of GSLS estimator as

   $$\hat{\beta}_{k,d} = PL_RP'\hat{\beta}_{OLS}, \tag{2.2.14}$$

   where $L_R = \Lambda(\Lambda + kI)^{-1}(I - d(\Lambda + kI)^{-1})$

In the Liu-type estimator $\hat{\beta}_{k,d}$, the first parameter k can be used to control the condition number of $X'X + kI$ to any desired level. The second parameter d can be used to improve the fit and statistical property. We will consider joint parameter restrictions of Liu estimator and Liu-type ridge estimator in Chapter 4.

2.2.5 $k-d$ class estimator

Sakallioğlu and Kaçiranlar (2008) introduce the $k-d$ class estimator which include OLS estimator, ORR estimator and Liu estimator

$$\hat{\beta}(k,d) = (X'X+I)^{-1}(X'Y + d\hat{\beta}(k)), \tag{2.2.15}$$

where $k > 0$, $-\infty < d < \infty$ and $\hat{\beta}(k) = (X'X+kI)^{-1}X'Y$.

The $k-d$ class estimator can be simplified to the following form,

$$\begin{aligned} \hat{\beta}(k,d) &= (X'X+I)^{-1}(I + d(X'X+kI)^{-1})X'Y \\ &= P(\Lambda+I)^{-1}(I + d(\Lambda+kI)^{-1})P'X'Y. \end{aligned} \tag{2.2.16}$$

Using equation (2.1.2), then

$$S = \Lambda(\Lambda+I)^{-1}(I + d(\Lambda+kI)^{-1}). \tag{2.2.17}$$

We can write the $k-d$ class estimator in the form of GSLS estimator as

$$\hat{\beta}(k,d) = PSP'\hat{\beta}_{OLS}. \tag{2.2.18}$$

If we want to use the $k-d$ class estimator, we can choose a large number k to reduce the condition number of $X'X+kI$ to a small number because we can adjust another parameter d. Sakallioğlu and Kaçiranlar prove that the $k-d$ class estimator has superior properties over the OLS estimators, ORR estimators and Liu estimators under $MMSE$ and MSE criteria. Since $k > 0$ and $-\infty < d < \infty$, we consider joint parameter restrictions of the $k-d$ class estimator in Chapter 4.

2.2.6 James-Stein estimator

For $Y \sim N(\theta, \sigma^2 I)$, the James-Stein (1961) estimator is defined as

$$\hat{Y}_{JS} = \left(1 - \frac{(p-2)\sigma^2}{Y'Y}\right) Y. \tag{2.2.19}$$

In a linear model,an extension of the James-Stein estimator is

$$\hat{\beta}_{JS} = \left(1 - \frac{(p-2)vs^2}{(v+2)\hat{\beta}'_{OLS}\hat{\beta}_{OLS}}\right) \hat{\beta}_{OLS}, \tag{2.2.20}$$

where s^2 is the estimate of variance with v degrees of freedom.

James-Stein estimator can be shown to dominate the OLS estimator under the MSE criterion when $p > 2$. That is to say, if we want to obtain an estimator outside linear unbiased estimators, we can find James-Stein estimator which is better than the OLS estimator under MSE criterion when $p > 2$.

The James-Stein estimator $\hat{\beta}_{JS}$ is non-linear biased estimator. It can be written in the form

$$\hat{\beta}_{JS} = PJP'\hat{\beta}_{OLS}, \text{ where} \tag{2.2.21}$$

$$J = a(Y)I_p = \left(1 - \frac{(p-2)vs^2}{(v+2)\hat{\beta}'_{OLS}\hat{\beta}_{OLS}}\right) I_p. \tag{2.2.22}$$

We should note, however, that $a(Y)$ is a function of Y. The main diagonal entries of $a(Y)I_p$ may not be in the interval [0,1]. Although, James-Stein estimator can be written in the form of GSLS estimator, it does not belong to the class of GSLS estimator.

Table 2.1: Summary of biased estimators

Biased estimators	Equations	GSLS estimators
Stein estimator (1956)	$\hat{\beta}_s = c\hat{\beta}_{OLS}$, where $0 < c < 1$.	$\hat{\beta}_s = PTP'\hat{\beta}_{OLS}$, where $T = cI$.
James-Stein estimator (1961)	$\hat{\beta}_{JS} = \left(1 - \frac{(p-2)vs^2}{(v+2)\hat{\beta}'_{OLS}\hat{\beta}_{OLS}}\right)\hat{\beta}_{OLS}$	$\hat{\beta}_{JS} = PJP'\hat{\beta}_{OLS}$, where $J = a(Y)I_p = \left(1 - \frac{(p-2)vs^2}{(v+2)\hat{\beta}'_{OLS}\hat{\beta}_{OLS}}\right)I_p$.
ORR estimator (1970)	$\hat{\beta}(k) = (X'X + kI)^{-1}X'Y$, where $k > 0$.	$\hat{\beta}(k) = PRP'\hat{\beta}_{OLS}$, where $R = \Lambda(\Lambda + kI)^{-1}$.
Liu estimator (1993)	$\hat{\beta}(d) = (X'X + I)^{-1}(X'Y + d\hat{\beta}_{OLS})$, where $0 < d < 1$.	$\hat{\beta}(d) = PLP'\hat{\beta}_{OLS}$, where $L = \Lambda(\Lambda + I)^{-1}(I + d\Lambda^{-1})$.
PC estimator (1994)	$\hat{\beta}_{PC}(W) = \sum_{i=1}^{p} w_i\tilde{\beta}^{(i)}$	$\hat{\beta}_{PC}(W) = PGP'\hat{\beta}_{OLS}$, where $G = \Lambda W$.
Liu-type OLS estimator (2003)	$\hat{\beta}_{k,d} = (X'X + kI)^{-1}(X'Y - d\hat{\beta}_{OLS})$, where $k > 0$, $-\infty < d < \infty$.	$\hat{\beta}_{k,d} = PL_{OLS}P'\hat{\beta}_{OLS}$, where $L_{OLS} = \Lambda(\Lambda + kI)^{-1}(I - d\Lambda^{-1})$.
Liu-type ridge estimator (2003)	$\hat{\beta}_{k,d} = (X'X + kI)^{-1}(X'Y - d\hat{\beta}(k))$, where $k > 0$, $-\infty < d < \infty$.	$\hat{\beta}_{k,d} = PL_RP'\hat{\beta}_{OLS}$, where $L_R = \Lambda(\Lambda + kI)^{-1}(I - d(\Lambda + kI)^{-1})$.
$k - d$ class estimator (2008)	$\hat{\beta}(k,d) = (X'X + I)^{-1}(X'Y + d\hat{\beta}(k))$, where $k > 0$, $-\infty < d < \infty$.	$\hat{\beta}(k,d) = PSP'\hat{\beta}_{OLS}$, where $S = \Lambda(\Lambda + I)^{-1}(I + d(\Lambda + kI)^{-1})$.

Chapter 3

Properties of the GSLS Estimators

3.1 Comparisons under matrix mean square error criterion

If $\hat{\beta}$ is an estimator of β, then the matrix mean square error ($MMSE$) of $\hat{\beta}$ is defined as

$$\begin{aligned} MMSE(\hat{\beta}) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] \\ &= Cov(\hat{\beta}) + Bias(\hat{\beta})(Bias(\hat{\beta}))'. \end{aligned}$$

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

$$\begin{aligned} Cov(\hat{\beta}_{GS}(A)) &= Cov(PAP'\hat{\beta}_{OLS}) \\ &= PAP'Cov(\hat{\beta}_{OLS})PAP' \\ &= \sigma^2 PAP'(X'X)^{-1}PAP' \\ &= \sigma^2 PA^2\Lambda^{-1}P', \end{aligned} \tag{3.1.1}$$

$$\begin{aligned} Bias(\hat{\beta}_{GS}(A)) &= E(\hat{\beta}_{GS}(A)) - \beta \\ &= PAP'(X'X)^{-1}X'E(Y) - \beta \\ &= PAP'\beta - \beta \\ &= P(A-I)P'\beta. \end{aligned} \tag{3.1.2}$$

Combining (3.1.1) and (3.1.2), we can find the $MMSE$ of $\hat{\beta}_{GS}(A)$

$$\begin{aligned} MMSE(\hat{\beta}_{GS}(A)) &= Cov(\hat{\beta}_{GS}(A)) + Bias(\hat{\beta}_{GS}(A))(Bias(\hat{\beta}_{GS}(A)))' \\ &= \sigma^2 PA\Lambda^{-1}AP' + P(I-A)P'\beta\beta'P(I-A)P'. \end{aligned} \tag{3.1.3}$$

Denote

$$\alpha = \frac{1}{\sigma}P'\beta$$

and

$$\begin{aligned} M(A) &= \frac{1}{\sigma^2}P'[MMSE(\hat{\beta}_{GS}(A))]P \\ &= A\Lambda^{-1}A + (I-A)\alpha\alpha'(I-A). \end{aligned}$$

If M and N are two matrices, then we will write $M > N$ or $N < M$ if $M - N$ is positive definite; and write $M \geq N$ or $N \leq M$ if $M - N$ is non-negative definite.

Let $\hat{\beta}_{GS}(A)$ and $\hat{\beta}_{GS}(B)$ be two GSLS estimators. Then $\hat{\beta}_{GS}(A)$ is said to be better than $\hat{\beta}_{GS}(B)$ under matrix mean square error, if

$$MMSE(\hat{\beta}_{GS}(A)) \leq MMSE(\hat{\beta}_{GS}(B))$$

for all (β, σ^2); or equivalently, if

$$M(\hat{\beta}_{GS}(A)) \leq M(\hat{\beta}_{GS}(B))$$

for all $\alpha = P'\beta/\sigma$.

For any two GSLS estimators, it is easy to see that

$$\begin{aligned} & M(\hat{\beta}_{GS}(B)) - M(\hat{\beta}_{GS}(A)) \\ = \; & B\Lambda^{-1}B + (I-B)\alpha\alpha'(I-B) - A\Lambda^{-1}A - (I-A)\alpha\alpha'(I-A) \\ = \; & (B^2 - A^2)\Lambda^{-1} + (I-B)\alpha\alpha'(I-B) - (I-A)\alpha\alpha'(I-A). \end{aligned} \tag{3.1.4}$$

Suppose $A < B$, then $(B^2 - A^2)\Lambda^{-1}$ is positive definite (p.d.) matrix. It is easy to see that

$$(B^2 - A^2)\Lambda^{-1} > 0 \quad \text{implies} \quad (B^2 - A^2)\Lambda^{-1} + (I-B)\alpha\alpha'(I-B) > 0.$$

Hence the problem to decide whether $MMSE(\hat{\beta}_{GS}(A)) \leq MMSE(\hat{\beta}_{GS}(B))$ reduces to that of deciding whether a matrix type $M - cc'$ is p.d. or nonnegative definite (n.n.d.) matrix when M is p.d. matrix. We need the following lemma.

Lemma 3.1.1 *Let M be a p.d. matrix and c be an nonzero vector. Then $M - cc'$ is p.d. (n.n.d.) matrix if and only if $c'M^{-1}c < 1$ ($c'M^{-1}c \leq 1$).*

Proof. see Farebrother (1976).

Now take $M = (B^2 - A^2)\Lambda^{-1} + (I-B)\alpha\alpha'(I-B)$ and $c = (I-B)\alpha$. From this we have the following result:

Theorem 3.1.1 *Suppose $A < B$. $MMSE(\hat{\beta}_{GS}(B)) - MMSE(\hat{\beta}_{GS}(A))$ is n.n.d. matrix if and only if*

$$\alpha'(I-A)\left[(B^2 - A^2)\Lambda^{-1} + (I-B)\alpha\alpha'(I-B)\right]^{-1}(I-A)\alpha \leq 1. \tag{3.1.5}$$

Comparisons between any two GSLS estimators and some special cases are studied under the $MMSE$ criterion by Wang (1990). The main results are as follows:

Lemma 3.1.2 *Let $\hat{\beta}_{GS}(A)$ and $\hat{\beta}_{GS}(B)$ be two GSLS estimators.*

(1) If $MSEM(\hat{\beta}_{GS}(A)) \leq MMSE(\hat{\beta}_{GS}(B))$ for all α, then $A = B$.

(2) If $A < B$, then $MMSE(\hat{\beta}_{GS}(A)) \leq MMSE(\hat{\beta}_{GS}(B))$ if and only if

$$(u_1 - 1)(1 + u_2) \leq v_1^2, \tag{3.1.6}$$

where

$$\begin{aligned} u_1 &= \alpha'\Lambda(B^2 - A^2)^+(I - A)^2\alpha, \\ u_2 &= \alpha'\Lambda(B^2 - A^2)^+(I - B)^2\alpha, \\ v_1 &= \alpha'\Lambda(B^2 - A^2)^+(I - A)(I - B)\alpha. \end{aligned}$$

(3) If $A \leq B$ and $A \neq B$, then $MMSE(\hat{\beta}_{GS}(A)) \leq MMSE(\hat{\beta}_{GS}(B))$ if and only if (3.1.6) and

$$(B - A)\alpha\alpha'(I - B)[I - (B - A)(B - A)^+] = 0 \tag{3.1.7}$$

hold.

(4) If there are two or more i's such that $a_i > b_i$, then $MMSE(\hat{\beta}_{GS}(A)) \leq MMSE(\hat{\beta}_{GS}(B))$ can never hold.

(Here M^+ denotes the Moore-Penrose inverse of the matrix M)

Proof. see Wang (1990).

3.2 Comparisons under mean square error criterion

Another measure of goodness of an estimator is mean square error (MSE) criterion. The main goal of this section is to discuss locally optimal estimator among the GSLS estimators, and then to compare any two GSLS estimators under the MSE criterion.

If $\hat{\beta}$ is an estimator of β, then the MSE of $\hat{\beta}$ is defined as

$$\begin{aligned} MSE(\hat{\beta}) &= E[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)] \\ &= trace(Cov(\hat{\beta})) + (Bias(\hat{\beta}))'Bias(\hat{\beta}). \end{aligned}$$

If $\hat{\beta}_{GS}(A)$ is a GSLS estimator, we can find the MSE of $\hat{\beta}_{GS}(A)$ by combining (3.1.1) and (3.1.2):

$$\begin{aligned} MSE(\hat{\beta}_{GS}(A)) &= trace(Cov(\hat{\beta}_{GS}(A))) + (Bias(\hat{\beta}_{GS}(A)))'Bias(\hat{\beta}_{GS}(A)) \\ &= \sigma^2 trace(A^2\Lambda^{-1}) + \beta' P(I-A)^2 P'\beta \\ &= \sigma^2 \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 (1-a_i)^2, \\ &= \sum_{i=1}^{p} \left[\frac{\sigma^2 a_i^2}{\lambda_i} + \delta_i^2 (1-a_i)^2 \right] \\ &= \sum_{i=1}^{p} D_i(a_i). \end{aligned} \tag{3.2.1}$$

where δ_i is the ith entries of vector $P'\beta$, $i = 1, 2, \cdots, p$.

3.2.1 Locally optimal estimator

To derive the locally optimal GSLS estimator at any (β, σ^2), we minimize $D_i(a_i)$ with respect to a_i. Since

$$\left.\frac{\partial D_i(a_i)}{\partial a_i}\right|_{a_i^{\text{opt}}} = \frac{2\sigma^2 a_i}{\lambda_i} - 2\delta_i^2(1-a_i) = 0, \tag{3.2.2}$$

we find that

$$a_i^{\text{opt}} = \frac{\lambda_i \delta_i^2}{\lambda_i \delta_i^2 + \sigma^2}. \tag{3.2.3}$$

Since the second order partial derivative to $D_i(a_i)$ with respect to a_i,

$$\frac{\partial^2 D_i(a_i)}{\partial^2 a_i} = \frac{\sigma^2}{\lambda_i} + \delta_i^2 > 0, \tag{3.2.4}$$

a_i^{opt} corresponds to the minimum of $D_i(a_i)$. However, a_i^{opt} depends on the unknown parameters β and σ^2. In practice, initial estimates $\hat{\beta}$, $\hat{\sigma}^2$ are needed. For example, the OLS estimates $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$ can be used as

$$\hat{a}_i^{\text{opt}} = \frac{\lambda_i \hat{\delta}_i^2}{\lambda_i \hat{\delta}_i^2 + \hat{\sigma}^2}, \tag{3.2.5}$$

where $\hat{\delta}_i$ is the ith entries of column matrix $P'\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS} = \frac{(Y - X\hat{\beta}_{OLS})'(Y - X\hat{\beta}_{OLS})}{n-p}$.

In some cases, $a_i \neq a_i^{\text{opt}}$ for some i, $i = 1, 2, \cdots, p$, then we have the following result.

Theorem 3.2.1 *For any GSLS estimator* $\hat{\beta}_{GS}(A) = PAP'\hat{\beta}_{OLS}$, $A = diag(a_1, a_2, \cdots, a_p)$, *if for some* i, $a_i \neq a_i^{opt}$, *then there exists an estimator* $PBP'\hat{\beta}_{OLS}$ *with* $0 \leq b_i \leq 1$, *such that* $MSE(PBP'\hat{\beta}_{OLS}) < MSE(\hat{\beta}_{GS}(A))$. *Further,* $PBP'\hat{\beta}_{OLS}$ *can be constructed as follows:*

1. *if some* $a_i \neq a_i^{opt}$, *then set* $b_i = a_i^{opt}$;
2. *otherwise set* $b_j = a_j$.

Proof. By using (3.2.1), we know that

$$\begin{aligned} E[(\hat{\beta}_{GS}(A) - \beta)'(\hat{\beta}_{GS}(A) - \beta)] &= \sigma^2 trace(A^2\Lambda^{-1}) + \beta' P(I - A)^2 P'\beta \\ &= \sigma^2(\frac{a_1^2}{\lambda_1} + \cdots + \frac{a_i^2}{\lambda_i} + \cdots + \frac{a_p^2}{\lambda_p}) \\ &\quad +\delta_1^2(1 - a_1)^2 + \cdots + \delta_i^2(1 - a_i)^2 \\ &\quad + \cdots + \delta_p^2(1 - a_p)^2, \end{aligned} \tag{3.2.6}$$

where δ_i is the ith entries of column matrix $P'\beta$.

For any A such that some $a_i \neq a_i^{\text{opt}}$, we can find that replacing a_i by choosing $b_i = a_i^{\text{opt}}$

$$\begin{aligned} E[(\hat{\beta}_{GS}(B) - \beta)'(\hat{\beta}_{GS}(B) - \beta)] &= \sigma^2(\frac{a_1^2}{\lambda_1} + \cdots + \frac{(a_i^{\text{opt}})^2}{\lambda_i} + \cdots + \frac{a_p^2}{\lambda_p}) \\ &\quad +\delta_1^2(1 - a_1)^2 + \cdots + \delta_i^2(1 - a_i^{\text{opt}})^2 \\ &\quad + \cdots + \delta_p^2(1 - a_p)^2, \end{aligned} \tag{3.2.7}$$

It is easy to see that

$$\sigma^2\frac{(a_i^{\text{opt}})^2}{\lambda_i}+\delta_i^2(1-a_i^{\text{opt}})^2<\sigma^2\frac{a_i^2}{\lambda_i}+\delta_i^2(1-a_i)^2.$$

Expression (3.2.7) leads to a smaller value than expression (3.2.6),

$$E[(\hat{\beta}_{GS}(B)-\beta)'(\hat{\beta}_{GS}(B)-\beta)]<E[(\hat{\beta}_{GS}(A)-\beta)'(\hat{\beta}_{GS}(A)-\beta)].$$

The conclusion follows.

Another way to obtain the optimal GSLS estimator is the unbiased estimator of MSE approach by Zhao (1995).

First, the unbiased estimator of $MSE(\hat{\beta}_{GS}(A))$ is

$$\begin{aligned}
\widehat{MSE}(\hat{\beta}_{GS}(A)) &= \hat{\sigma}^2 trace((2A-I)\Lambda^{-1})+\hat{\beta}'_{OLS}P(I-A)^2P'\hat{\beta}_{OLS}\\
&= \hat{\sigma}^2\sum_{i=1}^{p}\frac{2a_i-1}{\lambda_i}+\sum_{i=1}^{p}\hat{\delta}_i^2(1-a_i)^2\\
&= \sum_{i=1}^{p}\left[\frac{\hat{\sigma}^2(2a_i-1)}{\lambda_i}+\hat{\delta}_i^2(1-a_i)^2\right] \qquad (3.2.8)\\
&= \sum_{i=1}^{p}D_i^*(a_i). \qquad (3.2.9)
\end{aligned}$$

Then we can show that

$$
\begin{aligned}
& E(\widehat{MSE}(\hat{\beta}_{GS}(A))) \\
= \; & \sigma^2 trace((2A-I)\Lambda^{-1}) + trace(P(I-A)^2P'\sigma^2(X'X)^{-1}) + \beta' P(I-A)^2P'\beta \\
= \; & \sigma^2 trace((2A-I)\Lambda^{-1}) + \sigma^2 trace(P(I-A)^2\Lambda^{-1}P') + \beta' P(I-A)^2P'\beta \\
= \; & \sigma^2 trace(((2A-I)+(I-A)^2)\Lambda^{-1}) + \beta' P(I-A)^2P'\beta \\
= \; & \sigma^2 trace(A^2\Lambda^{-1}) + \beta' P(I-A)^2P'\beta \\
= \; & MSE(\hat{\beta}_{GS}(A)). \qquad (3.2.10)
\end{aligned}
$$

To minimize the unbiased estimator (3.2.8) with respect to a_i, we calculate

$$
\left.\frac{\partial D_i^*(a_i)}{\partial a_i}\right|_{\hat{a}_i^{\text{ue}}} = \frac{2\hat{\sigma}^2}{\lambda_i} - 2\hat{\delta}_i^2(1-a_i) = 0. \qquad (3.2.11)
$$

Then we can find that

$$
\hat{a}_i^{\text{ue}} = \frac{\lambda_i\hat{\delta}_i^2 - \hat{\sigma}^2}{\lambda_i\hat{\delta}_i^2} = 2 - \frac{1}{\hat{a}_i^{\text{opt}}}, \qquad (3.2.12)
$$

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

When $\frac{1}{2} \le \hat{a}_i^{\text{opt}} \le 1$, it is easy to see that the $\hat{a}_i^{\text{ue}}$ are in the interval $[0,1]$.

Since the second order partial derivative of $D_i(a_i)$ with respect to a_i,

$$
\frac{\partial^2 D_i^*(a_i)}{\partial^2 a_i} = 2\hat{\delta}_i^2 > 0, \qquad (3.2.13)
$$

we know that $\hat{a}_i^{\text{ue}}$ corresponds to a minimum of $D_i^*(a_i)$ when $\frac{1}{2} \le \hat{a}_i^{\text{opt}} \le 1$.

When $0 \leq \hat{a}_i^{\text{opt}} \leq \frac{1}{2}$, $\hat{a}_i^{\text{ue}}$ are not in the interval $[0, 1]$. In this case, we may define $D_i^*(a_i)$ to be zero, i.e.,

$$\frac{\hat{\sigma}^2(2a_i - 1)}{\lambda_i} + \hat{\delta}_i^2(1 - a_i)^2 = 0. \tag{3.2.14}$$

Then we find

$$\hat{a}_i^{\text{ue}} = 1 + \frac{\sqrt{1 - \frac{\lambda_i \hat{\delta}_i^2}{\hat{\sigma}^2}} - 1}{\frac{\lambda_i \hat{\delta}_i^2}{\hat{\sigma}^2}}. \tag{3.2.15}$$

Further, we can get $\hat{a}_i^{\text{ue}}$ in terms of $\hat{a}_i^{\text{opt}}$,

$$\hat{a}_i^{\text{ue}} = \frac{\sqrt{\frac{1-2\hat{a}_i^{\text{opt}}}{1-\hat{a}_i^{\text{opt}}}}}{1 + \sqrt{\frac{1-2\hat{a}_i^{\text{opt}}}{1-\hat{a}_i^{\text{opt}}}}}. \tag{3.2.16}$$

Now we have two methods to determining locally optimal estimator of the GSLS estimators,

1. Under the MSE criterion, we can determine locally optimal estimator as

$$\hat{a}_i^{\text{opt}} = \frac{\lambda_i \hat{\delta}_i^2}{\lambda_i \hat{\delta}_i^2 + \hat{\sigma}^2}, \tag{3.2.17}$$

where $\hat{\delta}_i$ is the ith entries of column matrix $P'\hat{\beta}$ and $\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n-p}$, $i = 1, 2, \cdots, p$.

2. Under the unbiased-estimated MSE criterion, we can determine locally optimal estimator as

   a. if $0 \leq \hat{a}_i^{\text{opt}} \leq \frac{1}{2}$, then

$$\hat{a}_i^{\text{ue}} = \frac{\sqrt{\frac{1-2\hat{a}_i^{\text{opt}}}{1-\hat{a}_i^{\text{opt}}}}}{1+\sqrt{\frac{1-2\hat{a}_i^{\text{opt}}}{1-\hat{a}_i^{\text{opt}}}}}; \tag{3.2.18}$$

   b. if $\frac{1}{2} \leq \hat{a}_i^{\text{opt}} \leq 1$, then

$$\hat{a}_i^{\text{ue}} = 2 - \frac{1}{\hat{a}_i^{\text{opt}}}, \tag{3.2.19}$$

   where $\hat{a}_i^{\text{opt}} = \frac{\lambda_i \hat{\delta}_i^2}{\lambda_i \hat{\delta}_i^2 + \hat{\sigma}^2}$, $i = 1, 2, \cdots, p$.

Both of these two locally optimal estimators depend on the unknown parameters β and σ^2. In practice, we suggest to use the ridge estimators $\hat{\beta}_R$ and $\hat{\sigma}_R^2$ as initial estimates. In Chapter 5, we use Monte Carlo simulation study to compare the performance of the GSLS estimators and other biased estimators.

3.2.2 Comparisons between any two GSLS estimators

In this subsection, we compare any two GSLS estimators under the MSE criterion. For any two GSLS estimators $\hat{\beta}_{GS}(A)$ and $\hat{\beta}_{GS}(B)$, the MSE are respectively

$$\begin{aligned} MSE(\hat{\beta}_{GS}(A)) &= \sigma^2 trace(A^2\Lambda^{-1}) + \beta' P(I-A)^2 P'\beta \\ &= \sigma^2 \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 (1-a_i)^2, \end{aligned} \tag{3.2.20}$$

$$
\begin{aligned}
MSE(\hat{\beta}_{GS}(B)) &= \sigma^2 trace(B^2\Lambda^{-1}) + \beta' P(I-B)^2 P'\beta \\
&= \sigma^2 \sum_{i=1}^{p} \frac{b_i^2}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 (1-b_i)^2. \qquad (3.2.21)
\end{aligned}
$$

Using (3.2.20) and (3.2.21), we find that

$$
\begin{aligned}
& MSE(\hat{\beta}_{GS}(B)) - MSE(\hat{\beta}_{GS}(A)) \\
= \; & \sigma^2 \sum_{i=1}^{p} \frac{b_i^2 - a_i^2}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 \left[(1-b_i)^2 - (1-a_i)^2\right] \\
= \; & \sum_{i=1}^{p} \left[\frac{\sigma^2}{\lambda_i}(b_i+a_i)(b_i-a_i) + \delta_i^2 (a_i-b_i)(2-a_i-b_i)\right] \\
= \; & \sum_{i=1}^{p} \left[(b_i-a_i)\left(\frac{\sigma^2}{\lambda_i}(b_i+a_i) - \delta_i^2(2-a_i-b_i)\right)\right] \\
= \; & \sum_{i=1}^{p} \left[\left(\frac{\lambda_i\delta_i^2+\sigma^2}{\lambda_i}\right)(b_i-a_i)(a_i+b_i-2a_i^{\text{opt}})\right]. \qquad (3.2.22)
\end{aligned}
$$

Thus we have the following result.

Theorem 3.2.2 *For any two GSLS estimators $\hat{\beta}_{GS}(A)$ and $\hat{\beta}_{GS}(B)$, $MSE(\hat{\beta}_{GS}(A)) \leq MSE(\hat{\beta}_{GS}(B))$ if either*

a.

$$2a_i^{\text{opt}} - b_i \leq a_i \leq b_i,$$

$i = 1, 2, \cdots, p$; or

b.

$$b_i \leq a_i \leq 2a_i^{\text{opt}} - b_i,$$

$i = 1, 2, \cdots, p$.

By taking special forms of A and $B = I$, the above theorem gives the comparison result between any GSLS estimator and the OLS estimator.

Corollary 3.2.1 *Let $\hat{\beta}_{GS}(A)$ be a GSLS estimator and $A \neq I$. If $\frac{\lambda_i \delta_i^2 - \sigma^2}{\lambda_i \delta_i^2 + \sigma^2} \leq a_i \leq 1$, $i = 1, 2, \cdots, p$, then $MSE(\hat{\beta}_{GS}(A)) \leq MSE(\hat{\beta}_{OLS})$.*

Chapter 4

Some Special GSLS Estimators

In this chapter, we study the properties of Liu estimators, Liu-type ridge estimators and the $k-d$ class estimators under the $MMSE$ and the MSE criteria, and compare them within the class of all GSLS estimators.

4.1 Optimality of Liu estimators

4.1.1 Admissibility of Liu estimators

First of all, we discuss the admissibility of Liu estimators $\hat{\beta}(d)$, and then compare $\hat{\beta}(d)$ and any $\hat{\beta}_{GS}(A)$ under the $MMSE$ and the MSE criteria.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be two estimators of β. Then $\hat{\beta}_1$ is said to be better than $\hat{\beta}_2$ under MSE, if for all β and σ^2,

$$E[(\hat{\beta}_1 - \beta)'(\hat{\beta}_1 - \beta)] \leq E[(\hat{\beta}_2 - \beta)'(\hat{\beta}_2 - \beta)] \tag{4.1.1}$$

and the inequality holds for at least one pair (β, σ^2).

An estimator $\hat{\beta}$ is said to be admissible if there does not exist another estimator which is better than $\hat{\beta}$. By the result of Rao (1976), we know that every GSLS estimator $\hat{\beta}_{GS}(A)$ is admissible.

The Liu estimator (1993) can be written in the form of GSLS estimator as

$$\hat{\beta}(d) = PLP'\hat{\beta}_{OLS}, \tag{4.1.2}$$

where $L = \Lambda(\Lambda + I)^{-1}(I + d\Lambda^{-1})$ and $0 < d < 1$.

To consider the parameter restriction of the Liu estimator, we first prove a lemma.

Lemma 4.1.1 *Let l_i be the main diagonal entries of L. Then the range of l_i is $\frac{\lambda_i}{\lambda_i+1} < l_i < 1$, $i = 1, 2, \cdots, p$.*

PROOF. It is not difficult to calculate the main diagonal entries l_i of the diagonal matrix L,

$$\begin{aligned} L &= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_p+1} \end{bmatrix} \begin{bmatrix} 1+\frac{d}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1+\frac{d}{\lambda_p} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1+d}{\lambda_1+1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\lambda_p+d}{\lambda_p+1} \end{bmatrix}. \end{aligned} \tag{4.1.3}$$

Since

$$0 < d < 1,$$

we have

$$\lambda_i < \lambda_i + d < 1 + \lambda_i,$$

and hence

$$\frac{\lambda_i}{\lambda_i + 1} < \frac{\lambda_i + d}{\lambda_i + 1} < 1.$$

From Theorem 3.2.1, we know that the main diagonal entries l_i should be in the interval $[0, 1]$. Then we can find that

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

and hence

$$-\lambda_i \leq d \leq 1, \ i = 1, 2, \cdots, p. \tag{4.1.4}$$

To consider the admissibility of Liu estimators $\hat{\beta}(d)$ in the class of all estimators of β, we need the following lemma.

Lemma 4.1.2 *A GSLS estimator $\hat{\beta}_{GS}(A)$ is admissible if and only if* $\text{rank}\,(I - A) \geq p - 2$.

Proof. see Wang (1990).

According to Lemma (4.1.2), we know that the necessary and sufficient conditions of admissibility of GSLS estimators are that at most two of a_i being 1. If $p = 2$, then $-\lambda_i \leq d \leq 1$ satisfy this condition. If $p > 2$ and $d = 1$, then l_i are all equal to 1, Liu estimators is not admissible. The interval for d should change to $-\lambda_i \leq d < 1$.

Now we have the main result of this section.

Theorem 4.1.1 *A Liu estimator $\hat{\beta}(d)$ is admissible if and only if*

1. $-\lambda_i \leq d \leq 1$ *when* $p = 2$, $i = 1, 2, \cdots, p$.

2. $-\lambda_i \leq d < 1$ *when* $p > 2$, $i = 1, 2, \cdots, p$.

4.1.2 Comparisons under $MMSE$ and MSE

By using (3.1.1) and (3.1.2), the covariance and bias of Liu estimators are given by, respectively

$$Cov(\hat{\beta}(d)) = \sigma^2 PL\Lambda^{-1}LP', \tag{4.1.5}$$

$$Bias(\hat{\beta}(d)) = P(L - I)P'\beta. \tag{4.1.6}$$

Combining (4.1.5) and (4.1.6), we find that

$$MMSE(\hat{\beta}(d)) = \sigma^2 PL\Lambda^{-1}LP' + P(I - L)P'\beta\beta'P(I - L)P', \tag{4.1.7}$$

$$MSE(\hat{\beta}(d)) = \sigma^2 \sum_{i=1}^{p} \frac{l_i^2}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 (1 - l_i)^2, \tag{4.1.8}$$

where δ_i is the ith entries of column matrix $P'\beta$ and $l_i = \frac{\lambda_i + d}{\lambda_i + 1}, i = 1, 2, \cdots, p$.

Using Theorem (3.1.1), we can show the optimality of Liu estimator within the GSLS estimator under $MMSE$ criterion.

Corollary 4.1.1 *Let $\hat{\beta}_{GS}(A)$ be a GSLS estimator and $\hat{\beta}_{GS}(L)$ be a Liu estimator. Suppose $A < L$. $MMSE(\hat{\beta}_{GS}(L)) \geq MMSE(\hat{\beta}_{GS}(A))$ if and only if*

$$\alpha'(I - A)\left[(L^2 - A^2)\Lambda^{-1} + (I - L)\alpha\alpha'(I - L)\right]^{-1}(I - A)\alpha \leq 1, \tag{4.1.9}$$

where

$$\begin{aligned}\alpha &= P'\beta/\sigma,\\ L &= \operatorname{diag}\left(\frac{\lambda_1+d}{\lambda_1+1},\frac{\lambda_2+d}{\lambda_2+1},\cdots,\frac{\lambda_p+d}{\lambda_p+1}\right).\end{aligned}$$

Now, we compare GSLS estimator and the Liu estimator under MSE criterion. The following Corollary follows from Theorem (3.2.2).

Corollary 4.1.2 *For any GSLS estimator $\hat{\beta}_{GS}(A)$ and Liu estimator $\hat{\beta}_{GS}(L)$, $MSE(\hat{\beta}_{GS}(A)) \leq MSE(\hat{\beta}_{GS}(L))$, if either*

a.

$$\frac{\lambda_i+d}{\lambda_i+1} \leq a_i \leq \frac{\lambda_i\delta_i^2(\lambda_i+2-d)-\sigma^2(\lambda_i+d)}{(\lambda_i+1)(\lambda_i\delta_i^2+\sigma^2)},$$

$i=1,2,\cdots,p$; *or*

b.

$$\frac{\lambda_i\delta_i^2(\lambda_i+2-d)-\sigma^2(\lambda_i+d)}{(\lambda_i+1)(\lambda_i\delta_i^2+\sigma^2)} \leq a_i \leq \frac{\lambda_i+d}{\lambda_i+1},$$

$i=1,2,\cdots,p$.

4.2 Optimality of Liu-type ridge estimators

4.2.1 Admissibility of Liu-type ridge estimators

Liu-type ridge estimators (2003) can be written in terms of the GSLS estimators

$$\hat{\beta}_{k,d} = PL_RP'\hat{\beta}_{OLS}, \tag{4.2.1}$$

where $L_R = \Lambda(\Lambda + kI)^{-1}(I - d(\Lambda + kI)^{-1})$, $k > 0$ and $-\infty < d < \infty$.

It is not difficult to calculate the main diagonal entries l_i^r of the diagonal matrix L_R,

$$\begin{aligned} L &= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+k} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_p+k} \end{bmatrix} \begin{bmatrix} 1-\frac{d}{\lambda_1+k} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1-\frac{d}{\lambda_p+k} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1(\lambda_1+k-d)}{(\lambda_1+k)^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\lambda_p(\lambda_p+k-d)}{(\lambda_p+k)^2} \end{bmatrix}. \end{aligned} \tag{4.2.2}$$

So,

$$l_i^r = \frac{\lambda_i(\lambda_i + k - d)}{(\lambda_i + k)^2}, \quad i = 1, 2, \cdots, p. \tag{4.2.3}$$

Since $k > 0$ and $-\infty < d < \infty$, the main diagonal entries l_i^r are not always in the interval $[0, 1]$. To consider joint parameters restriction of Liu-type ridge estimators, we first prove a theorem.

Theorem 4.2.1 *For any $PAP'\hat{\beta}_{OLS}$, $A = diag(a_1, a_2, \cdots, a_p)$ with some $a_i < 0$ or $a_i > 1$, there exists a GSLS estimator $\hat{\beta}_{GS}(B)$, $0 \le b_i \le 1$, that is better than $\hat{\beta}_{GS}(A)$. Further, $\hat{\beta}_{GS}(B)$ can be constructed as follows:*

1. *if some $a_i > 1$, then set $b_i = 1$;*
2. *if some $-1 \le a_i < 0$, then set $b_i = -a_i$;*
3. *if some $a_i < -1$, then set $b_i = 1$;*

4. otherwise set $b_j = a_j$.

PROOF. By using (3.2.1), we know that

$$E[(\hat{\beta}_{GS}(A)-\beta)'(\hat{\beta}_{GS}(A)-\beta)] = \sigma^2(\frac{a_1^2}{\lambda_1}+\cdots+\frac{a_p^2}{\lambda_p})+\delta_1^2(1-a_1)^2+\cdots+\delta_p^2(1-a_p)^2, \tag{4.2.4}$$

where δ_i is the ith entries of column matrix $P'\beta$.

For any A such that some $a_i > 1$, we can find that replacing a_i by choosing $b_i = 1$ leads to a smaller value in the expression (4.2.4).

For any A such that some $-1 < a_i < 0$, we can find that replacing a_i by $b_i = -a_i$ leads to a smaller value in the expression (4.2.4).

For any A such that some $a_i < -1$, we can find that replacing a_i by $b_i = 1$ leads to a smaller value in the expression (4.2.4).

By (4.1.1), for at least one pair (β, σ^2), it holds

$$E[(\hat{\beta}_{GS}(A)-\beta)'(\hat{\beta}_{GS}(A)-\beta)] > E[(\hat{\beta}_{GS}(B)-\beta)'(\hat{\beta}_{GS}(B)-\beta)].$$

The conclusion follows.

From Theorem (4.2.1), we know that the main diagonal entries l_i^r in the expression (4.2.3) should be in the interval $[0, 1]$. Since

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

we can find joint parameters restriction of Liu-type ridge estimators

$$-k\left(1+\frac{k}{\lambda_i}\right) \leq d \leq \lambda_i + k. \tag{4.2.5}$$

From Lemma (4.1.2), we obtain the following result.

Theorem 4.2.2 *A Liu-type ridge estimator $\hat{\beta}_{k,d}$ is admissible if and only if* $rank(I - L_R) \geq p-2$ *and* $-k(1+\frac{k}{\lambda_i}) \leq d \leq \lambda_i + k$ *hold, where* $L_R = \Lambda(\Lambda + kI)^{-1}(I - d(\Lambda + kI)^{-1})$, $k > 0$ *and* $i = 1, 2, \cdots, p$.

4.2.2 Comparisons under $MMSE$ and MSE

By using (3.1.1) and (3.1.2), the covariance and bias of the Liu-type ridge estimator are given by, respectively

$$Cov(\hat{\beta}_{k,d}) = \sigma^2 P L_R \Lambda^{-1} L_R P', \tag{4.2.6}$$

$$Bias(\hat{\beta}_{k,d}) = P(L_R - I)P'\beta. \tag{4.2.7}$$

Combining (4.2.6) and (4.2.7), we find that

$$MMSE(\hat{\beta}_{k,d}) = \sigma^2 P L_R \Lambda^{-1} L_R P' + P(I - L_R)P'\beta\beta' P(I - L_R)P', \tag{4.2.8}$$

$$MSE(\hat{\beta}_{k,d}) = \sigma^2 \sum_{i=1}^{p} \frac{l_i^{r2}}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 (1 - l_i^r)^2, \tag{4.2.9}$$

where δ_i is the ith entries of column matrix $P'\beta$ and $l_i^r = \frac{\lambda_i(\lambda_i + k - d)}{(\lambda_i + k)^2}, i = 1, 2, \cdots, p$.

Using Theorem (3.1.1), we show the optimality of Liu-type ridge estimator within the GSLS estimator under $MMSE$ criterion.

Corollary 4.2.1 *Let $\hat{\beta}_{GS}(A)$ be GSLS estimator and $\hat{\beta}_{GS}(L_R)$ be Liu-type ridge estimator. Suppose $A < L_R$. $MMSE(\hat{\beta}_{GS}(L_R)) \geq MMSE(\hat{\beta}_{GS}(A))$ if and only if*

$$\alpha'(I-A)\left[(L_R^2-A^2)\Lambda^{-1}+(I-L_R)\alpha\alpha'(I-L_R)\right]^{-1}(I-A)\alpha \leq 1, \qquad (4.2.10)$$

where

$$\begin{aligned} \alpha &= P'\beta/\sigma, \\ L &= \text{diag}\left(\frac{\lambda_1(\lambda_1+k-d)}{(\lambda_1+k)^2}, \frac{\lambda_2(\lambda_2+k-d)}{(\lambda_2+k)^2}, \cdots, \frac{\lambda_p(\lambda_p+k-d)}{(\lambda_p+k)^2}\right). \end{aligned}$$

Now, we compare GSLS estimator and the Liu-type ridge estimator under MSE criterion. The following Corollary follows from Theorem (3.2.2).

Corollary 4.2.2 *For any GSLS estimator $\hat{\beta}_{GS}(A)$ and Liu-type ridge estimator $\hat{\beta}_{GS}(L_R)$, $MSE(\hat{\beta}_{GS}(A)) \leq MSE(\hat{\beta}_{GS}(L_R))$, if either*

a.

$$\frac{\lambda_i(\lambda_i+k-d)}{(\lambda_i+k)^2} \leq a_i \leq \frac{\lambda_i\delta_i^2(\lambda_i^2+3k\lambda_i+2k^2+d\lambda_i)-\lambda_i\sigma^2(\lambda_i+k-d)}{(\lambda_i\delta_i^2+\sigma^2)(\lambda_i+k)^2},$$

$i = 1, 2, \cdots, p$; *or*

b.

$$\frac{\lambda_i\delta_i^2(\lambda_i^2+3k\lambda_i+2k^2+d\lambda_i)-\lambda_i\sigma^2(\lambda_i+k-d)}{(\lambda_i\delta_i^2+\sigma^2)(\lambda_i+k)^2} \leq a_i \leq \frac{\lambda_i(\lambda_i+k-d)}{(\lambda_i+k)^2},$$

$i = 1, 2, \cdots, p$.

4.3 Optimality of $k-d$ class estimators

4.3.1 Admissibility of the $k-d$ class estimators

The $k-d$ class estimators can be written in terms of the GSLS estimators

$$\hat{\beta}(k,d) = PSP'\hat{\beta}_{OLS}, \tag{4.3.1}$$

where $S = \Lambda(\Lambda+I)^{-1}(I+d(\Lambda+kI)^{-1})$, $k>0$ and $-\infty < d < \infty$.

It is not difficult to calculate the main diagonal entries s_i of the diagonal matrix S,

$$\begin{aligned} L &= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1+1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_p+1} \end{bmatrix} \begin{bmatrix} 1+\frac{d}{\lambda_1+k} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1+\frac{d}{\lambda_p+k} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\lambda_1(\lambda_1+k+d)}{(\lambda_1+1)(\lambda_1+k)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\lambda_p(\lambda_p+k+d)}{(\lambda_p+1)(\lambda_p+k)} \end{bmatrix}. \end{aligned} \tag{4.3.2}$$

So,

$$s_i = \frac{\lambda_i(\lambda_i+k+d)}{(\lambda_i+1)(\lambda_i+k)}, \quad i = 1,2,\cdots,p. \tag{4.3.3}$$

Since $k>0$ and $-\infty < d < \infty$, the main diagonal entries s_i are not always in the interval $[0,1]$.

According to Theorem (4.2.1), we know that the main diagonal entries s_i in the expression (4.3.3) should be all in the interval $[0, 1]$. Since

$$0 \leq \frac{\lambda_i(\lambda_i + k + d)}{(\lambda_i + 1)(\lambda_i + k)} \leq 1,$$

we can find that joint parameters restriction of $k - d$ class estimators

$$-(\lambda_i + k) \leq d \leq 1 + \frac{k}{\lambda_i}, \ i = 1, 2, \cdots, p. \tag{4.3.4}$$

From Lemma (4.1.2), we obtain the following result.

Theorem 4.3.1 *A $k-d$ class estimator $\hat{\beta}(k, d)$ is admissible if and only if $rank(I - S) \geq p - 2$ and $-(\lambda_i + k) \leq d \leq 1 + \frac{k}{\lambda_i}$ hold, where $S = \Lambda(\Lambda + I)^{-1}(I + d(\Lambda + kI)^{-1})$, $i = 1, 2, \cdots, p$.*

4.3.2 Comparisons under $MMSE$ and MSE

By using (3.1.1) and (3.1.2), the covariance and bias of the $k - d$ class estimator are given by, respectively

$$Cov(\hat{\beta}(k, d)) = \sigma^2 PS\Lambda^{-1}SP', \tag{4.3.5}$$

$$Bias(\hat{\beta}(k, d)) = P(S - I)P'\beta. \tag{4.3.6}$$

Combining (4.3.5) and (4.3.6), we find that

$$MMSE(\hat{\beta}(k, d)) = \sigma^2 PS\Lambda^{-1}SP' + P(I - S)P'\beta\beta'P(I - S)P', \tag{4.3.7}$$

$$MSE(\hat{\beta}(k, d)) = \sigma^2 \sum_{i=1}^{p} \frac{{s_i}^2}{\lambda_i} + \sum_{i=1}^{p} \delta_i^2 (1 - s_i)^2, \tag{4.3.8}$$

where δ_i is the ith entries of column matrix $P'\beta$ and $s_i = \frac{\lambda_i(\lambda_i+k+d)}{(\lambda_i+1)(\lambda_i+k)}, i = 1, 2, \cdots, p.$

Using Theorem (3.1.1), we show the optimality of $k-d$ class estimator within the GSLS estimator under $MMSE$ criterion.

Corollary 4.3.1 *Let $\hat{\beta}_{GS}(A)$ be GSLS estimator and $\hat{\beta}_{GS}(S)$ be $k-d$ class estimator. Suppose $A < S$. $MMSE(\hat{\beta}_{GS}(S)) - MMSE(\hat{\beta}_{GS}(A))$ is n.n.d. matrix if and only if*

$$\alpha'(I-A)\left[(S^2-A^2)\Lambda^{-1} + (I-S)\alpha\alpha'(I-S)\right]^{-1}(I-A)\alpha \leq 1, \tag{4.3.9}$$

where

$$\begin{aligned} \alpha &= P'\beta/\sigma, \\ S &= \text{diag}\left(\frac{\lambda_1(\lambda_1+k+d)}{(\lambda_1+1)(\lambda_1+k)}, \frac{\lambda_2(\lambda_2+k+d)}{(\lambda_2+1)(\lambda_2+k)}, \cdots, \frac{\lambda_p(\lambda_p+k+d)}{(\lambda_p+1)(\lambda_p+k)}\right). \end{aligned}$$

Now, we compare GSLS estimator and the $k-d$ class estimator under MSE criterion. The following Corollary follows from Theorem (3.2.2).

Corollary 4.3.2 *For any GSLS estimator $\hat{\beta}_{GS}(A)$ and $k-d$ class estimator $\hat{\beta}_{GS}(S)$, $MSE(\hat{\beta}_{GS}(A)) \leq MSE(\hat{\beta}_{GS}(S))$, if either*

a.

$$\frac{\lambda_i(\lambda_i+k+d)}{(\lambda_i+1)(\lambda_i+k)} \leq a_i \leq \frac{\delta_i^2\lambda_i[\lambda_i^2+(k-d+2)\lambda_i+2k] - \sigma^2\lambda_i(\lambda_i+k+d)}{(\lambda_i+1)(\lambda_i+k)(\lambda_i\delta_i^2+\sigma^2)},$$

for $i = 1, 2, \cdots, p$; or

b.

$$\frac{\delta_i^2\lambda_i[\lambda_i^2+(k-d+2)\lambda_i+2k]-\sigma^2\lambda_i(\lambda_i+k+d)}{(\lambda_i+1)(\lambda_i+k)(\lambda_i\delta_i^2+\sigma^2)} \leq a_i \leq \frac{\lambda_i(\lambda_i+k-d)}{(\lambda_i+k)^2},$$

for $i = 1, 2, \cdots, p$.

Chapter 5

Simulation Study

5.1 Simulation design

In this Chapter, we want to evaluate the performance of the GSLS estimators with a simulation study by considering different levels of multicollinearity and different levels of σ. This simulation study was adopted by McDonald and Galarneau (1975), Hemmerle and Brantle (1978), Wichern and Churchill (1978), Gibbons (1981), Liu (2003) and Yang and Xu (2009).

In this simulation study, the number of observations n is 100 and $p = 4$. The explanatory variables are generated using the following equations:

$$x_{ij} = (1 - \rho^2)^{1/2} z_{ij} + \rho z_{i5},\ i = 1, \cdots, 100,\ j = 1, \cdots, 4,$$

where z_{i1}, z_{i2}, z_{i3}, z_{i4} and z_{i5} are independent standard normal pseudo-random numbers and ρ^2 is the theoretical correlation between any two explanatory variables. Because we want to choose the explanatory variables with high correlation coefficients, four sets of correlation are considered in this study, $\rho = 0.9, 0.99, 0.999$ and

0.9999. The resulting condition numbers of $X'X$ are 4.75, 16.77, 47.51 and 160.51, respectively.

For each design matrix X we use two coefficient vectors: the first one is the normalized eigenvector corresponding to the largest eigenvalue of $X'X$ (β_L) and the other is the normalized eigenvector corresponding to the smallest eigenvalue of $X'X$ (β_S).

Observations on the dependent variable are then generated by the following equation.

$$y_i = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon_i,\ i = 1, \cdots, n,$$

where ϵ_i are independent normal pseudo-random numbers with mean zero and variance σ^2. We will investigate three values of σ: 0.5, 1, 1.5.

If we believe that our new biased estimator yields a better estimator, then we can replace $\hat{\beta}$ and $\hat{\sigma}$ by $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}$ to get an iterated estimator which was suggested by Hoerl and Kennard (1976).

In this simulation study, thirteen estimators are compared:

1. $\hat{\beta}_{OLS}$: Ordinary least squares estimator
2. $\hat{\beta}_{GS}(A^{\text{opt}})$: GSLS estimator with $\hat{a}_i = \hat{a}_i^{\text{opt}}$
3. $\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$: Iterated GSLS estimator with $\hat{a}_i = \hat{a}_i^{\text{opt}}$
4. $\hat{\beta}_{GS}(A^{\text{ue}})$: GSLS estimator with $\hat{a}_i = \hat{a}_i^{\text{ue}}$ (due to Zhao, 1995)
5. $\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$: Iterated GSLS estimator with $\hat{a}_i = \hat{a}_i^{\text{ue}}$

6. $\hat{\beta}(k)$: Ridge estimator with $k = (p\hat{\sigma}^2)/(\hat{\beta}'\hat{\beta})$

7. $\hat{\hat{\beta}}(k)$: Iterated Ridge estimator

8. $\hat{\beta}(d)$: Liu estimator with $\hat{d} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^{p}(1/(\lambda_i(\lambda_i+1)))}{\sum_{i=1}^{p}(\hat{\delta}_i^2/(\lambda_i+1)^2)}\right]$ (due to Liu, 1993)

9. $\hat{\hat{\beta}}(d)$: Iterated Liu estimator

10. $\hat{\beta}_{k,d}$: Liu-type ridge estimator with $\hat{k} = \frac{\lambda_1 - 100\lambda_p}{99}$ and $\hat{d} = \frac{\sum_{i=1}^{p}((\hat{\sigma}^2 - \hat{k}\hat{\delta}_i^2)/(\lambda_i+\hat{k})^2)}{\sum_{i=1}^{p}((\lambda_i\hat{\delta}_i^2+\hat{\sigma}^2)/\lambda_i(\lambda_i+\hat{k})^2)}$ (due to Liu, 2003)

11. $\hat{\hat{\beta}}_{k,d}$: Iterated Liu-type ridge estimator

12. $\hat{\beta}(k,d)$: $k-d$ class estimator with $\hat{k} = p\hat{\sigma}_{OLS}^2/(\hat{\beta}'_{OLS}\hat{\beta}_{OLS})$ and $\hat{d} = \frac{\sum_{i=1}^{p}(\lambda_i(\hat{\delta}_i^2 - \hat{\sigma}^2))/((\lambda_i+1)^2(\lambda_i+k))}{\sum_{i=1}^{p}(\lambda_i(\lambda_i\hat{\delta}_i^2+\hat{\sigma}^2))/((\lambda_i+1)^2(\lambda_i+k)^2)}$ (due to Sakallioğlu and Kaçiranlar, 2008)

13. $\hat{\hat{\beta}}(k,d)$: Iterated $k-d$ class estimator

For each choice of ρ and σ, the experiment is replicated 1000 times by generating new error terms ϵ_i while X and β vector are fixed. After 1000 samples are generated, the estimated mean squared error ($EMSE$) is computed for each of the above estimators. $EMSE$ is defined by

$$EMSE(\hat{\beta}) = \frac{1}{1000}\sum_{j=1}^{1000}\sum_{i=1}^{4}(\hat{\beta}_{ij} - \beta_i)^2$$

where $\hat{\beta}_{ij}$ denotes the estimate of the ith parameter in jth replication and β_i are the true parameter values.

5.2 Numerical results

Our computations here and below are all performed by using the Matlab 7.0. In Table 5.1, we use true values of β and σ^2 to compare the GSLS estimator with the OLS estimator, Liu estimator, Liu-type ridge estimator and $k-d$ class estimator. For practical purposes, we should replace unknown parameters β and σ^2 by some suitable estimates $\hat{\beta}$ and $\hat{\sigma}^2$. In Table 5.2, we substitute the OLS estimators $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$ for β and σ^2. Then we substitute the ridge estimators $\hat{\beta}_R$ and $\hat{\sigma}^2_R$ for β and σ^2 in Table 5.3.

Table 5.1: EMSE using β and σ^2 under strong multicollinearity

ρ	σ	β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GS}(A^{\text{opt}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	$\hat{\beta}_{GS}(A^{\text{ue}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	$\hat{\beta}(k)$	$\hat{\hat{\beta}}(k)$	$\hat{\beta}(d)$	$\hat{\hat{\beta}}(d)$	$\hat{\beta}_{k,d}$	$\hat{\hat{\beta}}_{k,d}$	$\hat{\beta}(k,d)$	$\hat{\hat{\beta}}(k,d)$
0.9	0.5	β_L	0.0379	0.0007	0.0007	0.0100	0.0100	0.0314	0.0346	0.0344	0.0344	0.0042	0.0043	0.0344	0.0344
		β_S	0.0366	0.0158	0.0165	0.0211	0.0218	0.0433	0.0424	0.0372	0.0381	0.0372	0.0381	0.0372	0.0372
	1	β_L	0.1469	0.0027	0.0027	0.0387	0.0388	0.1028	0.1070	0.1334	0.1334	0.0099	0.0102	0.1334	0.1334
		β_S	0.1483	0.0571	0.0732	0.0790	0.1000	0.1416	0.1904	0.1378	0.1380	0.1347	0.1608	0.1377	0.1377
	1.5	β_L	0.3313	0.0062	0.0064	0.0875	0.0876	0.1986	0.1837	0.3007	0.3007	0.0191	0.0195	0.3007	0.3007
		β_S	0.3465	0.1264	0.1867	0.1776	0.2723	0.2844	0.3936	0.3193	0.3194	0.2919	0.3620	0.3192	0.3193
0.99	0.5	β_L	0.4278	0.0006	0.0006	0.1074	0.1074	0.0966	0.1944	0.1765	0.1765	0.0261	0.0338	0.1765	0.1765
		β_S	0.4251	0.1609	0.2410	0.2539	0.3632	0.4266	0.5701	0.3390	0.3473	0.4300	0.4904	0.3394	0.3392
	1	β_L	1.6675	0.0026	0.0026	0.4188	0.4188	0.1648	0.2159	0.6889	0.6889	0.1017	0.1298	0.6889	0.6889
		β_S	1.6631	0.3821	0.5644	0.7639	0.8174	0.6865	0.8413	0.8391	0.8425	0.6757	0.6960	0.8375	0.8385
	1.5	β_L	3.4635	0.0054	0.0055	0.8699	0.8700	0.1914	0.1527	1.4254	1.4254	0.2096	0.2652	1.4254	1.4254
		β_S	3.5618	0.6235	0.7770	1.4338	1.2019	0.8541	0.9281	1.6557	1.6569	0.9638	0.9139	1.6510	1.6537
0.999	0.5	β_L	3.6681	0.0006	0.0006	0.9174	0.9174	0.0345	0.1175	0.1135	0.1135	0.0086	0.0086	0.1135	0.1135
		β_S	3.7094	0.6175	0.7710	1.1419	1.1199	0.8625	0.9785	0.8099	0.8099	0.9151	0.9234	0.8101	0.8099
	1	β_L	13.7354	0.0023	0.0024	3.4356	3.4356	0.0374	0.0390	0.4241	0.4241	0.0320	0.0322	0.4241	0.4241
		β_S	13.9141	0.8381	0.9085	3.6107	3.6674	0.9473	0.9955	1.1295	1.1295	0.9476	0.9489	1.1280	1.1291
	1.5	β_L	32.4126	0.0053	0.0054	8.1071	8.1073	0.0436	0.0244	1.0034	1.0034	0.0754	0.0760	1.0034	1.0034
		β_S	31.3365	0.9186	0.9466	7.8522	7.9760	0.9757	0.9997	1.6480	1.6480	0.9857	0.9835	1.6445	1.6469
0.9999	0.5	β_L	39.6284	0.0007	0.0007	9.9077	9.9077	0.0040	0.0131	0.0137	0.0137	0.0018	0.0018	0.0137	0.0137
		β_S	42.4099	0.9638	0.9961	10.8294	10.8865	0.9913	1.0001	0.9888	0.9888	0.9946	0.9947	0.9888	0.9888
	1	β_L	167.4800	0.0026	0.0026	41.8719	41.8720	0.0061	0.0061	0.0567	0.0567	0.0067	0.0067	0.0567	0.0567
		β_S	160.2666	0.9853	0.9990	40.2100	40.3194	0.9993	1.0006	1.0281	1.0281	0.9993	0.9993	1.0279	1.0281
	1.5	β_L	370.0555	0.0059	0.0060	92.5183	92.5184	0.0094	0.0082	0.1250	0.1250	0.0149	0.0149	0.1250	0.1250
		β_S	381.4082	0.9966	0.9994	95.4394	95.6253	1.0051	1.0013	1.1027	1.1027	1.0082	1.0081	1.1022	1.1026

Table 5.2: EMSE using $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$ under strong multicollinearity

ρ	σ	β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GS}(A^{\text{opt}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	$\hat{\beta}_{GS}(A^{\text{ue}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	$\hat{\beta}(k)$	$\hat{\hat{\beta}}(k)$	$\hat{\beta}(d)$	$\hat{\hat{\beta}}(d)$	$\hat{\beta}_{k,d}$	$\hat{\hat{\beta}}_{k,d}$	$\hat{\beta}(k,d)$	$\hat{\hat{\beta}}(k,d)$
0.9	0.5	β_L	0.0382	0.0184	0.0128	0.0145	0.0114	0.0349	0.0348	0.0347	0.0347	0.0164	0.0107	0.0347	0.0347
		β_S	0.0363	0.0255	0.0224	0.0234	0.0216	0.0392	0.0404	0.0371	0.0373	0.0371	0.0373	0.0362	0.0362
	1	β_L	0.1495	0.0735	0.0520	0.0589	0.0470	0.1117	0.1097	0.1356	0.1356	0.0607	0.0358	0.1356	0.1356
		β_S	0.1522	0.1172	0.1102	0.1121	0.1137	0.1746	0.2146	0.1420	0.1417	0.1618	0.1777	0.1414	0.1414
	1.5	β_L	0.3245	0.1520	0.1031	0.1177	0.0950	0.1953	0.1813	0.2948	0.2948	0.1234	0.0681	0.2948	0.2948
		β_S	0.3239	0.2536	0.2559	0.2575	0.2743	0.3260	0.4129	0.2965	0.2964	0.3213	0.3740	0.2963	0.2963
0.99	0.5	β_L	0.4292	0.2046	0.1395	0.1596	0.1199	0.2393	0.2177	0.1842	0.1777	0.0962	0.0532	0.1766	0.1766
		β_S	0.4140	0.3296	0.3353	0.3319	0.3531	0.3931	0.4880	0.3770	0.3720	0.3923	0.4691	0.3342	0.3341
	1	β_L	1.6224	0.7485	0.4998	0.5782	0.4375	0.5992	0.3806	0.7000	0.6753	0.3584	0.1974	0.6710	0.6710
		β_S	1.6243	1.0105	0.9142	0.9346	0.8533	0.8675	0.8196	0.9289	0.8660	0.7323	0.7060	0.8413	0.8409
	1.5	β_L	3.6022	1.6454	1.0854	1.2540	0.9677	1.1298	0.5458	1.5429	1.4888	0.7860	0.4296	1.4810	1.4810
		β_S	3.9251	2.2381	1.8404	1.9260	1.6161	1.6557	1.1995	1.9936	1.8630	1.2890	1.0319	1.7874	1.7864
0.999	0.5	β_L	3.4145	1.5856	1.0734	1.2378	0.9719	1.1032	0.5577	0.6092	0.1967	0.0196	0.0082	0.1057	0.1057
		β_S	3.5975	2.0287	1.6770	1.7730	1.5067	1.4979	1.1025	1.3288	0.9780	0.8868	0.9247	0.8133	0.8132
	1	β_L	13.6935	6.4256	4.3829	5.0121	4.0060	4.1110	1.7842	2.6810	1.0481	0.0791	0.0327	0.4236	0.4236
		β_S	13.4550	6.4218	4.5256	5.0915	3.9854	4.1081	2.0213	2.9241	1.5096	0.9576	0.9552	1.1398	1.1398
	1.5	β_L	30.9262	14.3735	9.6899	11.1458	8.6890	8.8046	3.4097	5.6664	1.9745	0.1764	0.0733	0.9527	0.9527
		β_S	29.5627	13.5737	9.1423	10.5112	8.1254	8.2885	3.3476	5.4535	2.2881	1.0428	0.9922	1.6345	1.6355
0.9999	0.5	β_L	42.5460	20.0652	13.6474	15.6730	11.9777	12.1804	4.5627	7.6771	1.4657	0.0033	0.0018	0.0144	0.0144
		β_S	42.0176	20.0422	13.8274	15.6869	12.2733	12.4899	5.3509	8.2190	2.9688	0.9920	0.9947	0.9895	0.9890
	1	β_L	172.5951	84.0485	59.0274	67.0551	52.5900	51.0084	20.0408	33.6744	7.9912	0.0132	0.0069	0.0581	0.0581
		β_S	172.5547	83.3447	58.0842	65.7431	52.6907	51.0537	20.5714	34.3558	9.5435	0.9996	0.9984	1.0226	1.0273
	1.5	β_L	381.9151	181.7188	125.3091	143.2388	112.0917	109.3140	40.8455	68.3473	14.1799	0.0297	0.0155	0.1309	0.1310
		β_S	370.6922	175.2020	120.5930	137.6712	111.2718	105.9536	40.5804	67.7037	16.4391	1.0249	1.0130	1.1057	1.1184

Table 5.3: EMSE using $\hat{\beta}_R$ and $\hat{\sigma}^2_R$ under strong multicollinearity

ρ	σ	β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GS}(A^{\text{opt}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	$\hat{\beta}_{GS}(A^{\text{ue}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	$\hat{\beta}(k)$	$\hat{\hat{\beta}}(k)$	$\hat{\beta}(d)$	$\hat{\hat{\beta}}(d)$	$\hat{\beta}_{k,d}$	$\hat{\hat{\beta}}_{k,d}$	$\hat{\beta}(k,d)$	$\hat{\hat{\beta}}(k,d)$
0.9	0.5	β_L	0.0365	0.0174	0.0120	0.0138	0.0107	0.0333	0.0333	0.0332	0.0332	0.0156	0.0101	0.0332	0.0332
		β_S	0.0381	0.0275	0.0244	0.0254	0.0236	0.0417	0.0432	0.0393	0.0395	0.0392	0.0395	0.0382	0.0382
	1	β_L	0.1507	0.0731	0.0511	0.0582	0.0454	0.1127	0.1108	0.1368	0.1368	0.0604	0.0351	0.1368	0.1368
		β_S	0.1456	0.1098	0.1020	0.1040	0.1052	0.1696	0.2102	0.1365	0.1363	0.1553	0.1702	0.1361	0.1361
	1.5	β_L	0.3468	0.1695	0.1188	0.1351	0.1036	0.2109	0.1957	0.3146	0.3146	0.1387	0.0790	0.3146	0.3146
		β_S	0.3334	0.2646	0.2678	0.2670	0.2891	0.3432	0.4340	0.3065	0.3065	0.3375	0.3961	0.3064	0.3064
0.99	0.5	β_L	0.4140	0.0425	0.0429	0.0744	0.0903	0.1809	0.2023	0.1704	0.1704	0.0434	0.0388	0.1704	0.1704
		β_S	0.4056	0.4091	0.4257	0.5175	0.4265	0.6307	0.7608	0.3212	0.3305	0.5047	0.5146	0.3215	0.3214
	1	β_L	1.6243	0.1538	0.1474	0.2955	0.3510	0.2452	0.2490	0.6724	0.6724	0.1667	0.1481	0.6723	0.6723
		β_S	1.6341	0.7605	0.8968	0.6206	0.6984	0.8066	0.9256	0.8352	0.8413	0.7033	0.7129	0.8341	0.8346
	1.5	β_L	3.8252	0.3878	0.3922	0.6946	0.8407	0.3063	0.2292	1.5747	1.5778	0.3945	0.3515	1.5747	1.5747
		β_S	3.8390	1.0334	1.2361	0.9264	1.1234	0.8813	0.9530	1.7025	1.7095	0.9150	0.9034	1.6992	1.7004
0.999	0.5	β_L	3.4145	0.0008	0.0006	0.8512	0.8539	0.1023	0.1355	0.1057	0.1057	0.0081	0.0081	0.1057	0.1057
		β_S	3.5975	0.9862	0.9988	1.1398	1.1424	0.9944	0.9999	0.8132	0.8132	0.9252	0.9252	0.8132	0.8132
	1	β_L	13.6935	0.0035	0.0024	3.4131	3.4242	0.0400	0.0412	0.4236	0.4236	0.0322	0.0322	0.4236	0.4236
		β_S	13.4550	0.9921	1.0002	3.6400	3.6495	0.9972	1.0001	1.1403	1.1403	0.9552	0.9552	1.1400	1.1398
	1.5	β_L	31.6093	0.0069	0.0051	7.8811	7.9049	0.0256	0.0223	0.9739	0.9739	0.0735	0.0734	0.9739	0.9739
		β_S	30.8039	0.9936	1.0007	7.9701	7.9933	0.9988	1.0004	1.6820	1.6820	0.9973	0.9973	1.6812	1.6809
0.9999	0.5	β_L	41.3570	0.0007	0.0007	10.3394	10.3398	0.0131	0.0140	0.0139	0.0139	0.0018	0.0018	0.0139	0.0139
		β_S	40.9149	1.0001	1.0002	10.5229	10.5232	1.0001	1.0000	0.9887	0.9887	0.9947	0.9947	0.9887	0.9887
	1	β_L	170.1838	0.0029	0.0029	42.5467	42.5481	0.0064	0.0064	0.0575	0.0575	0.0071	0.0071	0.0575	0.0575
		β_S	168.0705	1.0013	1.0010	42.3475	42.3487	1.0007	1.0002	1.0346	1.0346	1.0006	1.0006	1.0346	1.0346
	1.5	β_L	374.7944	0.0063	0.0063	93.7000	93.7034	0.0088	0.0088	0.1283	0.1283	0.0154	0.0154	0.1283	0.1283
		β_S	372.4012	1.0028	1.0019	93.2923	93.2950	1.0015	1.0003	1.0961	1.0961	1.0067	1.0067	1.0961	1.0961

Based on above Tables, we get the following results.

1. From table (5.1), we find that $\hat{\beta}_{GS}(A^{\text{opt}})$ always gives better performance than the other biased estimators by using the true values of β and σ^2. In particular, we can observe that in the presence of strong collinearity, $\hat{\beta}_{GS}(A^{\text{opt}})$ is much superior than $\hat{\beta}_{OLS}$. All of these results agree with the theory (3.2.2).

2. From table (5.2), we find that $\hat{\beta}_{GS}(A^{\text{opt}})$ does not have smaller $EMSE$ than the other biased estimators by replacing β and σ by $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$. $\hat{\beta}_{GS}(A^{\text{ue}})$ can improve the performance of $\hat{\beta}_{GS}(A^{\text{opt}})$ which agree with the theoretical results. The GSLS estimators highly depend on the choice of the starting values of β and σ^2. The OLS estimator is unstable and sometimes gives misleading information. From now on, we suggest to replace the starting values of β and σ^2 by their respective ridge estimators $\hat{\beta}_R$ and $\hat{\sigma}^2_R$ to make the GSLS estimators more reliable.

3. In table (5.3), we replace β and σ^2 by ridge estimators $\hat{\beta}_R$ and $\hat{\sigma}^2_R$. $\hat{\beta}_{GS}(A^{\text{opt}})$ generally gives better performance than the corresponding biased estimators.

When there exists only very weak multicollinearity, we still believe that our new biased estimator yields a better estimator. Because we want to choose the explanatory variables with weak correlation coefficients, four sets of correlation are considered in this study, $\rho = 0.7, 0.5, 0.3$ and 0. The resulting condition numbers of $X'X$ are 3.13, 1.80, 1.54 and 1.33, respectively. Due to the low levels of multicollinearity, every biased estimator gives very good performance.

Table 5.4: EMSE using β and σ^2 under weak multicollinearity

ρ	σ	β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GS}(A^{\text{opt}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	$\hat{\beta}_{GS}(A^{\text{ue}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	$\hat{\beta}(k)$	$\hat{\hat{\beta}}(k)$	$\hat{\beta}(d)$	$\hat{\hat{\beta}}(d)$	$\hat{\beta}_{k,d}$	$\hat{\hat{\beta}}_{k,d}$	$\hat{\beta}(k,d)$	$\hat{\hat{\beta}}(k,d)$
0.7	0.5	β_L	0.0203	0.0009	0.0010	0.0058	0.0058	0.0184	0.0193	0.0193	0.0193	0.0118	0.0113	0.0193	0.0193
		β_S	0.0201	0.0097	0.0100	0.0123	0.0126	0.0227	0.0220	0.0200	0.0204	0.0199	0.0204	0.0202	0.0202
	1	β_L	0.0768	0.0036	0.0037	0.0219	0.0220	0.0632	0.0641	0.0729	0.0729	0.0231	0.0242	0.0729	0.0729
		β_S	0.0771	0.0368	0.0433	0.0468	0.0542	0.0832	0.1095	0.0753	0.0756	0.0755	0.0847	0.0753	0.0753
	1.5	β_L	0.1807	0.0084	0.0087	0.0515	0.0518	0.1353	0.1271	0.1714	0.1714	0.0381	0.0398	0.1714	0.1714
		β_S	0.1804	0.0809	0.1092	0.1041	0.1433	0.1693	0.2486	0.1724	0.1725	0.1670	0.2024	0.1724	0.1724
0.5	0.5	β_L	0.0124	0.0014	0.0014	0.0041	0.0041	0.0118	0.0121	0.0121	0.0121	0.0120	0.0117	0.0121	0.0121
		β_S	0.0121	0.0044	0.0045	0.0063	0.0064	0.0127	0.0125	0.0121	0.0122	0.0122	0.0122	0.0121	0.0121
	1	β_L	0.0495	0.0052	0.0053	0.0163	0.0164	0.0449	0.0454	0.0482	0.0482	0.0376	0.0405	0.0482	0.0482
		β_S	0.0499	0.0175	0.0189	0.0255	0.0270	0.0492	0.0543	0.0488	0.0488	0.0484	0.0512	0.0488	0.0488
	1.5	β_L	0.1099	0.0122	0.0128	0.0366	0.0372	0.0946	0.0920	0.1069	0.1069	0.0670	0.0731	0.1069	0.1069
		β_S	0.1147	0.0403	0.0471	0.0585	0.0662	0.1064	0.1295	0.1118	0.1118	0.1076	0.1195	0.1118	0.1118
0.3	0.5	β_L	0.0101	0.0015	0.0015	0.0036	0.0036	0.0098	0.0099	0.0099	0.0099	0.0100	0.0098	0.0099	0.0099
		β_S	0.0104	0.0033	0.0034	0.0051	0.0051	0.0107	0.0105	0.0103	0.0104	0.0104	0.0104	0.0103	0.0103
	1	β_L	0.0401	0.0059	0.0060	0.0144	0.0146	0.0372	0.0378	0.0392	0.0392	0.0350	0.0368	0.0392	0.0392
		β_S	0.0406	0.0137	0.0145	0.0204	0.0212	0.0399	0.0431	0.0398	0.0398	0.0395	0.0413	0.0398	0.0398
	1.5	β_L	0.0934	0.0125	0.0132	0.0327	0.0334	0.0830	0.0820	0.0913	0.0913	0.0716	0.0763	0.0913	0.0913
		β_S	0.0942	0.0314	0.0359	0.0469	0.0519	0.0898	0.1064	0.0925	0.0925	0.0901	0.0987	0.0925	0.0925
0	0.5	β_L	0.0107	0.0020	0.0020	0.0041	0.0041	0.0105	0.0106	0.0106	0.0106	0.0108	0.0107	0.0106	0.0106
		β_S	0.0109	0.0038	0.0038	0.0056	0.0056	0.0111	0.0111	0.0108	0.0110	0.0109	0.0110	0.0108	0.0108
	1	β_L	0.0449	0.0083	0.0086	0.0175	0.0178	0.0425	0.0435	0.0441	0.0441	0.0419	0.0440	0.0441	0.0441
		β_S	0.0399	0.0118	0.0125	0.0188	0.0195	0.0389	0.0417	0.0391	0.0391	0.0387	0.0404	0.0391	0.0391
	1.5	β_L	0.0962	0.0167	0.0180	0.0365	0.0378	0.0864	0.0878	0.0941	0.0941	0.0814	0.0872	0.0941	0.0941
		β_S	0.1010	0.0319	0.0362	0.0490	0.0537	0.0946	0.1087	0.0990	0.0990	0.0954	0.1039	0.0990	0.0990

Table 5.5: EMSE using $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$ under weak multicollinearity

ρ	σ	β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GS}(A^{\text{opt}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	$\hat{\beta}_{GS}(A^{\text{ue}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	$\hat{\beta}(k)$	$\hat{\hat{\beta}}(k)$	$\hat{\beta}(d)$	$\hat{\hat{\beta}}(d)$	$\hat{\beta}_{k,d}$	$\hat{\hat{\beta}}_{k,d}$	$\hat{\beta}(k,d)$	$\hat{\hat{\beta}}(k,d)$
0.7	0.5	β_L	0.0200	0.0099	0.0070	0.0079	0.0062	0.0190	0.0190	0.0189	0.0189	0.0137	0.0129	0.0189	0.0189
		β_S	0.0194	0.0142	0.0127	0.0132	0.0124	0.0213	0.0217	0.0198	0.0199	0.0198	0.0198	0.0199	0.0199
	1	β_L	0.0813	0.0404	0.0289	0.0325	0.0266	0.0683	0.0678	0.0771	0.0771	0.0443	0.0361	0.0771	0.0771
		β_S	0.0804	0.0621	0.0569	0.0582	0.0563	0.0973	0.1135	0.0779	0.0777	0.0848	0.0873	0.0772	0.0772
	1.5	β_L	0.1809	0.0896	0.0635	0.0715	0.0580	0.1300	0.1256	0.1715	0.1715	0.0891	0.0658	0.1715	0.1715
		β_S	0.1773	0.1450	0.1431	0.1433	0.1521	0.2170	0.2828	0.1691	0.1691	0.1927	0.2142	0.1690	0.1690
0.5	0.5	β_L	0.0123	0.0066	0.0049	0.0054	0.0045	0.0121	0.0121	0.0120	0.0120	0.0117	0.0117	0.0120	0.0120
		β_S	0.0119	0.0079	0.0067	0.0070	0.0064	0.0122	0.0122	0.0120	0.0120	0.0120	0.0120	0.0119	0.0119
	1	β_L	0.0479	0.0255	0.0191	0.0211	0.0176	0.0441	0.0440	0.0466	0.0466	0.0401	0.0404	0.0466	0.0466
		β_S	0.0479	0.0323	0.0277	0.0291	0.0266	0.0516	0.0535	0.0470	0.0470	0.0491	0.0496	0.0470	0.0470
	1.5	β_L	0.1120	0.0603	0.0456	0.0504	0.0419	0.0946	0.0934	0.1089	0.1089	0.0806	0.0792	0.1089	0.1089
		β_S	0.1078	0.0757	0.0667	0.0693	0.0659	0.1196	0.1353	0.1051	0.1051	0.1123	0.1175	0.1051	0.1051
0.3	0.5	β_L	0.0104	0.0058	0.0044	0.0048	0.0041	0.0102	0.0102	0.0102	0.0102	0.0101	0.0101	0.0102	0.0102
		β_S	0.0104	0.0068	0.0058	0.0061	0.0055	0.0106	0.0107	0.0105	0.0105	0.0105	0.0105	0.0104	0.0104
	1	β_L	0.0419	0.0234	0.0182	0.0198	0.0170	0.0398	0.0397	0.0411	0.0411	0.0389	0.0393	0.0411	0.0411
		β_S	0.0421	0.0282	0.0241	0.0253	0.0232	0.0441	0.0452	0.0413	0.0413	0.0427	0.0430	0.0413	0.0413
	1.5	β_L	0.0950	0.0525	0.0404	0.0442	0.0379	0.0844	0.0840	0.0929	0.0929	0.0792	0.0808	0.0929	0.0929
		β_S	0.0911	0.0616	0.0532	0.0557	0.0514	0.0971	0.1054	0.0890	0.0890	0.0929	0.0957	0.0890	0.0890
0	0.5	β_L	0.0107	0.0061	0.0048	0.0052	0.0045	0.0106	0.0106	0.0105	0.0105	0.0106	0.0106	0.0105	0.0105
		β_S	0.0110	0.0071	0.0060	0.0063	0.0057	0.0111	0.0111	0.0110	0.0110	0.0110	0.0110	0.0109	0.0109
	1	β_L	0.0419	0.0238	0.0186	0.0202	0.0174	0.0402	0.0402	0.0410	0.0410	0.0401	0.0405	0.0410	0.0410
		β_S	0.0429	0.0278	0.0235	0.0248	0.0225	0.0443	0.0451	0.0421	0.0421	0.0433	0.0436	0.0421	0.0421
	1.5	β_L	0.0975	0.0567	0.0449	0.0485	0.0425	0.0900	0.0904	0.0955	0.0955	0.0894	0.0924	0.0955	0.0955
		β_S	0.0974	0.0657	0.0564	0.0590	0.0547	0.1044	0.1129	0.0956	0.0956	0.1007	0.1046	0.0956	0.0956

Table 5.6: EMSE using $\hat{\beta}_R$ and $\hat{\sigma}^2_R$ under weak multicollinearity

ρ	σ	β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GS}(A^{\text{opt}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	$\hat{\beta}_{GS}(A^{\text{ue}})$	$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	$\hat{\beta}(k)$	$\hat{\hat{\beta}}(k)$	$\hat{\beta}(d)$	$\hat{\hat{\beta}}(d)$	$\hat{\beta}_{k,d}$	$\hat{\hat{\beta}}_{k,d}$	$\hat{\beta}(k,d)$	$\hat{\hat{\beta}}(k,d)$
0.7	0.5	β_L	0.0201	0.0099	0.0070	0.0079	0.0063	0.0191	0.0191	0.0190	0.0190	0.0138	0.0131	0.0190	0.0190
		β_S	0.0202	0.0150	0.0134	0.0139	0.0130	0.0221	0.0225	0.0207	0.0207	0.0207	0.0207	0.0209	0.0209
	1	β_L	0.0800	0.0401	0.0288	0.0323	0.0263	0.0673	0.0668	0.0759	0.0759	0.0440	0.0361	0.0759	0.0759
		β_S	0.0789	0.0615	0.0568	0.0581	0.0564	0.0951	0.1112	0.0762	0.0760	0.0831	0.0857	0.0755	0.0755
	1.5	β_L	0.1803	0.0884	0.0619	0.0701	0.0552	0.1288	0.1243	0.1708	0.1708	0.0873	0.0634	0.1708	0.1708
		β_S	0.1780	0.1455	0.1441	0.1443	0.1565	0.2171	0.2821	0.1698	0.1698	0.1936	0.2167	0.1697	0.1697
0.5	0.5	β_L	0.0125	0.0066	0.0050	0.0055	0.0046	0.0122	0.0122	0.0122	0.0122	0.0118	0.0118	0.0122	0.0122
		β_S	0.0124	0.0085	0.0074	0.0077	0.0071	0.0128	0.0129	0.0126	0.0126	0.0126	0.0126	0.0125	0.0125
	1	β_L	0.0514	0.0276	0.0208	0.0229	0.0189	0.0473	0.0472	0.0501	0.0501	0.0425	0.0426	0.0501	0.0501
		β_S	0.0506	0.0343	0.0294	0.0309	0.0281	0.0541	0.0561	0.0495	0.0495	0.0517	0.0522	0.0495	0.0495
	1.5	β_L	0.1132	0.0610	0.0463	0.0511	0.0431	0.0957	0.0945	0.1101	0.1101	0.0813	0.0798	0.1101	0.1101
		β_S	0.1101	0.0767	0.0673	0.0699	0.0659	0.1196	0.1341	0.1073	0.1073	0.1133	0.1181	0.1073	0.1073
0.3	0.5	β_L	0.0101	0.0054	0.0041	0.0045	0.0037	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100
		β_S	0.0098	0.0063	0.0053	0.0056	0.0052	0.0099	0.0099	0.0098	0.0098	0.0098	0.0098	0.0098	0.0098
	1	β_L	0.0403	0.0221	0.0169	0.0186	0.0159	0.0380	0.0380	0.0394	0.0394	0.0370	0.0374	0.0394	0.0394
		β_S	0.0422	0.0282	0.0241	0.0254	0.0232	0.0446	0.0457	0.0416	0.0416	0.0431	0.0434	0.0416	0.0416
	1.5	β_L	0.0950	0.0525	0.0401	0.0437	0.0372	0.0845	0.0841	0.0929	0.0929	0.0794	0.0809	0.0929	0.0929
		β_S	0.0951	0.0654	0.0567	0.0592	0.0552	0.1028	0.1116	0.0933	0.0933	0.0981	0.1012	0.0933	0.0933
0	0.5	β_L	0.0109	0.0062	0.0048	0.0052	0.0045	0.0108	0.0108	0.0107	0.0107	0.0108	0.0108	0.0107	0.0107
		β_S	0.0112	0.0072	0.0061	0.0064	0.0058	0.0113	0.0113	0.0112	0.0112	0.0112	0.0112	0.0111	0.0111
	1	β_L	0.0399	0.0221	0.0170	0.0186	0.0159	0.0382	0.0382	0.0391	0.0391	0.0381	0.0384	0.0391	0.0391
		β_S	0.0416	0.0273	0.0231	0.0244	0.0225	0.0438	0.0448	0.0409	0.0409	0.0425	0.0429	0.0409	0.0409
	1.5	β_L	0.0960	0.0551	0.0432	0.0468	0.0405	0.0878	0.0881	0.0940	0.0940	0.0865	0.0891	0.0940	0.0940
		β_S	0.0982	0.0664	0.0571	0.0598	0.0557	0.1057	0.1144	0.0964	0.0964	0.1018	0.1057	0.0964	0.0964

Chapter 6

Applications

6.1 Portland cement

6.1.1 The proposed procedure

To illustrate our theoretical results, we use an example in Chapter 1 which is commonly used in literature to compare our proposed estimator with ridge estimator, Liu estimator, Liu-type ridge estimator and $k-d$ class estimator.

Following the original analysis by Woods et al. (1932), we fit a homogeneous linear model (1.1.1) without intercept to the data. Under this model, we have $n = 13$ observations and $p = 4$ unknown regression coefficients. We find the eigenvalues of $X'X$: $44663.3027, 5965.3394, 809.9521, 105.4058$. The condition number of $X'X$ is 20.5846. $X'X$ may be considered as well-conditioned, so there exists only very weak multicollinearity. For practical purposes, we have to replace the unknown parameters β and σ^2 by some suitable estimators. In Table (6.1), we replace the unknown parameters β and σ^2 with their respective OLS estimators $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS} = 5.8455$. In Table (6.2), we replace the unknown parameters β and σ^2 by

their respective ridge estimators $\hat{\beta}_R$ and $\hat{\sigma}_R^2 = 5.8848$. For this data, we can find the following results under the homogeneous linear regression model.

Table 6.1: Summary of the homogeneous model using $\hat{\beta}_{OLS}$ and $\hat{\sigma}^2_{OLS}$

Estimators	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_{OLS}$	2.1930	1.1533	0.7585	0.4863
$\hat{\beta}_{GS}(A^{\text{opt}})$	2.1655	1.1563	0.7463	0.4931
$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	2.1646	1.1564	0.7458	0.4934
$\hat{\beta}_{GS}(A^{\text{ue}})$	2.1651	1.1564	0.7461	0.4932
$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	2.1641	1.1565	0.7455	0.4936
$\hat{\beta}(k)$	2.1478	1.1638	0.7261	0.4931
$\hat{\hat{\beta}}(k)$	2.1461	1.1642	0.7249	0.4934
$\hat{\beta}(d)$	2.1793	1.1565	0.7486	0.4884
$\hat{\hat{\beta}}(d)$	2.1793	1.1565	0.7486	0.4884
$\hat{\beta}_{k,d}$	1.3212	1.3547	0.1498	0.6148
$\hat{\hat{\beta}}_{k,d}$	1.2154	1.3668	0.1324	0.6217
$\hat{\beta}(k,d)$	2.1793	1.1565	0.7486	0.4884
$\hat{\hat{\beta}}(k,d)$	2.1793	1.1565	0.7486	0.4884

Table 6.2: Summary of the homogeneous model using $\hat{\beta}_R$ and $\hat{\sigma}_R^2$

Estimators	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_{OLS}$	2.1930	1.1533	0.7585	0.4863
$\hat{\beta}_{GS}(A^{\text{opt}})$	2.1639	1.1567	0.7451	0.4934
$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	2.1645	1.1565	0.7458	0.4934
$\hat{\beta}_{GS}(A^{\text{ue}})$	2.1634	1.1568	0.7449	0.4935
$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	2.1641	1.1565	0.7455	0.4936
$\hat{\beta}(k)$	2.1461	1.1642	0.7249	0.4934
$\hat{\hat{\beta}}(k)$	2.1460	1.1642	0.7248	0.4934
$\hat{\beta}(d)$	2.1793	1.1565	0.7486	0.4884
$\hat{\hat{\beta}}(d)$	2.1793	1.1565	0.7486	0.4884
$\hat{\beta}_{k,d}$	1.3207	1.3548	0.1497	0.6148
$\hat{\hat{\beta}}_{k,d}$	1.2152	1.3668	0.1324	0.6217
$\hat{\beta}(k,d)$	2.1793	1.1565	0.7486	0.4884
$\hat{\hat{\beta}}(k,d)$	2.1793	1.1565	0.7486	0.4884

To further identify the MSE performances of the GSLS estimator, the Liu estimator, the Liu-type ridge estimator and the $k-d$ class estimator, we following the strategy adopted by Hald (1952), Gorman and Toman (1966) and Daniel and Wood (1980), fit a linear model with intercept (inhomogeneous model) to the data. Under this model, the condition number of $X'X$ is 6056.3443, so $X'X$ may now be considered as ill-conditioned. Since there exists very strong multicollinearity, the OLS estimator may not be reliable.

According to the simulation study results, we use the ridge estimates as the starting values β and σ^2,

$$\hat{\beta}_R = (8.5870, 2.1046, 1.0648, 0.6681, 0.3996)'$$

and $\hat{\sigma}^2_R = 6.4241$ in Table (6.3).

Table 6.3: Summary of the inhomogeneous model using $\hat{\beta}_R$ and $\hat{\sigma}_R^2$

Estimators	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\hat{\beta}_{OLS}$	62.4054	1.5511	0.5102	0.1019	-0.1441
$\hat{\beta}_{GS}(A^{\text{opt}})$	0.8978	2.1536	1.1474	0.7357	0.4847
$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	0.0553	2.1604	1.1563	0.7436	0.4937
$\hat{\beta}_{GS}(A^{\text{ue}})$	31.1174	1.8422	0.8360	0.4175	0.1796
$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	31.2249	1.8393	0.8351	0.4154	0.1791
$\hat{\beta}(k)$	0.2356	2.1851	1.1522	0.7520	0.4848
$\hat{\hat{\beta}}(k)$	0.0625	2.1297	1.1671	0.7129	0.4951
$\hat{\beta}(d)$	0.1230	2.1781	1.1552	0.7473	0.4871
$\hat{\hat{\beta}}(d)$	0.1230	2.1781	1.1552	0.7473	0.4871
$\hat{\beta}_{k,d}$	0.0309	0.8623	1.4051	0.0792	0.6436
$\hat{\hat{\beta}}_{k,d}$	0.0308	0.8574	1.4055	0.0789	0.6439
$\hat{\beta}(k,d)$	0.1230	2.1781	1.1552	0.7473	0.4871
$\hat{\hat{\beta}}(k,d)$	0.1230	2.1781	1.1552	0.7473	0.4871

6.1.2 Results

Comparing with the other estimators, there are two problems associated with the OLS estimator under the inhomogeneous model. First, the intercept of the OLS estimator $\hat{\beta}_0 = 62.4054$ is too large. Second, the sign of the correlation coefficient $\hat{\beta}_4 = -0.1441$ of X_4 is negative.

Numerical results are summarized in Tables (6.3) under the inhomogeneous linear regression model to compare the GSLS estimator with the OLS estimator, the Liu estimator, the Liu-type ridge estimator and the $k - d$ class estimator.

Based on Table (6.3), we have the following comments.

1. We replace β and σ^2 by ridge estimator $\hat{\beta}_R$ and $\hat{\sigma}^2_R$ in Table (6.3). $\hat{\beta}_{GS}(A^{\text{opt}})$ gives better performance than the corresponding biased estimators. We can see that our proposed estimator corrects all problems. The sign of the correlation coefficient $\hat{\beta}_4 = 0.4847$ of X_4 is positive and the intercept of the GSLS estimator $\hat{\beta}_0 = 0.8978$ is small. Our proposed estimator $\hat{\beta}_{GS}(A^{\text{opt}})$ under the inhomogeneous linear regression model is very close to the OLS estimator under homogeneous linear regression model. These agree with the original analysis by Woods et al. (1932).

2. Because the condition number (> 6000) is so large, the intercept of the starting value of ridge estimator $\hat{\beta}_0 = 8.5870$ is still large. In this situation, we suggest to use the iterated GSLS estimator $\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$ which gives a very small intercept $\hat{\beta}_0 = 0.0553$.

6.2 Air pollution in U.S. cities

6.2.1 The proposed procedure

We fit a linear model with intercept (inhomogeneous model) to the air pollution data. The condition number of $X'X$ is 22798.5515. $X'X$ may be considered as quite ill-conditioned.

First, we replace the starting values β and σ^2 by their respective ridge estimator

$$\hat{\beta}_R = (49.5356, -0.5161, 0.0687, -0.0432, -1.7058, 0.1956, 0.1100)'$$

and $\hat{\sigma}_R^2 = 225.0979$. The thirteen biased estimators are given in Table (6.4).

6.2.2 Results

Based on Table (6.4), we have the following comments.

1. We suggest to use the iterated GSLS estimator $\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$ due to extremely large condition number (= 22798.5515). Since there is very strong multicollinearity, we can not trust the OLS estimator. The sign of the correlation coefficient $\hat{\beta}_6 = -0.0521$ of the OLS estimator is negative. We can see that our proposed estimator corrects this problem. The sign of the correlation coefficient $\hat{\beta}_6 = 0.0010$ is positive.

2. Our proposed estimator $\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$ reduces the intercept of the OLS estimator without over shrunken it. And also, the signs of $\hat{\beta}_1$ and $\hat{\beta}_5$ are consistent under

both the OLS estimator and the GSLS estimator. Based on the simulation study, we have sufficient confidence that the GSLS estimator gives better performance than the other biased estimators.

Table 6.4: Summary of the air pollution using $\hat{\beta}_R$ and $\hat{\sigma}^2_R$

Estimators	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
$\hat{\beta}_{OLS}$	111.7285	-1.2679	0.0649	-0.0393	-3.1814	0.5124	-0.0521
$\hat{\beta}_{GS}(A^{\text{opt}})$	57.0555	-0.6285	0.0687	-0.0435	-1.4082	0.3314	0.0311
$\hat{\hat{\beta}}_{GS}(A^{\text{opt}})$	64.9043	-0.7197	0.0688	-0.0435	-1.5286	0.3817	0.0010
$\hat{\beta}_{GS}(A^{\text{ue}})$	4.6416	0.0184	0.0715	-0.0462	-0.3598	0.0222	0.1892
$\hat{\hat{\beta}}_{GS}(A^{\text{ue}})$	31.1000	-0.3024	0.0698	-0.0444	-0.9974	0.1552	0.1218
$\hat{\beta}(k)$	14.4860	-0.0926	0.0708	-0.0454	-0.8718	0.0172	0.2013
$\hat{\hat{\beta}}(k)$	1.3094	0.0625	0.0716	-0.0462	-0.5235	-0.0475	0.2340
$\hat{\beta}(d)$	9.7399	-0.0355	0.0711	-0.0457	-0.7574	-0.0068	0.2135
$\hat{\hat{\beta}}(d)$	9.7399	-0.0355	0.0711	-0.0457	-0.7574	-0.0068	0.2135
$\hat{\beta}_{k,d}$	0.0006	0.0306	0.0450	-0.0127	0.0055	0.0241	0.0838
$\hat{\hat{\beta}}_{k,d}$	0.0005	0.0234	0.0383	-0.0046	0.0042	0.0184	0.0635
$\hat{\beta}(k,d)$	9.7399	-0.0355	0.0711	-0.0457	-0.7574	-0.0068	0.2135
$\hat{\hat{\beta}}(k,d)$	9.7399	-0.0355	0.0711	-0.0457	-0.7574	-0.0068	0.2135

Bibliography

[1] Akdeniz, F. (2001). The examination and analysis of residuals for some biased estimators in linear regression. *Communications in Statistics: Theory and Methods*, 30, 1171-1183.

[2] Akdeniz, F., 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, 1789-1797.

[3] Akdeniz, F., Kaçiranlar, S. (2001). More on the new biased estimator in linear regression. *Indian Journal of Statistics*, 63, 321-325.

[4] Arslan, O., Billor, N. (2000). Robust Liu estimator for regression based on an M-estimator. *Journal of Applied Statistics*, 27, 39-47.

[5] Belsley, D.A., Kuh, E., Welsch, R.E. (1980). Regression Diagnostics. *John Wiley Sons*, New York.

[6] Cohen, A. (1966). All admissible linear estimates of the mean vector. *Ann. Math. Statist.*, 37, 458-463.

[7] Draper, N. and Van N. (1979). Ridge regression and James-Stein estimation: review and comments. *Technometrics*, 21, 451-466.

[8] Farebrother R. (1976). Further results on the mean square error of ridge regression. *J R Stat Soc B*, 38, 248-250.

[9] Fisher, R.A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, Vol. 85, No. 4, 597-612.

[10] Foong, N., Chin, L. and Hoe and Hoe, Q. (2007). An overview of biased estimators. *Journal of Physical Science*, Vol. 18. 2, 89-106.

[11] Frisch R. (1934). Statistical confluence analysis by means of complete regression systems. *Univeristy Institute of Economics, Oslo*, No. 5, 192.

[12] Galton, Francis (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, Vol. 15, 246-263.

[13] Galilei, Galileo (1632). Dialogue concerning the two chief world systems - ptolemaic and copernican. *University of California Press.*

[14] Gauss, Carl Friedrich (1809). Theoria combinationis observationum erroribus minimis obnoxiae.

[15] Gibbons, D. (1981). A simulation study of some ridge estimators. *Journal of American Statistical Association*, 76, 131-139.

[16] Groß, J. (2003). Restricted ridge estimation. *Statistics and Probability Letters*, 65, 57-64.

[17] Gui, Q. (1994). A class of multivariate principal components estimates. *Mathematical Statistics and Applied Probability*, Vol. 9. 3, 1-13.

[18] Hand, D., Daly, F., McConway, K., Lunn, D. (1994). A Handbook of Small Data Sets. *Chapman and Hall*, London.

[19] Hemmerle, W. and Brantle, T. (1978). Explicit and constrained generalized ridge estimation. *Technometrics*, Vol. 20, No, 2, 109-120.

[20] Hoerl, A. and Kennard, R. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67.

[21] Hoerl, A., Kennard, R. and Baldwin, K. (1975). Ridge regression: some simulations. *Communications in Statistics*, 4, 105-123.

[22] Hoerl, A. and Kennard, R. (1976). Ridge regression: iterative estimation of the biasing parameter. *Communications in Statistics: Theory and Methods*, 5, 77-88.

[23] Hubert, M. and Wijekoon, P. (2006). Improvement of the Liu estimator in linear regression model. *Statistical Papers*, 47, 471-479.

[24] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fouth Berkeley Symp. Math. Statist. and Prob.*, 1, 361-379.

[25] Kaçiranlar, S., Sakallioğlu, S., Akdeniz, F., Styan., G. P. H., Werner, H. J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on portland cement. *Sankhya, Ser. B, Indian J. Statist.*, 61: 443-459.

[26] Kaçiranlar, S., Sakallioğlu, S. (2001). Combining the Liu estimator and the principal component regression estimator. *Communications in Statistics: Theory and Methods*, 30, 2699-2706.

[27] Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics*, Vol. 22, No. 2, 393-402.

[28] Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics*, Vol. 32, No. 5, 1009-1020.

[29] Liu, K. (2004). More on Liu-type estimator in linear regression. *Communications in Statistics*, Vol. 33, No. 11, 2723-2733.

[30] McDonald, G., Galarneau, D. (1975). A monte Carlo Evaluation of some ridge-type estimators. *Journal of American Statistical Association*, 70, 407-416.

[31] Montgomery, D., Peck, E. and Vining G. (2006). Introduction to linear regression analysis. 4th ed., *Wiley*, New York.

[32] Özkale, M., and Kaçiranlar, S. (2007). A prediction-oriented criterion for choosing the biasing parameter in Liu estimation. *Communicaitons in Statistics*, 36, 1889-1903.

[33] Pearson, K., Yule, G.U., Blanchard, N., Lee,A. (1903). The law of ancestral heredity. *Biometrika*, Vol. 2, No. 2, 211-236.

[34] Rabe-Hesketh, S., Everitt, B. (2004). A handbook of statistical analyses using Stata. *Chapman and Hall/CRC*, Boca Raton, Florida.

[35] Rao, C.R. (1976). Estimation of parameters in a linear model. *The Annals of Statistics*, Vol. 4, No. 6, 1023-1037.

[36] Sakallioğlu, S. and Kaçiranlar, S. (2008). A new biased estimator based on ridge estimation. *Stat Papers*, 49, 669-689.

[37] Sokal, R., and Rohlf, F. (1981). Biometry. *W.H.Freeman*, New York.

[38] Stain, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*,1, 103-106.

[39] Trenkler, G. and Toutenbury, H. (1990). Mean squared error matrix comparisons between biased estimators - an overview of recent results. *Statistical Papers*, 31, 165-179

[40] Vinod, H.D. and Ullah A. (1981). Recent advances in regression methods. *Marcel Dekker*, New York.

[41] Wang, L. (1990). Generalized shrunken least squares estimators. *Chinese Journal of Applied Probability and Statistics*, Vol. 6. 3, 225-232.

[42] Wichern, D. and Churchill, G. (1978). A comparison of ridge estimators. *Technometrics*, Vol. 20, No. 3, 301-311.

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

[44] Yang H. and Xu, J. (2009). An alternative stochastic restricted Liu estimator in linear regression. *Stat Papers*, 50, 639-647.

[45] Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, Vol. 60, No. 4, 812-854.

[46] Zhao, Z. (1995). A note on the generalized shrunken least squares estimators. *Mathematica Applicata*, 8, 90-95.