On the Modified Almost Unbiased Ridge Estimator in Linear Regression Model

Enas Gawdat Yehia

Department of Statistics, Faculty of Commerce (Girls' Branch)
Al-Azhar University, Tafahna Al-Ashraf, Egypt.

Abstract

In order to overcome the negative effects caused by multicollinearity between the explanatory variables in the linear regression model, a new estimator namely modified almost unbiased ridge estimator is presented with its statistical characteristics in this paper. Also, the matrix mean squared error and squared bias criteria are adopted as a basis for comparisons between the new estimator and the ordinary least squares estimator, ridge estimator, and almost unbiased ridge estimator. Further, selection of the biasing parameter is discussed. Moreover, to check the performance of the new estimator versus the other estimators considered in this paper in the sense of scalar mean squared error, a study of Monte Carlo simulation and a real data example are conducted. The results indicate that in terms of scalar mean squared error, the new estimator, modified almost unbiased ridge estimator outperforms the others in use. So, it can be safely used when multicollinearity exists in a linear regression model.

***Keywords**: Multicollinearity, Ridge estimator, Almost unbiased ridge estimator, Matrix mean squared error, Monte Carlo simulation.*

1. Introduction

In linear regression model, the problem of multicollinearity occurs in the existence of linear dependencies between the explanatory variables. It is well known that *ordinary least squares* (OLS) estimation is the preferred method for estimating the parameters in linear regression model since it gives an unbiased estimator with minimum variance [Johnson and

Wichern (2007)]. However, when the problem of multicollinearity exists, the *ordinary least squares estimator* (OLSE) will be unstable with high variance [Vinod and Ullah (1981)].

To circumvent the multicollinearity problem in linear regression, many popular estimators of biased estimation methods have been introduced. These estimators include the Stein estimator by Stein (1956), the principal component estimator by Massy (1965), the ridge estimator by Hoerl and Kennard (1970), the Liu estimator by Liu (1993), and the Liu-type estimator by Liu (2003). Also, two versions of the two-parameter estimator by Özkale and Kaçiranlar (2007) and Yang and Chang (2010), the ridge-type estimator by Kibria and Lukman (2020), the modified one-parameter Liu estimator by Lukman *et al*. (2020), the generalized Kibria-Lukman estimator by Dawoud *et al*. (2022), and the new two-parameter estimator by Owolabi *et al*. (2022) were introduced for the linear regression model.

Another set of suggested estimators for dealing with multicollinearity are the almost unbiased estimators. For the linear regression model, the almost unbiased ridge estimator (AURE) by Singh *et al*. (1986), the almost unbiased Liu estimator by Alheety and Kibria (2009), the almost unbiased ridge-type principal component estimator and the almost unbiased Liu-type principal component estimator by Li and Yang (2014), the modified almost unbiased Liu estimator by Armairajan and Wijekoon (2017), and the almost unbiased Liu principal component estimator by Ahmed *et al*. (2021) were presented.

In this paper, a new estimator namely *modified almost unbiased ridge estimator* (MAURE) is proposed for overcoming the effects of multicollinearity in linear regression.

This paper is structured as follows. In Section 2, for linear regression model, the OLSE, RE, and AURE and their statistical characteristics are discussed. In Section 3, the new estimator, MAURE is presented. In Section 4, the superiority of the MAURE over the OLSE, RE, AURE based on the criteria of *matrix mean squared error* (MMSE) and *squared bias* (SB) is provided. In Section 5, selection of the biasing parameter is given. In Section 6, a study of Monte Carlo simulation is performed to compare the performance of the new estimator, MAURE with other considered estimators: OLSE, RE, and AURE in terms of *scalar mean squared error* (SMSE). Also, a real data example is included in Section 7. Finally, in Section 8, the conclusion is given.

2. Statistical Methodology

The linear regression model has the following standard form:

$$y = X\,\beta + \epsilon, \tag{1}$$

where y is a $n \times 1$ vector of response variable, X is a $n \times m$ full rank matrix of n observations on m explanatory variables, β is a $m \times 1$ vector of unknown regression coefficients, and ϵ is a $n \times 1$ vector of random error with mean vector $E(\epsilon) = 0$ and covariance matrix $Cov(\epsilon) = \sigma^2 I_n$, I_n is an $n \times n$ identity matrix.

By considering the OLS method for estimating the regression coefficients, the OLSE of β can be obtained as follows:

$$\hat{\beta}_{OLSE} = V^{-1}\,X'y, \tag{2}$$

where $V = X'X$.

The $\hat{\beta}_{OLSE}$ is an unbiased estimator and its covariance matrix is given as

$$Cov(\hat{\beta}_{OLSE}) = \sigma^2 V^{-1}. \tag{3}$$

The MMSE of $\hat{\beta}_{OLSE}$ is as follows:

$$\begin{aligned} MMSE(\hat{\beta}_{OLSE}) &= Cov(\hat{\beta}_{OLSE}) + B(\hat{\beta}_{OLSE})B'(\hat{\beta}_{OLSE}) \\ &= \sigma^2 V^{-1}, \end{aligned} \tag{4}$$

where $B(.)$denotes the bias vector.

Also, the SMSE of $\hat{\beta}_{OLSE}$ is given as follows:

$$SMSE(\hat{\beta}_{OLSE}) = tr[MMSE(\hat{\beta}_{OLSE})] \quad (5)$$

$$= \sigma^2 \sum_{j=1}^{m} \frac{1}{\eta_j},$$

where η_j are the eigenvalues of V.

When the model (1) is suffering from multicollinearity because of correlated explanatory variables, the OLSE becomes biased and has high variance, which leads to unstable parameters estimates.

For tackling the effect of multicollinearity in linear regression model, Hoerl and Kennard (1970) introduced the RE which is defined as follows:

$$\hat{\beta}_{RE} = (V + k\,I)^{-1} V\, \hat{\beta}_{OLSE} \quad (6)$$

$$= M_k\, \hat{\beta}_{OLSE},$$

where $M_k = (V + k\,I)^{-1} V$, and k is the biasing parameter called the ridge parameter, $k > 0$.

The following statistical characteristics belong to the RE:

$$E(\hat{\beta}_{RE}) = M_k\, \beta, \quad (7)$$

$$B(\hat{\beta}_{RE}) = (M_k - I)\beta \quad (8)$$

$$= -k(V + k\,I)^{-1}\beta$$

$$= \zeta_1, \text{(say)}$$

$$Cov(\hat{\beta}_{RE}) = \sigma^2 M_k V^{-1} M_k', \quad (9)$$

$$MMSE(\hat{\beta}_{RE}) = Cov(\hat{\beta}_{RE}) + B(\hat{\beta}_{RE}) B'(\hat{\beta}_{RE}) \quad (10)$$

$$= \sigma^2 M_k V^{-1} M_k' + (M_k - I)\beta\beta'(M_k - I)',$$

and

$$SMSE(\hat{\beta}_{RE}) = tr[MMSE(\hat{\beta}_{RE})] \quad (11)$$

$$= \sigma^2 \sum_{j=1}^{m} \frac{\eta_j}{(k+\eta_j)^2} + \sum_{j=1}^{m} \frac{k^2\alpha_j^2}{(k+\eta_j)^2} \ ,$$

where α_j is the jth element of $Q'\hat{\beta}_{OLSE}$ and Q is an orthogonal matrix defined as $QHQ' = V, H = \text{diag}(\eta_1, \eta_2, \dots, \eta_m)$.

Based on the following definition, the AURE is proposed by Singh *et al*. (1986) in linear regression model.

Definition 1. Assume that $\hat{\beta}^*$ is a biased estimator of β and $B(\hat{\beta}^*) = E(\hat{\beta}^*) - \beta = A\beta$ is the bias vector of $\hat{\beta}^*$. Then, the almost unbiased estimator of β is $\hat{\beta} = \hat{\beta}^* - A\hat{\beta}^* = (I - A)\hat{\beta}^*$. [Kadiyala (1984)]

The AURE is defined as follows:

$$\hat{\beta}_{AURE} = G_k \, \hat{\beta}_{OLSE}, \tag{12}$$

where $G_k = I - k^2(V + k\, I)^{-2}$.

The statistical characteristics of the AURE are given as follows:

$$E(\hat{\beta}_{AURE}) = G_k \, \beta, \tag{13}$$

$$\begin{aligned} B(\hat{\beta}_{AURE}) &= (G_k - I)\beta \\ &= -k^2(V + k\, I)^{-2}\beta \\ &= \zeta_2, \text{(say)} \end{aligned} \tag{14}$$

$$Cov(\hat{\beta}_{AURE}) = \sigma^2 G_k V^{-1} G_k', \tag{15}$$

$$\begin{aligned} MMSE(\hat{\beta}_{AURE}) &= Cov(\hat{\beta}_{AURE}) + B(\hat{\beta}_{AURE})B'(\hat{\beta}_{AURE}) \\ &= \sigma^2 G_k V^{-1} G_k' + (G_k - I)\beta\beta'(G_k - I)', \end{aligned} \tag{16}$$

and

$$SMSE(\hat{\beta}_{AURE}) = tr[MMSE(\hat{\beta}_{AURE})] \tag{17}$$

$$= \sigma^2 \sum_{j=1}^{m} \frac{(\eta_j^2 + 2k\eta_j)^2}{(k+\eta_j)^4 \eta_j} + \sum_{j=1}^{m} \frac{k^4 \alpha_j^2}{(k+\eta_j)^4}.$$

3. The New Estimator

In this section, a new almost unbiased estimator, MAURE is proposed based on the RE and AURE as follows:

$$\hat{\beta}_{MAURE} = G_k \, \hat{\beta}_{RE} \tag{18}$$

$$= G_k M_k \, \hat{\beta}_{OLSE}.$$

The new estimator, MAURE has the following characteristics:

$$E(\hat{\beta}_{MAURE}) = G_k M_k \, \beta, \tag{19}$$

$$B(\hat{\beta}_{MAURE}) = (G_k M_k - I)\beta \tag{20}$$

$$= -k[(V + k\,I)^2 + kV](V + k\,I)^{-3}\beta$$

$$= \zeta_3, \text{ (say)}$$

$$Cov(\hat{\beta}_{MAURE}) = \sigma^2 G_k M_k V^{-1} G_k' M_k', \tag{21}$$

$$MMSE(\hat{\beta}_{MAURE}) = Cov(\hat{\beta}_{MAURE}) + B(\hat{\beta}_{MAURE}) B'(\hat{\beta}_{MAURE}) \tag{22}$$

$$= \sigma^2 G_k M_k V^{-1} G_k' M_k' + (G_k M_k - I)\beta\beta'(G_k M_k - I)',$$

and

$$SMSE(\hat{\beta}_{MAURE}) = tr[MMSE(\hat{\beta}_{MAURE})] \tag{23}$$

$$= \sigma^2 \sum_{j=1}^{m} \frac{[(k+\eta_j)^2 - k^2]^2 \eta_j}{(k+\eta_j)^6} + \sum_{j=1}^{m} [\frac{[(k+\eta_j)^2 - k^2]\eta_j}{(k+\eta_j)^3} - 1]^2 \alpha_j^2 .$$

4. Superiority of the MAURE

Based on the MMSE and SB, the following comparisons are performed.

4.1 MMSE comparisons

When the comparison between any two estimators $\hat{\beta}_1$and $\hat{\beta}_2$ of β is performed by the criterion of MMSE, the following Lemmas can be used:

Lemma 1. *Suppose that F and $D$are $n \times n$ matrices such that $F > 0$, and $D \geq 0$. Then, $F > D$ if and only if $\eta_{max}(DF^{-1}) < 1$.* [Rao *et al*. (2008)]

Lemma 2. *For two estimators $\hat{\beta}_1$and $\hat{\beta}_2$ of β, suppose that $S = Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ is positive definite. Then, $MMSE(\hat{\beta}_1) - MMSE(\hat{\beta}_2)$ is non-negative definite if and only if $B_2'[S + B_1B_1']^{-1}B_2 \leq 1$, where B_1 and B_2 denote the bias vectors of $\hat{\beta}_1$and $\hat{\beta}_2$ respectively.* [Trenkler and Toutenburg (1990)]

The following comparisons are performed between the new estimator, MAURE and the OLSE, RE, and AURE by the MMSE criterion.

4.1.1 The MMSE comparison between the OLSE and MAURE

Using (4) and (22), the MMSE difference of the OLSE and MAURE is given by

$$
\begin{aligned}
&MMSE(\hat{\beta}_{OLSE}) - MMSE(\hat{\beta}_{MAURE}) \\
&\quad = \sigma^2[V^{-1} - G_kM_kV^{-1}G_k'M_k'] - (G_kM_k - I)\beta\beta'(G_kM_k - I)' \qquad (24) \\
&\quad = \sigma^2A_1 - (G_kM_k - I)\beta\beta'(G_kM_k - I)' \\
&\quad = \sigma^2A_1 - \zeta_3\zeta_3',
\end{aligned}
$$

where $A_1 = V^{-1} - G_kM_kV^{-1}G_k'M_k'$.

Then, according to Lemma 2, Theorem 1 is stated as follows:

Theorem 1. *When $\eta_{max}(G_kM_kV^{-1}G_k'M_k') < 1$, the MAURE is superior to OLSE based on the MMSE criterion if and only if $\zeta_3'[\sigma^2A_1]^{-1}\zeta_3 \leq 1$.*

Proof: Since V^{-1} and $G_kM_kV^{-1}G_k'M_k'$ are positive definite matrices, then, $A_1 = V^{-1} - G_kM_kV^{-1}G_k'M_k' > 0$ according to Lemma 1 if and only if

$\eta_{max}(G_k M_k V^{-1} G_k' M_k') < 1$. Consequently, from Lemma 2, $MMSE(\hat{\beta}_{OLSE}) - MMSE(\hat{\beta}_{MAURE})$ is a non-negative definite matrix if and only if $\beta'(G_k M_k - I)'[\sigma^2 A_1]^{-1}(G_k M_k - I)\beta \leq 1$.

4.1.2 The MMSE comparison between the RE and MAURE

Using (10) and (22), the MMSE difference of the RE and MAURE is as follows:

$$\begin{aligned} MMSE(\hat{\beta}_{RE}) - MMSE(\hat{\beta}_{MAURE}) & \\ &= \sigma^2[M_k V^{-1} M_k' - G_k M_k V^{-1} G_k' M_k'] + (M_k - I)\beta\beta'(M_k - I)' \\ &-(G_k M_k - I)\beta\beta'(G_k M_k - I)' \qquad (25) \\ &= \sigma^2[F_1 - D_1] + \zeta_1\zeta_1' - \zeta_3\zeta_3' \end{aligned}$$

where $F_1 = M_k V^{-1} M_k'$, and $D_1 = G_k M_k V^{-1} G_k' M_k'$.

Then, Theorem 2 is stated as follows:

Theorem 2. *When $\eta_{max}(D_1 F_1^{-1}) < 1$, the MAURE is superior to RE based on the MMSE criterion if and only if* $\zeta_3'[\sigma^2(F_1 - D_1) + \zeta_1\zeta_1']^{-1}\zeta_3 \leq 1$.

Proof: Since F_1 and D_1 are positive definite matrices, then, $F_1 - D_1 > 0$ according to Lemma 1 if and only if $\eta_{max}(D_1 F_1^{-1}) < 1$. Consequently, from Lemma 2, $MMSE(\hat{\beta}_{RE}) - MMSE(\hat{\beta}_{MAURE})$ is a non-negative definite matrix if and only if

$$\beta'(G_k M_k - I)'[\sigma^2(F_1 - D_1) + (M_k - I)\beta\beta'(M_k - I)']^{-1}(G_k M_k - I)\beta \leq 1.$$

4.1.3 The MMSE comparison between the AURE and MAURE

Using (16) and (22), the MMSE difference of the AURE and MAURE is given as follows:

$$\begin{aligned} MMSE(\hat{\beta}_{AURE}) - MMSE(\hat{\beta}_{MAURE}) & \\ &= \sigma^2[G_k V^{-1} G_k' - G_k M_k V^{-1} G_k' M_k'] + (G_k - I)\beta\beta'(G_k - I)' \end{aligned}$$

$$-(G_kM_k - I)\beta\beta'(G_kM_k - I)' \tag{26}$$

$$= \sigma^2[F_2 - D_1] + \zeta_2\zeta_2' - \zeta_3\zeta_3',$$

where $F_2 = G_kV^{-1}G_k'$.

Then, Theorem 3 is stated as follows:

Theorem 3. *When* $\eta_{max}(D_1F_2^{-1}) < 1$*, the MAURE is superior to AURE based on the MMSE criterion if and only if* $\zeta_3'[\sigma^2(F_2 - D_1) + \zeta_2\zeta_2']^{-1}\zeta_3 \leq 1$.

Proof: Since F_2 and D_1 are positive definite matrices, then, from Lemma 1, $F_2 - D_1 > 0$ if and only if $\eta_{max}(D_1F_2^{-1}) < 1$. Consequently, by Lemma 2, $MMSE(\hat{\beta}_{AURE}) - MMSE(\hat{\beta}_{MAURE})$ is a non-negative definite matrix if and only if

$$\beta'(G_kM_k - I)'[\sigma^2(F_2 - D_1) + (G_k - I)\beta\beta'(G_k - I)']^{-1}(G_kM_k - I)\beta \leq 1.$$

4.2 Squared bias comparisons

Based on the SB criterion, the following comparisons are discussed between the MAURE and the RE and AURE.

4.2.1 The SB comparison between the RE and MAURE

From (8) and (20), the difference of SB between the RE and MAURE is given as follows:

$$\left\|B(\hat{\beta}_{RE})\right\|^2 - \left\|B(\hat{\beta}_{MAURE})\right\|^2 = \beta'k^2(V + kI)^{-2}\beta$$
$$-\beta'k^2(V + kI)^{-6}[(V + kI)^2 + kV]^2\beta. \tag{27}$$

Then, Theorem 4 is given as follows:

Theorem 4. *Under SB criterion,* $\left\|B(\hat{\beta}_{RE})\right\|^2 - \left\|B(\hat{\beta}_{MAURE})\right\|^2 > 0$ *for* $k > 0$.

Proof: Since $\alpha = Q'\beta$, the difference of SB between the RE and MAURE is

$$\left\|B\left(\hat{\beta}_{RE}\right)\right\|^2 - \left\|B\left(\hat{\beta}_{MAURE}\right)\right\|^2 = \alpha' k^2 (H + kI)^{-2} \alpha$$

$$-\alpha' k^2 (H + kI)^{-6} [(H + kI)^2 + kH]^2 \alpha$$

$$= \alpha' R_1 \alpha,$$

where $R_1 = k^2[(H + kI)^{-2} - (H + kI)^{-6}[(H + kI)^2 + kH]^2]$.

Therefore, for $k > 0$, $\alpha' R_1 \alpha > 0$.

4.2.2 The SB comparison between the AURE and MAURE

From (14) and (20), the difference of SB between the AURE and MAURE is given by

$$\left\|B\left(\hat{\beta}_{AURE}\right)\right\|^2 - \left\|B\left(\hat{\beta}_{MAURE}\right)\right\|^2 = \beta' k^4 (V + kI)^{-4} \beta$$

$$-\beta' k^2 (V + kI)^{-6} [(V + kI)^2 + kV]^2 \beta. \qquad (28)$$

Then, Theorem 5 is stated as follows:

Theorem 5. *Under SB criterion,* $\left\|B\left(\hat{\beta}_{AURE}\right)\right\|^2 - \left\|B\left(\hat{\beta}_{MAURE}\right)\right\|^2 > 0$ *for* $k > 0$.

Proof: Since $\alpha = Q'\beta$, the difference of SB between the AURE and MAURE is

$$\left\|B\left(\hat{\beta}_{AURE}\right)\right\|^2 - \left\|B\left(\hat{\beta}_{MAURE}\right)\right\|^2 = \alpha' k^4 (H + kI)^{-4} \alpha$$

$$-\alpha' k^2 (H + kI)^{-6} [(H + kI)^2 + kH]^2 \alpha$$

$$= \alpha' R_2 \alpha,$$

where $R_2 = k^2[k^2(H + kI)^{-4} - (H + kI)^{-6}[(H + kI)^2 + kH]^2]$.

Therefore, for $k > 0$, $\alpha' R_2 \alpha > 0$.

5. Selection of k Biasing Parameter

For determining the biasing parameter k, many methods have been proposed. Some of the more popular of these methods are considered as follows:

Hoerl and Kennard (1970) suggested the following estimator of k:

$$k_1 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{max}^2}, \tag{29}$$

where $\hat{\sigma}^2 = \frac{1}{n-m}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, and $\hat{\alpha}_{max}^2$ is the maximum value of α_j^2.

Hoerl *et al*. (1975) suggested an alternative estimator as follows:

$$k_2 = \frac{m\,\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}. \tag{30}$$

In Khalaf and Shukur (2005), the following estimator was proposed:

$$k_3 = \frac{\hat{\sigma}^2\,\eta_{max}}{\hat{\sigma}^2(n-m)+\eta_{max}\hat{\alpha}_{max}^2}. \tag{31}$$

In addition, Alkhamisi *et al*. (2006) proposed an estimator of k as follows:

$$k_4 = median\left[\frac{\hat{\sigma}^2\,\eta_j}{\hat{\sigma}^2(n-m)+\eta_j\hat{\alpha}_j^2}\right], j = 1,2,\dots,m \tag{32}$$

Further, Muniz and Kibria (2009) suggested the following estimator:

$$k_5 = \left[\prod_{j=1}^{m}\left[\frac{\hat{\sigma}^2\,\eta_j}{\hat{\sigma}^2(n-m)+\eta_j\hat{\alpha}_j^2}\right]\right]^{\frac{1}{m}}. \tag{33}$$

6. Monte Carlo Simulation Study

A Monte Carlo simulation study is conducted using the R 4. 0. 3 programme to evaluate the performance of the new estimator, MAURE in the linear regression model against the OLSE, RE, and AURE in the sense of SMSE.

The explanatory variables are generated following McDonald and Galarneau (1975) as follows:

$$x_{ij} = [1-\rho^2]^{1/2} u_{ij} + \rho u_{im}, i = 1,2,\dots,n\,, j = 1,2,\dots,m, \qquad (34)$$

where ρ^2 represents the correlation degree between any two explanatory variables, and $u_{ij} \sim N(0,1)$ is the independent pseudo-random variable.

The response variable is obtained as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im} + \epsilon_i, i = 1,2,\dots,n, \qquad (35)$$

where $\epsilon_i \sim N(0, \sigma^2 I_n)$, and the coefficient β_0 is chosen so that $\beta_0 = 0$, and $\sum_{j=1}^{m} \beta_j^2 = 1$, where $\beta_1 = \beta_2 = \cdots = \beta_m$ following Kibria (2003).

For $m = 5$ and 9 explanatory variables, different values of $\rho = 0.80, 0.90$, and 0.99, different sample sizes $n = 50{,}100$, and 200, and different values of error variance $\sigma^2 = 1{,}3$, and 5 are considered.

For a combination of the values of m, ρ, n, and σ^2, the generated data are repeated 1000 times and the SMSE is computed as follows:

$$SMSE(\hat{\beta}) = \frac{\sum_{r=1}^{1000} (\hat{\beta}_r - \beta)' (\hat{\beta}_r - \beta)}{1000}, \qquad (36)$$

where $\hat{\beta}_r$ is the estimated value of β by any estimator in the *r*th replication.

The estimated SMSE values for all estimators with $k_1 - k_5$ and the combination of n, σ^2, m, and ρ respectively are given in Tables 1-9.

As Tables 1-9 show, the values of the estimated SMSE of all estimators: OLSE, RE, AURE, and MAURE increase as ρ increases. Also, regarding σ^2, it is clear that there is an increase in the estimated values of SMSE for all estimators when σ^2 increases. Regarding the number of explanatory variables m, there is an increase in the estimated SMSE of all estimators as m increases. While, regarding the sample size n, the estimated values of SMSE decrease for all estimators as n increases. Additionally, the new estimator, MAURE has the best performance among the OLSE, RE, and AURE in all cases in the sense of SMSE, and the RE has better performance than the AURE. Furthermore, for different selection formulas of k estimators for RE, AURE, and MAURE, the k_2 outperforms the other estimators as it has the lowest SMSE values. Moreover, the MAURE with k_2 achieves the best performance compared to the RE, AURE, and OLSE in terms of SMSE.

Table 1. Estimated SMSE values when $n = 50$ and $\sigma^2 = 1$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	0.351048	0.673147	7.570923	0.851854	1.640154	15.913152
k_1	RE	0.277766	0.455258	4.146347	0.595955	1.047394	9.065484
	AURE	0.339372	0.615007	6.137674	0.779581	1.423234	13.005796
	MAURE	0.269706	0.424826	3.684082	0.559983	0.967068	8.167809
k_2	RE	0.188942	0.283050	2.038223	0.296838	0.471601	3.270827
	AURE	0.288936	0.478880	3.958932	0.507166	0.848831	6.384525
	MAURE	0.164448	0.228192	1.427405	0.235147	0.356344	2.260312
k_3	RE	0.344251	0.641357	5.574217	0.797512	1.420743	10.274806
	AURE	0.350962	0.672130	7.156683	0.849215	1.616724	13.983278
	MAURE	0.344167	0.640408	5.322501	0.795135	1.402247	9.500092
k_4	RE	0.350135	0.671090	7.540886	0.846091	1.626873	15.763641
	AURE	0.351046	0.673143	7.570847	0.851825	1.640077	15.912171
	MAURE	0.350133	0.671086	7.540811	0.846062	1.626796	15.762673
k_5	RE	0.349974	0.670427	7.512269	0.846179	1.625615	15.695905
	AURE	0.351046	0.673140	7.570631	0.851826	1.640062	15.911075
	MAURE	0.349971	0.670420	7.511980	0.846151	1.625524	15.693869

Table 2. Estimated SMSE values when $n = 50$ and $\sigma^2 = 3$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	3.039852	6.330734	69.122354	7.482948	15.439034	145.920237
k_1	RE	1.720351	3.483515	37.395874	4.364396	8.896725	82.888974
	AURE	2.522118	5.152129	55.578869	6.218038	12.736390	118.704602
	MAURE	1.536798	3.096136	33.137072	3.951658	8.026685	74.689508
k_2	RE	0.966706	1.793912	17.340391	1.791992	3.293641	29.347271
	AURE	1.781511	3.424272	34.179466	3.419519	6.399360	57.438825
	MAURE	0.715599	1.279011	11.920441	1.275172	2.291857	20.231654
k_3	RE	2.979897	6.020608	50.722076	6.999737	13.269624	93.981254
	AURE	3.039070	6.320577	65.247012	7.458809	15.199968	127.847661
	MAURE	2.979137	6.011147	48.382537	6.978083	13.082144	86.839344
k_4	RE	3.032093	6.311004	68.852863	7.433542	15.311411	144.560133
	AURE	3.039839	6.330694	69.121672	7.482704	15.438289	145.911325
	MAURE	3.032080	6.310964	68.852185	7.433301	15.310676	144.551342
k_5	RE	3.030598	6.304222	68.586192	7.433669	15.297053	143.905224
	AURE	3.039834	6.330662	69.119650	7.482705	15.438111	145.900621
	MAURE	3.030580	6.304150	68.583517	7.433429	15.296143	143.886001

Table 3. Estimated SMSE values when $n = 50$ and $\sigma^2 = 5$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	8.630817	17.899826	194.851426	21.288998	43.234012	400.941836
k_1	RE	4.783869	9.845366	107.513769	12.402166	25.129168	227.590711
	AURE	7.086631	14.542416	158.262542	17.651156	35.879476	326.970650
	MAURE	4.255089	8.756691	95.814029	11.230123	22.705321	204.764437
k_2	RE	2.463440	4.743140	49.689387	4.754952	9.215763	79.683001
	AURE	4.712497	9.223908	97.519096	9.287703	17.956712	155.978852
	MAURE	1.749650	3.310729	34.336293	3.290024	6.391755	54.896390
k_3	RE	8.460256	17.006468	143.538171	19.894838	37.265187	257.908949
	AURE	8.628584	17.870437	184.244358	21.219670	42.581609	351.607296
	MAURE	8.458086	16.979093	137.079943	19.832607	36.753015	238.120529
k_4	RE	8.608849	17.842800	194.072116	21.144586	42.882064	397.214581
	AURE	8.630780	17.899709	194.849440	21.288282	43.231966	400.917430
	MAURE	8.608813	17.842684	194.070142	21.143876	42.880042	397.190504
k_5	RE	8.604589	17.823086	193.298586	21.144802	42.841971	395.410324
	AURE	8.630765	17.899614	194.843527	21.288284	43.231472	400.887925
	MAURE	8.604537	17.822876	193.290772	21.144095	42.839465	395.357493

Table 4. Estimated SMSE values when $n = 100$ and $\sigma^2 = 1$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	0.197499	0.421659	4.389635	0.533123	1.151260	11.165997
k_1	RE	0.173133	0.325312	2.429037	0.422386	0.801352	6.437291
	AURE	0.195383	0.405501	3.607418	0.512868	1.051568	9.253197
	MAURE	0.171399	0.314325	2.157773	0.409258	0.753415	5.806203
k_2	RE	0.128588	0.216762	1.309512	0.235296	0.399197	2.566514
	AURE	0.179118	0.342233	2.504849	0.380990	0.699534	5.070289
	MAURE	0.118753	0.185593	0.936808	0.196278	0.312340	1.760954
k_3	RE	0.196425	0.415713	3.862609	0.522184	1.094061	8.457010
	AURE	0.197495	0.421606	4.345748	0.532947	1.149028	10.570899
	MAURE	0.196421	0.415661	3.826253	0.522014	1.092013	8.113700
k_4	RE	0.197391	0.421378	4.386405	0.532663	1.150074	11.152656
	AURE	0.197499	0.421659	4.389633	0.533123	1.151259	11.165984
	MAURE	0.197390	0.421377	4.386404	0.532662	1.150073	11.152643
k_5	RE	0.197375	0.421297	4.383190	0.532581	1.149772	11.145165
	AURE	0.197499	0.421659	4.389628	0.533122	1.151259	11.165967
	MAURE	0.197374	0.421296	4.383184	0.532580	1.149771	11.145135

Table 5. Estimated SMSE values when $n = 100$ and $\sigma^2 = 3$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	1.836521	3.721732	41.420301	4.697253	9.744537	106.245569
k_1	RE	1.088320	2.059638	23.262280	2.799464	5.690927	61.708367
	AURE	1.582763	3.073866	34.162841	3.991788	8.142378	88.450082
	MAURE	0.980847	1.825556	20.794584	2.541842	5.146614	55.734592
k_2	RE	0.644187	1.126178	11.224147	1.236926	2.391282	23.732535
	AURE	1.178257	2.150797	21.976346	2.386128	4.667407	46.945341
	MAURE	0.482374	0.805348	7.795781	0.875519	1.667799	16.266963
k_3	RE	1.826005	3.669841	36.389417	4.598294	9.265204	80.357800
	AURE	1.836483	3.721261	41.003378	4.695617	9.725614	100.557588
	MAURE	1.825967	3.669380	36.043237	4.596718	9.247873	77.064373
k_4	RE	1.835496	3.719337	41.388529	4.693235	9.734803	106.118039
	AURE	1.836521	3.721730	41.420286	4.697250	9.744529	106.245452
	MAURE	1.835495	3.719336	41.388514	4.693232	9.734795	106.117922
k_5	RE	1.835334	3.718626	41.356306	4.692484	9.732228	106.044581
	AURE	1.836520	3.721730	41.420238	4.697249	9.744524	106.245278
	MAURE	1.835333	3.718624	41.356244	4.692480	9.732216	106.044291

Table 6. Estimated SMSE values when $n = 100$ and $\sigma^2 = 5$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	4.961537	10.363294	104.835006	13.043441	27.648155	284.209160
k_1	RE	2.755104	5.707858	56.151382	7.647085	16.322096	164.782091
	AURE	4.109348	8.492974	84.567274	10.950440	23.168192	235.838522
	MAURE	2.443945	5.067522	49.515441	6.921420	14.817160	148.951511
k_2	RE	1.517399	2.828932	25.404208	3.184078	6.514651	61.970643
	AURE	2.887524	5.569688	51.575793	6.276990	12.873790	124.164654
	MAURE	1.087532	1.949507	16.826291	2.195238	4.469673	41.748776
k_3	RE	4.934403	10.214606	91.781732	12.770174	26.263339	215.154084
	AURE	4.961440	10.361928	103.725956	13.038933	27.593654	269.020124
	MAURE	4.934307	10.213267	90.864164	12.765828	26.213394	206.403098
k_4	RE	4.958895	10.356420	104.755305	13.032356	27.619745	283.863684
	AURE	4.961536	10.363291	104.834967	13.043434	27.648132	284.208842
	MAURE	4.958894	10.356417	104.755266	13.032349	27.619723	283.863366
k_5	RE	4.958476	10.354372	104.674354	13.030276	27.612206	283.664253
	AURE	4.961535	10.363289	104.834848	13.043431	27.648119	284.208368
	MAURE	4.958475	10.354368	104.674197	13.030266	27.612170	283.663463

Table 7. Estimated SMSE values when $n = 200$ and $\sigma^2 = 1$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	0.098297	0.189498	1.820089	0.200284	0.419871	4.255156
k_1	RE	0.091767	0.164256	1.028742	0.179225	0.333126	2.445291
	AURE	0.097999	0.187089	1.523711	0.198416	0.404147	3.499996
	MAURE	0.091498	0.162308	0.916877	0.177694	0.322692	2.210073
k_2	RE	0.075470	0.120889	0.575493	0.116733	0.185585	1.038406
	AURE	0.094398	0.170006	1.080374	0.167233	0.294187	1.969621
	MAURE	0.072824	0.110878	0.420883	0.104561	0.155906	0.750040
k_3	RE	0.098165	0.188863	1.751707	0.199511	0.415245	3.802195
	AURE	0.098297	0.189496	1.818359	0.200282	0.419832	4.215825
	MAURE	0.098164	0.188861	1.750070	0.199509	0.415207	3.769677
k_4	RE	0.098284	0.189470	1.819776	0.200240	0.419749	4.253699
	AURE	0.098297	0.189498	1.820089	0.200284	0.419871	4.255155
	MAURE	0.098283	0.189469	1.819775	0.200237	0.419748	4.253698
k_5	RE	0.098281	0.189460	1.819392	0.200234	0.419728	4.252935
	AURE	0.098297	0.189498	1.820088	0.200284	0.419871	4.255155
	MAURE	0.098280	0.189457	1.819390	0.200233	0.419726	4.252934

Table 8. Estimated SMSE values when $n = 200$ and $\sigma^2 = 3$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	0.861388	1.657872	17.218996	1.786755	3.707291	40.984078
k_1	RE	0.586259	0.976489	9.348854	1.153346	2.190208	23.693803
	AURE	0.790002	1.419442	13.937248	1.572577	3.113270	33.835942
	MAURE	0.548799	0.879685	8.285206	1.067440	1.988487	21.429132
k_2	RE	0.369010	0.564186	4.484363	0.550239	0.943431	8.734934
	AURE	0.625578	1.037606	8.880905	1.002341	1.795227	17.309635
	MAURE	0.297791	0.421426	3.073438	0.415131	0.678498	5.958965
k_3	RE	0.860214	1.652257	16.579394	1.779666	3.665643	36.538292
	AURE	0.861387	1.657860	17.202946	1.786733	3.706934	40.604536
	MAURE	0.860213	1.652244	16.564196	1.779645	3.665293	36.223355
k_4	RE	0.861274	1.657624	17.216060	1.786334	3.706212	40.969344
	AURE	0.861388	1.657872	17.218996	1.786755	3.707290	40.984074
	MAURE	0.861273	1.657622	17.216059	1.786333	3.706211	40.969340
k_5	RE	0.861247	1.657517	17.212426	1.786301	3.706007	40.961513
	AURE	0.861388	1.657872	17.218994	1.786755	3.707290	40.984068
	MAURE	0.861246	1.657516	17.212424	1.786300	3.706006	40.961504

Table 9. Estimated SMSE values when $n = 200$ and $\sigma^2 = 5$

k	Estimator	$m = 5$			$m = 9$		
		ρ			ρ		
		0.80	0.90	0.99	0.80	0.90	0.99
	OLSE	2.380186	4.431373	49.868205	4.962694	11.010611	107.951265
k_1	RE	1.370403	2.367388	27.488886	2.988804	6.480510	61.518124
	AURE	2.015829	3.570195	40.706469	4.215498	9.230612	88.625519
	MAURE	1.226222	2.082414	24.443784	2.727965	5.875739	55.329943
k_2	RE	0.764650	1.245046	13.157329	1.312507	2.564345	22.921840
	AURE	1.434322	2.419482	25.807951	2.522152	5.011295	45.083946
	MAURE	0.556001	0.872252	9.116656	0.932153	1.781586	15.811685
k_3	RE	2.376862	4.416384	48.002718	4.943094	10.883740	96.471783
	AURE	2.380184	4.431340	49.821361	4.962634	11.009521	106.960320
	MAURE	2.376859	4.416351	47.958358	4.943033	10.882671	95.651511
k_4	RE	2.379865	4.430710	49.859615	4.961528	11.007315	107.914116
	AURE	2.380186	4.431373	49.868204	4.962694	11.010610	107.951255
	MAURE	2.379864	4.430709	49.859614	4.961527	11.007314	107.914106
k_5	RE	2.379788	4.430425	49.848977	4.961438	11.006688	107.894352
	AURE	2.380186	4.431373	49.868200	4.962694	11.010610	107.951242
	MAURE	2.379787	4.430424	49.848972	4.961437	11.006687	107.894330

7. Real Data Example

In this section, a real data set of Total National Research and Development Expenditures is considered as a percent of Gross National Product by Country: 1972-1986 according to Gruber (1998) and later analyzed by Li and Yang (2011) and Arumairajan and Wijekoon (2017). This data set involves 10 observations. The response variable y as well as the four explanatory variables x_1, x_2, x_3, and x_4 are defined as follows: y is the percentage spent by the United States, x_1 is the percent spent by France, x_2 is the percent spent by West Germany, x_3 is the percent spent by Japan, and x_4 is the percent spent by the former Soviet Union.

The statistic value of the Shapiro-Wilk normality test equals to 0.91333 with $p - value = 0.3047$, which indicates the normality of the response variable y at 5% significance level.

The correlation matrix of the explanatory variables is as follows:

$$\begin{bmatrix} 1 & 0.89 & 0.92 & 0.31 \\ 0.89 & 1 & 0.96 & 0.16 \\ 0.92 & 0.96 & 1 & 0.33 \\ 0.31 & 0.16 & 0.33 & 1 \end{bmatrix}$$

It is obvious that there are correlations greater than 0.80 between $x_1 and\ x_2$, $x_1 and\ x_3$, and $x_2 and\ x_3$ which indicates the existence of high relationship between the explanatory variables.

Also, for checking the presence of multicollinearity, the condition number (CN) of the data is computed by

$$CN = (\max(\eta_j)/\min(\eta_j))^{1/2}, j = 1,2,\dots,m, \tag{37}$$

where $\max(\eta_j)$ and $\min(\eta_j)$ are the largest and smallest eigenvalues of $X'X$ respectively.
Since the eigenvalues of $X'X$ matrix are obtained as $\eta_1 = 302.962606$, $\eta_2 = 0.728305$, $\eta_3 = 0.044569$, and $\eta_4 = 0.034520$, the value of CN is 93.682340 shows the presence of severe multicollinearity in this data.

The estimated coefficients and SMSE of the OLSE, RE, AURE, and MAURE for $k_1 - k_5$ are given in Table 10.

Table 10. Estimated coefficients and SMSE of the estimators

Estimator		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	SMSE
OLSE		0.64546	0.08959	0.14356	0.15262	0.08079
RE	k_1	0.58264	0.10653	0.16818	0.15972	0.06280
	k_2	0.54318	0.11792	0.18291	0.16415	0.05885
	k_3	0.58265	0.10653	0.16818	0.15972	0.06280
	k_4	0.58814	0.10498	0.16608	0.15910	0.06380
	k_5	0.57288	0.10930	0.17188	0.16082	0.06130
AURE	k_1	0.63641	0.08736	0.14075	0.15230	0.07698
	k_2	0.62154	0.08378	0.13640	0.15180	0.07162
	k_3	0.63642	0.08736	0.14075	0.15230	0.07699
	k_4	0.63793	0.08773	0.14121	0.15235	0.07759
	k_5	0.63339	0.08663	0.13985	0.15219	0.07581
MAURE	k_1	0.57436	0.10401	0.16535	0.15939	0.06133
	k_2	0.52256	0.11086	0.17564	0.16328	0.05845
	k_3	0.57436	0.10401	0.16535	0.15939	0.06133
	k_4	0.58119	0.10290	0.16371	0.15882	0.06240
	k_5	0.56199	0.10588	0.16813	0.16037	0.05981

Table 10 shows that the new estimator, MAURE has the smallest SMSE values than other estimators for all values of k.

8. Conclusion

In this paper, for overcoming multicollinearity in linear regression model, a new estimator, MAURE was presented with its statistical characteristics. By considering the criteria of MMSE and SB, the comparisons between the new estimator, MAURE and the OLSE, RE, and AURE were provided. Further, a study of Monte Carlo simulation and a real data example were conducted to evaluate the performance of the MAURE versus the other existing estimators under the SMSE criterion. The results showed the superiority of the new estimator, MAURE over all existing estimators in terms of SMSE. So, the MAURE can be safely used when multicollinearity exists in a linear regression model.

References

1. Ahmed, A. E., Ali, H. M., and Amal, H. A. (2021). Almost Unbiased Liu Principal Component Estimator in the Presence of Multicollinearity and Autocorrelation. *The Egyptian Statistical Journal*, 64(1), 21-33.
2. Alheety, M. I., and Kibria, B. M. G. (2009). On the Liu and Almost Unbiased Liu Estimators in the Presence of Multicollinearity with Heteroscedastic or Correlated Errors. *Surveys in Mathematics and its Applications*, 4, 155-167.
3. Alkhamisi, M., Khalaf, G., and Shukur, G. (2006). Some Modifications for Choosing Ridge Parameters. *Communications in Statistics-Theory and Methods*, 35(11), 2005-2020.
4. Arumairajan, S., and Wijekoon, P. (2017). Modified Almost Unbiased Liu Estimator in Linear Regression Model. *Communications in Mathematics and Statistics*, 5(3), 261-276.
5. Dawoud, I., Abonazel, M. R., and Awwad, F. (2022). Generalized Kibria-Lukman Estimator: Method, Simulation, and Application. *Frontiers in Applied Mathematics and Statistics*, 8, 1-6.
6. Gruber, M. H. J. (1998). *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. Dekker, Inc., New York.
7. Hoerl, A. E., and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
8. Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975). Ridge Regression: Some Simulations. *Communications in Statistics*, 4(2), 105-123.
9. Johnson, R. A., and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Sixth Edition, Pearson Prentice Hall, New Jersey.

10. Kadiyala, K. (1984). A Class of Almost Unbiased and Efficient Estimators of Regression Coefficients. *Economics Letters*, 16, 293-296.
11. Khalaf, G., and Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics-Theory and Methods*, 34(5), 1177-1182.
12. Kibria, B. M. G. (2003). Performance of Some New Ridge Regression Estimators. *Communications in Statistics-Simulation and Computation*, 32(2), 419-435.
13. Kibria, B. M. G., and Lukman, A. F. (2020). A New Ridge-Type Estimator for the Linear Regression Model: Simulations and Applications. *Scientifica*, 2020(1), 1-16.
14. Li, Y., and Yang, H. (2011). Two Kinds of Restricted Modified Estimators in Linear Regression Model. *Journal of Applied Statistics*, 38(7), 1447-1454.
15. Li, Y., and Yang, H. (2014). Two Classes of Almost Unbiased Type Principal Component Estimators in Linear Regression Model. *Journal of Applied Mathematics*, 2014(1), 1-6.
16. Liu, K. (1993). A New Class of Biased Estimate in Linear Regression. *Communications in Statistics-Theory and Methods*, 22(2), 393-402.
17. Liu, K. (2003). Using Liu-Type Estimator to Combat Collinearity. *Communications in Statistics-Theory and Methods*, 32(5), 1009-1020.
18. Lukman, A. F., Kibria, B. M. G., Ayinde, K., and Jegede, S. L. (2020). Modified One-Parameter Liu Estimator for the Linear Regression Model. *Modelling and Simulation in Engineering*, 2020(1), 1-17.
19. Massy, W. F. (1965). Principal Components Regression in Exploratory Statistical Research. *Journal of the American Statistical Association*, 60(309), 234-256.
20. McDonald, G. C., and Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*, 70(350), 407-416.

21. Muniz, G., and Kibria, B. M. G. (2009). On Some Ridge Regression Estimators: An Empirical Comparisons. *Communications in Statistics-Simulation and Computation*, 38(3), 621-630.
22. Owolabi, A. T., Ayinde, K., Idowu, J. I., Oladapo, O. J., and Lukman, A. F. (2022). A New Two-Parameter Estimator in the Linear Regression Model with Correlated Regressors. *Journal of Statistics Applications & Probability*, 11(2), 499-512.
23. Özkale, M. R., and Kaçiranlar, S. (2007). The Restricted and Unrestricted Two-Parameter Estimators. *Communications in Statistics-Theory and Methods*, 36(15), 2707-2725.
24. Rao, C. R., Toutenburg, H., Shalabh, and Heumann, C. (2008). *Linear Models and Generalizations: Least Squares and Alternatives*. Third Edition, Springer-Verlag, Berlin.
25. Singh, B., Chaubey, Y. P., and Dwivedi, T. D. (1986). An Almost Unbiased Ridge Estimator. *The Indian Journal of Statistics*, 48(3), 342-346.
26. Stein, 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*, Berkeley, University of California Press, 1, 197-206.
27. Trenkler, G., and Toutenburg, H. (1990). Mean Squared Error Matrix Comparisons between Biased Estimators-An Overview of Recent Results. *Statistical Papers*, 31(1), 165-179.
28. Vinod, H. D., and Ullah, A. (1981). *Recent Advances in Regression Methods*. Dekker, New York.
29. Yang, H., and Chang, X. (2010). A New Two-Parameter Estimator in Linear Regression. *Communications in Statistics-Theory and Methods*, 39(6), 923-934.