ISSN: 2469-5831

Dawoud and Kibria. Int J Clin Biostat Biom 2020, 6:031
DOI: 10.23937/2469-5831/1510031
Volume 6 | Issue 2
Open Access

International Journal of
Clinical Biostatistics and Biometrics

ORIGINAL ARTICLE

A New Two Parameter Biased Estimator for the Unrestricted Linear Regression Model: Theory, Simulation and Application

Issam Dawoud[1] and BM Golam Kibria[2*]

[1]*Department of Mathematics, Al-Aqsa University, Gaza, Palestine*

[2]*Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA*

***Corresponding author:** *BM Golam Kibria, Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA*

Abstract

This paper proposes a new biased estimator for estimating the regression parameters for the multiple linear regression models when the regressors are correlated. Theoretical comparisons and simulations results show that, the proposed estimator performs better than other existing estimators under some conditions in the smaller mean squares error sense. A real-life dataset is analyzed to illustrate the findings of the paper.

Keywords

Biased estimator, Liu estimator, Monte Carlo simulation, Multicollinearity, MSE, Ridge regression estimator

Introduction

The ordinary least squares (OLS) estimator is the best linear unbiased estimator and has been used to estimate the parameters of the linear regression model since its inception. One of the important assumptions for the linear regression model is that the regressors (independent variables) are independent. However, in practice the regressors may or may not be independent, which causes the problem of multicollinearity. In the presence of the multicollinearity, the OLS estimator is inefficient and gives wrong sign of the parameters in the multiple linear regression models [1]. To handle these problems, many authors have given different types of estimators, to mention a few, [1-7], and recently [8], for one parameter biased estimator. However, to refer for two parameters mode, the following authors are notable, [9-15], among others.

The main objective of this paper is to propose a new two parameter biased estimator for the regression coefficients and then to compare the performance of the new estimator with the OLS, the ordinary ridge regression (ORR) of Hoerl and Kennard [1], the Liu of Liu [6] and the Kibria-Lukman (KL) of Kibria and Lukman [8] estimators. The rest of the paper is organized as follows: Some estimators and their statistical properties are given in Section 2. In section 3, the theoretical comparisons among the proposed estimator and existing estimators and the biasing parameters k and d are given. A Monte Carlo simulation study is performed in section 4. A real-life data are analyzed in section 5. Finally, Some Conclusions are given in section 6.

The Model and Estimators

Consider the following linear regression model:

$$y = X\beta + \varepsilon, \tag{2.1}$$

where y is an $n \times 1$ vector of the response variable, X is a known $n \times p$ full rank matrix of the explanatory variables, β is an $p \times 1$ vector of unknown regression coefficients, ε is a $n \times 1$ vector of disturbance assumed to be distributed with mean vector 0 and variance covariance matrix $\sigma^2 I$, and I is an identity matrix of order $n \times n$. To define various estimators, canonical form of the model (2.1) is given by:

$$y = Z\alpha + \varepsilon,\ \ \varepsilon \sim N(0, \sigma^2 I_n) \tag{2.2}$$

where, $Z = XC$, $\alpha = C'\beta$, and C is an orthogonal matrix such that $Z'Z = C'X'XC = U = diag(u_1, u_2, \ldots, u_p)$.

Citation: Dawoud I, Kibria BMG (2020) A New Two Parameter Biased Estimator for the Unrestricted Linear Regression Model: Theory, Simulation and Application. Int J Clin Biostat Biom 6:031. doi.org/10.23937/2469-5831/1510031

Accepted: November 09, 2020: **Published:** November 11, 2020

Then, the OLS estimator of α is given as:

$$\hat{\alpha} = U^{-1}Z'y, \quad (2.3)$$

and then the mean squared error matrix (MSEM) of $\hat{\alpha}$ is given by

$$MSEM(\hat{\alpha}) = \sigma^2 U^{-1}. \quad (2.4)$$

The ORR of α [1] is

$$\hat{\alpha}_k = NU\hat{\alpha}, \quad (2.5)$$

where, $N = [U + kI_p]^{-1}$ and the MSEM of $\hat{\alpha}_k$ would be

$$MSEM(\hat{\alpha}_k) = \sigma^2 NUN' + (NU - I_p)\alpha\alpha'(NU - I_p)'. \quad (2.6)$$

Hoerl, et al. [16] defined the biasing parameter *k* for $\hat{\alpha}_k$ as follows:

$$\hat{k}_{HM} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\hat{\alpha}_i^2}. \quad (2.7)$$

The Liu estimator of α [6] is

$$\hat{\alpha}_d = F\hat{\alpha}, \quad (2.8)$$

where, $F = [U + I_p]^{-1}[U + dI_p]$ and the biasing parameter *d* of $\hat{\alpha}_d$ is given by

$$\hat{d}_{opt} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^{p} 1/(u_i(u_i+1))}{\sum_{i=1}^{p} (\hat{\alpha}_i^2/(u_i+1)^2)} \right], \quad (2.9)$$

$$MSEM(\hat{\alpha}_d) = \sigma^2 FU^{-1}F' + (1-d)^2(U+I_p)^{-1}\alpha\alpha'(U+I_p)^{-1}. \quad (2.10)$$

If $\hat{d}_{opt}$ is negative, Ozkale and Kaciranlar [9] adopt the alternative biasing parameter below:

$$\hat{d}_{alt} = \min\left[\frac{\hat{\alpha}_i^2}{(\hat{\sigma}^2/u_i) + \hat{\alpha}_i^2}\right]_{i=1}^{p}. \quad (2.11)$$

The recently proposed KL estimator by Kibria and Lukman [8] is defined as,

$$\hat{\alpha}_{KL} = NM\hat{\alpha}, \quad (2.12)$$

where, $M = (U - k\,\mathrm{I}_p)$ and the biasing parameter *k* of the KL estimator is defined by

$$\hat{k}_{\min}(KL) = \min\left[\frac{\hat{\sigma}^2}{2\hat{\alpha}_i^2 + (\hat{\sigma}^2/u_i)}\right], \quad (2.13)$$

$$MSEM(\hat{\alpha}_{KL}) = \sigma^2 NMU^{-1}M'N' + [NM - I_p]\alpha\alpha'[NM - I_p]'. \quad (2.14)$$

The new biased (NB) estimator of α is obtained by minimizing $(y - Z\alpha)'(y - Z\alpha)$ subject to $(\sqrt{k}\,\alpha + \sqrt{k}\,(1+d)\hat{\alpha})'(\sqrt{k}\,\alpha + \sqrt{k}\,(1+d)\hat{\alpha}) = c$ where *c* is a constant,

$$(y - Z\alpha)'(y - Z\alpha) + [(\sqrt{k}\,\alpha + \sqrt{k}\,(1+d)\hat{\alpha})'(\sqrt{k}\,\alpha + \sqrt{k}\,(1+d)\hat{\alpha}) - c]. \quad (2.15)$$

Here, *k* and *d* are the Lagrangian multipliers.

The solution to (2.15) gives the new estimator as follows:

$$\hat{\alpha}_{NB} = (U + k\mathrm{I}_p)^{-1}(U - k(1+d)\mathrm{I}_p)\hat{\alpha} = NB\hat{\alpha} \quad (2.16)$$

where $B = (U - k(1+d)\mathrm{I}_p)$, $k \ge 0$ and $0 < d < 1$.

Moreover, the proposed NB estimator is also obtained by augmenting $-\sqrt{k}\,(1+d)\hat{\alpha} = \sqrt{k}\,\alpha + \varepsilon'$ to equation (2.2) and then using the OLS estimate.

The proposed NB estimator is a general estimator which includes the OLS and the KL estimator:

If *k* = 0 in $\hat{\alpha}_{NB}$ becomes $\hat{\alpha}$.

If $d = 0$ in $\hat{}_{NB}$ becomes $\hat{\alpha}_{KL}$.

The MSEM of the proposed NB estimator of α is given by

$$MSEM(\hat{\alpha}_{NB}) = \sigma^2 NBU^{-1}B'N' + [NB - I_p]\,\alpha\alpha'[NB - I_p]'. \quad (2.17)$$

The lemmas below will be used for theoretical comparisons among estimators in the next section.

Lemma 2.1: [17] Let E be an $n \times n$ positive definite matrix, that is $E > 0$ and α be some vector; then, $E - \alpha\alpha' > 0$ if and only if $\alpha' E^{-1}\alpha < 1$.

Lemma 2.2: [18] Let $\alpha_i = A_i\, y$, $i = 1,2$ be two linear estimators of α. Suppose that $D = Cov(\hat{\alpha}_1) - Cov(\hat{\alpha}_2) > 0$, where $Cov(\hat{\alpha}_i)\, i = 1,2$ be the covariance matrix of $\hat{\alpha}_i$ and $b_i = Bias(\hat{\alpha}_i) = (A_i X - I)\alpha$, $i = 1,2$. Consequently,

$$\Delta(\hat{\alpha}_1 - \hat{\alpha}_2) = MSEM(\hat{\alpha}_1) - MSEM(\hat{\alpha}_2) = \sigma^2 D + b_1 b_1' - b_2 b_2' > 0 \quad (2.18)$$

if and only if $b_2'[\sigma^2 D + b_1' b_1]^{-1} b_2 < 1$ where $MSEM(\hat{\alpha}_i) = Cov\ (\hat{\alpha}_i) + b_i b_i'$.

Comparison among the Estimators

Comparison between $\hat{\alpha}$ and $\hat{\alpha}_{NB}$

Theorem 3.1: $MSEM\ (\hat{\alpha}) - MSEM\ (\hat{\alpha}_{NB}) > 0$ if and only if

$$\alpha'[NB - I_p]'\,[\sigma^2 (U^{-1} - NBU^{-1}B'N')][NB - I_p]\alpha < 1 \quad (3.1)$$

Proof:

$$D(\hat{\alpha}) - D(\hat{\alpha}_{NB}) = \sigma^2 (U^{-1} - NBU^{-1}B'N')$$

$$= \sigma^2\, diag\left\{\frac{1}{u_i} - \frac{(u_i - k(1+d))^2}{u_i(u_i + k)^2}\right\}_{i=1}^{p} \quad (3.2)$$

where $U^{-1} - NBU^{-1}B'N'$ will be positive definite (pd) if and only if $(u_i + k)^2 - (u_i - k(1+d))^2 > 0$ or $(u_i + k) - (u_i - k(1+d)) > 0$. Clearly, for $k > 0$ and $0 < d < 1$, $(u_i + k) - (u_i - k(1+d)) = k(2+d) > 0$. By Lemma 2.2. The proof is completed.

Comparison between $\hat{\alpha}_k$ and $\hat{\alpha}_{NB}$

Theorem 3.2: $MSEM\ (\hat{\alpha}_k) - MSEM\ (\hat{\alpha}_{NB}) > 0$ if and only if

$$\alpha'[NB - I_p]'\,[V_1 + (NU - I_p)\alpha\alpha'(NU - I_p)'][NB - I_p]\alpha < 1 \quad (3.3)$$

where, $V_1 = \sigma^2 (NUN' - NBU^{-1}B'N')$

Proof:

$$V_1 = \sigma^2 (NUN' - NBU^{-1}B'N')$$

$$= \sigma^2\, diag\left\{\frac{u_i}{(u_i + k)^2} - \frac{(u_i - k(1+d))^2}{u_i(u_i + k)^2}\right\}_{i=1}^{p} \quad (3.4)$$

where, $NUN' - NBU^{-1}B'N'$ will be pd if and only if $u_i^2 - (u_i - k(1+d))^2 > 0$ or $u_i - (u_i - k(1+d)) > 0$. Clearly, for $k > 0$ and $0 < d < 1$, $u_i - (u_i - k(1+d)) = k(1+d) > 0$. By Lemma 2.2. The proof is completed.

Comparison between $\hat{\alpha}_d$ and $\hat{\alpha}_{NB}$

Theorem 3.3: $MSEM\ (\hat{\alpha}_d) - MSEM\ (\hat{\alpha}_{NB}) > 0$ if and only if

$$\alpha'[NB - I_p]'\,[V_2 + (F - I_p)\alpha\alpha'(F - I_p)'][NB - I_p]\alpha < 1 \quad (3.5)$$

where, $V_2 = \sigma^2 (FU^{-1}F - NBU^{-1}B'N')$

Proof:

$$V_2 = \sigma^2 (FU^{-1}F - NBU^{-1}B'N')$$

$$= \sigma^2\, diag\left\{\frac{(u_i + d)^2}{u_i(u_i + 1)^2} - \frac{(u_i - k(1+d))^2}{u_i(u_i + k)^2}\right\}_{i=1}^{p} \quad (3.6)$$

where, $FU^{-1}F - NBU^{-1}B'N'$ will be pd if and only if $(u_i + d)^2 (u_i + k)^2 - (u_i + 1)^2 (u_i - k(1+d))^2 > 0$ or $(u_i + d)(u_i + k) - (u_i + 1)(u_i - k(1+d)) > 0.$ Clearly, for $k > 0$ and $0 < d < 1$, $(u_i + d)(u_i + k) - (u_i + 1)(u_i - k(1+d)) = u_i(2k + kd + d - 1) + k(2d + 1) > 0.$ By Lemma 2.2. The proof is completed.

Comparison between $\hat{\alpha}_{KL}$ and $\hat{\alpha}_{NB}$

Theorem 3.4: $MSEM(\hat{\alpha}_{KL}) - MSEM(\hat{\alpha}_{NB}) > 0$ if and only if

$$\alpha'[NB - I_p]'[V_3 + (NM - I_p)\alpha\alpha'(NM - I_p)'][NB - I_p]\alpha < 1 \quad (3.7)$$

where, $V_3 = \sigma^2 (NMU^{-1}M'N' - NBU^{-1}B'N')$

Proof:

$$V_2 = \sigma^2 (NMU^{-1}M'N' - NBU^{-1}B'N')$$

$$= \sigma^2 \, diag\left\{ \frac{(u_i - k)^2}{u_i(u_i + k)^2} - \frac{(u_i - k(1+d))^2}{u_i(u_i + k)^2} \right\}_{i=1}^{p} \quad (3.8)$$

where, $NMU^{-1}M'N' - NBU^{-1}B'N'$ will be pd if and only if $(u_i - k)^2 - (u_i - k(1+d))^2 > 0$ or $(u_i - k) - (u_i - k(1+d)) > 0.$ Clearly, for $k > 0$ and $0 < d < 1$, $(u_i - k) - (u_i - k(1+d)) = kd > 0.$ By Lemma 2.2. The proof is completed.

Selection of the parameters *k* and *d*

Different biasing parameters estimators of *k* and *d* are proposed in different studies, for example, Hoerl and Kennard [1], Liu [6], [16,19-23], among others.

So the optimal values of *k* and *d* for the proposed NB estimator is going to be found. At first, by minimizing the equation *m*, we obtain the optimal value of *k* when *d* is fixed as

$$MSEM(\hat{\alpha}_{NB}) = E((\hat{\alpha}_{NB} - \alpha)'(\hat{\alpha}_{NB} - \alpha)),$$

$$m = tr(MSEM(\hat{\alpha}_{NB})),$$

$$m = \sigma^2 \sum_{i=1}^{p} \frac{(u_i - k(1+d))^2}{u_i(u_i + k)^2} + k^2(2+d)^2 \sum_{i=1}^{p} \frac{\alpha_i^2}{(u_i + k)^2} \quad (3.9)$$

Differentiating *m* with respect to *k* and setting $(\partial m/\partial k) = 0,$ we get

$$k = \frac{u_i \sigma^2}{(d+1)(u_i \alpha_i^2 + \sigma^2) + u_i \alpha_i^2}. \quad (3.10)$$

Then, the estimated optimal value of *k* is given as follows:

$$\hat{k} = \frac{u_i \hat{\sigma}^2}{(d+1)(u_i \hat{\alpha}_i^2 + \hat{\sigma}^2) + u_i \hat{\alpha}_i^2}, \quad (3.11)$$

and,

$$\hat{k}_{\min}(NB) = \min\left\{ \frac{u_i \hat{\sigma}^2}{(d+1)(u_i \hat{\alpha}_i^2 + \hat{\sigma}^2) + u_i \hat{\alpha}_i^2} \right\}_{i=1}^{p}. \quad (3.12)$$

Also, the optimal value of *d* will be found by differentiating *m* with respect to *d* when *k* is fixed and setting $(\partial m/\partial d) = 0,$ we get

$$d = \frac{u_i \sigma^2 - (2u_i \alpha_i^2 + \sigma^2)k}{k(u_i \alpha_i^2 + \sigma^2)}. \quad (3.13)$$

Then, the estimated optimal *d* with the unbiased estimators is given by

$$\hat{d} = \frac{u_i \hat{\sigma}^2 - (2u_i \hat{\alpha}_i^2 + \hat{\sigma}^2)\hat{k}}{\hat{k}(u_i \hat{\alpha}_i^2 + \hat{\sigma}^2)}, \quad (3.14)$$

and

$$\hat{d}_{\min}(NB) = \min\left\{\frac{u_i\hat{\sigma}^2 - (2u_i\hat{\alpha}_i^2 + \hat{\sigma}^2)\hat{k}_{\min}(NB)}{\hat{k}_{\min}(NB)(u_i\hat{\alpha}_i^2 + \hat{\sigma}^2)}\right\}_{i=1}^{p}. \quad (3.15)$$

The parameters *k* and *d* estimators selection in $\hat{\alpha}_{NB}$ are found iteratively as follows:

Find an initial estimate of *d* using $\hat{d} = \min(\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2})$.

Determine $\hat{k}_{\min}(NB)$ from (3.12) using $\hat{d}$ in step 1.

Estimate $\hat{d}_{\min}(NB)$ in (3.15) by using $\hat{k}_{\min}(NB)$ in step 2.

If $\hat{d}_{\min}(NB) > 1$ or $\hat{d}_{\min}(NB) < 0$, use $\hat{d}_{\min}(NB) = \hat{d}$.

Simulation Study

A Monte Carlo simulation study is performed to show the performance of the NB estimator over some existing estimators. It contains two parts: (i) Simulation technique and (ii) Results discussion.

Simulation technique

Using the equation below, we generate the explanatory variables (see, Gibbons [24] and Kibria [19]):

$$x_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{i,p+1}, \quad i = 1,2,...,n, \quad j = 1,2,...,p \quad (4.1)$$

where z_{ij} are independent standard normal pseudo-random numbers and *ρ* is the correlation between any two explanatory variables and *ρ* here has two values 0.9 and 0.99. The *n* observations for the response variable *y* are gotten by the following equation:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i, \quad i = 1,2,\ldots,n \quad (4.2)$$

where e_i are $i.i.d\ N(0,\sigma^2)$. The values of *β* are obtained as $\beta'\beta = 1$ [25]. Also, we choose the values of the biasing parameters of the estimators as *k* = *d* = 0.1, 0.2, ..., 0.9 by following the work of Wichern and Churchill [26] and Kan, et al. [27] is that when k lies between 0 and 1, the ORR estimator performs better. The Monte Carlo simulation study replication is 1000 times for *n* = 50 and 100 and $\sigma^2 = 1, 25, and\ 100$. For each replicate, we calculate the mean square error (MSE) of the estimators using the next equation:

$$MSE(\alpha^*) = \frac{1}{1000}\sum_{j=1}^{1000}(\alpha_{ij}^* - \alpha_i)'(\alpha_{ij}^* - \alpha_i) \quad (4.3)$$

where α_{ij}^* is the estimator and α_i is the true parameter. The estimated MSEs of the estimators are shown in Table 1, Table 2, Table 3 and Table 4 for (ρ = 0.90 and n = 50), (ρ = 0.99 and n = 50), (ρ = 0.90 and n = 100), and (ρ = 0.99 and n = 100), respectively.

Simulation results discussions

From Table 1, Table 2, Table 3 and Table 4, we observed that, when the factors *σ* and *ρ* are going to increase, the estimated MSE values are also going to increase, while *n* is going to increase, the estimated MSE values are going to decrease. Also, the OLS estimator is performing the worst for all cases in presence of the multicollinearity. Moreover, the simulation results show that, the proposed NB estimator is performing better than the other estimators for most of cases. The Liu estimator gives better results in the MSE values when *k* = *d* = 0.1, 0.2 i.e. when the biasing parameters are near to zero. For ρ = 0.9: The condition number (CN) is approximately around 5 and the variance inflation factors (VIFs) are around 4 to 6 such that it is observed that a close agreement of the proposed NB (although better results) with the KL estimator for low values of k = d and a better performance as k = d increases for a fixed value of σ. This improvement increases with the increase of the value of σ and an even better improvements is observed if ρ = 0.99 where the CN and the VIFs become larger and are approximately around 15 and around 36 to 61, respectively. So, the proposed NB estimator works better for the strong correlated explanatory variables. So, the performance of the proposed NB estimator almost depends on the value of ρ, σ, the biasing parameters k and d, and the true parameter. Thus, simulation results are consistent with the theoretical results.

Application

To illustrate the theoretical and simulation results of this paper, we consider a real life data in this section. The Portland cement data was originally adopted by Woods, et al. [28]. This data was also analyzed by many researchers, for examples, [29,30]; Lukman, et al. [14] and recently by Kibria and Lukman [8], among others. And this data is analyzed here to explain the performance of the proposed NB estimator and the other existing estimators. The

Table 1: Estimated MSE for OLS, ORR, Liu, KL and the proposed NB.

$\rho = 0.90,\ n = 50$						
Eigenvalues of $X'X$ are: 167.6671, 12.0130, and 6.3459						
CN of $X'X$ is 5.1402						
VIFs are: 4.5049, 4.5496, and 6.3642						
σ	$k = d$	OLS	ORR	Liu	KL	NB
1	0.1	0.215736	0.211191	**0.180083**	0.206646	0.206242
	0.2		0.206747	**0.183921**	0.198061	0.196344
	0.3		0.202505	0.187658	0.189779	**0.186143**
	0.4		0.198465	0.191597	0.181901	**0.175538**
	0.5		0.194425	0.195536	0.174427	**0.164832**
	0.6		0.190587	0.199475	0.167155	**0.153924**
	0.7		0.18685	0.203414	0.160287	**0.143016**
	0.8		0.183214	0.207454	0.153722	**0.132108**
	0.9		0.17978	0.211595	0.147359	**0.121301**
5	0.1	5.392794	5.279472	**4.501368**	5.167362	5.156151
	0.2		5.169786	**4.596207**	4.951626	4.908600
	0.3		5.063635	4.692157	4.745182	**4.651858**
	0.4		4.960817	4.789117	4.547626	**4.387642**
	0.5		4.861231	4.887188	4.358554	**4.117972**
	0.6		4.764776	4.986168	4.177461	**3.844464**
	0.7		4.671351	5.086259	4.003943	**3.569037**
	0.8		4.580653	5.187461	3.837798	**3.293307**
	0.9		4.492783	5.289572	3.678622	**3.019193**
10	0.1	21.57118	21.11779	**18.00547**	20.66945	20.62491
	0.2		20.67914	**18.38503**	19.80671	19.63450
	0.3		20.25454	18.76883	18.98103	**18.60743**
	0.4		19.84337	19.15667	18.19081	**17.55077**
	0.5		19.44523	19.54875	17.43422	**16.47179**
	0.6		19.05941	19.94498	16.70974	**15.37776**
	0.7		18.6854	20.34524	16.01577	**14.27564**
	0.8		18.32281	20.74974	15.35109	**13.17282**
	0.9		17.97123	21.15839	14.71419	**12.07607**

Note: Smaller MSE value is bolded in each case.

Table 2: Estimated MSE for OLS, ORR, Liu and the proposed NB.

$\rho = 0.99,\ n = 50$						
Eigenvalues of $X'X$ are: 182.2172, 1.2633, and 0.6648						
CN of $X'X$ is 16.5554						
VIFs are: 40.6681, 42.3133, and 61.4367						
σ	$k = d$	OLS	ORR	Liu	KL	NB

1	0.1	1.984104	1.624860	**0.580584**	1.301724	1.271328
	0.2		1.355478	**0.694926**	0.847110	0.759900
	0.3		1.148316	0.819570	0.541416	**0.403920**
	0.4		0.985626	0.954720	0.335580	**0.173094**
	0.5		0.855372	1.100172	0.198390	**0.046818**
	0.6		0.749598	1.256130	0.109344	**0.010404**
	0.7		0.662490	1.422492	0.054366	**0.005324**
	0.8		0.589866	1.599258	0.023868	**0.001673**
	0.9		0.528768	1.786428	0.011118	**0.000347**
5	0.1	49.60148	40.62089	**14.51470**	32.54249	31.78422
	0.2		33.88736	**17.37162**	21.17836	18.99719
	0.3		28.70841	20.48894	13.53397	**10.09678**
	0.4		24.63953	23.86667	8.388276	**4.326840**
	0.5		21.38450	27.50481	4.959444	**1.168818**
	0.6		18.73975	31.40335	2.731764	**0.256734**
	0.7		16.56164	35.5623	1.356702	**0.132498**
	0.8		14.74645	39.98165	0.593538	**0.041769**
	0.9		13.21767	44.66141	0.273156	**0.008660**
10	0.1	198.4053	162.4829	**58.05881**	130.1693	127.1369
	0.2		135.5488	**69.48658**	84.71345	75.98878
	0.3		114.8336	81.95598	54.13609	**40.38731**
	0.4		98.55832	95.46690	33.55341	**17.30726**
	0.5		85.53832	110.0192	19.83788	**4.675476**
	0.6		74.95919	125.6130	10.92716	**1.026936**
	0.7		66.24665	142.2492	5.426808	**0.529992**
	0.8		58.98578	159.9258	2.374152	**0.167072**
	0.9		52.87078	178.6448	1.092318	**0.034653**

Note: Smaller MSE value is bolded in each case.

Table 3: Estimated MSE for OLS, ORR, Liu, KL and the proposed NB.

$\rho = 0.90,\ n = 100$						
Eigenvalues of $X'X$ are: 273.8443, 18.5471, and 14.8552						
CN of $X'X$ is 4.2935						
VIFs are: 4.2216, 3.9332, and 4.7453						
σ	$k = d$	OLS	ORR	Liu	KL	NB
1	0.1	0.107464	0.106353	**0.098172**	0.105242	0.105141
	0.2		0.105242	**0.099182**	0.103121	0.102616
	0.3		0.104232	0.100192	0.101000	**0.099990**
	0.4		0.103121	0.101202	0.098879	**0.097162**
	0.5		0.102111	0.102212	0.096859	**0.094233**
	0.6		0.101101	0.103323	0.094839	**0.091203**
	0.7		0.100091	0.104333	0.092920	**0.088072**

	0.8		0.099081	0.105343	0.091001	**0.084840**
	0.9		0.098071	0.106454	0.089082	**0.081507**
5	0.1	2.687711	2.659734	**2.452785**	2.631858	2.629131
	0.2		2.632161	**2.478338**	2.577318	2.566410
	0.3		2.605093	2.503992	2.523889	**2.499851**
	0.4		2.57853	2.529848	2.471672	**2.429656**
	0.5		2.55227	2.555805	2.420566	**2.356128**
	0.6		2.526515	2.581863	2.370571	**2.27957**
	0.7		2.501063	2.608123	2.321586	**2.200184**
	0.8		2.476116	2.634484	2.273712	**2.118374**
	0.9		2.451573	2.661047	2.226848	**2.034443**
10	0.1	10.75074	10.63884	**9.810837**	10.52753	10.51642
	0.2		10.52875	**9.913049**	10.30917	10.26554
	0.3		10.42037	10.01587	10.09556	**9.999101**
	0.4		10.31392	10.11909	9.886587	**9.718321**
	0.5		10.20908	10.22302	9.682062	**9.424209**
	0.6		10.10586	10.32745	9.482082	**9.117775**
	0.7		10.00435	10.43239	9.286243	**8.800433**
	0.8		9.904464	10.53794	9.094646	**8.473193**
	0.9		9.80609	10.64409	8.90719	**8.137267**

Note: Smaller MSE value is bolded in each case.

Table 4: Estimated MSE for OLS, ORR, Liu, KL and the proposed NB.

$\rho = 0.99,\ n = 100$						
Eigenvalues of $X'X$ are: 308.1041, 1.9490, and 1.5571						
CN of $X'X$ is 14.0666						
VIFs are: 40.7553, 36.2165, and 45.1145						
σ	$k = d$	OLS	ORR	Liu	KL	NB
1	0.1	1.011126	0.914532	**0.491028**	0.823038	0.814062
	0.2		0.831504	**0.539376**	0.669936	0.639744
	0.3		0.759492	0.590172	0.544782	**0.487560**
	0.4		0.696660	0.643314	0.442170	**0.357102**
	0.5		0.641478	0.698700	0.357612	**0.247962**
	0.6		0.592722	0.756534	0.288048	**0.159834**
	0.7		0.549372	0.816612	0.230724	**0.092718**
	0.8		0.510714	0.879138	0.183498	**0.046308**
	0.9		0.476136	0.943908	0.144738	**0.020604**
5	0.1	25.27764	22.86401	**12.27468**	20.57483	20.35277
	0.2		20.78791	**13.48522**	16.74850	15.99462
	0.3		18.98812	14.75430	13.62016	**12.18941**
	0.4		17.41701	16.08193	11.05303	**8.926428**
	0.5		16.03685	17.46811	8.941014	**6.197520**
	0.6		14.81744	18.91294	7.200996	**3.996156**
	0.7		13.73461	20.41622	5.767182	**2.317440**
	0.8		12.76826	21.97814	4.586634	**1.157496**

	0.9		11.90207	23.59862	3.617124	**0.513060**
10	0.1	101.1106	91.45606	**49.09852**	82.29921	81.41089
	0.2		83.15142	**53.94056**	66.99391	63.97858
	0.3		75.95236	59.01700	54.48044	**48.75773**
	0.4		69.66794	64.32763	44.2119	**35.70581**
	0.5		64.14729	69.87245	35.76395	**24.78977**
	0.6		59.26985	75.65156	28.80398	**15.98442**
	0.7		54.93822	81.66497	23.06842	**9.269556**
	0.8		51.07283	87.91258	18.34643	**4.629474**
	0.9		47.60840	94.39447	14.46829	**2.051730**

Note: Smaller MSE value is bolded in each case.

regression model for this data is given by

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i . \quad (5.1)$$

For knowing more details about this data, we refer Woods, et al. [28].

To show the existence of the multicollinearity, different measures are calculated as the VIFs which are 38.50, 254.42, 46.87 and 282.51, Eigenvalues of $X'X$ are 44676.206, 5965.422, 809.952 and 105.419 and the CN of $X'X$ approximately equals 20.58. The VIFs, the eigenvalues and the CN tell us that there is a severe multicollinearity in the data. Also, $\hat{\sigma}$ is equal to 2.446. The correlation coefficients matrix of the explanatory variables are presented in Table 5 such that there is a significant and strong relationship among the following explanatory variables: X_1 and X_3, X_2 and X_4. Then, the estimated parameters and the MSE values of the estimators are presented in Table 6. It appears from Table 6 that the proposed NB estimator is performing the best where it is giving an obvious improvement over all existing estimators and a little improvement over the KL estimator in which this is consistent with the simulation results because $\hat{\sigma}$ here is small and not all the explanatory variables are significantly or strongly correlated at the same degree of correlation even though the data has high CN and VIFs.

Some Concluding Remarks

In this paper, we proposed a new biased (NB) estimator for handling the multicollinearity problem in the multiple linear regression models. Some existing estimators are the special case of the proposed estimator. The proposed NB estimator is compared theoretically with the Ordinary least squares (OLS) estimator, the Ordinary ridge regression (ORR) estimator, the Liu estimator and the Kibria-Lukman (KL) estimator, and then the biasing parameters *d* and *k* of the NB estimator are derived. A Monte Carlo simulation study is performed for comparing the performance of the OLS, ORR, Liu, KL

Table 5: The correlation coefficients matrix of the explanatory variables.

Variables	X_1	X_2	X_3	X_4
X_1	1	0.229	-0.824*	-0.245
X_2	0.229	1	-0.139	-0.973*
X_3	-0.824*	-0.139	1	0.030
X_4	-0.245	-0.973*	0.030	1

Note: *Correlation is significant at 0.05.

Table 6: The results of regression coeffcients and the corresponding MSE values.

Coef.	$\hat{\alpha}$	$\hat{\alpha}(\hat{k}_{\min})$	$\hat{\alpha}(\hat{d}_{alt})$	$\hat{\alpha}_{KL}(\hat{k}_{\min})$	$\hat{\alpha}_{NB}(\hat{k}_{\min}, \hat{d}_{\min})$
α_0	62.4053	8.58715	27.6657	27.6270	27.6004
α_1	1.55110	2.10461	1.90080	1.90884	1.9091
α_2	0.51016	1.06484	0.86996	0.86859	0.8689
α_3	0.10190	0.66808	0.46192	0.46782	0.4681
α_4	-0.14406	0.39959	0.20801	0.20724	0.2075
MSE	4912.0902	2989.8202	2170.9669	2170.9604	2170.9596
(k, d)	-------	0.007676	0.442224	0.000471	(0.000471, 0.001536)

and the proposed NB estimators. The main finding of this simulation is that the proposed NB estimator performed better than the above mentioned estimators under some conditions. A real-life data is analyzed to support the findings of the paper.

Acknowledgments

Authors are grateful to referees for their valuable comments and suggestions, which certainly improved the quality and presentation of the paper. Author B M Golam Kibria would like to dedicate this paper to his Ph. D. supervisor, Late Professor M Safiul Haq for his guidance and constant inspiration during his stay at The University of Westen Ontario and finally to achieve this present position.

References

1. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12: 55-67.
2. Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkley Symposium on Mathematical Statistics and Probability 1: 197-206.
3. Massy WF (1965) Principal components regression in exploratory statistical research. Journal of the American Statistical Association 60: 234-256.
4. Mayer LS, Willke TA (1973) On biased estimation in linear models. Technometrics 15: 497-508.
5. Swindel BF (1976) Good ridge estimators based on prior information. Communications in Statistics-Theory and Methods 5: 1065-1075.
6. Liu K (1993) A new class of biased estimate in linear regression. Communication in Statistics-Theory and Methods 22: 393-402.
7. 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.
8. Kibria BMG, Lukman AF (2020) A new ridge-type estimator for the linear regression model: Simulations and applications. Hindawi Scientifica.
9. Ozkale MR, Kaçiranlar S (2007) The restricted and unrestricted two-parameter estimators. Communications in Statistics-Theory and Methods 36: 2707-2725.
10. Sakallioglu S, Kaçiranlar S (2008) A new biased estimator based on ridge estimation. Statistical Papers 49: 669-689.
11. Yang H, Chang X (2010) A new two-parameter estimator in linear regression. Communications in Statistics-Theory and Methods 39: 923-934.
12. Roozbeh M (2018) Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. Computational Statistics & Data Analysis 117: 45-61.
13. Akdeniz F, Roozbeh M (2019) Generalized difference-based weighted mixed almost unbiased ridge estimator in partially linear models. Statistical Papers 60: 1717-1739.
14. Lukman AF, Ayinde K, Binuomote S, Clement OA (2019) Modified ridge-type estimator to combat multicollinearity: Application to chemical data. Journal of Chemometrics 33: e3125.
15. Lukman AF, Ayinde K, Sek SK, Adewuyi E (2019) A modified new two-parameter estimator in a linear regression model. Modelling and Simulation in Engineering.
16. Hoerl AE, Kannard RW, Baldwin KF (1975) Ridge regression: Some simulations. Communications in Statistics 4: 105-123.
17. Farebrother RW (1976) Further results on the mean square error of ridge regression. Journal of the Royal Statistical Society: Series B (Methodological) 38: 248-250.
18. Trenkler G, Toutenburg H (1990) Mean squared error matrix comparisons between biased estimators-an overview of recent results. Statistical Papers 31: 165-179.
19. Kibria BMG (2003) Performance of some new ridge regression estimators. Communications in Statistics-Simulation and Computation 32: 419-435.
20. Kibria BMG, Banik S (2016) Some ridge regression estimators and their performances. Journal of Modern Applied Statistical Methods 15: 206-238.
21. Lukman AF, Ayinde K (2017) Review and classifications of the ridge parameter estimation techniques. Hacettepe Journal of Mathematics and Statistics 46: 953-967.
22. Mansson K, Kibria BMG, Shukur G (2015) Performance of some weighted Liu estimators for logit regression model: An application to Swedish accident data. Communications in Statistics Theory and Methods 44: 363-375.
23. Khalaf G, Shukur G (2005) Choosing ridge parameter for regression problems. Communications in Statistics Theory and Methods 34: 1177-1182.
24. Gibbons DG (1981) A simulation study of some ridge estimators. Journal of the American Statistical Association 76: 131-139.
25. Newhouse JP, Oman SD (1971) An evaluation of ridge estimators. A report prepared for United States air force project RAND.
26. Wichern DW, Churchill GA (1978) A comparison of ridge estimators. Technometrics 20: 301-311.
27. Kan B, Alpu O, Yazici B (2013) Robust ridge and robust liu estimator for regression based on the LTS estimator. Journal of Applied Statistics 40: 644-655.
28. Woods H, Steinour HH, Starke HR (1932) Effect of composition of Portland cement on heat evolved during hardening. Ind Eng Chem 24: 1207-1214.
29. Kaciranlar S, Sakallioglu S, Akdeniz F, Styan GPH, Werner HJ (1999) A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on portland cement. Sankhya: The Indian Journal of Statistics, Series B 61: 443-459.
30. Li Y, Yang H (2012) A new Liu-type estimator in linear regression model. Statistical Papers 53: 427-437.