frontiers | Frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 20 April 2022
doi: 10.3389/fams.2022.880086

Generalized Kibria-Lukman Estimator: Method, Simulation, and Application

Issam Dawoud[1], Mohamed R. Abonazel[2] and Fuad A. Awwad[3]*

[1] Department of Mathematics, Al-Aqsa University, Gaza, Palestine, [2] Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt, [3] Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh, Saudi Arabia

In the linear regression model, the multicollinearity effects on the ordinary least squares (OLS) estimator performance make it inefficient. To solve this, several estimators are given. The Kibria-Lukman (KL) estimator is a recent estimator that has been proposed to solve the multicollinearity problem. In this paper, a generalized version of the KL estimator is proposed, along with the optimal biasing parameter of our proposed estimator derived by minimizing the scalar mean squared error. Theoretically, the performance of the proposed estimator is compared with the OLS, the generalized ridge, the generalized Liu, and the KL estimators by the matrix mean squared error. Furthermore, a simulation study and the numerical example were performed for comparing the performance of the proposed estimator with the OLS and the KL estimators. The results indicate that the proposed estimator is better than other estimators, especially in cases where the standard deviation of the errors was large and when the correlation between the explanatory variables is very high.

Keywords: generalized liu estimator, multicollinearity, generalized ridge estimator, biasing parameter, ridge-type estimator

OPEN ACCESS

Edited by:
*Min Wang,
University of Texas at San Antonio,
United States*

Reviewed by:
*Muhammad Suhail,
University of Agriculture,
Peshawar, Pakistan
Zakariya Yahya Algamal,
University of Mosul, Iraq*

****Correspondence:***
*Mohamed R. Abonazel
mabonazel@cu.edu.eg*

Specialty section:
*This article was submitted to
Statistics and Probability,
a section of the journal
Frontiers in Applied Mathematics and
Statistics*

Received: *21 February 2022*
Accepted: *14 March 2022*
Published: *20 April 2022*

Citation:
*Dawoud I, Abonazel MR and
Awwad FA (2022) Generalized
Kibria-Lukman Estimator: Method,
Simulation, and Application.
Front. Appl. Math. Stat. 8:880086.
doi: 10.3389/fams.2022.880086*

INTRODUCTION

The statistical consequences of multicollinearity are well-known in statistics for a linear regression model. Multicollinearity is known as the approximately linear dependency among the columns of the matrix X in the following linear model

$$y = X\beta + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I_n\right) \tag{1}$$

where y is an $n \times 1$ vector of the given dependent variable, X is a known $n \times p$ matrix of the given explanatory variables, β is an $p \times 1$ vector of given unknown regression parameters, and ε is described as an $n \times 1$ vector of the disturbances. Then, the ordinary least squares (OLS) estimator of β for the model (1) is given as:

$$\hat{\beta} = (X'X)^{-1}X'y$$

The multicollinearity problem effects on the behavior of the OLS estimator make it inefficient. Sometimes, it produces wrong signs [1, 2]. Many studies were conducted to handle this. For example, Hoerl and Kennard [2] proposed the ordinary ridge and the generalized ridge (GR) estimators, while Liu [3] introduced the popular Liu and the generalized Liu (GL), and very recently, Kibria and Lukman [1] proposed a ridge-type estimator called the Kibria–Lukman (KL) estimator which is defined by

$$\hat{\beta}_{KL} = (X'X + kI_p)^{-1}(X'X - kI_p)\hat{\beta}, k > 0$$

This estimator has been extended for use in different generalized linear models, such as Lukman et al. [4, 5], Akram et al. [6], and Abonazel et al. [7].

According to recent papers [8–10], we can say that the efficiency of any bias estimator will increase if the estimator is modified or generalized using bias parameters that vary from observation to observation in the sample (k_i and/or d_i) rather than in fixed bias parameters (k and/or d). Hence, the main purpose of this paper is to develop a general form of the KL estimator to combat the multicollinearity in the linear regression model.

The rest of the discussion in this paper is structured as follows: Section Statistical Methodology presents the statistical methodology. In Section Superiority of the Proposed GKL Estimator, we theoretically compare the proposed general form of the KL estimator with each of the mentioned estimators. In Section The Biasing Parameter Estimator of the GKL Estimator, we give the estimation of the biasing parameter of the proposed estimator. Different scenarios of the Monte Carlo simulation are done in Section A Monte Carlo Simulation Study. A real data is used in Section Empirical Application. Finally, Section Conclusion presents some conclusions.

STATISTICAL METHODOLOGY

Canonical Form

The canonical form of the model in equation (1) is used as follows:

$$y = Z\alpha + \varepsilon \tag{2}$$

where $Z = XR, \alpha = R^{'}\beta$, and R is an orthogonal matrix such that $Z'Z = R'X'XR = G = diag(g_1, g_2, \ldots, g_p)$. Then, the OLS of α is as:

$$\hat{\alpha} = G^{-1}Z'y \tag{3}$$

and the matrix mean squared error (MMSE) is given as,

$$MMSE\left(\hat{\alpha}\right) = \sigma^2 G^{-1} \tag{4}$$

Ridge Regression Estimators

The OR and the GR of $\alpha$are, respectively, defined as follows [2]:

$$\hat{\alpha}_{OR} = W_1 G\hat{\alpha} \tag{5}$$

$$\hat{\alpha}_{GR} = W_2 G\hat{\alpha} \tag{6}$$

where $W_1 = [G + kI_p]^{-1}$, $k > 0$ and $W_2 = [G + K]^{-1}$, with $K = diag(k_1, k_2, ..., k_p)$, $k_i > 0$, and $i = 1, 2, ..., p$.

The MMSE of the OR and the GR are given respectively as:

$$MMSE(\hat{\alpha}_{OR}) = \sigma^2 W_1 G W_1{}' + (W_1 G - I_p)\alpha\alpha'(W_1 G - I_p)' \tag{7}$$

$$MMSE(\hat{\alpha}_{GR}) = \sigma^2 W_2 G W_2{}' + (W_2 G - I_p)\alpha\alpha'(W_2 G - I_p)' \tag{8}$$

Liu Regression Estimators

The Liu and the GL of $\alpha$are respectively defined as follows [3]:

$$\hat{\alpha}_{Liu} = F_1\hat{\alpha} \tag{9}$$

$$\hat{\alpha}_{GL} = F_2\hat{\alpha} \tag{10}$$

where

$$F_1 = [G + I_p]^{-1}[G + dI_p], 0 < d < 1 \text{ and } F_2 = [G + I_p]^{-1} [G + D], \text{with } D = diag(d_1, d_2, ..., d_p) \text{ and } 0 < d_i < 1.$$

The MMSE of the Liu and the GL are, respectively, given as:

$$MMSE(\hat{\alpha}_{Liu}) = \sigma^2 F_1 G^{-1} F_1{}' + (F_1 - I_p)\alpha\alpha'(F_1 - I_p)' \tag{11}$$

$$MMSE(\hat{\alpha}_{GL}) = \sigma^2 F_2 G^{-1} F_2{}' + (F_2 - I_p)\alpha\alpha'(F_2 - I_p)' \tag{12}$$

Kibria–Lukman Estimator

The KL estimator of $\alpha$is given as Kibria and Lukman [1]:

$$\hat{\alpha}_{KL} = W_1 M_1\hat{\alpha} \tag{13}$$

where $M_1 = [G - kI_p]$ and the MMSE of this estimator is given as:

$$MMSE(\hat{\alpha}_{KL}) = \sigma^2 W_1 M_1 G^{-1} M_1{}' W_1{}' + [W_1 M_1 - I_p]\alpha\alpha'[W_1 M_1 - I_p]' \tag{14}$$

The Proposed GKL Estimator

Now, by replacing W_1 with W_2 and M_1 with $M_2 = [G - K]$ in the KL estimator, we obtain the general form of the GKL estimator as follows:

$$\hat{\alpha}_{GKL} = W_2 M_2\hat{\alpha} \tag{15}$$

then, the MMSE of the proposed GKL estimator is computed by,

$$MMSE(\hat{\alpha}_{GKL}) = \sigma^2 W_2 M_2 G^{-1} M_2{}' W_2{}' + [W_2 M_2 - I_p] \alpha\alpha'[W_2 M_2 - I_p]' \tag{16}$$

SUPERIORITY OF THE PROPOSED GKL ESTIMATOR

In this section, we make a comparison of the proposed GKL estimator with each of OLS, GR, GL, and KL estimators. First, we offer some useful lemmas for our comparisons of estimators.

Lemma 1: Wang et al. [11]: Suppose M and N are $n \times n$ positive definite matrices, then $M > N$ if and only if (iff) $\lambda - 1_{\max}$, where $\lambda - 1_{\max}$ is the maximum eigenvalue of NM^{-1} matrix.

Lemma 2: Farebrother [12]: Let S be an $n \times n$ positive definite matrix. That is, $S > 0$ and α be some vector. Then, $S - \alpha\alpha^{'} > 0$ iff $\alpha' S^{-1}\alpha < 1$.

Lemma 3: Trenkler and Toutenburg [13]: Let $\alpha_i = U_i w$, $i = 1, 2$ be any two linear estimators of α. Suppose that $Q = Cov(\hat{\alpha}_1) - Cov(\hat{\alpha}_2) > 0$, where $Cov(\hat{\alpha}_i), i = 1, 2$ be the covariance matrix of $\hat{\alpha}_i$ and $b_i = Bias(\hat{\alpha}_i) = (U_i X - I)\alpha$. Then,

$$\Delta\left(\hat{\alpha}_1 - \hat{\alpha}_2\right) = MMSE\left(\hat{\alpha}_1\right) - MMSE\left(\hat{\alpha}_2\right) = \sigma^2 Q + b_1 b_1{}' - b_2 b_2{}' > 0 \tag{17}$$

iff $b_2'[\sigma^2 Q + b_1' b_1]^{-1} b_2 < 1$ where $MMSE(\hat{\alpha}_i) = Cov(\hat{\alpha}_i) + b_i b_i'$

Theorem 1: $\hat{\alpha}_{GKL}$ is superior to $\hat{\alpha}$ iff

$$\alpha'[W_2 M_2 - I_p]'[\sigma^2(G^{-1} - W_2 M_2 G^{-1} M_2' W_2')] [W_2 M_2 - I_p]\alpha < 1 \quad (18)$$

Proof: The covariance matrices difference is written as

$$Difference = \sigma^2 \left(G^{-1} - W_2 M_2 G^{-1} M_2' W_2'\right) = \sigma^2 diag\left\{\frac{1}{g_i} - \frac{(g_i - k_i)^2}{g_i (g_i + k_i)^2}\right\}_{i=1}^{p} \quad (19)$$

where $G^{-1} - W_2 M_2 G^{-1} M_2' W_2'$ becomes positive definite iff $(g_i + k_i)^2 - (g_i - k_i)^2 > 0$ or $(g_i + k_i) - (g_i - k_i) > 0$. It is clear that for $k_i > 0, i = 1, 2, ..., p$, $(g_i + k_i) - (g_i - k_i) = 2k_i > 0$. Therefore, this is done using Lemma 3.

Theorem 2: When $\lambda - 1_{\max}$, $\hat{\alpha}_{GKL}$ is superior to $\hat{\alpha}_{GR}$ iff

$$\alpha'[W_2 M_2 - I_p]'[V_1 + (W_2 G - I_p)\alpha\alpha'(W_2 G - I_p)'] [W_2 M_2 - I_p]\alpha < 1 \quad (20)$$

$$\lambda - 1_{\max} \quad (21)$$

where $V_1 = \sigma^2(W_2 G W_2' - W_2 M_2 G^{-1} M_2' W_2')$, $N = W_2 K G^{-1} K W_2'$, and $M = 2 W_2 K K W_2'$.

Proof:

$$V_1 = \sigma^2 \left(W_2 G W_2' - W_2 M_2 G^{-1} M_2' W_2'\right)$$
$$= \sigma^2 \left(W_2 G W_2' - W_2 \left(G - K\right) G^{-1} \left(G - K\right) W_2'\right)$$
$$= \sigma^2 \left(2 W_2 K K W_2' - W_2 K G^{-1} K W_2'\right)$$
$$= \sigma^2 (M - N)$$

For $k_i > 0$, it is obvious that $M > 0$ and $N > 0$. Then, $M - N > 0$ iff $\lambda - 1_{\max}$, where $\lambda - 1_{\max}$ is the maximum eigenvalue of NM^{-1}. So, this is done by Lemma 1.

Theorem 3: $\hat{\alpha}_{GKL}$ is superior to $\hat{\alpha}_{GL}$ iff

$$\alpha'[W_2 M_2 - I_p]'[V_2 + (F_2 - I_p)\alpha\alpha'(F_2 - I_p)'][W_2 M_2 - I_p] \alpha < 1 \quad (22)$$

where $V_2 = \sigma^2(F_2 G^{-1} F_2' - W_2 M_2 G^{-1} M_2' W_2')$.

Proof: The covariance matrices difference is written as

$$V_2 = \sigma^2 \left(F_2 G^{-1} F_2' - W_2 M_2 G^{-1} M_2' W_2'\right) = \sigma^2 diag\left\{\frac{(g_i + d_i)^2}{g_i (g_i + 1)^2} - \frac{(g_i - k_i)^2}{g_i (g_i + k_i)^2}\right\}_{i=1}^{p} \quad (23)$$

where $F_2 G^{-1} F_2' - W_2 M_2 G^{-1} M_2' W_2'$ becomes positive definite iff $(g_i + k_i)^2 (g_i + d_i)^2 - (g_i - k_i)^2 (g_i + 1)^2 > 0$ or $(g_i + k_i)(g_i + d_i) - (g_i - k_i)(g_i + 1) > 0$. So, if $k_i > 0$ and $0 < d_i < 1$, $(g_i + k_i)(g_i + d_i) - (g_i - k_i)(g_i + 1) = k_i(2g_i + d_i + 1) + g_i(d_i - 1) > 0$. So, this is done by Lemma 3.

Theorem 4: $\hat{\alpha}_{GKL}$ is superior to $\hat{\alpha}_{KL}$ iff

$$\alpha'[W_2 M_2 - I_p]'[V_3 + (W_1 M_1 - I_p)\alpha\alpha'(W_1 M_1 - I_p)'] [W_2 M_2 - I_p]\alpha < 1 \quad (24)$$

where $V_3 = \sigma^2(W_1 M_1 G^{-1} M_1' W_1' - W_2 M_2 G^{-1} M_2' W_2')$.

Proof: The covariance matrices difference is written as

$$V_3 = \sigma^2 \left(W_1 M_1 G^{-1} M_1' W_1' - W_2 M_2 G^{-1} M_2' W_2'\right) = \sigma^2 diag\left\{\frac{(g_i - k)^2}{g_i (g_i + k)^2} - \frac{(g_i - k_i)^2}{g_i (g_i + k_i)^2}\right\}_{i=1}^{p} \quad (25)$$

where $W_1 M_1 G^{-1} M_1' W_1' - W_2 M_2 G^{-1} M_2' W_2'$ becomes positive definite iff $(g_i + k_i)^2 (g_i - k)^2 - (g_i - k_i)^2 (g_i + k)^2 > 0$ or $(g_i + k_i)(g_i - k) - (g_i - k_i)(g_i + k) > 0$. So, if $k_i > 0$ and $k_i > k$, $(g_i + k_i)(g_i - k) - (g_i - k_i)(g_i + k) = 2g_i(k_i - k) > 0$. So, this is done by Lemma 3.

THE BIASING PARAMETER ESTIMATOR OF THE GKL ESTIMATOR

The performance of any estimator depends on its biasing parameter. Therefore, the determination of the biasing parameter of an estimator is an important issue. Different studies analyzed this issue (e.g., [2, 3, 8–10, 14–24]).

Kibria and Lukman [1] proposed the biasing parameter estimator of the KL estimator as follows:

$$\hat{k} = \min\left\{\frac{\hat{\sigma}^2}{[(\hat{\sigma}^2/g_i) + 2\hat{\alpha}_i^2]}\right\}_{i=1}^{p} \quad (26)$$

Here, we find the estimation of the optimal values of k_i for the proposed GKL estimator. The optimal values of k_i are obtained by minimizing

$$MMSE(\hat{\alpha}_{GKL}) = E[(\hat{\alpha}_{GKL} - \alpha)' \left(\hat{\alpha}_{GKL} - \alpha\right)],$$
$$m(k_1, k_2, ..., k_p) = tr(MMSE(\hat{\alpha}_{GKL}), \text{and}$$
$$m(k_1, k_2, ..., k_p) = \sigma^2 \sum_{i=1}^{p} \frac{(g_i - k_i)^2}{g_i (g_i + k_i)^2} + \sum_{i=1}^{p} \frac{4k_i^2 \alpha_i^2}{(g_i + k_i)^2} \quad (27)$$

Differentiating $m(k_1, k_2, ..., k_p)$ with respect to k_i and setting $[\frac{\partial m(k_1, k_2, ..., k_p)}{\partial k_i}] = 0$, the optimal values of k_i after replacing σ^2 and α_i^2 by their unbiased estimators become as follows:

$$\hat{k}_i = \frac{\hat{\sigma}^2}{((\hat{\sigma}^2/g_i) + 2\hat{\alpha}_i^2)}, i = 1, 2, ..., p \quad (28)$$

A MONTE CARLO SIMULATION STUDY

The explanatory variables are generated as follows [25–27]:

$$x_{ji} = (1 - \rho^2)^{\frac{1}{2}} a_{ji} + \rho a_{jp}, j = 1, 2, ..., n, i = 1, 2, ..., p \quad (29)$$

where a_{ji} are the independent pseudo-random numbers that have the standard normal distribution and ρ is known that the correlation between two given explanatory variables. The dependent variable y are given by:

$$y_j = \beta_1 x_{j1} + \beta_2 x_{j2} + \ldots + \beta_p x_{jp} + \varepsilon_j, j = 1, 2, \ldots, n \quad (30)$$

TABLE 1 | The factors' values of the simulation study.

Factor	Symbol	Levels
Sample size	n	50, 100, 150
Standard deviation	σ	1, 5, 10
Degree of correlation	ρ	0.8, 0.9, 0.99
Explanatory variables number	p	3, 7
Replicates number	MCN	5,000

TABLE 2 | Estimated mean squared error (EMSE) values of the estimators when $p = 3$.

n	σ	ρ	OLS	KL	GKL
50	1	0.8	0.1249	**0.1094**	0.1548
		0.9	0.2260	**0.1829**	0.2738
		0.99	2.0641	1.1439	**1.1208**
	5	0.8	3.1235	1.7550	**1.6052**
		0.9	5.6491	2.8600	**2.4774**
		0.99	51.6036	22.2378	**17.6275**
	10	0.8	12.4940	6.2898	**5.3865**
		0.9	22.5965	10.5775	**8.7621**
		0.99	206.4144	87.8850	**69.2762**
100	1	0.8	0.0605	**0.0557**	0.0701
		0.9	0.1107	**0.0964**	0.1373
		0.99	1.0308	**0.6454**	0.7558
	5	0.8	1.5118	**0.9306**	0.9509
		0.9	2.7663	1.5097	**1.4056**
		0.99	25.7697	11.3736	**8.9376**
	10	0.8	6.0471	3.1436	**2.7244**
		0.9	11.0651	5.2952	**4.3648**
		0.99	103.0788	44.4958	**34.4270**
150	1	0.8	0.0420	**0.0393**	0.0469
		0.9	0.0768	**0.0687**	0.0928
		0.99	0.7125	**0.4700**	0.6113
	5	0.8	1.0497	**0.6763**	0.7487
		0.9	1.9189	1.0893	**1.0826**
		0.99	17.8124	7.7631	**6.1352**
	10	0.8	4.1988	2.2214	**1.9830**
		0.9	7.6756	3.6905	**3.1029**
		0.99	71.2496	29.9827	**23.1604**

For each case, the smallest EMSE value is bolded.

where ε_j are the $i.i.dN(0, \sigma^2)$. The values of β are given such that $\beta'\beta = 1$ as discussed in Dawoud and Abonazel [28], Algamal and Abonazel [29], Abonazel et al. [7, 30], and Awwad et al. [31]. Also, all factors that used in the simulation are given in **Table 1**.

In order to see the performance of the OLS, KL, and the proposed GKL estimators with their biasing parameters estimators presented in Section Statistical Methodology, the estimated mean squared error (EMSE) are calculated for each replicate with different values of σ, ρ, n, and p using the following formula:

$$EMSE(\alpha^*) = \frac{1}{MCN} \sum_{l=1}^{MCN} (\alpha_l^* - \alpha)'(\alpha_l^* - \alpha) \tag{31}$$

where α_l^* is the estimated vector of α at the lth experiment of the simulation.

The EMSE values of the OLS, KL, and GKL estimators are presented in **Tables 2**, **3**. We can conclude the following based on the simulation results:

1. When the standard deviation (σ), the degree of multicollinearity (ρ), and the explanatory variables number

TABLE 3 | EMSE values of the estimators when $p = 7$.

n	σ	ρ	OLS	KL	GKL
50	1	0.8	0.4143	**0.3129**	0.4302
		0.9	0.6792	**0.5399**	0.6831
		0.99	7.3867	3.9941	**3.0983**
	5	0.8	10.3568	5.5139	**4.1882**
		0.9	19.4796	10.0849	**7.4658**
		0.99	184.6673	92.8175	**66.6994**
	10	0.8	41.4272	21.1839	**15.5082**
		0.9	77.9186	39.4124	**28.5547**
		0.99	738.6690	370.3048	**265.3667**
100	1	0.8	0.1766	**0.1529**	0.2137
		0.9	0.3322	**0.2702**	0.3652
		0.99	3.1561	1.9888	**1.7020**
	5	0.8	4.4159	2.7275	**2.2455**
		0.9	8.3060	4.8911	**3.8358**
		0.99	78.9019	43.6091	**32.3890**
	10	0.8	17.6638	10.1544	**7.7808**
		0.9	33.2240	18.6747	**14.0582**
		0.99	315.6077	173.4003	**128.2151**
150	1	0.8	0.1105	**0.0992**	0.1341
		0.9	0.2081	**0.1773**	0.2504
		0.99	1.9769	1.3108	**1.2036**
	5	0.8	2.7632	1.7804	**1.5371**
		0.9	5.2014	3.1588	**2.5389**
		0.99	49.4224	27.3769	**20.2601**
	10	0.8	11.0529	6.4542	**4.9732**
		0.9	20.8054	11.8006	**8.8790**
		0.99	197.6896	108.306	**79.6545**

For each case, the smallest EMSE value is bolded.

TABLE 4 | Estimated coefficients and mean squared error (MSE) values of the estimators.

Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	MSE
OLS	2.1930	1.1533	0.7585	0.4863	0.0638
KL	2.1764	1.1572	0.7465	0.4888	0.0629
GKL	2.1653	1.1613	0.7312	0.4904	0.0620

(p) get an increase, the EMSE values of estimators get an increase.

2. The EMSE values of estimators get a decrease in case of the sample size gets an increase.
3. The GKL is better than the OLS estimator in all different values of factors except when $\sigma = 1$ and $\rho = 0.80,\ 0.90$ with the considered values of p and n.
4. The GKL is better than the KL estimator in all different values of factors except the following cases: (i) for $n = 50$ when $\sigma = 1$ and $\rho = 0.80,\ 0.90$ with $p = 3$ or 7, (ii) for $n = 100, 150$ when $\sigma = 1$ in all presented values of ρ with $p = 3$ or when $\sigma = 5$ and $\rho = 0.80$ with $p = 3$, and (iii) for $n = 100, 150$ when $\sigma = 1$ and $\rho = 0.80,\ 0.90$ with $p = 7$.
5. Finally, we see that the proposed GKL estimator is obviously efficient in case of the standard deviation getting large and when the correlation among the explanatory variables are very high.

EMPIRICAL APPLICATION

For clarifying the performance of the proposed GKL estimator, the dataset of the Portland cement that was originally due to Woods et al. [32], which was considered in Kibria and Lukman [1], where the dependent variable is the heat evolved after 180 days of curing and measured in calories per gram of cement. In this study, the first explanatory variable is tricalcium aluminate, the second explanatory variable is tricalcium silicate, the third explanatory variable is tetracalcium aluminoferrite, and the fourth explanatory variable is β-dicalcium silicate. The eigenvalues of $X'X$ matrix are 44,676.21, 5,965.42, 809.95, and 105.42. Then, the condition number is 20.58. Therefore, multicollinearity exists among the predictors. The estimated error variance is $\hat{\sigma}^2 = 5.84$, which shows high noise in the data. The estimated values of the optimal parameters in the GKL estimator are calculated as derived in Section Statistical Methodology. Also, the equation proposed by Kibria and Lukman [1] for estimating the biasing parameter of the KL estimator is used. Consequently, the mean square error (MSE) of the OLS, KL, and GKL estimators are presented in **Table 4**. From **Table 4**, we can note that the KL estimator is better than the OLS estimator, and the GKL estimator is better than the OLS and KL estimators.

CONCLUSION

In this paper, we proposed the GKL estimator. The performance of the proposed GKL estimator is theoretically compared with the OLS, GR, GL, and KL estimators in terms of known matrix mean squared error. Moreover, the optimal shrinkage parameter of the proposed GKL estimator is presented. A simulation study and the numerical example were performed for comparing the performance of the proposed GKL estimator with the OLS and KL estimators based on the estimated mean squared error criterion. The results indicate that the proposed estimator is better than other estimators, in particular, in the case the standard deviation of the errors was large and when the correlation between the explanatory variables is very high.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

ID, MA, and FA contributed to conception and structural design of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank the Deanship of Scientific Research at King Saud University, represented by the Research Center, at CBA for supporting this research financially.

REFERENCES

1. Kibria BMG, Lukman AF. A new ridge-type estimator for the linear regression model: simulations and applications. *Hindawi Sci.* (2020) 2020:9758378. doi: 10.1155/2020/9758378
2. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. (1970) 12:55–67.
3. Liu K. A new class of biased estimate in linear regression. *Commun Stat Theory Methods.* (1993) 22:393–402.
4. Lukman AF, Algamal ZY, Kibria BG, Ayinde K. The KL estimator for the inverse Gaussian regression model. *Concurr Comput Prac Exp.* (2021) 33:e6222. doi: 10.1002/cpe.6222
5. Lukman AF, Dawoud I, Kibria BM, Algamal ZY, Aladeitan B. A new ridge-type estimator for the gamma regression model. *Scientifica*. (2021) 2021:5545356. doi: 10.1155/2021/5545356
6. Akram MN, Kibria BG, Abonazel MR, Afzal N. On the performance of some biased estimators in the gamma regression model: simulation and applications. *J Stat Comput Simul.* (2022) 1–23. doi: 10.1080/00949655.2022.2032059
7. Abonazel MR, Dawoud I, Awwad FA, Lukman AF. Dawoud–Kibria estimator for beta regression model: simulation and application. *Front Appl Math Stat.* (2022) 8:775068. doi: 10.3389/fams.2022.775068
8. Rashad NK, Hammood NM, Algamal ZY. Generalized ridge estimator in negative binomial regression model. *J Phys.* (2021) 1897:012019. doi: 10.1088/1742-6596/1897/1
9. Farghali RA, Qasim M, Kibria BM, Abonazel MR. Generalized two-parameter estimators in the multinomial logit regression model: methods, simulation and application. *Commun Stat Simul Comput.* (2021) 1–16. doi: 10.1080/03610918.2021.1934023
10. Abdulazeez QA, Algamal ZY. Generalized ridge estimator shrinkage estimation based on particle swarm optimization algorithm. *Electro J Appl Stat Anal.* (2021) 14:254–65. doi: 10.1285/I20705948V14N1P254
11. Wang SG, Wu MX, Jia ZZ. *Matrix Inequalities*. Beijing: Chinese Science Press (2006).
12. Farebrother RW. Further results on the mean square error of ridge regression. *J R Stat Soc Ser B.* (1976) 38:248–50.

13. Trenkler G, Toutenburg H. Mean squared error matrix comparisons between biased estimators-an overview of recent results. *Stat Pap.* (1990) 31: 165–79.
14. Hoerl AE, Kannard RW, Baldwin KF. Ridge regression: some simulations. *Commun. Stat.* (1975) 4:105–23.
15. Khalaf G, Shukur G. Choosing ridge parameter for regression problems. *Commun Stat Theory Methods.* (2005) 34:1177–82. doi: 10.1081/STA-200056836
16. Khalaf G, Månsson K, Shukur G. Modified ridge regression estimators. *Commun Stat Theory Methods.* (2013) 42:1476–87. doi: 10.1080/03610926.2011.593285
17. Månsson K, Kibria BMG, Shukur G. Performance of some weighted Liu estimators for logit regression model: an application to Swedish accident data. *Commun Stat Theory Methods.* (2015) 44:363–75. doi: 10.1080/03610926.2012.745562
18. Kibria BMG, Banik S. Some ridge regression estimators and their performances. *J Mod Appl Stat Methods.* (2016) 15:206–38. doi: 10.22237/jmasm/1462075860
19. Algamal ZY. A new method for choosing the biasing parameter in ridge estimator for generalized linear model. *Chemometr Intell Lab Syst.* (2018) 183:96–101. doi: 10.1016/j.chemolab.2018.10.014
20. Abonazel MR, Farghali RA. Liu-type multinomial logistic estimator. *Sankhya B.* (2019) 81:203–25. doi: 10.1007/s13571-018-0171-4
21. Qasim M, Amin M, Omer T. Performance of some new Liu parameters for the linear regression model. *Commun Stat Theory Methods.* (2020) 49:4178–96. doi: 10.1080/03610926.2019.1595654
22. Suhail M, Chand S, Kibria BG. Quantile based estimation of biasing parameters in ridge regression model. *Commun Stat Simul Comput.* (2020) 49:2732–44. doi: 10.1080/03610918.2018.1530782
23. Babar I, Ayed H, Chand S, Suhail M, Khan YA, Marzouki R. Modified Liu estimators in the linear regression model: an application to tobacco data. *PLoS ONE.* (2021) 16:e0259991. doi: 10.1371/journal.pone.0259991
24. Abonazel MR, Taha IM. Beta ridge regression estimators: simulation and application. *Commun Stat Simul Comput.* (2021) 1–13. doi: 10.1080/03610918.2021.1960373
25. McDonald GC, Galarneau DI. A Monte Carlo evaluation of some ridge-type estimators. *J Am Stat Assoc.* (1975) 70:407–16. doi: 10.2307/2285832
26. Gibbons DG. A simulation study of some ridge estimators. *J Am Stat Assoc.* (1981) 76:131–9.
27. Kibria BMG. Performance of some new ridge regression estimators. *Commun Stat Simul Comput.* (2003) 32:419–35. doi: 10.1081/SAC-120017499
28. Dawoud I, Abonazel MR. Robust Dawoud–Kibria estimator for handling multicollinearity and outliers in the linear regression model. *J Stat Comput Simul.* (2021) 91:3678–92. doi: 10.1080/00949655.2021.1945063
29. Algamal ZY, Abonazel MR. Developing a Liu-type estimator in beta regression model. *Concurr Comput Pract Exp.* (2022) 34:e6685. doi: 10.1002/cpe.6685
30. Abonazel MR, Algamal ZY, Awwad FA, Taha IM. A new two-parameter estimator for beta regression model: method, simulation, and application. *Front Appl Math Stat.* (2022) 7:780322. doi: 10.3389/fams.2021.780322
31. Awwad FA, Dawoud I, Abonazel MR. Development of robust Özkale–Kaçiranlar and Yang–Chang estimators for regression models in the presence of multicollinearity and outliers. *Concurr Comput Pract Exp.* (2022) 34:e6779. doi: 10.1002/cpe.6779
32. Woods H, Steinour HH, Starke HR. Effect of composition of Portland cement on heat evolved during hardening. *Indust Eng Chem.* (1932) 24:1207–14. doi: 10.1021/ie50275a002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Dawoud, Abonazel and Awwad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.