Al-Aqsa University Journal - Natural Sciences Series, Volume 25, Issue 1, January 2023, pp. 33-46
ISSN 2070-3155

مُقدِّر انحدار متحيز مقترح لمعالجة وجود الارتباط الذاتي والتداخل الخطي المتعدد

د. عصام عاطف داود

أستاذ الإحصاء المشارك، قسم الرياضيات، جامعة الأقصى، غزة، فلسطين

(تاريخ الاستلام 2023/01/01، تاريخ القبول 2023/01/26)

A Proposed Biased Regression Estimator for Treating the Existence of Autocorrelation and Multicollinearity

Dr. Issam Dawoud

Associate Professor of Statistics, Mathematics Department, Al-Aqsa University, Gaza, Palestine

(Received 01/01/2023, Accepted 26/01/2023)

د. عصام عاطف داود - أستاذ الإحصاء المشارك، قسم الرياضيات، جامعة الأقصى، غزة، فلسطين- E-mail address: **ia.dawoud@alaqsa.edu.ps**

الملخص:

تهدف هذه الدراسة إلى اقتراح مُقدِّر انحدار متحيز لمعالجة كل من الارتباط الذاتي بين الأخطاء والتداخل الخطي المتعدد بين المتغيرات المستقلة معاً. تم ذكر خصائص المُقدّر المتحيز المعمم المقترح (GB)، وتم التحقق من أدائه باستخدام معيار متوسط مربع الأخطاء بمقارنته بمُقدِّر المربعات الصغرى المعمم و مُقدِّر ridge المعمم وكذلك مُقدِّر Liu المعمم. أيضاً، تم اختيار مقدرات للمعلمتين k و d لمُقدِّر GB المقترح والمقدرات الأخرى. قد أجريت دراسة محاكاة مكثفة، حيث أن النتائج تشير إلى أن أداء المُقدِّر المقترح GB أفضل بالمقارنة مع المُقدرّات المعممة المذكورة تحت شروط معينة. وأخيرًا، تم استخدام قاعدتين من البيانات الحقيقية لتوضيح تلك النتائج.

Abstract:

This study aims at suggesting a biased regression estimator for treating both the autocorrelation between errors and the multicollinearity among the explanatory variables together. The properties of the suggested generalized biased (GB) estimator are stated, and its performance is investigated using the mean square error criterion over the generalized least squares estimator, the generalized ridge estimator, and the generalized Liu estimator. Also, the estimators for the two parameters k and d of the suggested GB estimator and the others are selected. A massive study of simulation is performed, and the results indicate that the suggested GB estimator performed well compared to the other generalized estimators under specific conditions. Finally, two real datasets are used to explain these findings.

Keywords: Generalized Biased Estimator, Generalized Liu Estimator, Multicollinearity, Generalized Ridge Estimator, Correlated errors

1. Introduction:

The popular linear regression model with uncorrelated errors is considered as

$$y = X\beta + v, \quad (1)$$

where y is indicated as an $n \times 1$ vector of the dependent variable, X is defined as an $n \times p$ matrix of known explanatory variables with a full rank, β is indicated as a $p \times 1$ vector of unknown regression parameters, and v is indicated as an $n \times 1$ vector of errors with mean is equal to zero and the variance-covariance, $Cov(v) = \sigma^2 I_n$, σ^2 is the variance, and I_n is known as an $n \times n$ identity matrix. The popular ordinary least squares (OLS) of unknown β in (1) is as follows:

$$\hat{\beta} = (X'X)^{-1} X' y. \quad (2)$$

In linear regression models, it is known that the independence assumption of explanatory variables holds. But there may be a linear relationship between explanatory variables in real-life situations in which this causes a problem called multicollinearity (Lukman and Ayinde, 2017; Lukman et al., 2019; Qasim et al., 2020; Dawoud, 2021a). In the case of multicollinearity, the OLS estimator is inefficient and produces wrong signs sometimes (Hoerl and Kennard, 1970). To handle these problems, some studies have introduced estimators with one parameter, such as the ridge estimator, which is $\hat{\beta}_R = (X'X + kI_p)^{-1} X' y,\ k > 0$ (Hoerl and Kennard, 1970), the Liu estimator, which is $\hat{\beta}_L = (X'X + I_p)^{-1}(X'X + dI_p)\hat{\beta},\ 0 < d < 1$ (Liu, 1993), as well as the recent estimator of Kibria and Lukman (2020). Then, some studies combined the two parameters (k, d) in one estimator, as in Ozkale and Kaçiranlar (2007), Dawoud and Kibria (2020), Dawoud (2021b), Dawoud et al. (2022), as well as the new type of regression estimator recently given by Aslam and Ahmad (2022) and called the modified Liu-ridge type estimator, in which it is defined as $\hat{\beta}_{MLRT} = (X'X + I_p)^{-1}(X'X + dI_p) \times (X'X + k(1+d)I_p)^{-1} X'y$, where the parameters k and d are defined above.

Time series data happen frequently in many fields, such as business, economics, and engineering, where the uncorrelated errors assumption of these data does not often hold, and in this case, the OLS estimator becomes inefficient. To solve this, the generalized least squares (GLS) estimator is defined, which is unbiased and whose variance is less than that of the OLS estimator (Güler and Kaçiranlar, 2009). However, the multicollinearity problem may exist even after getting the GLS estimator because of a large variance, which yields untrusted estimates. As a result, Trenkler (1984) proposed the generalized ridge (GR) estimator to extend the ridge regression idea for reducing multicollinearity effects with correlated errors, and Kaciranlar (2003) introduced the generalized Liu (GL) estimator. Moreover, there are many studies about solving the above problems, for example, (Bayhan and Bayhan, 1998), (Alheety and Kibria, 2009), (Güler and Kaçiranlar, 2009), (Alkhamisi, 2010), (Huang and Yang, 2015), (Eledum and Alkhalifa, 2012), (Şiray et al., 2014), (Chandra and Sarkar, 2016), (Chandra and Tyagi, 2017), and (Tyagi and Chandra, 2017).

Consider the linear regression model with correlated errors as:

$$y = X\beta + \varepsilon, \tag{3}$$

where y, X, β are as defined in model (1) and ε is considered as an $n \times 1$ errors vector with $E(\varepsilon) = 0$, and $Cov(\varepsilon) = \sigma^2 V$. Since $\sigma^2 V$ is the errors variance-covariance matrix, V is known as positive definite (p.d.), so there is an $n \times n$ matrix E, where $E'E = V$, then model (3) is rewritten as

$$E^{-1}y = E^{-1}X\beta + E^{-1}\varepsilon, \tag{4}$$

let $y_* = E^{-1}y$, $X_* = E^{-1}X$ as well as $\varepsilon_* = E^{-1}\varepsilon$, so $E(\varepsilon_*) = 0$ as well as $Cov(\varepsilon_*) = \sigma^2 I_n$, see Alheety and Kibria (2009).

So, the transformed model is given as

$$y_* = X_*\beta + \varepsilon_*, \tag{5}$$

which it satisfies the error assumption, $\varepsilon_* \sim N(0, \sigma^2 I_n)$. So, the OLS estimator using (5) is as follows:

$$\begin{aligned} \tilde{\beta}_{GLS} &= (X_*'X_*)^{-1}X_*'y_* \\ &= (X'V^{-1}X)^{-1}X'V^{-1}y, \\ &= D^{-1}X'V^{-1}y \end{aligned} \tag{6}$$

where $D = (X'V^{-1}X)$ and $\tilde{\beta}_{GLS}$ is called the Aitken estimator or the GLS estimator of β, where it is the best linear unbiased estimator of the parameter β and $Cov(\tilde{\beta}_{GLS}) = \sigma^2 D^{-1}$, (Aitken, 1936).

The mean squared error matrix (MSEM) as well as the mean squared error (MSE) of an estimator $\tilde{\beta}$ are given as:

$$MSEM(\tilde{\beta}) = Cov(\tilde{\beta}) + \left(Bias(\tilde{\beta})\right)\left(Bias(\tilde{\beta})\right)', \tag{7}$$

$$MSE(\tilde{\beta}) = Trace\left(MSEM(\tilde{\beta})\right). \tag{8}$$

The GLS estimator in the canonical form is given as follows:

$$\tilde{\alpha}_{GLS} = Q^{-1}W'y_*, \tag{9}$$

the MSEM and the MSE are given respectively as:

$$MSEM\,(\tilde{\alpha}_{GLS}) = \sigma^2\, Q^{-1}, \tag{10}$$

$$MSE(\tilde{\alpha}_{GLS}) = \sigma^2 \sum_{j=1}^{p} \frac{1}{q_j}, \tag{11}$$

where $W = X_* P = E^{-1} XP$, P is known as an orthogonal matrix of D, with columns representing the eigenvectors of D, and $W'W = Q = diag\{q_1, q_2, ..., q_p\}$ is the diagonal matrix, with elements representing the eigenvalues of D, and $\alpha = P'\beta$, where $q_1 \geq q_2 \geq ... \geq q_p \geq 0$.

The objective of this study is to suggest a generalized biased estimator for the regression parameter when both autocorrelation and multicollinearity problems occur, and then to study the performance and compare the suggested generalized biased estimator with the GLS, the GR, and the GL estimators.

This study is planned as follows: In Section 2, we give the available estimators and the suggested estimator. We give some theorems about the superiority of the suggested generalized biased estimator in Section 3. In Section 4, the selected estimators of the parameters of all estimators are given. A simulation is conducted in Section 5. Two real-life applications are illustrated in Section 6. Finally, a conclusion is summarized in Section 7.

2. The Available Estimators and the Suggested Estimator

2.1 Generalized Ridge Estimator

Since the GLS estimator is still affected by multicollinearity, the GR estimator is proposed by Trenkler (1984) for α as follows:

$$\tilde{\alpha}_{GR} = F^{-1} W' y_*, \tag{12}$$

and the MSEM as well as the MSE of the GR estimator are given as

$$MSEM\,(\tilde{\alpha}_{GR}) = \sigma^2 F^{-1} Q F^{-1} + (F^{-1}Q - I)\,\alpha\,\alpha'\,(F^{-1}Q - I)', \tag{13}$$

$$MSE(\tilde{\alpha}_{GR}) = \sigma^2 \sum_{j=1}^{p} \frac{q_j}{F_j^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{F_j^2}, \tag{14}$$

where $F = (Q + k\, I_p)$ and $F_j = (q_j + k)$.

2.2 Generalized Liu Estimator

The GL estimator is proposed by Kaciranlar (2003) for α as:

$$\tilde{\alpha}_{GL} = R^{-1} N\, \tilde{\alpha}_{GLS}, \tag{15}$$

and the MSEM as well as the MSE of the GL estimator are given as

$$MSEM\,(\tilde{\alpha}_{GL}) = \sigma^2\, R^{-1} N\, Q^{-1} N R^{-1} + (R^{-1}N - I)\,\alpha\,\alpha'\,(R^{-1}N - I)', \tag{16}$$

$$MSE(\hat{\beta}_{GL}) = \sigma^2 \sum_{j=1}^{p} \frac{N_j^2}{q_j R_j^2} + (1-d)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{R_j^2}, \tag{17}$$

where $R = (Q + I_p)$, $R_j = (q_j + 1)$,

$N = (Q + d\, I_p)$, and $N_j = (q_j + d)$.

2.3 The Suggested Generalized Biased Estimator

The efficiency of the modified Liu-ridge type estimator given by Aslam and Ahmad (2022) over previous and recent estimators for linear regression models encourages us to proposed the generalized form of it for handling both autocorrelation and multicollinearity together. Therefore, we suggest the generalized biased (GB) estimator for α as follows:

$$\tilde{\alpha}_{GB} = R^{-1}NT^{-1}W'\,y_*, \qquad (18)$$

where $T = (Q + k(1+d)I_p)$.

The GB estimator is a biased regression estimator of α. The bias amount and covariance are given respectively,

$$Bias(\tilde{\alpha}_{GB}) = [R^{-1}NT^{-1}Q - I_p]\alpha \ , \qquad (19)$$

$$Cov\ (\tilde{\alpha}_{GB}) = \sigma^2\, R^{-1}NT^{-1}QT^{-1}NR^{-1}. \qquad (20)$$

So, the MSEM and the MSE are defined as

$$\begin{aligned}MSEM(\tilde{\alpha}_{GB}) &= \sigma^2\, R^{-1}NT^{-1}QT^{-1}NR^{-1} \\ &+ [R^{-1}NT^{-1}Q - I_p]\,\alpha\alpha'[R^{-1}NT^{-1}Q - I_p]'\end{aligned} \qquad (21)$$

$$MSE(\tilde{\alpha}_{GB}) = \sigma^2\sum_{j=1}^{p}\frac{q_j\,N_j^2}{R_j^2T_j^2} + \sum_{j=1}^{p}\frac{M_j^2\,\alpha_j^2}{R_j^2T_j^2}, \qquad (22)$$

where $T_j = (q_j + k(1+d))$ and $M_j = q_j(1-d) + k(1+d)R_j$.

3. The Superiority of the GB Estimator

Theorem 1:

If $\sum_{j=1}^{p}\frac{M_j^2\,\alpha_j^2}{R_j^2T_j^2} < \sum_{j=1}^{p}\frac{\sigma^2\left(R_j^2T_j^2 - q_j^2\,N_j^2\right)}{q_jR_j^2T_j^2}$, then $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GLS})$.

Proof: The difference between $MSE(\tilde{\alpha}_{GLS})$ and $MSE(\tilde{\alpha}_{GB})$ is as:

$$\begin{aligned}\Delta_1 &= MSE(\tilde{\alpha}_{GB}) - MSE(\tilde{\alpha}_{GLS}) \\ &= \sum_{j=1}^{p}\left[\frac{\sigma^2\left(q_j^2\,N_j^2 - R_j^2T_j^2\right) + q_jM_j^2\,\alpha_j^2}{q_jR_j^2T_j^2}\right].\end{aligned} \qquad (23)$$

Δ_1 in equation (23) will be negative if $\sum_{j=1}^{p}\frac{M_j^2\,\alpha_j^2}{R_j^2T_j^2} < \sum_{j=1}^{p}\frac{\sigma^2\left(R_j^2T_j^2 - q_j^2\,N_j^2\right)}{q_jR_j^2T_j^2}$, then $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GLS})$. Therefore, the GB estimator is better than the GLS estimator if $\sum_{j=1}^{p}\frac{M_j^2\,\alpha_j^2}{R_j^2T_j^2} < \sum_{j=1}^{p}\frac{\sigma^2\left(R_j^2T_j^2 - q_j^2\,N_j^2\right)}{q_jR_j^2T_j^2}$.

Theorem 2:

If $\sum_{j=1}^{p}\frac{\alpha_j^2\left(M_j^2F_j^2 - k^2R_j^2T_j^2\right)}{R_j^2T_j^2F_j^2} < \sum_{j=1}^{p}\frac{\sigma^2q_j\left(R_j^2T_j^2 - N_j^2F_j^2\right)}{R_j^2T_j^2F_j^2}$, then $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GR})$.

Proof: The difference between $MSE(\tilde{\alpha}_{GR})$ and $MSE(\tilde{\alpha}_{GB})$ is as:

$$\begin{aligned}\Delta_2 &= MSE(\tilde{\alpha}_{GB}) - MSE(\tilde{\alpha}_{GR}) \\ &= \sum_{j=1}^{p}\left[\frac{\sigma^2q_j\left(N_j^2F_j^2 - R_j^2T_j^2\right) + \alpha_j^2\left(M_j^2F_j^2 - k^2R_j^2T_j^2\right)}{R_j^2T_j^2F_j^2}\right].\end{aligned} \qquad (24)$$

Δ_2 in equation (24) will be negative if $\sum_{j=1}^{p}\frac{\alpha_j^2\left(M_j^2F_j^2 - k^2R_j^2T_j^2\right)}{R_j^2T_j^2F_j^2} < \sum_{j=1}^{p}\frac{\sigma^2q_j\left(R_j^2T_j^2 - N_j^2F_j^2\right)}{R_j^2T_j^2F_j^2}$, then $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GR})$. Therefore, the GB estimator is better than the GR estimator if $\sum_{j=1}^{p}\frac{\alpha_j^2\left(M_j^2F_j^2 - k^2R_j^2T_j^2\right)}{R_j^2T_j^2F_j^2} < \sum_{j=1}^{p}\frac{\sigma^2q_j\left(R_j^2T_j^2 - N_j^2F_j^2\right)}{R_j^2T_j^2F_j^2}$.

Theorem 3:

If

$$\sum_{j=1}^{p}\frac{\alpha_j^2\left(M_j^2-(1-d)^2T_j^2\right)}{R_j^2T_j^2}<\sum_{j=1}^{p}\frac{\sigma^2N_j^2\left(T_j^2-q_j^2\right)}{q_jR_j^2T_j^2},$$

then $MSE(\tilde{\alpha}_{GB})<MSE(\tilde{\alpha}_{GL})$.

Proof: The difference between $MSE(\tilde{\alpha}_{GL})$ and $MSE(\tilde{\alpha}_{GB})$ is as:

$$\Delta_3=MSE(\tilde{\alpha}_{GB})-MSE(\tilde{\alpha}_{GL})$$
$$=\sum_{j=1}^{p}\left[\frac{\sigma^2N_j^2\left(q_j^2-T_j^2\right)+\alpha_j^2q_j\left(M_j^2-(1-d)^2T_j^2\right)}{q_jR_j^2T_j^2}\right]. \quad (25)$$

Δ_3 in equation (25) will be negative if

$$\sum_{j=1}^{p}\frac{\alpha_j^2\left(M_j^2-(1-d)^2T_j^2\right)}{R_j^2T_j^2}<\sum_{j=1}^{p}\frac{\sigma^2N_j^2\left(T_j^2-q_j^2\right)}{q_jR_j^2T_j^2},$$

then $MSE(\tilde{\alpha}_{GB})<MSE(\tilde{\alpha}_{GL})$. Therefore, the GB estimator is better than the GL estimator if

$$\sum_{j=1}^{p}\frac{\alpha_j^2\left(M_j^2-(1-d)^2T_j^2\right)}{R_j^2T_j^2}<\sum_{j=1}^{p}\frac{\sigma^2N_j^2\left(T_j^2-q_j^2\right)}{q_jR_j^2T_j^2}.$$

4. Selected estimators for the biasing parameters k and d

We choose the estimators of the biasing parameters (k,d) for all above estimators as:

$$\tilde{k}=\min\left(\frac{\tilde{\sigma}^2}{\tilde{\alpha}_{GLS(j)}^2}\right)_{j=1}^{p}, \quad (26)$$

$$\tilde{d}=\max\left(\frac{\tilde{\alpha}_{GLS(j)}^2}{\frac{\tilde{\sigma}^2}{q_j}+\tilde{\alpha}_{GLS(j)}^2}\right)_{j=1}^{p}, \quad (27)$$

where $\tilde{\sigma}^2$ is the estimated variance based on the GLS estimator; see for other details: Hoerl and Kennard (1970), Kibria (2003), Ozkale and Kaciranlar (2007), and Aslam and Ahmad (2022).

5. A massive Simulation Study

A massive simulation study is carried out to compare the performance of the proposed GB estimator with that of the existing estimators. The computational procedures are done via MATLAB software. According to Kibria (2003) and Dawoud and Kibria (2020), the following equation is used for generating the explanatory variables:

$$x_{ij}=(1-\gamma^2)^{1/2}z_{ij}+\gamma z_{i,p+1}, \quad i=1,2,...,n, \quad j=1,2,...,p \quad (28)$$

where z_{ij}'s are considered independent pseudo-random numbers that follow a standard normal distribution, and γ is determined, where γ^2 is known as the correlation between two explanatory variables. According to Kibria (2003), the coefficient vector is the largest eigenvalue of D. The following is the dependent variable y:

$$y_i=\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_px_{ip}+e_i, \quad i=1,2,\ldots,n \quad (29)$$

and the error processes for (29) are chosen as:

$$e_i=u_i, \quad \text{White Noise errors} \quad (30)$$

$$e_i=\rho e_{i-1}+u_i, \quad \text{AR(1) errors} \quad (31)$$

$$e_i=u_i+\tau u_{i-1}, \quad \text{MA(1) errors} \quad (32)$$

where u_i are known as $i.i.d\ N(0,\sigma^2)$. The values of β are selected, where $\beta'\beta=1$, (Dawoud and Kibria, 2020). The V matrices and the variances of e_i for (30), (31) and (32) are given in Trenkler (1984) and Ozkale (2009), respectively,

$$V = I_n,\ \sigma^2 = \sigma_u^2 \tag{33}$$

$$V = (v_{rs}),\ \ v_{rs} = \rho^{|r-s|}, \quad \sigma^2 = \sigma_u^2/(1-\rho^2), \quad r,s = 1,2,...,n \tag{34}$$

$$V = \frac{1}{1+\tau^2}\begin{pmatrix} 1+\tau^2 & \tau & 0 & \cdots & 0 & 0 \\ \tau & 1+\tau^2 & \tau & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\tau^2 & \tau \\ 0 & 0 & 0 & \cdots & \tau & 1+\tau^2 \end{pmatrix},\ \sigma^2 = \sigma_u^2(1+\tau^2). \tag{35}$$

Also, all factors used in this simulation are stated in Table 1.

Table 1. The values of the factors that are considered in simulation

Factor	Symbol	Values
Autocorrelation of the error	ρ , τ	0.3, 0.7
Sample size	n	50, 150
Standard deviation	σ	1, 5
Degree of correlation	γ	0.7, 0.9
Number of explanatory variables	p	3, 7
Number of replicates	MCN	1000

To investigate the performances of the GLS, GR, GL, and the suggested GB estimator, we compute the estimated MSE values by using the following formula:

$$MSE(\delta^*) = \frac{1}{MCN}\sum_{j=1}^{MCN}(\delta_{ij}^* - \delta_i)'(\delta_{ij}^* - \delta_i), \tag{36}$$

where δ_{ij}^* is estimator output and δ_i is true parameter output.

The estimators' estimated MSEs are shown in Tables 2-4. The smallest MSE value is bolded in each row.

Table 2. The Estimated MSE of Estimators for White Noise Errors

n	p	σ	γ	GLS	GR	GL	GB
50	3	1	0.70	0.0882	0.0801	0.0880	**0.0742**
			0.90	0.2136	0.1727	0.2117	**0.1444**
		5	0.70	2.2058	1.3463	2.1786	**1.0364**
			0.90	5.3394	2.7537	5.2069	**1.9064**
	7	1	0.70	0.2822	0.2238	0.2809	**0.1863**
			0.90	0.7436	0.5195	0.7354	**0.3983**
		5	0.70	7.0556	3.8890	6.9586	**2.7580**
			0.90	18.5888	9.6747	18.1893	**6.5686**
150	3	1	0.70	0.0302	0.0288	0.0301	**0.0277**
			0.90	0.0751	0.0672	0.0750	**0.0607**

		5	0.70	0.7551	0.5332	0.7528	**0.4319**
			0.90	1.8766	1.0818	1.8615	**0.7728**
	7	1	0.70	0.0790	0.0726	0.0790	**0.0671**
			0.90	0.2089	0.1778	0.2086	**0.1540**
		5	0.70	1.9751	1.3220	1.9693	**1.0081**
			0.90	5.2216	3.1817	5.1911	**2.2812**

Table 3. The Estimated MSE of Estimators for AR(1) Errors

ρ	n	p	σ	γ	**GLS**	**GR**	**GL**	**GB**
0.3	50	3	1	0.70	0.0848	0.0771	0.0846	**0.0713**
				0.90	0.2079	0.1697	0.2067	**0.1427**
			5	0.70	2.1193	1.3102	2.0964	**1.0116**
				0.90	5.1976	2.7310	5.0807	**1.9002**
		7	1	0.70	0.2826	0.2224	0.2818	**0.1843**
				0.90	0.7444	0.5147	0.7383	**0.3941**
			5	0.70	7.0658	3.8532	6.9881	**2.7261**
				0.90	18.6110	9.5815	18.2710	**6.5111**
	150	3	1	0.70	0.0279	0.0267	0.0278	**0.0257**
				0.90	0.0696	0.0629	0.0695	**0.0572**
			5	0.70	0.6971	0.4965	0.6954	**0.4029**
				0.90	1.7407	1.0261	1.7299	**0.7405**
		7	1	0.70	0.0739	0.0683	0.0738	**0.0634**
				0.90	0.1949	0.1676	0.1948	**0.1463**
			5	0.70	1.8469	1.2608	1.8426	**0.9718**
				0.90	4.8728	3.0227	4.8500	**2.1887**
0.7	50	3	1	0.70	0.0760	0.0703	0.0759	**0.0657**
				0.90	0.1887	0.1599	0.1884	**0.1384**
			5	0.70	1.8993	1.2484	1.8919	**0.9832**
				0.90	4.7186	2.6755	4.6789	**1.9241**
		7	1	0.70	0.2713	0.2166	0.2710	**0.1816**
				0.90	0.7137	0.5050	0.7113	**0.3945**
			5	0.70	6.7820	3.8421	6.7490	**2.7752**
				0.90	17.8424	9.5541	17.7020	**6.6833**
	150	3	1	0.70	0.0218	0.0210	0.0217	**0.0204**
				0.90	0.0545	0.0503	0.0543	**0.0466**
			5	0.70	0.5438	0.4076	0.5434	**0.3370**
				0.90	1.3626	0.8559	1.3595	**0.6335**
		7	1	0.70	0.0585	0.0548	0.0584	**0.0516**
				0.90	0.1541	0.1357	0.1540	**0.1209**
			5	0.70	1.4620	1.0300	1.4607	**0.8064**
				0.90	3.8518	2.4739	3.8447	**1.8293**

Table 4. The Estimated MSE of Estimators for MA(1)Errors

τ	n	p	σ	γ	**GLS**	**GR**	**GL**	**GB**
0.3	50	3	1	0.70	0.0867	0.0785	0.0865	**0.0723**
				0.90	0.2141	0.1734	0.2128	**0.1450**

			5	0.70	2.1663	1.3277	2.1419	**1.0197**
				0.90	5.3515	2.7563	5.2280	**1.8998**
		7	1	0.70	0.2790	0.2208	0.2782	**0.1835**
				0.90	0.7342	0.5112	0.7288	**0.3924**
			5	0.70	6.9738	3.8294	6.8975	**2.7111**
				0.90	18.3559	9.5278	18.0433	**6.4801**
	150	3	1	0.70	0.0274	0.0263	0.0273	**0.0253**
				0.90	0.0685	0.0618	0.0684	**0.0563**
			5	0.70	0.6856	0.4885	0.6840	**0.3966**
				0.90	1.7116	1.0077	1.7009	**0.7274**
		7	1	0.70	0.0725	0.0670	0.0724	**0.0623**
				0.90	0.1912	0.1644	0.1910	**0.1436**
			5	0.70	1.8117	1.2360	1.8074	**0.9524**
				0.90	4.7797	2.9574	4.7570	**2.1402**
0.7	50	3	1	0.70	0.0670	0.0612	0.0669	**0.0564**
				0.90	0.1713	0.1419	0.1710	**0.1206**
			5	0.70	1.6757	1.0443	1.6660	**0.7998**
				0.90	4.2831	2.2736	4.2255	**1.5837**
		7	1	0.70	0.2224	0.1807	0.2222	**0.1525**
				0.90	0.5812	0.4163	0.5793	**0.3248**
			5	0.70	5.5599	3.0948	5.5270	**2.1985**
				0.90	14.5298	7.6308	14.3962	**5.2299**
	150	3	1	0.70	0.0174	0.0169	0.0173	**0.0164**
				0.90	0.0440	0.0408	0.0437	**0.0379**
			5	0.70	0.4347	0.3261	0.4344	**0.2690**
				0.90	1.0997	0.6883	1.0968	**0.5077**
		7	1	0.70	0.0461	0.0433	0.0460	**0.0409**
				0.90	0.1217	0.1070	0.1216	**0.0953**
			5	0.70	1.1520	0.8093	1.1509	**0.6333**
				0.90	3.0418	1.8833	3.0356	**1.3688**

Tables 2-4 illustrate that the increases of γ, σ and p lead to higher estimated MSE values, while the increases of n lead to lower estimated MSE values for white noise, MA(1), and, AR(1) errors. According to AR(1), and MA(1) errors, as the values of (ρ,τ) increase, the MSE values for all different values of γ, σ, p, and n decrease. As expected, when multicollinearity and correlated errors exist, the GLS estimator has the lowest performance of all estimators. By selecting these estimators for the parameters (k,d), the results indicated that the GR estimator is better than the GL estimator. In addition, the suggested GB estimator has the best performance by giving the smallest MSE values.

6. Real-life Applications:

The GLS, GR, GL, and the suggested GB estimators' performances in the MSE sense are studied in this section using two real-life datasets.

6.1 Application 1:

We first use the real-life US gross domestic product (GDP) data that is given in Gujarati (2002). And, this data was investigated by Chandra and Sarkar (2016) and Tyagi and

Chandra (2017). The quarterly US data variables are defined as GDP growth (Y), personal disposable income (X_1), personal consumption expenditure (X_2), corporate tax after profits (X_3) and net corporate dividend payments (X_4) for the years 1970–1991. The eigenvalues of $X'X$ for the standardized variables are 324.6527396, 21.8742084, 1.2365875, and 0.2364645, indicating the presence of multicollinearity in this data. Also, the value of the Durbin-Watson test of this data is calculated as 0.4784, which gives an indication of the existence of a positive autocorrelation at $\alpha = 0.05$, where the critical values are $d_L = 1.429$ and $d_U = 1.611$ for $n = 88$. The error term follows AR(1) with $\hat{\rho} = 0.7530$. So, the matrix V is constructed using (34). Also, $\hat{\sigma}^2_{GLS}$ equals 0.00292. As a result, the eigenvalues of D are $q_1 = 56.855746$, $q_2 = 7.238663$, $q_3 = 0.578735$ and $q_4 = 0.230750$. And the condition number of D is calculated as 246.3951 which means multicollinearity occurs.

The given estimated parameters as well as the estimators' MSEs are stated in Table 5.

Table 5. The regression coefficients as well as their MSEs

Coef.	**GLS**	**GR**	**GL**	**GB**
α_1	0.2073	0.2119	0.2074	0.2163
α_2	0.5930	0.5827	0.5928	0.5728
α_3	0.1054	0.1064	0.1054	0.1074
α_4	0.1087	0.1134	0.1089	0.1179
MSE	**0.0182**	**0.0173**	**0.0181**	**0.0168**
k	------	0.0083	------	0.0083
d	------	------	0.9989	0.9989

Table 5 clarifies that the GLS estimator performs the worst in the presence of multicollinearity. Also, the GR estimator is better than the GL estimator. Moreover, the lowest MSE value is for the suggested GB estimator. So, the suggested GB estimator is the best among the mentioned estimators in the presence of both autocorrelation and multicollinearity problems simultaneously.

Through our application, we verify the theoretical results with the estimated values of the parameters as follows:

-In theorem 1, the necessary condition is

$$\sum_{j=1}^{p} \frac{M_j^2 \tilde{\alpha}^2_{GLS(j)}}{R_j^2 T_j^2} = 0.0000655 < \sum_{j=1}^{p} \frac{\tilde{\sigma}^2 \left(R_j^2 T_j^2 - q_j^2 N_j^2 \right)}{q_j R_j^2 T_j^2} = 0.0019323 \; ;$$

then, $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GLS})$. That is, the suggested GB estimator outperforms the GLS estimator.

- **In theorem 2**, the necessary condition is $\sum_{j=1}^{p} \frac{\tilde{\alpha}_{GLS(j)}^2 \left(M_j^2 F_j^2 - \tilde{k}^2 R_j^2 T_j^2\right)}{R_j^2 T_j^2 F_j^2} = 0.0000486 < \sum_{j=1}^{p} \frac{\tilde{\sigma}^2 q_j \left(R_j^2 T_j^2 - N_j^2 F_j^2\right)}{R_j^2 T_j^2 F_j^2} = 0.0009335$; then, $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GR})$. That is, the suggested GB estimator outperforms the GR estimator.

- **In theorem 3**, the necessary condition is $\sum_{j=1}^{p} \frac{\tilde{\alpha}_{GLS(j)}^2 \left(M_j^2 - (1-\tilde{d})^2 T_j^2\right)}{R_j^2 T_j^2} = 0.0000655 < \sum_{j=1}^{p} \frac{\tilde{\sigma}^2 N_j^2 \left(T_j^2 - q_j^2\right)}{q_j R_j^2 T_j^2} = 0.0019028$; then, $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GL})$. That is, the suggested GB estimator outperforms the GL estimator.

6.2 Application 2:

Here, we use the data on Egypt's GDP that is taken from the World Bank, where this data is stated and investigated by Youssef (2022). The dependent variable is GDP (Y), and the explanatory variables are nitrous oxide emissions (NOE) (X_1), and methane emissions (ME) (X_2) for the years 1990–2022. The eigenvalues of $X'X$ are calculated as 1.3994e+10, and 0.0018e+10, which clarify the occurrence of multicollinearity. In addition, the Durbin-Watson value is calculated as 0.395, giving an indication of the existence of a positive autocorrelation at $\alpha = 0.05$, where $d_L = 1.321$ and $d_U = 1.577$ for $n = 33$. So, the error term follows AR(1) with $\hat{\rho} = 0.8025$. Then, the matrix V is constructed using equation (34). Hence, $\hat{\sigma}_{GLS}^2$ is equal to 1.1494e+22. The eigenvalues of D are $q_1 = 1.9235\text{e}+9$, and $q_2 = 0.0351\text{e}+9$. And the condition number of D is computed as 54.8, which means the multicollinearity still exists.

Table 6 shows the estimated parameters as well as the MSEs of the estimators.

Table 6. The regression coefficients as well as their MSEs

Coef.	**GLS**	**GR**	**GL**	**GB**
α_1	8.6804e+6	6.2889e+6	8.6804e+6	5.7257e+6
α_2	4.9690e+6	6.1661e+6	4.9691e+6	5.9177e+6
MSE	**3333.2e+11**	**237.43e+11**	**3332.9e+11**	**178.56e+11**
k	-----	1.5254e+08	-----	1.5254e+08
d	-----	-----	0.9265	0.9265

Table 6 clarifies that the GLS estimator performs the worst in the presence of multicollinearity. Also, the GR estimator is better than the GL estimator. Moreover, the lowest MSE value is for the suggested GB estimator. So, the suggested GB estimator is the best among the mentioned estimators in the presence of both autocorrelation and multicollinearity problems simultaneously.

Through our application, we verify the theoretical results with the estimated values of the parameters as follows:

- **In theorem 1**, the necessary condition is $\sum_{j=1}^{p} \frac{M_j^2 \tilde{\alpha}_{GLS(j)}^2}{R_j^2 T_j^2} = 21.03e+12 < \sum_{j=1}^{p} \frac{\tilde{\sigma}^2 \left(R_j^2 T_j^2 - q_j^2 N_j^2\right)}{q_j R_j^2 T_j^2} = 325.22e+12$;

then, $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GLS})$. That is, the suggested GB estimator outperforms the GLS estimator.

-**In theorem 2**, the necessary condition is

$$\sum_{j=1}^{p} \frac{\tilde{\alpha}^2_{GLS(j)}\left(M_j^2 F_j^2 - \tilde{k}^2 R_j^2 T_j^2\right)}{R_j^2 T_j^2 F_j^2} = 4.31e+12 < \sum_{j=1}^{p} \frac{\tilde{\sigma}^2 q_j \left(R_j^2 T_j^2 - N_j^2 F_j^2\right)}{R_j^2 T_j^2 F_j^2} = 8.36e+12$$

; then, $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GR})$. That is, the suggested GB estimator outperforms the GR estimator.

-**In theorem 3**, the necessary condition is

$$\sum_{j=1}^{p} \frac{\tilde{\alpha}^2_{GLS(j)}\left(M_j^2 - (1-\tilde{d})^2 T_j^2\right)}{R_j^2 T_j^2} = 21.02e+12 < \sum_{j=1}^{p} \frac{\tilde{\sigma}^2 N_j^2 \left(T_j^2 - q_j^2\right)}{q_j R_j^2 T_j^2} = 325.21e+12$$

; then, $MSE(\tilde{\alpha}_{GB}) < MSE(\tilde{\alpha}_{GL})$. That is, the suggested GB estimator outperforms the GL estimator.

7. Conclusion

A generalized biased (GB) regression estimator is introduced when both correlated errors and multicollinearity exist. The suggested GB estimator is investigated versus the generalized least squares (GLS), the generalized ridge (GR), and the generalized Liu (GL) estimators via mean squared error. A massive simulation has been made to show the performance of the suggested GB estimator by comparing it to other existing estimators. Clearly, under specific conditions, the suggested GB estimator has performed well when compared to other existing estimators. Two real-life datasets are also used to illustrate these ideas. The findings of this paper could be generalized to other error terms (variance-covariance structures) in other studies. Finally, the recommendation for practitioners is to use the suggested GB estimator in the presence of both autocorrelation and multicollinearity problems at the same time by selecting appropriate estimators of the parameters (k, d).

Conflicts of interest

No conflict of interest was declared by the author.

References

Aitken A.C. 1936: On least squares and linear combination of observations, Proceedings of Royal Statistical Society, Edinburgh, 55, 42–48.

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.

Alkhamisi M.A. 2010: Ridge estimation in linear models with autocorrelated errors, Communications in Statistics-Theory and Methods, 39(14), 2630–2644.

Aslam M. and Ahmad S. 2022: The modified Liu-ridge-type estimator: a new class of biased estimators to address multicollinearity, Communications in Statistics - Simulation and Computation, 51(11), 6591-6609.

Bayhan G.M. and Bayhan M. 1998: Forecasting using autocorrelated errors and multicollinear predictor variables, Computers and Industrial Engineering, 34(2), 413–421.

Chandra S. and Sarkar N. 2016: A restricted r-k class estimator in the mixed regression model with autocorrelated disturbances, Statistical Papers, 57(2), 429–449.

Chandra S. and Tyagi G. 2017: A general class of biased estimators in the presence of multicollinearity with autocorrelated errors, International Journal of Mathematics and Statistics, 18(2), 30-49.

Dawoud I. and Kibria, B.M.G. 2020: A New Biased Estimator to Combat the

Multicollinearity of the Gaussian Linear Regression Model, Stats, 3(4), 526-541.

Dawoud I. 2021a: Performance of Some New Biasing Parameter Estimates of Kibria-Lukman Estimator in Linear Regression Model, Al-Aqsa University Journal (Natural Sciences Series), 23(1), 31-49.

Dawoud I. 2021b: A new improved estimator for reducing the multicollinearity effects, Communications in Statistics –Simulation and Computation, 10.1080/03610918.2021.1939374.

Dawoud I., Lukman A.F. and Haadi A.R. 2022: A new biased regression estimator: theory, simulation and application. Scientific African, 15 (2022), p. e01100.

Eledum H.Y.A. and Alkhalifa A.A. 2012: Generalized two stages ridge regression estimator GTR for multicollinearity and autocorrelated errors, Canadian Journal on Science and Engineering Mathematics, 3, 79–85.

Gujarati D.N. 2002: Basic Econometrics, Mc-Graw Hill, NY(USA).

Güler H. and Kaçiranlar S. 2009: A comparison of mixed and ridge estimators of linear models, Communications in Statistics-Simulation and Computation, 38(2), 368–401.

Hoerl A.E. and Kennard R.W. 1970: Ridge regression: biased estimation for nonorthogonal problems, Technometrics, 12 (1), 55-67.

Huang J. and Yang H. 2015: On a principal component two-parameter estimator in linear model with autocorrelated errors, Statistical Papers, 56(1), 217-230.

Kaciranlar S. 2003: Liu estimator in the general linear regression model, Journal of Applied Statistical Science, 13, 229-234.

Kibria B.M.G. 2003: Performance of some new ridge regression estimators, Communication in Statistics- Simulation and Computation, 32(2), 419–435.

Kibria B.M.G. and Lukman A.F. 2020. A New Ridge-Type Estimator for the Linear Regression Model: Simulations and Applications, Scientifica, 2020, 1-16.

Lukman A.F. and Ayinde K. 2017: Review and classications of the ridge parameter estimation techniques, Hacettepe Journal of Mathematics and Statistics, 46(5), 953-967.

Lukman A.F., Ayinde K., Binuomote S. and Clement O.A. 2019: Modified ridge-type estimator to combat multicollinearity: Application to chemical data, Chemometrics, DOI:10.1002/cem.3125.

Liu K. 1993: A new class of biased estimate in linear regression, Communications in Statistics-Theory and Methods, 22, 393–402.

Ozkale M.R. and Kaçiranlar S. 2007: The restricted and unrestricted two-parameter estimators, Communications in Statistics-Theory and Methods, 36(15), 2707–2725.

Ozkale M.R. 2009: A stochastic restricted ridge regression estimator, Journal of Multivariate Analysis, 100(8), 1706-1716.

Qasim M., Amin M. and Omer T. 2020: Performance of some new Liu parameters for the linear regression model, Communications in Statistics - Theory and Methods, 49(17), 4178-4196.

Şiray G.U., Kaçiranlar S. and Sakallıoğlu S. 2014: r–k class estimator in the linear regression model with correlated errors, Statistical Papers, 55, 393–407.

Trenkler G. 1984: On the performance of biased estimators in the linear regression model with correlated or heteroscedastic errors, Journal of Economics, 25, 179–190.

Tyagi G. and Chandra S. 2017: A Note on the Performance of Biased Estimators with Autocorrelated Errors, International Journal of Mathematics and Mathematical Sciences, 2017, 1-12.

Youssef A.M.A. 2022: Detecting of Multicollinearity, Autocorrelation and Heteroscedasticity in Regression Analysis, Advances, 3(3), 140-152.