Asian Journal of Probability and Statistics

Volume 26, Issue 9, Page 61-83, 2024; Article no.AJPAS.122417
ISSN: 2582-0230

Addressing Autocorrelation, Multicollinearity, and Heavy-Tail Errors in the Linear Regression Model

Abdulrasheed Bello Badawaire [a*], Kayode Ayinde [b] and S. O. Olanrewaju [c]

[a] Department of Mathematics and Statistics, Federal University Wukari, Nigeria.
[b] Department of Mathematics and Statistics, Northwest Missouri State University, Maryville, MO 64468, USA.
[c] Department of Statistics, University of Abuja, Abuja, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author SOO managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ajpas/2024/v26i9646

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/122417

Original Research Article

Received: 20/06/2024
Accepted: 22/08/2024
Published: 28/08/2024

Abstract

The most popular estimator for estimating parameters of linear regression models is the Ordinary Least Squares (OLS) Estimator. The OLS is considered the best linear unbiased estimator when certain assumptions are not violated. However, when autocorrelation, multicollinearity, and heavy-tail error are jointly present in the dataset, the OLS estimator is inefficient and imprecise. In this paper, we developed an estimator of linear regression model parameters that jointly handle multicollinearity, autocorrelation, and heavy tail errors. The new estimator, LADHLKL, was derived by combining the Hildreth-Lu (HL), the Kibria Lukman (KL), and the Least Absolute Deviation (LAD) estimators. The LADHLKL poses both the

---

**Corresponding author: Email: abdulrasheed@fuwukari.edu.ng, abdulrasheedbello313@gmail.com;*

Cite as: *Badawaire, Abdulrasheed Bello, Kayode Ayinde, and S. O. Olanrewaju. 2024. "Addressing Autocorrelation, Multicollinearity, and Heavy-Tail Errors in the Linear Regression Model". Asian Journal of Probability and Statistics 26 (9):61-83. https://doi.org/10.9734/ajpas/2024/v26i9646.*

characteristics of the LAD, HL, and KL estimators which makes it resistant to both problems. We examined the properties of the proposed estimator and compared its performance with other existing estimators in terms of mean square error. An application to real-life data and simulation study revealed that the proposed estimator dominates other estimators in all the considered conditions in terms of mean square error.

Keywords: LAD estimator; Hildreth-Lu estimator; Kibria-Lukman estimator; MSE; multicollinearity; autocorrelation; heavy-tail residuals; simulation Study.

1 Introduction

Multiple linear regression is a technique in statistics that describes the relationship between the response variable and at least two other variables called explanatory variables. The OLS is the most popular estimator used in estimating the linear regression model parameters. In the linear regression model, both the assumptions of uncorrelated regressors and uncorrelated residuals can be violated jointly [1]. If these two assumptions are jointly violated, the OLS as the usual estimator used in estimating the parameters of the linear regression model will be inefficient. To jointly combat these problems, Ayinde et al., [2] Proposed some combined estimators that efficiently estimate the parameters of linear regression models in the presence of multicollinearity and autocorrelated errors. H. Y. A. Eledum & Alkhalifa, [3], H. Eledum & Zahri, [4] proposed Two Stages Ridge estimator when the assumptions of uncorrelated predictors and uncorrelated errors in the linear regression model are jointly violated. Zubair & Adenomon, [5] combined the Praise-Winston estimator with the Kibria-Lukman estimator to handle the joint effects of multicollinearity and serially correlated residuals in the linear regression model.

Multicollinearity and outliers are among the problems that are jointly encountered in regression analysis. The usual OLS estimator gives unfavorable results when multicollinearity and outliers jointly exist in data [6]. In the joint presence of outliers and multicollinearity, the ridge regression, Liu, principal components, stein estimators, etc, and other robust estimators such as M, MM, S, LTS, LMS, and LAD yield inefficient results [7]. Lukman et al., [6] proposed a two-parameter estimator called Ridge-Type Modified M-Estimator to circumvent the joint effect of multicollinearity and outliers in the linear regression model. Suhail et al., [7] also proposed some quantile-based ridge M-estimators to combat y-direction outliers and multicollinearity simultaneously under various distributions of error terms.

The threat posed by the autocorrelation becomes more complicated when this assumption violation comes together with outliers [8]. Researchers who erroneously assumed independent residuals and the complete absence of outliers obtained inefficient estimates as a consequence [8]. In the joint presence of autocorrelation and outliers, the OLS and some of its modifications are inefficient [9,10]. In regression models, it is commonly found that the response variable contains a significant number of outliers, and the residual term is serially correlated. These assumptions violations unfavorably affect model estimation (Yang et al., 2023). To address these problems, Kucuk & Asikgil, [10] proposed a robust modified two-stage least squares to handle the simultaneous effects of autocorrelation and outliers in non-linear regression.

This study aims to develop an estimator of linear regression model parameters that handle multicollinearity, autocorrelation, and heavy-tail errors jointly with the following objectives:

I. To propose an estimator that jointly handles multicollinearity, autocorrelation, and heavy-tail outliers in the residuals.
II. To examine the properties of the proposed estimator.
III. To identify the most efficient estimator among the proposed and considered estimators when autocorrelation, multicollinearity, and heavy-tail errors jointly exist.
IV. To apply the proposed estimator to real-life data.

2 Materials and Methods

Consider the linear regression model given by:

$$y = X\beta + u \tag{1}$$

Where y is a nx1 random vector of a response variable, X is a nxp matrix with full rank, β is a px1 vector of estimable parameters, and u is a nx1 random vector of residuals, distributed as $N(0, \sigma^2 I_n)$.

Let T be an orthogonal matrix, satisfying $T^T X^T XT = \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ where Λ is a diagonal matrix of order p x p with diagonal elements $\lambda_1, \lambda_2, ..., \lambda_p$ as the eigenvalues of $X^T X$. T and Λ are the matrices of eigenvectors and eigenvalues, respectively. Hence, the canonical form of model (1) is:

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

Where $Z = XT$ and $\alpha = T^T \beta$.

The OLS estimator of α is:

$$\hat{\alpha}_{OLS} = \Lambda^{-1} Z^T y \tag{3}$$

$$MSE(\hat{\alpha}_{OLS}) = \sigma^2 \Lambda^{-1} \tag{4}$$

If the residuals of model (2) above are generated by a first-order autoregressive process given by:

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{5}$$

Where $|\rho| < 1,\ t = 1, 2, 3, ..., n$, and ε_t are not necessarily normal but are independent and identically distributed.

Here, $E(u_t) = 0,$ and $E(u_t u_t^T) = \delta_u^2 \Omega \neq \delta_u^2 I$ where Ω is a positive definite symmetric matrix. Under this situation, $\hat{\alpha}_{OLS}$, although linear unbiased, is found to be inefficient.

The feasible Generalized Least Squares (FGLS) estimator suggested by (Hildreth & John, 1960) is among the appropriate estimators for estimating linear regression model parameters when the residuals are correlated. The procedure is:

$$\text{From } E(u_t u_t^{'}) = \delta_u^2 \Omega = \delta_u^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-1} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-1} & \rho^{n-1} & \cdots & 1 \end{pmatrix} \tag{6}$$

Let V be an n×n non-singular matrix such that $\Omega^{-1} = V^{'} V$ defined by:

$$V = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{pmatrix} \tag{7}$$

The variables in equation (2) above are transformed as:

$$y^* = Vy, \; Z^* = VZ, \; u^* = Vu \tag{8}$$

Hence,

$$y^* = Z^* \alpha + u^* \tag{9}$$

Using the OLS estimator, estimate the coefficients of the model in equation (9). The resulting estimator $\hat{\alpha}_{GLS}$ of α_{GLS} is called the HL estimator and is given by

$$\hat{\alpha}_{GLS} = (Z^T \Omega^{-1} Z)^{-1} Z^T \Omega^{-1} y$$
$$\hat{\alpha}_{GLS} = \Lambda^{*-1} Z^{*T} y^* \tag{10}$$

MSE of GLS:

$$MSE(\hat{\alpha}_{GLS}) = \frac{\delta^2}{\Lambda^*} \tag{11}$$

Often, the regressors involved in regression analysis are found to be linearly related. This condition renders the OLS estimator inefficient by inflating the coefficient estimates' variances, leading to erroneous conclusions. To address this problem, (B. M. Kibria & Lukman, 2020) introduced a biased estimator defined as:

$$\hat{\alpha}_{KL} = (\Lambda + kI_p)^{-1} (\Lambda - kI_p) \hat{\alpha}_{OLS}$$

Where p is the number of estimable parameters and k>0, is the biasing parameter whose estimator is given by:

$$\hat{k}_{OLS} = \min \left(\frac{\hat{\delta}^2_{OLS}}{2\hat{\alpha}^2_{OLS_j} + (\hat{\delta}^2_{OLS} / \lambda_j)} \right)$$

$$\hat{\alpha}_{KL} = R(k) H(k) \hat{\alpha}_{OLS} \tag{12}$$

Where $R(k) = (\Lambda + kI_p)^{-1}$ and $H(k) = (\Lambda - kI_p)$

Bias, variance-covariance, and Mean Square Error matrix of the KL estimator are:

$$Bias(\hat{\alpha}_{KL}) = (R(k) H(k) - I_P) \alpha \tag{13}$$

$$\text{cov}(\hat{\alpha}_{KL}) = \delta^2 R(k)H(k)\Lambda^{-1}(H(k)R(k))^T \quad (14)$$

$$MSEM(\hat{\alpha}_{KL}) = \delta^2 R(k)H(k)\Lambda^{-1}(H(k)R(k))^T + (R(k)H(k) - I_P)\alpha\alpha^T(R(k)H(k) - I_P)^T \quad (15)$$

When the residuals in Equation (2) above have a heavy-tail distribution, the usual OLS estimator will be affected. Dielman, [11] proposed the LAD estimator that minimizes the sum of the absolute deviations of the residuals. This estimator is less affected by the heavy-tailed errors than the OLS estimator in Equation (3), the FGLS estimator in Equation (10), and the KL estimator in Equation (12). The LAD estimator is defined as:

$$\hat{\alpha}_{LAD} = \min \sum_{t=1}^{n} |u_t| \quad (16)$$

The MSE of the LAD estimator is given by

$$MSE(\hat{\alpha}_{LAD}) = tr(\Omega) \text{ where } \Omega = \text{cov}(\hat{\alpha}_{LAD}) = \delta_{LAD}^2 \Lambda^{-1} \quad (17)$$

But apart from the autocorrelated residuals, if at least two of the independent variables in Equation (2) above are linearly related, then the OLS estimator in Equation (3), the FGLS estimator in Equation (10), and the KL estimator in Equation (12) would be inefficient. To simultaneously combat these two problems, (Zubair & Adenomon, 2021) suggested the generalized KL estimator given by

$$\hat{\alpha}_{GLKL} = R^*(k)H^*(k)\hat{\alpha}_{GLS}, k > 0 \quad (18)$$

Where $R^*(k) = (\Lambda^* + kI_p)^{-1}$ and $H^*(k) = (\Lambda^* - kI_p)$

Their estimator is also biased. The Bias, Covariance, and MSE matrices of their estimator are:

$$Bias(\hat{\alpha}_{GLKL}) = (R^*(k)H^*(k) - I_P)\alpha \quad (19)$$

$$\text{cov}(\hat{\alpha}_{GLKL}) = \delta^2 R^*(k)H^*(k)\Lambda^{*-1}(R^*(k)H^*(k))^T \quad (20)$$

$$MSEM(\hat{\alpha}_{GLKL}) = \delta^2 R^*(k)H^*(k)\Lambda^{*-1}(R^*(k)H^*(k))^T + (R^*(k)H^*(k) - I_P)\alpha\alpha^T(R^*(k)H^*(k) - I_P)^T \quad (21)$$

If, from the model (2), the independent variables involved are correlated and the residual term has heavy-tail outliers, the $\hat{\alpha}_{OLS}$, $\hat{\alpha}_{KL}$ and $\hat{\alpha}_{LAD}$ in Equations (3), (12), and (16) respectively will be inefficient. Majid et al., [12]. suggested a robust KL M estimator to jointly address outliers in the response variable and multicollinearity given by:

$$\hat{\alpha}_{MKL} = (\Lambda + kI)^{-1}(\Lambda - kI)\hat{\alpha}_M \quad (22)$$

Where $\hat{\alpha}_M$ is the robust-M estimator.

Bias, covariance, and the MSE matrices of their Robust KL M estimator are:

$$Bias(\hat{\alpha}_{MKL}) = (R(k)H(k) - I_p)\alpha \quad (23)$$

$$\text{cov}(\hat{\alpha}_{MKL}) = R(k)H(k)\Omega(R(k)H(k))^T \quad (24)$$

where $\Omega = \text{var}(\hat{\alpha}_M)$

$$MSEM(\hat{\alpha}_{MKL}) = R(k)H(k)\Omega(R(k)H(k))^T + (R(k)H(k) - I_P)\alpha\alpha^T (R(k)H(k) - I_P)^T \quad (25)$$

Hence, the robust KL base on LAD estimator is used to combat multicollinearity and heavy-tail outliers in the residuals simultaneously, and is given by:

$$\hat{\alpha}_{LADKL} = (\Lambda + kI)^{-1}(\Lambda - kI)\hat{\alpha}_{LAD}$$

$$\hat{\alpha}_{LADKL} = R(k)H(k)\hat{\alpha}_{LAD} \quad (26)$$

The Bias, variance-covariance, and MSE matrices of the LAD KL estimator are:

$$Bias(\hat{\alpha}_{LADKL}) = (R(k)H(k) - I_P)\alpha \quad (27)$$

$$\text{cov}(\hat{\alpha}_{LADKL}) = R(k)H(k)\Omega(R(k)H(k))^T \quad (28)$$

Where $\Omega = \text{cov}(\hat{\alpha}_{LAD}) = \delta^2_{LAD}\Lambda^{-1}$

$$MSEM(\hat{\alpha}_{LADKL}) = R(k)H(k)\Omega(R(k)H(k))^T + (R(k)H(k) - I_P)\alpha\alpha^T (R(k)H(k) - I_P)^T \quad (29)$$

However, in many reported cases where the residuals, apart from following the autoregressive of first order, have heavy-tail outliers, the $\hat{\alpha}_{OLS}$, $\hat{\alpha}_{GLS}$ and $\hat{\alpha}_{LAD}$ were found to be inefficient and imprecise. To overcome these two challenges, Robust LAD estimator is applied to the transformed data (autocorrelated free data). The robust LAD estimator was used to estimate the parameters of the model in equations (9). The resulting estimator is called the LADHL estimator and is given by:

$$\hat{\alpha}_{LADHL} = \min \sum_{t=1}^{n} \left| u_t^* \right| \quad (30)$$

Where u_t^* is the autocorrelated residual in equation (9).

The MSE of the LADHL estimator is given by

$$MSE(\hat{\alpha}_{LADHL}) = tr\left(\Omega^*\right) \text{ where } \Omega^* = \text{cov}(\hat{\alpha}_{LADHL}) = \frac{\delta^2_{LADHL}}{\Lambda^*} \quad (31)$$

2.1 The proposed estimator

$\hat{\alpha}_{LADKL}$ is resistant to both heavy-tail outliers and multicollinearity, though sensitive to autocorrelation. $\hat{\alpha}_{HLKL}$ is sensitive to heavy-tail outliers but resistant to autocorrelation and multicollinearity. $\hat{\alpha}_{LADHL}$ is resistant to heavy-tail outliers and autocorrelation but sensitive to multicollinearity. To jointly address the problems of heavy-tail errors, autocorrelation, and multicollinearity, we proposed the LAD HL KL estimator given by

$$\hat{\alpha}_{LADHLKL} = (\Lambda^* + kI_p)^{-1}(\Lambda^* - kI_p)\hat{\alpha}_{LADHL}, k > 0 \quad (32)$$

$$\hat{\alpha}_{LADHLKL} = R^{*}(k)H^{*}(k)\hat{\alpha}_{LADHL} \tag{33}$$

Where $\hat{\alpha}_{LADHL}$ is the LAD Hildreth-Lu estimator in equation (30).

2.1.1 Properties of the proposed estimator

Bias, variance-covariance, and Mean Square Error matrix of the LAD HL KL estimator are:

Bias:

$$\begin{aligned} Bias(\hat{\alpha}_{LADHLKL}) &= E[\hat{\alpha}_{LADHLKL}] - \alpha \\ &= R^{*}(k)H^{*}(k)E(\hat{\alpha}_{LADHL}) - \alpha \\ &= R^{*}(k)H^{*}(k)\alpha - \alpha \\ Bias(\hat{\alpha}_{LADHLKL}) &= (R^{*}(k)H^{*}(k) - I_{P})\alpha \end{aligned} \tag{34}$$

Variance:

$$\begin{aligned} \text{cov}(\hat{\alpha}_{LADHLKL}) &= \text{cov}(R^{*}(k)H^{*}(k)\hat{\alpha}_{LADHL}) \\ \text{cov}(\hat{\alpha}_{LADHLKL}) &= (R^{*}(k)H^{*}(k)\,\text{cov}(\hat{\alpha}_{LADHL})((R^{*}(k)H^{*}(k))^{T} \\ \text{cov}(\hat{\alpha}_{LADHLKL}) &= R^{*}(k)H^{*}(k)\Omega^{*}(R^{*}(k)H^{*}(k))^{T} \end{aligned} \tag{35}$$

Where $\Omega^{*} = \text{cov}(\hat{\alpha}_{LADHL}) = \dfrac{\delta^{2}_{LADHL}}{\Lambda^{*}}$

Mean Square Error Matrix:

The mean square error matrix is given by

$$MSEM(\hat{\alpha}_{LADHLKL}) = E((\hat{\alpha}_{LADHLKL} - \alpha)(\hat{\alpha}_{LADHLKL} - \alpha)^{T})$$

$$MSEM(\hat{\alpha}_{LADHLKL}) = \text{var}(\hat{\alpha}_{LADHLKL}) + Bias(\hat{\alpha}_{LADHLKL})Bias(\hat{\alpha}_{LADHLKL})^{T})$$

$$MSEM(\hat{\alpha}_{LADHLKL}) = R^{*}(k)H^{*}(k)\Omega^{*}(R^{*}(k)H^{*}(k))^{T} + (R^{*}(k)H^{*}(k) - I_{P})\alpha\alpha^{T}((R^{*}(k)H^{*}(k) - I_{P}))^{T} \tag{36}$$

Hence, the scalar MSE is given by

$$MMSE(\hat{\alpha}_{LADPWKL}) = tr\left[\text{var}(\hat{\alpha}_{LADHLKL}) + Bias(\hat{\alpha}_{LADHLKL})(Bias(\hat{\alpha}_{LADHLKL}))^{T}\right]$$

$$SMSE(\hat{\alpha}_{LADHLKL}) = \sum_{j=1}^{p} \frac{(\lambda_{j}^{*} - k)^{2}}{(\lambda_{j}^{*} + k)^{2}} \Omega^{*}{}_{jj} + 4k^{2} \sum_{j=1}^{p} \frac{\alpha_{j}^{2}}{(\lambda_{j}^{*} + k)^{2}} \tag{37}$$

2.2 Robust choice of the biasing parameter

Using the procedure of optimization to derive the biasing parameter of an estimator became a tradition [13]. Rewriting equation (37) as a function of k.

$$f(k) = \sum_{j=1}^{p} \frac{(\lambda_j^* - k)^2}{(\lambda_j^* + k)^2} \Omega^*_{jj} + 4k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j^* + k)^2} \tag{38}$$

The procedure is to minimize equation (38). This is achieved by setting $\frac{\partial f(k)}{\partial k} = 0$ and by proceeding will yield an estimation for the biasing parameter k which is complex. Hence, we proposed to use a robust equivalence of the biasing parameter used for the KL estimator [14]. The estimate of the shrinkage parameter used for the KL estimator is presented in equation (39).

$$\hat{k} = \min\left(\frac{\hat{\delta}^2}{2\hat{\alpha}_j^2 + (\hat{\delta}^2 / \lambda_j)} \right) \tag{39}$$

The robust equivalence of the biasing parameter in equation (39) is given by

$$\hat{k}_{LADHL} = \min\left(\frac{\hat{\sigma}^2_{LADHL}}{2\hat{\alpha}^2_{j,LADHL} + \hat{\delta}^2_{LADHL} / \lambda_j^*} \right) \tag{40}$$

Where $\hat{\alpha}_{LADHL}$ is the estimator of α_{LADHL} as given in equation (30), p is the number of explanatory variables, λ_j^* is the jth diagonal element of Λ^*, and $\hat{\sigma}^2_{LADHL} = \frac{\sum_{i=1}^{n} u^2_{LADHL}}{n-p}$, with the assumption that $\hat{\alpha}_{LADHL} \sim N\left(\alpha, \omega^2 \left(\frac{1}{n}\Lambda^*\right)^{-1}\right)$, and it holds since $\sqrt{n}(\hat{\alpha}_{LADHL} - \alpha) \rightarrow N\left(0, \omega^2 \left(\frac{1}{n}\Lambda^*\right)^{-1}\right)$ where $\omega = [2f(0)]^{-1}$ is a nuisance parameter, and $f(0)$ is the height of the density of the errors at zero, their median [15].

2.3 Monte carlo simulation

2.3.1 Simulation technique

A simulation study was performed using R software to evaluate the performances of OLS, HL, LAD, KL, LADHL, LADKL, HLKL, and LADHLKL. The regressors were generated using the procedure adopted by (Badawaire et al., [16], Gibbons, [17], B. M. G. Kibria, [18], B. M. Kibria & Lukman, [14]. The procedure is:

$$x_{tj} = \left(1 - \gamma^2\right)^{\frac{1}{2}} c_{tj} + \gamma c_{t,p+1} \tag{41}$$

$$t = 1, 2, \ldots, n, \; j = 1, 2, \ldots, p,$$

Where c_{tj} are the pseudo-random numbers from the standard normal distribution and γ denote the correlation between any two explanatory variables which are set as γ = 0.85, 0.95, and 0.99, p is the number of the regressors considered, which is 3. The variables are standardized such that $X^T X$ and $X^T y$ are in correlation forms. The n observations of the response variable are generated by

$$y_i = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_p x_{tp} + u_t \tag{42}$$

$$t = 1, 2, ..., n,$$
$$u_t = \rho u_{t-1} + \varepsilon_t \text{ such that } u_t \sim N(0, \delta^2 \Omega) \tag{43}$$

Where ρ is the autocorrelation coefficient and was set to ρ = 0.85, 0.95, and 0.99, and the error variance was set as δ_u^2 =0.5, 1, and 4. For the model, we assume zero intercept, and the values of the coefficients are selected such that $\beta^T \beta = 1$.

Outliers are being introduced by increasing the magnitude of the tail values of ε_t. The procedure is

$$\varepsilon_{(t:1 \le t < (n - n_{outlier}))} + \left(25 * \max\left(\varepsilon_t\right) + \varepsilon_{(t:(n - n_{outlier}) \le t \le n)}\right) \tag{44}$$

Where $n_{outlietr}$ is the number of outliers along the tail of ε_t.

Three (3) sample sizes were used: n = 10, 30, and 50, and three (3) levels of percentages of outliers were considered ($n_{outlier}$%= 10%, 20%, 30%).

By implication, 1000 replications of these experiments are carried out and the MSE is estimated as

$$MSE(\hat{\alpha}) = \frac{1}{1000} \sum_{j=1}^{1000} (\hat{\alpha}_{tj} - \alpha_t)^T (\hat{\alpha}_{tj} - \alpha_t) \tag{45}$$

3 Results and Discussion

3.1 Simulation results and discussion

The Monte Carlo simulation's results of the estimators are presented in Tables 1-9.

The EMSE of the OLS, HL, LAD, LADHL, KL, HLKL, LADKL, and LADHLKL estimators for the various factors taken into consideration in this investigation are shown in Tables 1–9. The error variance values, sample sizes, autocorrelation levels, multicollinearity, and outlier percentages are among the variables taken into account. It has been noted that these variables have an impact on the simulation design. In particular, we found that:

1. When autocorrelation and heavy tail outliers in the residuals coexist, the increase in the level of multicollinearity negatively affects the performances of OLS and HL, while having the least impact on LADHL, LADKL, and the proposed LADHLKL. But in terms of minimal MSE, the suggested LADHLKL outperformed all the estimators that were taken into consideration.
2. Except the LADHL and the proposed LADHLKL estimators, whose MSE decreases as the level of autocorrelation rises from 0.85 to 0.95 across sample sizes and percentages of outliers considered, and the HL estimator, whose MSE decreases as the level of autocorrelation increases from 0.85, 0.95, to 0.99 when the sample size is 50 and the percentage of outliers is 30, the estimators' performances deteriorate as the level of autocorrelation rises (from 0.85 to 0.95, to 0.99) in the joint presence of multicollinearity and heavy-tail outliers.
3. All estimators perform worse as the percentage of outliers rises, though not always. This is true for all levels of multicollinearity, autocorrelation, and error variance values. When it came to the minimal MSE, the suggested LADHLKL fared better than alternative estimators, with the LADKL estimator being a close second.

4. The estimators' Performance improved as the sample size increased from 10, 30, and 50, except for the proposed LADHLKL estimator when the values of error variance are 0.5 and 1 and the level of multicollinearity is 0.85, whose EMSE increases as sample size increases from 30 to 50 across all the percentages of outliers and levels of autocorrelation. However, in comparison with other estimators, the proposed LADHLKL was still better.
5. An increase in the error variance values generally hurts the estimators' performances. As the error variance goes from 0.5, 1 to 4, the EMSE of each estimator increases. The least impacted are the LADKL, LADHL, and the proposed LADHLKL. However, the suggested LADHLKL has shown its superiority in terms of lower MSE values.

3.2 Numerical application

We used the dataset that was originally used by (H. Y. A. Eledum & Alkhalifa, 2012). Though we considered the 1960 to 1975 period for the data, (Awwad et al., [19], H. Eledum & Zahri, [4], Lukman et al., [20], Lukman, Ayinde, Kun, et al., [21], Lukman, Ayinde, Olatayo, et al., [22] used the same dataset for the period from 1960 to 1990. The regression model is given as follows:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \tag{46}$$

Where y_t represents the product in the manufacturing sector, X_{1t} is the imported intermediate commodities, X_{2t} is the imported capital commodities, and X_{3t} is the imported raw materials.

The method of Ordinary Least Squares was used to estimate the parameters of model (46). Based on the fitted model, the following statistics were obtained:

The correlation matrix of the predictors is:

$$r_X = \begin{pmatrix} 1 & 0.9617 & 0.9660 \\ & 1 & 0.8721 \\ & & 1 \end{pmatrix}$$

From the correlation matrix above, it can be seen that the estimated correlations between the independent variables γ_j are more than 0.870 and the Variance Inflation Factor VIF_j for j = 1, 2, 3, as can be seen from Table 10 above are all greater than 30. This implies that the data suffer from multicollinearity.

Also, the Durbin-Watson statistic value DW (0.803) <dl (0.860), with a p-value of 0.001. This implies that the model suffers from autoregressive of order 1 (AR (1)).

Table 10 above also shows that the value of the Kurtosis for the residuals, K is greater than three (K>3), which indicate that the residuals is heavily skewed to the right.

Therefore, apart from having a heavy-tail residual, the data suffers from both multicollinearity and autocorrelation.

We used the proposed LADHLKL and seven other considered estimators to estimate the coefficients of the model (46). The estimates of the regression coefficients, their corresponding AMSE values, and ranks are presented in Table 11.

The results of the real-life application show that our proposed LADHLKL estimator had the smallest MSE value and therefore, performed better than other considered estimators. This result is consistent with the simulation results [23].

Table 1. Estimated MSE values when the error variance = 0.5, n = 10

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	171.850	167.313	3.948	2.979	57.120	52.581	0.854	0.545
	0.85	0.950	521.054	515.607	11.095	8.751	170.142	159.604	2.027	1.376
		0.990	2662.446	2640.594	56.205	46.216	854.159	805.721	9.329	7.371
		0.850	172.782	167.744	10.752	7.225	56.735	51.954	3.358	2.351
10%	0.95	0.950	524.316	514.052	26.886	19.562	170.030	155.752	7.022	5.963
		0.990	2682.937	2628.849	133.217	95.701	859.987	787.690	29.946	28.565
		0.850	173.945	172.518	53.383	29.718	53.625	51.701	21.682	14.212
	0.99	0.950	528.623	523.409	127.379	80.150	163.743	152.859	45.126	37.759
		0.990	2720.707	2660.798	544.818	390.588	851.880	759.942	167.828	185.367
		0.850	687.952	197.826	4.305	1.659	223.138	28.763	0.801	0.150
	0.85	0.950	2003.240	508.808	12.276	5.101	622.531	59.613	2.043	0.332
		0.990	9945.791	2283.137	61.016	25.098	2986.876	223.022	9.461	1.263
		0.850	758.495	233.403	11.567	1.601	247.299	35.392	2.916	0.144
20%	0.95	0.950	2214.230	611.247	31.043	4.819	693.740	76.874	7.165	0.315
		0.990	11014.820	2778.822	150.442	23.366	3343.373	298.453	32.651	1.279
		0.850	787.724	248.173	56.008	2.021	253.732	38.246	18.573	0.212
	0.99	0.950	2304.366	654.002	145.678	6.105	719.248	84.430	45.017	0.392
		0.990	11493.084	2986.618	702.271	29.184	3507.063	331.958	209.842	1.561
		0.850	1421.815	391.260	4.753	2.545	416.174	86.622	1.015	0.238
	0.85	0.950	4235.762	1184.063	12.399	12.064	1217.035	261.653	2.246	0.806
		0.990	21416.884	6008.384	61.372	92.610	6056.272	1317.118	10.300	5.616
		0.850	1724.258	498.788	12.788	2.344	514.410	114.816	3.382	0.234
30%	0.95	0.950	5134.572	1497.830	30.468	13.066	1502.639	341.403	7.018	1.126
		0.990	25950.784	7554.185	143.468	98.629	7471.364	1696.965	31.518	8.427
		0.850	1858.946	548.156	60.894	2.657	558.709	127.776	19.151	0.306
	0.99	0.950	5539.193	1641.920	141.751	24.052	1639.743	377.947	40.013	3.142
		0.990	28020.038	8263.977	653.303	172.310	8194.601	1870.526	180.424	25.883

Table 2. Estimated MSE values when the error variance = 0.5, n = 30

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	106.719	8.108	1.265	0.155	35.820	1.175	0.112	0.033
	0.85	0.950	201.888	23.009	3.373	0.411	45.893	3.302	0.233	0.023
		0.990	752.410	110.860	15.865	1.943	120.280	15.718	0.880	0.077
		0.850	125.960	10.229	5.255	0.149	40.749	1.578	1.115	0.032
10%	0.95	0.950	247.390	29.720	14.075	0.401	58.534	4.882	2.378	0.021
		0.990	951.318	145.659	65.821	1.916	180.921	25.252	9.704	0.071
		0.850	138.080	11.381	35.785	0.180	46.359	2.020	16.982	0.035
	0.99	0.950	280.813	33.300	90.733	0.493	75.824	6.327	36.016	0.026
		0.990	1111.871	164.108	397.612	2.378	251.031	32.913	144.675	0.093
		0.850	316.520	11.455	1.288	0.185	93.776	1.652	0.071	0.037
	0.85	0.950	475.720	33.685	3.402	0.485	106.222	4.438	0.119	0.022
		0.990	1669.627	168.706	16.217	2.300	236.746	22.500	0.425	0.063
		0.850	491.612	14.656	5.504	0.179	146.191	1.884	0.312	0.036
20%	0.95	0.950	760.786	43.963	14.313	0.470	167.560	5.033	0.557	0.020
		0.990	2665.315	222.332	68.524	2.229	395.432	25.812	2.105	0.058
		0.850	591.930	16.742	33.414	0.229	179.059	2.137	4.513	0.040
	0.99	0.950	936.987	50.667	89.597	0.609	205.221	5.678	9.848	0.025
		0.990	3315.553	257.444	394.421	2.888	496.005	29.182	32.877	0.076
		0.850	997.066	2.946	2.147	0.203	456.666	0.198	0.139	0.028
	0.85	0.950	2507.966	7.638	6.358	0.537	671.345	0.480	0.224	0.018
		0.990	12984.931	36.108	31.029	2.537	3104.076	1.611	0.571	0.061
		0.850	1564.191	7.528	6.349	0.194	737.149	0.420	0.604	0.028
30%	0.95	0.950	3604.272	20.622	18.923	0.512	1051.750	1.261	0.880	0.018
		0.990	18017.443	98.269	90.898	2.424	3354.627	5.685	1.809	0.057
		0.850	1926.157	12.207	28.125	0.232	909.472	0.678	4.596	0.029
	0.99	0.950	4270.783	34.247	102.255	0.621	1287.733	2.047	9.733	0.021
		0.990	20926.636	164.958	500.156	2.947	4039.963	9.707	33.906	0.074

Table 3. Estimated MSE values when the error variance = 0.5, n = 50

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	38.623	1.488	0.413	0.060	3.906	0.265	0.063	0.036
	0.85	0.950	109.631	3.961	1.112	0.162	10.967	0.278	0.039	0.014
		0.990	535.311	18.845	5.531	0.778	53.133	0.710	0.081	0.018
		0.850	57.477	2.030	1.378	0.055	6.745	0.338	0.161	0.042
10%	0.95	0.950	163.408	5.438	4.122	0.148	18.900	0.362	0.208	0.013
		0.990	799.318	26.169	21.668	0.718	91.837	0.966	0.798	0.014
		0.850	69.634	2.339	11.048	0.063	9.488	0.376	2.595	0.047
	0.99	0.950	197.702	6.297	37.641	0.168	26.334	0.409	8.371	0.014
		0.990	968.013	30.487	201.833	0.794	128.207	1.126	45.023	0.017
		0.850	44.017	0.439	0.450	0.069	12.604	0.027	0.057	0.051
	0.85	0.950	86.516	1.404	1.224	0.188	24.467	0.033	0.032	0.012
		0.990	308.820	7.369	6.113	0.911	35.699	0.182	0.065	0.014
		0.850	54.261	1.357	1.355	0.063	18.454	0.056	0.138	0.053
20%	0.95	0.950	104.599	4.259	4.451	0.173	34.586	0.074	0.122	0.011
		0.990	371.003	22.282	23.383	0.830	48.796	0.433	0.316	0.012
		0.850	67.573	2.267	10.716	0.072	21.568	0.090	0.936	0.053
	0.99	0.950	138.918	7.037	40.476	0.196	39.657	0.127	1.827	0.012
		0.990	532.447	36.671	220.367	0.942	55.551	0.727	9.034	0.013
		0.850	474.990	1.916	0.608	0.078	69.733	0.285	0.044	0.057
	0.85	0.950	1205.431	4.757	1.742	0.215	148.007	0.236	0.034	0.014
		0.990	5438.885	21.543	8.584	1.029	458.181	0.643	0.089	0.014
		0.850	696.372	0.753	1.531	0.074	99.619	0.340	0.140	0.058
30%	0.95	0.950	1712.754	1.362	6.097	0.205	194.927	0.159	0.146	0.013
		0.990	7547.278	5.282	28.265	1.000	445.363	0.127	0.349	0.013
		0.850	793.216	0.431	8.685	0.087	121.058	0.426	1.183	0.060
	0.99	0.950	1911.602	0.450	47.519	0.239	230.985	0.194	1.239	0.015
		0.990	8291.114	1.555	213.142	1.157	454.620	0.095	3.456	0.015

Table 4. Estimated MSE values when the error variance = 1, n = 10

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
10%	0.85	0.850	343.699	334.626	7.896	5.958	114.240	105.132	1.707	1.090
		0.950	1,042.108	1,031.215	22.190	17.502	340.283	319.193	4.053	2.752
		0.990	5,324.892	5,281.187	112.411	92.432	1,708.318	1,611.434	18.658	14.742
	0.95	0.850	345.564	335.488	21.505	14.449	113.469	103.879	6.717	4.703
		0.950	1,048.632	1,028.104	53.772	39.124	340.060	311.489	14.045	11.928
		0.990	5,365.874	5,257.699	266.435	191.402	1,719.973	1,575.373	59.892	57.132
	0.99	0.850	347.890	345.036	106.766	59.435	107.250	103.371	43.366	28.428
		0.950	1,057.246	1,046.818	254.757	160.300	327.486	305.704	90.253	75.521
		0.990	5,441.413	5,321.595	1,089.635	781.176	1,703.759	1,519.877	335.657	370.736
20%	0.85	0.850	1,375.904	395.653	8.609	3.317	446.291	57.524	1.605	0.302
		0.950	4,006.481	1,017.617	24.553	10.201	1,245.067	119.224	4.086	0.665
		0.990	19,891.582	4,566.275	122.033	50.196	5,973.754	446.043	18.922	2.526
	0.95	0.850	1,516.990	466.806	23.134	3.203	494.613	70.777	5.835	0.291
		0.950	4,428.459	1,222.494	62.086	9.638	1,387.484	153.742	14.332	0.631
		0.990	22,029.640	5,557.644	300.884	46.732	6,686.747	596.904	65.301	2.558
	0.99	0.850	1,575.449	496.346	112.015	4.041	507.479	76.482	37.155	0.427
		0.950	4,608.731	1,308.003	291.356	12.209	1,438.502	168.853	90.037	0.784
		0.990	22,986.168	5,973.235	1,404.542	58.368	7,014.128	663.914	419.684	3.122
30%	0.85	0.850	2,843.630	782.519	9.506	5.091	832.366	173.200	2.031	0.471
		0.950	8,471.523	2,368.125	24.798	24.127	2,434.076	523.278	4.492	1.608
		0.990	42,833.768	12,016.768	122.743	185.220	12,112.544	2,634.218	20.599	11.230
	0.95	0.850	3,448.516	997.576	25.577	4.688	1,028.840	229.585	6.765	0.463
		0.950	10,269.144	2,995.660	60.936	26.132	3,005.284	682.777	14.036	2.248
		0.990	51,901.567	15,108.369	286.936	197.259	14,942.728	3,393.912	63.037	16.852
	0.99	0.850	3,717.893	1,096.313	121.788	5.315	1,117.439	255.503	38.309	0.611
		0.950	11,078.386	3,283.840	283.502	48.104	3,279.493	755.863	80.027	6.280
		0.990	56,040.075	16,527.954	1,306.606	344.620	16,389.203	3,741.034	360.849	51.762

Table 5. Estimated MSE values when the error variance = 1, n = 30

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	213.438	16.216	2.531	0.310	71.912	2.343	0.224	0.045
	0.85	0.950	403.777	46.018	6.747	0.823	91.903	6.612	0.472	0.044
		0.990	1504.819	221.720	31.731	3.887	240.585	31.446	1.761	0.155
		0.850	251.919	20.458	10.509	0.299	81.791	3.143	2.250	0.043
10%	0.95	0.950	494.780	59.440	28.151	0.803	117.188	9.774	4.767	0.039
		0.990	1902.636	291.319	131.642	3.831	361.867	50.517	19.410	0.141
		0.850	276.160	22.762	71.570	0.360	93.018	4.025	34.020	0.048
	0.99	0.950	561.625	66.600	181.466	0.987	151.764	12.666	72.052	0.050
		0.990	2223.741	328.216	795.224	4.755	502.086	65.842	289.352	0.187
		0.850	633.040	22.909	2.576	0.370	188.579	3.223	0.140	0.050
	0.85	0.950	951.440	67.370	6.805	0.970	212.799	8.844	0.242	0.040
		0.990	3339.255	337.412	32.433	4.601	473.558	44.984	0.851	0.126
		0.850	983.224	29.311	11.008	0.358	293.548	3.679	0.647	0.049
20%	0.95	0.950	1521.572	87.926	28.626	0.940	335.536	10.038	1.126	0.037
		0.990	5330.629	444.663	137.049	4.459	790.943	51.614	4.211	0.115
		0.850	1183.860	33.483	66.828	0.457	359.325	4.180	9.106	0.055
	0.99	0.950	1873.974	101.334	179.193	1.218	410.880	11.331	19.729	0.047
		0.990	6631.106	514.889	788.841	5.776	992.094	58.359	65.757	0.151
		0.850	1994.133	5.892	4.293	0.405	914.695	0.451	0.313	0.038
	0.85	0.950	5015.932	15.277	12.717	1.074	1343.140	0.988	0.464	0.036
		0.990	25969.863	72.216	62.059	5.073	6208.221	3.233	1.145	0.122
		0.850	3128.382	15.056	12.697	0.387	1476.102	0.884	1.302	0.037
30%	0.95	0.950	7208.544	41.244	37.846	1.023	2104.096	2.554	1.797	0.034
		0.990	36034.887	196.538	181.795	4.848	6709.366	11.386	3.626	0.114
		0.850	3852.314	24.414	56.250	0.464	1820.974	1.388	9.461	0.039
	0.99	0.950	8541.566	68.493	204.511	1.242	2576.137	4.127	19.562	0.042
		0.990	41853.273	329.916	1000.313	5.893	8080.052	19.435	67.831	0.148

Table 6. Estimated MSE values when the error variance = 1, n = 50

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	77.247	2.977	0.826	0.119	7.802	0.440	0.085	0.056
	0.85	0.950	219.262	7.923	2.223	0.323	21.931	0.527	0.073	0.021
		0.990	1070.621	37.689	11.063	1.556	106.265	1.415	0.163	0.035
		0.850	114.955	4.060	2.756	0.109	13.459	0.577	0.290	0.059
10%	0.95	0.950	326.817	10.876	8.243	0.296	37.791	0.695	0.415	0.019
		0.990	1598.636	52.339	43.335	1.435	183.673	1.929	1.597	0.027
		0.850	139.269	4.679	22.097	0.126	18.941	0.648	5.194	0.061
	0.99	0.950	395.405	12.595	75.282	0.335	52.658	0.787	16.743	0.021
		0.990	1936.026	60.974	403.667	1.588	256.412	2.248	90.047	0.033
		0.850	88.035	0.878	0.900	0.138	26.097	0.028	0.076	0.056
	0.85	0.950	173.032	2.808	2.448	0.376	49.383	0.075	0.065	0.017
		0.990	617.640	14.738	12.226	1.822	71.509	0.370	0.132	0.027
		0.850	108.521	2.713	2.711	0.126	37.989	0.077	0.246	0.056
20%	0.95	0.950	209.197	8.517	8.902	0.346	69.706	0.153	0.249	0.016
		0.990	742.007	44.565	46.766	1.659	97.721	0.871	0.635	0.023
		0.850	135.147	4.534	21.432	0.143	44.280	0.135	1.865	0.057
	0.99	0.950	277.836	14.074	80.951	0.392	79.873	0.257	3.660	0.017
		0.990	1064.894	73.342	440.733	1.884	111.238	1.460	18.072	0.026
		0.850	949.980	3.832	1.216	0.156	140.636	0.471	0.060	0.063
	0.85	0.950	2410.863	9.514	3.484	0.430	296.561	0.446	0.072	0.020
		0.990	10877.771	43.086	17.168	2.058	916.494	1.280	0.181	0.028
		0.850	1392.745	1.506	3.063	0.148	201.126	0.540	0.256	0.063
30%	0.95	0.950	3425.508	2.724	12.193	0.410	390.759	0.283	0.306	0.019
		0.990	15094.555	10.564	56.530	1.999	890.946	0.249	0.703	0.025
		0.850	1586.432	0.863	17.370	0.175	244.427	0.690	2.336	0.067
	0.99	0.950	3823.204	0.900	95.038	0.478	463.092	0.347	2.525	0.022
		0.990	16582.227	3.110	426.284	2.314	909.514	0.185	6.927	0.030

Table 7. Estimated MSE values when the error variance = 4, n = 10

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	1374.797	1338.503	31.584	23.831	456.955	420.378	6.827	4.358
	0.85	0.950	4168.430	4124.858	88.758	70.010	1361.132	1276.700	16.214	11.011
		0.990	21299.568	21124.750	449.642	369.730	6833.273	6445.699	74.632	58.972
		0.850	1382.255	1341.952	86.020	57.797	453.872	415.369	26.871	18.822
10%	0.95	0.950	4194.528	4112.418	215.087	156.496	1360.239	1245.887	56.181	47.722
		0.990	21463.498	21030.794	1065.738	765.607	6879.894	6301.458	239.567	228.536
		0.850	1391.561	1380.145	427.066	237.742	428.995	413.336	173.476	113.733
	0.99	0.950	4228.984	4187.274	1019.029	641.201	1309.946	1222.746	361.017	302.098
		0.990	21765.654	21286.381	4358.542	3124.703	6815.038	6079.475	1342.628	1482.957
		0.850	5503.615	1582.610	34.437	13.269	1785.234	230.089	6.429	1.218
	0.85	0.950	16025.923	4070.466	98.211	40.805	4980.290	476.886	16.347	2.662
		0.990	79566.328	18265.099	488.131	200.783	23895.019	1784.165	75.690	10.105
		0.850	6067.959	1867.223	92.536	12.811	1978.524	283.076	23.360	1.175
20%	0.95	0.950	17713.837	4889.978	248.343	38.551	5549.956	614.946	57.332	2.530
		0.990	88118.561	22230.575	1203.538	186.927	26746.992	2387.602	261.206	10.232
		0.850	6301.794	1985.386	448.061	16.165	2029.989	305.887	148.663	1.728
	0.99	0.950	18434.925	5232.012	1165.422	48.836	5754.030	675.386	360.158	3.141
		0.990	91944.672	23892.941	5618.168	233.472	28056.516	2655.638	1678.738	12.488
		0.850	11374.521	3130.077	38.024	20.363	3329.553	692.588	8.127	1.863
	0.85	0.950	33886.093	9472.502	99.193	96.509	9736.329	2092.980	17.967	6.417
		0.990	171335.071	48067.071	490.973	740.880	48450.183	10536.789	82.397	44.910
		0.850	13794.064	3990.305	102.306	18.753	4115.452	918.107	27.073	1.834
30%	0.95	0.950	41076.574	11982.639	243.745	104.527	12021.163	2730.966	56.145	8.974
		0.990	207606.268	60433.476	1147.746	789.034	59770.918	13575.558	252.147	67.394
		0.850	14871.570	4385.251	487.151	21.259	4469.855	1021.769	153.275	2.432
	0.99	0.950	44313.543	13135.361	1134.008	192.417	13117.999	3023.303	320.116	25.098
		0.990	224160.301	66111.817	5226.425	1378.480	65556.817	14964.044	1443.396	207.034

Table 8. Estimated MSE values when the error variance = 4, n = 30

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	853.753	64.865	10.123	1.239	288.966	9.360	0.945	0.116
	0.85	0.950	1615.107	184.074	26.988	3.292	368.180	26.490	1.917	0.169
		0.990	6019.277	886.882	126.923	15.547	962.463	125.826	7.049	0.621
		0.850	1007.677	81.834	42.038	1.195	328.586	12.528	9.135	0.108
10%	0.95	0.950	1979.121	237.762	112.603	3.210	469.336	39.149	19.125	0.151
		0.990	7610.545	1165.275	526.567	15.326	1447.590	202.134	77.649	0.567
		0.850	1104.639	91.048	286.278	1.441	373.532	16.042	136.371	0.128
	0.99	0.950	2246.501	266.401	725.865	3.946	607.619	50.719	288.306	0.193
		0.990	8894.966	1312.862	3180.895	19.021	2008.460	263.440	1157.422	0.751
		0.850	2532.158	91.638	10.305	1.480	759.294	12.513	0.597	0.123
	0.85	0.950	3805.760	269.479	27.218	3.880	852.919	35.221	0.993	0.150
		0.990	13357.019	1349.648	129.733	18.402	1894.552	179.857	3.410	0.502
		0.850	3932.897	117.245	44.031	1.430	1179.838	14.306	2.748	0.118
20%	0.95	0.950	6086.290	351.705	114.505	3.758	1344.155	40.023	4.568	0.138
		0.990	21322.518	1778.653	548.195	17.835	3164.150	206.410	16.851	0.462
		0.850	4735.439	133.934	267.313	1.829	1443.153	16.290	36.851	0.138
	0.99	0.950	7495.895	405.337	716.773	4.874	1645.641	45.208	79.079	0.176
		0.990	26524.426	2059.555	3155.366	23.102	3968.776	233.411	263.044	0.605
		0.850	7976.531	23.566	17.174	1.621	3665.370	2.107	1.472	0.097
	0.85	0.950	20063.727	61.106	50.866	4.296	5374.728	4.090	1.938	0.140
		0.990	103879.450	288.865	248.235	20.294	24833.214	12.988	4.597	0.488
		0.850	12513.530	60.223	50.789	1.550	5913.133	3.780	5.710	0.093
30%	0.95	0.950	28834.174	164.975	151.384	4.093	8419.261	10.373	7.369	0.132
		0.990	144139.548	786.150	727.181	19.392	26838.000	45.630	14.537	0.459
		0.850	15409.257	97.654	224.999	1.856	7293.708	5.745	39.187	0.108
	0.99	0.950	34166.264	273.973	818.043	4.968	10307.789	16.672	78.719	0.165
		0.990	167413.091	1319.664	4001.251	23.572	32320.817	77.843	271.415	0.594

Table 9. Estimated MSE values when the error variance = 4, n = 50

% of Outlier	ρ	γ	OLS	HL	LAD	LADHL	KL	HLKL	LADKL	LADHLKL
		0.850	308.986	11.906	3.303	0.478	31.183	1.376	0.228	0.086
	0.85	0.950	877.050	31.691	8.893	1.293	87.718	1.982	0.287	0.060
		0.990	4282.485	150.756	44.251	6.223	425.057	5.638	0.653	0.140
		0.850	459.819	16.240	11.025	0.437	53.712	1.882	1.082	0.084
10%	0.95	0.950	1307.266	43.502	32.974	1.184	151.123	2.642	1.663	0.052
		0.990	6394.542	209.355	173.342	5.742	734.682	7.696	6.390	0.107
		0.850	557.076	18.715	88.386	0.502	75.620	2.149	20.858	0.090
	0.99	0.950	1581.620	50.379	301.128	1.342	210.583	3.009	66.985	0.061
		0.990	7744.104	243.897	1614.668	6.352	1025.636	8.976	360.192	0.128
		0.850	352.139	3.512	3.599	0.550	108.772	0.070	0.217	0.076
	0.85	0.950	692.130	11.233	9.793	1.504	199.716	0.356	0.269	0.048
		0.990	2470.559	58.952	48.906	7.288	286.574	1.505	0.535	0.109
		0.850	434.085	10.854	10.843	0.505	157.267	0.208	0.928	0.074
20%	0.95	0.950	836.789	34.068	35.608	1.385	281.413	0.649	1.029	0.044
		0.990	2968.028	178.258	187.066	6.636	391.511	3.509	2.551	0.091
		0.850	540.588	18.135	85.730	0.573	182.732	0.399	7.513	0.078
	0.99	0.950	1111.345	56.295	323.806	1.568	322.203	1.050	14.677	0.050
		0.990	4259.578	293.370	1762.933	7.535	445.604	5.866	72.310	0.104
		0.850	3799.922	15.326	4.864	0.623	568.242	1.470	0.196	0.090
	0.85	0.950	9643.451	38.055	13.936	1.719	1188.895	1.667	0.321	0.055
		0.990	43511.083	172.345	68.672	8.231	3666.614	5.092	0.735	0.109
		0.850	5570.979	6.025	12.250	0.593	813.682	1.577	1.002	0.088
30%	0.95	0.950	13702.031	10.895	48.773	1.641	1567.413	0.976	1.309	0.051
		0.990	60378.220	42.256	226.121	7.996	3564.847	0.971	2.838	0.099
		0.850	6345.727	3.451	69.480	0.699	988.942	2.077	9.293	0.099
	0.99	0.950	15292.814	3.600	380.153	1.911	1857.800	1.207	10.343	0.062
		0.990	66328.908	12.439	1705.136	9.258	3639.382	0.713	27.781	0.116

Table 10. The value of the Kurtosis for the residuals

α	K	n	p	$F_{(3,12)}$	R^2	$\hat{\sigma}^2$	dl	du	DW	VIF_1	VIF_2	VIF_3
0.05	3.248	16	3	19.950	0.833	7564.42	0.860	1.730	0.803	133.709	37.380	41.928

Table 11. Coefficient estimates, MSE, and ranks of the proposed and considered estimators

Estimator	Coefficients Estimates				AMSE	Rank
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$		
OLS	151.916796	1.9391019	-0.3409638	0.09901955	3223.63517	8
HL	44.0793292	0.27069863	0.40093283	0.0379899	57.8128197	3
KL	132.205018	2.65452496	-0.6078002	-0.9097097	2439.3762	7
LAD	53.91659	5.992997	-1.63648	-6.89525	481.907	6
HLKL	42.8024969	0.28411587	0.39876059	0.03093152	213.532204	4
LADHL	33.78182	0.110542	0.469087	0.398365	49.703	2
LADKL	19.7427914	7.22685362	-2.0970903	-8.631091	352.453154	5
HLLADKL	32.5811165	0.11802228	0.46912318	0.39889461	40.1687232	1

4 Conclusion

The assumptions of the absence of multicollinearity among the regressors, independent and normally distributed residuals are sometimes jointly violated in multiple linear regression models, yet have not been sufficiently studied.

In this research, we introduced a Robust Feasible Generalized Kibria-Lukman (LADHLKL) estimator to mitigate the joint effect of multicollinearity, autocorrelation, and heavy-tail outliers in residuals in the linear regression models. We derived the statistical properties of the proposed estimator. We compared the performance of the proposed LADHLKL estimator with other existing estimators based on the mean square error (MSE) criterion through simulation and real-life application. Both the results from the numerical example and the simulation study showed that our proposed estimator outperformed other estimators. The proposed estimator is, therefore, recommended for researchers in handling the joint effect of multicollinearity, autocorrelation, and heavy-tail errors in a linear regression model. For future research, we plan to extend this study to generalized linear models.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during the writing or editing of manuscripts.

Competing Interests

The authors have declared that no competing interests exist.

References

[1] Firinguetti LL. A simulation study of ridge regression estimators with autocorrelated errors. Communications in Statistics-Simulation and Computation. 1989;18(2):673–702.

[2] Ayinde K, Lukman AF, Arowolo OT. Combined parameters estimation methods of linear regression model with multicollinearity and autocorrelation. Journal of Asian Scientific Research. 2015;5(5):243–250.

[3] Eledum HYA, Alkhalifa AA. Generalized two stages ridge regression estimator GTR for multicollinearity and autocorrelated errors. Canadian Journal on Science and Engineering Mathematics. 2012;3(3): 79–83.

[4] Eledum H, Zahri M. Relaxation method for two stages ridge regression estimator. International Journal of Pure and Applied Mathematics. 2013;85(4):653–667.

[5] Zubair MA, Adenomon MO. Comparison of estimators efficiency for linear regressions with joint presence of autocorrelation and multicollinearity. Science World Journal. 2021;16(2):103–109.

[6] Lukman AF, Ayinde K, Kibria G, Jegede SL. Two-parameter modified ridge-type m-estimator for linear regression model. The Scientific World Journal; 2020.

[7] Suhail M, Chand S, Babar I. Pak. J. Statist. On some new ridge m-estimators for linear regression models under various error distributions. Pak. J. Statist. 2021;37(4): 369–391.

[8] Riazoshams H, Midi HB, Sharipov OS. The performance of robust two-stage estimator in nonlinear regression with autocorrelated error. Communications in Statistics-Simulation and Computation. 2010;39(6), 1251–1268.

[9] Kucuk S, Asikgil B. A novel hybrid robust tapering approach for nonlinear regression in the presence of autocorrelation and outliers. Communications in Statistics-Simulation and Computation. 2021;1–17.

[10] Kucuk S, Asikgil, B. A novel hybrid robust tapering approach for nonlinear regression in the presence of autocorrelation and outliers. Communications in Statistics-Simulation and Computation. 2023;52(11):5550–5566.

[11] Dielman TE. Least absolute value estimation in regression models: An annotated bibliography. Communications in Statistics-Theory and Methods. 1984;13(4):513–541.

[12] Majid A, Ahmad S, Aslam M, Kashif M. A robust Kibria–Lukman estimator for linear regression model to combat multicollinearity and outliers. Concurrency and Computation: Practice and Experience. 2023; 35(4):e7533.

[13] Arum KC, Ugwuowo FI, Oranye HE, Alakija TO, Ugah TE, Asogwa OC. Combating outliers and multicollinearity in linear regression model using robust Kibria-Lukman mixed with principal component estimator, simulation and computation. Scientific African. 2023;19:e01566.

[14] Kibria BM, Lukman AF. A new ridge-type estimator for the linear regression model: Simulations and applications. Scientifica; 2020.

[15] Furno M. Estimating the variance of the LAD regression coefficients. Computational Statistics & Data Analysis. 1998;27(1):11–26.

[16] Badawaire AB, Dawoud I, Lukman AF, Laoye V, Olatunji A. Biasing Estimator to Mitigate Multicollinearity in Linear Regression Model. Al-Bahir Journal for Engineering and Pure Sciences. 2023;2(1):1.

[17] Gibbons DG. A simulation study of some ridge estimators. Journal of the American Statistical Association. 1991;76(373), 131–139.

[18] Kibria BMG. Performance of some new ridge regression estimators. Communications in Statistics-Simulation and Computation. 2003;32(2):419–435.

[19] 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. Concurrency and Computation: Practice and Experience. 2022;34(6):e6779.

[20] Lukman AF, Osowole OI, Ayinde K. Two stage robust ridge method in a linear regression model. Journal of Modern Applied Statistical Methods. 2015;14(2):8.

[21] Lukman AF, Ayinde K, Kun SS, Adewuyi ETA modified new two-parameter estimator in a linear regression model. Modelling and Simulation in Engineering; 2019.
Available:https://doi.org/10.1155/2019/6342702

[22] Lukman AF, Ayinde K, Olatayo A, Bamidele R, Aladeitan BB, Adagunodo TA. Developing a New Estimator in Linear Regression Model. Journal of Physics: Conference Series. 2019;1299(1), 12128.

[23] Hildreth C, John YL. Demand relations with autocorrelated disturbances. Technical Bulletin. Michigan State University Agricultural Experiment Station. 1960;276.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle5.com/review-history/122417