International Journal of Development Mathematics Vol 3 Issue 3 Page No. 323 - 337

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

A New Robust LQS-NTP Estimator for Mitigating Correlated Endogenous Variables and Extreme Observations in Linear Regression Models: Theoretical Development and Applications

Adewale A. Titilola[a*], Timothy O. Olatayo[a] and Abass I. Taiwo[a]

[a]Department of Statistics, Olabisi Onanbanjo University Ago-Iwoye

ARTICLE INFO

Article history:
Received 02 July2026
Received in revised form 10 August 2026
Accepted 30 August 2026

Keywords:
Robust Regression, LQS-NTP Estimator, Multicollinearity, Extreme Observations, Mean Squared Error

MSC 2020 Subject classification: 62P20

ABSTRACT

This study proposes a Robust Least Quantile of Squares–New Two-Parameter (LQS-NTP) estimator for addressing multicollinearity and extreme observations in linear regression models. The proposed estimator combines the high-breakdown robustness of the Least Quantile of Squares (LQS) method with the shrinkage properties of the New Two-Parameter (NTP) estimator. The Mean Squared Error (MSE) of the proposed estimator was derived, and its biasing parameters were obtained by minimizing the corresponding MSE. The performance of the proposed estimator was assessed using Boston Housing and Portland Cement datasets and compared with some already existing methods. The data sets exhibited substantial multicollinearity, together with several outlying observations. The proposed LQS-NTP estimator achieved the lowest MSE across both data employed. These results demonstrate that the proposed estimator provides an effective alternative for regression estimation in the simultaneous presence of multicollinearity and extreme observations.

1 Introduction

Linear regression is a fundamental statistical technique widely used to assess the association between a response variable and one or more predictor variables. The Ordinary Least Squares (OLS) estimator, developed from the foundational contributions of Gauss and Legendre, remains a standard approach to parameter estimation because of its simplicity, computational efficiency, and well-established statistical properties. Within the framework of the Classical Linear Regression Model (CLRM), the Gauss–Markov theorem demonstrates that OLS is the Best Linear Unbiased Estimator (BLUE), achieving the smallest variance among all linear unbiased estimators when the underlying assumptions are satisfied (Greene, 2018; Wooldridge, 2020; Montgomery *et al*., 2021).

However, the efficiency and optimality of OLS rely on assumptions such as linearity, exogeneity, homoscedasticity, error independence, and the absence of perfect multicollinearity. In empirical studies, these conditions may not always hold, and their violation can adversely affect the accuracy, stability, and reliability of OLS-based estimates. Modern datasets arising from economics, finance, medicine, engineering, agriculture, environmental sciences, and the social sciences often exhibit complex characteristics such as multicollinearity, endogeneity, and extreme observations. These problems reduce the reliability of classical regression methods by producing unstable parameter estimates, inflated standard errors, poor predictive performance, and misleading statistical inference (Kutner *et al*., 2005; Montgomery *et al*., 2021).

Consequently, developing estimation procedures that simultaneously address these challenges has become one of the major areas of research in regression analysis. Multicollinearity occurs when explanatory variables are highly correlated, causing the design matrix to become nearly singular. Under this condition, the variances of the estimated

* Corresponding author. Tel.: +2348066160296
E-mail address: adewale.titilola6565@gmail.com (Titilola, A. A.)
https://doi.org/10.62054/ijdm/0303.16

regression coefficients increase substantially, rendering hypothesis testing unreliable and making parameter estimates extremely sensitive to minor perturbations in the data (Belsley *et al*., 1980). Several diagnostic measures, including the Variance Inflation Factor (VIF), condition indices, tolerance values, and eigenvalue decomposition, have been developed to detect multicollinearity.

To mitigate its adverse effects, researchers have proposed numerous biased estimation techniques that intentionally introduce a small amount of bias in exchange for a considerable reduction in estimator variance. Among the earliest and most influential biased estimators is ridge regression introduced by Hoerl and Kennard (1970), which stabilizes coefficient estimation by incorporating a positive ridge parameter into the information matrix. Although ridge regression considerably improves estimation in the presence of multicollinearity, its dependence on the least squares criterion makes it highly sensitive to influential observations.

To overcome some of these limitations, Liu (1993) proposed the Liu estimator, which introduced an additional biasing parameter that provides greater flexibility in controlling the bias–variance trade-off. Building upon these estimators, several generalized ridge, modified ridge, Liu-type, and New Two-Parameter (NTP) estimators have been proposed to improve estimation efficiency under varying degrees of multicollinearity (Kibria, 2003; Özkale and Kaçiranlar, 2007; Lukman *et al*., 2019).

In particular, the NTP estimator has demonstrated superior performance over many existing biased estimators because its two shrinkage parameters provide greater flexibility in minimizing the Mean Squared Error (MSE). Besides multicollinearity, another important source of model misspecification is endogeneity. Endogeneity arises when one or more explanatory variables are correlated with the random disturbance term, violating the exogeneity assumption of the classical linear regression model. This problem may result from omitted variable bias, simultaneity, reverse causality, measurement error, or dynamic feedback mechanisms.

Unlike multicollinearity, which primarily inflates estimator variance, endogeneity produces biased and inconsistent parameter estimates regardless of sample size, thereby invalidating statistical inference and prediction (Greene, 2018; Wooldridge, 2020). In empirical studies, endogeneity frequently coexists with multicollinearity, making parameter estimation considerably more difficult. Regression analysis is further complicated by the presence of outliers and leverage points. Outliers may arise because of recording errors, measurement inaccuracies, data contamination, natural variability, or structural changes in the underlying population. Since the OLS estimator minimizes the sum of squared residuals, observations with unusually large residuals receive disproportionately large weights during estimation. Consequently, even a single influential observation may substantially distort regression coefficients, inflate residual variance, and reduce predictive accuracy.

This sensitivity gives OLS a breakdown point of zero, indicating that one extreme observation can render the estimator unreliable (Rousseeuw and Leroy, 1987). The limitations of classical least squares estimation motivated the development of robust regression techniques designed to reduce the influence of atypical observations. Early contributions by Huber (1964, 1981), Hampel (1974), and Rousseeuw (1984) laid the theoretical foundation of robust statistics.

Among the most widely used robust estimators are M-estimators, S-estimators, MM-estimators, Least Median of Squares (LMS), Least Trimmed Squares (LTS), and Least Quantile of Squares (LQS). While M-estimators effectively reduce the influence of vertical outliers through bounded loss functions, they remain susceptible to high-leverage observations. High-breakdown estimators such as LMS, LTS, and particularly LQS overcome this limitation by minimizing ordered residual statistics instead of the conventional residual sum of squares, thereby providing substantially greater resistance to leverage points and severe contamination. Recent developments in regression methodology have increasingly focused on integrating robust estimation procedures with biased regression estimators to simultaneously address multicollinearity and outlier contamination.

Studies by Dawoud and Abonazel (2021), Awwad and Abonazel (2022), Erisoğlu *et al*. (2024), Altukhaes *et al.* (2024), and Oyeleke *et al*. (2025) demonstrated that robust ridge-type, Liu-type, and two-parameter estimators consistently outperform conventional OLS and classical biased estimators under contaminated datasets. Likewise, Adejumo *et al*. (2023) proposed the Robust-M New Two-Parameter (RM-NTP) estimator, which successfully integrated robust M-estimation with the New Two-Parameter shrinkage framework proposed by Yang and Chang (2010). Simulation studies showed that the RM-NTP estimator achieved lower Mean Squared Errors than several competing estimators under varying degrees of multicollinearity and outlier contamination. Nevertheless, because M-estimation still assigns positive weights to all observations, the estimator remains susceptible to high-leverage observations. Despite these advances, relatively little attention has been devoted to developing estimators capable of

simultaneously mitigating correlated endogenous variables and extreme observations using high-breakdown robust estimation procedures.
Existing robust biased estimators rely predominantly on M-estimation or Least Trimmed Squares (LTS), whereas the Least Quantile of Squares (LQS) estimator offers stronger protection against leverage points through its high-breakdown properties. Furthermore, most existing studies have concentrated on handling multicollinearity among explanatory variables without explicitly addressing the combined effects of correlated endogenous variables and outlier contamination.
This gap motivates the development of the proposed New Robust Least Quantile of Squares–New Two-Parameter (LQS-NTP) estimator. The proposed methodology integrates the high-breakdown robustness of the Least Quantile of Squares estimator with the bias-reduction capability of the New Two-Parameter estimator to simultaneously mitigate correlated endogenous variables and extreme observations. The performance of the proposed estimator will be evaluated using real-life applications. It is anticipated that the proposed LQS-NTP estimator will provide more stable parameter estimates, lower Mean Squared Errors, and improved predictive performance than existing ridge-type, Liu-type, robust M-estimators, and conventional New Two-Parameter estimators, thereby contributing significantly to the growing body of literature on robust biased estimation for linear regression models.

Consider the classical linear regression model (CLRM) as given in (1).

$$y_i = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \ldots \gamma_p x_{ip} + v_{ij}, i = 1,2,\ldots,n, j = 0,1,2,\ldots p \tag{1}$$

where n is the number of observations, p is the number of regressors such that $n>p$. The y_i denotes the ith observed endogenous variable, x_{ij} is the ith value of the exogenous variable. Also, γ_j, $j = 0,1,\ldots,p$, is the regression coefficients, and v_i is the error term such that $E(v_i) = 0$ and $Var(v_i) = \sigma^2 I_n$ and I_n is an n-identity matrix (Birkes and Dodge, 1993).
The Matrix form of (1) is given in (2);

$$Y = X\gamma + v_i \tag{2}$$

The CLR estimator also known as OLS estimator is defined as in equation (3):

$$\hat{\gamma} = G^{-1} X^T Y \tag{3}$$

where $G = (X^T X)$

2 Review of some Biased-Based Estimators

2.1 Ordinary Ridge Regression Estimator

Hoerl and Kennard (1970) proposed Ridge regression estimator in order to deal with the problem of multicollinearity in regression analysis. The Ridge parameter was added to the *G* matrix to reduce the collinearity effect. The $\hat{\gamma}_{ORR}$ estimator was defined as;

$$\hat{\gamma}_{ORR} = (G + kI)^{-1} X^T y. \tag{4}$$

where *I* is the *(pxp)* identity matrix and $k = \frac{P\sigma^2}{\sum_{i=1}^{p} \hat{\alpha}_i^2}$ is the ridge parameter such that *p* is the number of explanatory variable, σ^2 is the Mean Square Error (MSE) and $\hat{\alpha}_i$ is an unbiased estimator of $\hat{\gamma}$.

i. Liu Estimator

To mitigate the high correlation in the data set, Liu (1993) came up with the Liu Estimator by combining the Stein Estimator with Ridge estimator to form;

$$\hat{\gamma}_L = (G + I)^{-1} (G + dI) \hat{\gamma}_{OLS}. \tag{5}$$

where $d = 1 - \sigma^2 \left[\frac{\sum_{i=1}^{p} \frac{1}{\lambda_i(\lambda_i+1)}}{\sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i+1)^2}} \right]$ such that λ_i is the ith Eigen value G and $\hat{\alpha}_i = Q^T \hat{\gamma}$

ii. KL-Esimator

Kibria and Lukman (2020) proposed another one parameter estimator that can combat the problem of multicollinearity in linear regression models aside from the already existing ones. They defined the estimator as:

$$\hat{\gamma}_{KL} = (G + kI_p)^T (G - kI_p)\hat{\gamma}_{OLS} \quad (6)$$

where $k = \frac{\hat{\sigma}^2}{2\hat{\alpha}_i^2 + (\hat{\sigma}^2/\lambda_i)}$ and $\hat{\gamma}_{OLS}$ is the OLS estimator

iii. Ridge-Liu estimator

Two–parameter Ridge-Liu estimator was introduced by Özkale and Kaciranlar (2007) and defined as in (7):

$$\hat{\gamma}_{TP} = (G + kI)^{-1}(X^T y + kd\hat{\gamma}_{OLS}). \quad (7)$$

iv. Modified Ridge-Type Estimator

In CLR, Lukman *et al.* (2019) proposed the modified ridge-type estimator (MRT) as an alternative estimator to handle the problem of multicollinearity as it can be expressed in (8).

$$\hat{\gamma}^{MRT}(k,d) = (G + k(1+d))^{-1} X^T y. \quad (8)$$

The MRT estimator will give the same results with the OLS and the ridge estimator majorly when $k = 0$ and $d = 0$.

v. DK-Estimator

As an alternative method of reducing the impact of multicollinearity in linear regression model, Dawoud and Kibria (2020) proposed a new biased estimator denoted as; $\hat{\beta}_{DK}$and defined as:

$$\hat{\beta}_{DK} = (G + k(1+d)I_p)^{-1}(G - k(1+d)I_p)\hat{\gamma}_{OLS}. \quad (9)$$

The new ridge-type estimator proposed by Owolabi *et al.* (2022) as an alternative method of dealing with the problem of multicollinearity in regression analysis aside already existing ones. They expressed the estimator as:

$$\hat{\alpha}_{p1}(k,d) = (G + (k+d)I)^{-1} X^T Y \quad (10)$$

where $k = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} - d$ and $d = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} - k$

2.2 The Robust Regression Methods

Below are review of various robust estimators that can handle extreme observations,

i. M-Estimator

Huber (1964) introduced the M-estimator as a generalization of maximum likelihood estimation, with the primary objective of reducing the influence of extreme observations on parameter estimates. Unlike ordinary OLS, which minimizes the sum of squared residuals and is highly sensitive to outliers, M-estimators minimize a robust loss function that assigns progressively less weight to large residuals. The loss function is typically symmetric and designed to limit the influence of extreme observations, thereby improving the robustness of parameter estimates. Consequently, M-estimators are particularly suitable for regression analysis involving outliers, heavy-tailed error distributions, or departures from the normality assumption.

$$\hat{\gamma}_M = min \sum_{i=1}^{n} \rho\left(\frac{y_i - x_{ij}\hat{\beta}}{s}\right). \quad (11)$$

where s is a robust estimate of scale expressed as $s = \frac{median|e_i - median(e_i)|}{0.6745}$.

ii. MM-estimator

Yohai (1987) extended the M-estimation framework by introducing the MM-estimator, which combines a high breakdown point with strong statistical efficiency. The MM-estimator achieves robustness through a multi-stage procedure that begins with an initial high-breakdown robust estimator and is subsequently refined using an M-estimation re-weighting scheme. This approach limits the influence of outlying observations while maintaining high efficiency when the underlying errors are normally distributed. The estimator is defined as:

$$\hat{\gamma}_{MM} = \sum \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\hat{\beta}_j}{s_{MM}}\right) x_{ij} \quad . \quad (12)$$

where $s_{MM}$is the standard deviation obtained from the residual of S-estimation and $\rho$is a Tukey's biweight function. (Susanti *et al.*, 2014).

iii. The Least Absolute Deviation (LAD Estimator

The Least Absolute Deviation (LAD) estimator, also referred to as the Least Absolute Error (LAE**)** estimator, was first introduced by Boscovich in 1757 (Birkes and Dodge, 1993). The LAD method estimates the regression coefficients by minimizing the sum of the absolute values of the residuals. Unlike ordinary least squares, which minimizes the squared residuals, LAD places less emphasis on large deviations and is therefore more robust to the presence of outliers. The mathematical formulation of the LAD estimator is given in Equation (13):

$$\hat{\gamma}_{LAD} = min\sum_{i=1}^{n}\left|y_i - x_i\hat{\beta}\right|. \tag{13}$$

This method is very useful when the distribution of the residual does not follow a normal distribution. Also, LAD is robust when the contamination is in the y-direction. Ahmed and Maha, (2016).

iv. Least Quantile of Square (LQS) estimator

The Least Quantile of Squares (LQS) estimator is a robust regression method based on the order statistics of squared residuals which was developed by Rousseeuw (1984). It generalizes the Least Median of Squares (LMS) approach by replacing the median of the squared residuals with a specified quantile. The LQS estimator is based on selecting a specified quartile of the ordered squared residuals rather than relying solely on the median. Specifically, it minimizes a chosen quantile of the squared residuals, thereby reducing the influence of observations that deviate substantially from the fitted model. The method is defined in terms of a proportion (q) of observations, such that the remaining *(1-q)* proportion is regarded as potentially extreme or contaminated observations. Consequently, the LQS estimator provides a highly robust approach to parameter estimation in the presence of outliers. The estimator is defined as:

$$\hat{\gamma}_{LQS} \in \arg\min_{\gamma}\left|r_{(q)}\right| \tag{14}$$

where $r_{(q)}$denotes the residual corresponding to the qth ordered absolute residual: $\left|r_{(1)}\right| \leq \left|r_{(2)}\right| \leq\ldots\leq \left|r_{(n)}\right|$.

3 The proposed Estimator

As an alternative to some of the already existing methods of dealing with high correlation among the explanatory variables in CLRM, Yang and Chang (2010) proposed NTP estimator defines as:

$$\hat{\gamma}_{NTP} = (G+I)^{-1}(G+d^{NTP}I)(G+k^{NTP}I)^{-1}G\hat{\gamma}_{OLS} \tag{15}$$

where d^{NTP} and k^{NTP} are the biasing parameters for NTP estimator.

It has been noted that the NTP estimator ($\hat{\gamma}_{NTP}$) is underperforms in the presence of extreme values. However, there is need to develop the robust version that can mitigate the two challenges concurrently by incorporating the robust LQS method into (15) in exchange for $\hat{\gamma}_{OLS}$.. The proposed estimator can be defined as:

$$\hat{\gamma}_{NTP}^{LQS} = (G+I)^{-1}\left(G+d_{NTP}^{LQS}I\right)\left(G+k_{NTP}^{LQS}I\right)^{-1}G\,\hat{\gamma}_{LQS} \tag{16}$$

where k_{NTP}^{LQS} and d_{NTP}^{LQS} are the estimated robust biasing parameters for the $\hat{\gamma}_{NTP}^{LQS}$ estimator respectively.

3.1 The Canonical form of the robust LQS-NTP Estimator

The general form of linear regression model as given in equation (1), therefore its canonical form is hereby defined as follows:

$$y = D\hat{\theta} + e \tag{17}$$

where $D = XZ$ and $\hat{\theta} = Z^T\beta$, Z is the orthogonal matrix such that $D^TD = Z^TGZ = \tau = diag(\tau_1, \tau_2, \ldots, \tau_p)$ where $\tau_1, \tau_2, \ldots, \tau_p > 0$ are the ordered eigenvalues of G.Let $\hat{\theta}_{LQS}$ be the any robust regression estimator of the equations $\sum_{i=1}^{n}\psi.\left({}^{e_i}/_{s}\right) = 0$and $\sum_{i=1}^{n}\psi\left({}^{e_i}/_{s}\right)m_i = 0$ such that $e_i = y_i - m_i^T\hat{\theta}_{LQS}$ where m is a scale parameter and $\psi$is a selected useful function. Hampel *et al*. (1986) and Huber (1981).

Therefore, Canonical form of (17) and some of the already existing estimators are as follows:

$$\hat{\theta} = H^{-1}D^Ty, \tag{18}$$

where $D = XZ$, $H = D^TD$

$\hat{\theta}(k) = \left(H + kI_p\right)^{-1} D^T y,$ (19)
$\hat{\theta}_M(k) = \left(I_p + kD^{-1}\right)\hat{\theta}_M,$ (20)
$\hat{\theta}_M(d) = \left(H + I_p\right)^{-1}\left(H + dI_p\right)\hat{\theta}_M,$ (21)
$\hat{\theta}_M(KL) = \left(H + kI_p\right)^{-1}\left(H - kI_p\right)\hat{\theta}_M,$ (22)
$\hat{\theta}_M(TP) = \left(H + kI_p\right)^{-1}\left(H + kdI_p\right)\hat{\theta}_M,$
(23)
$\hat{\theta}_M(DK) = \left(H + k(1+d)I_p\right)^{-1}\left(H - k(1+d)I_p\right)\hat{\theta}_M,$
(24)
The canonical form of the Robust LQS- NTP estimator of θ is defined as:
$\hat{\theta}_{LQS}(NTP) = (H + I)^{-1}(H + dI)(H + kI)^{-1}H\hat{\theta}_{LQS}$ (25)

3.1.1 Determination of the MSE for the Robust LQS-NTP Estimator

The MSE of OLS estimator $\hat{\theta}$ can be expressed as:
$MSE\left(\hat{\theta}\right) = E(\hat{\theta} - \theta)^T(\hat{\theta} - \theta),$ (26)
$= tr(cov(\hat{\theta})) + bias(\hat{\theta})^T bias(\hat{\theta}).$
For the purpose of application, (26) can be written as:
$MSE\left(\hat{\theta}\right) = \sum_{i=1}^{p} \sigma^2 / \tau_i.$ (27)
Equation (27) can further be expressed as:
$MSE\left(\hat{\theta}_M\right) = \sum_{i=1}^{p} \Omega_{ii}.$ (28)
where Ω_{ii} is the diagonal element for $cov(\hat{\theta}_M)$ which is equivalent toΩ and is finite.
Hoerl and Kennard (1970) derived the MSE for Ridge estimator $\hat{\theta}(k)$ as follows:
$MSE\left(\hat{\theta}(k)\right) = \sigma^2 \sum_{i=1}^{p} \frac{\tau_i}{(\tau_i + k)^2} + \sum_{i=1}^{p} \frac{k^2\theta_i^2}{(\tau_i + k)^2}$ (29)
Liu (1993), estimated the MSE for the Liu estimator $\hat{\theta}_d$ as follows:
$MSE\left(\hat{\theta}_d\right) = \sigma^2 \sum_{i=1}^{p} \frac{(\tau_i + d)^2}{(\tau_i + 1)^2 \tau_i} + (d-1)^2 \sum_{i=1}^{p} \frac{\hat{\theta}_i^2}{(\tau_i + 1)^2}$ (30)
MSE for Robust Ridge estimator $\hat{\theta}_R(k)$is defined as:
$MSE\left(\hat{\theta}_M(k)\right) = \sum_{i=1}^{p} \frac{\tau_i^2\Omega_{ii} + k^2\hat{\theta}_i^2}{(\tau_i + k)^2}$ (31)
MSE for robust Liu estimator $\hat{\theta}_M(d)$ denoted as $MSE\left(\hat{\theta}_M(d)\right)$ is given as:
$MSE\left(\hat{\theta}_M(d)\right) = \sum_{i=1}^{p} \frac{(\tau_i + d)^2\Omega_{ii} + (1-d)^2\hat{\theta}_i^2}{(\tau_i + 1)^2}$ (32)
MSE of robust Dawoud-Kibria is expressed as:
$MSE(\hat{\theta}_M(DK)) = \sum_{i=1}^{p} \frac{(\tau_i - k(1-d))^2\Omega_{ii}}{(\tau_i + k(1+d))^2} + \sum_{i=1}^{p} \frac{4k^2(1-d)^2\hat{\theta}_i^2}{(\tau_i + k(1+d))^2}$ (33)
Therefore, following the submission of Yang and Chang (2010), the MSE of the robust LQS-NTP can be expressed as:
$MSE\left(\hat{\theta}_{NTP}^{LQS}\right) = \sum_{i=1}^{p} \frac{(\tau_i + d)^2\tau_i\Omega_{ii}}{(\tau_i + 1)^2(\tau_i + k)^2} + \sum_{i=1}^{p} \frac{\left((k+1-d)\tau_i + k\right)^2\hat{\theta}_i^2}{(\tau_i + 1)^2(\tau_i + k)^2}.$ (34)

3.1.2 Conditions for Establishing the Superiority of the Proposed Estimator LQS-NTP and theoretical comparison with some already existing estimators

The superiority of the proposed estimator relative to some existing estimators is established under the following conditions:
Conditions:
(i) The function ψ is skew-symmetric and non-decreasing**.**
(ii) The error terms are symmetrically distributed with zero mean and finite variance.
(iii) The diagonal element of the relevant matrix, denoted by $cov(\hat{\theta}_R)$ which is Ω_{ii} is finite**.**

i. Comparison of $MSE\left(\hat{\boldsymbol{\theta}}_{NTP}^{LQS}\right)$ with $MSE\left(\hat{\boldsymbol{\theta}}_{NTP}\right)$,

$MSE\left(\hat{\theta}_{NTP}^{LQS}\right) < MSE\left(\hat{\theta}_{NTP}\right)$, if and only if $\sum_{i=1}^{p} \Omega_{ii} < \sum_{i=1}^{p} \sigma^2$.

$U_{NTP}^{LQS} = MSE\left(\hat{\theta}_{NTP}^{LQS}\right) - MSE\left(\hat{\theta}_{NTP}\right)$

$$U_{NTP}^{LQS} = \sum_{i=1}^{p}\left[\frac{(\tau_i+d)^2\tau_i\Omega_{ii}+\left((k+1-d)\tau_i+k\right)^2\hat{\theta}_i^2}{(\tau+d)^2(\tau_i+k)^2}\right] - \sum_{i=1}^{p}\left[\frac{(\tau_i+d)^2\tau_i\sigma^2+\left((k+1-d)\tau_i+k\right)^2\hat{\theta}_i^2}{(\tau_i+d)^2(\tau+k)^2}\right] \tag{35}$$

$$= \sum_{i=1}^{p}\left[\frac{t_1\Omega_{ii}+t_2-t_1\sigma^2-t_2}{(\tau_i+d)^2(\tau_i+k)^2}\right] \tag{36}$$

Where $t_1 = \left(\tau_i + d\right)^2 \tau_i$, $t_2 = \left((k+1-d)\tau_i + k\right)^2 \hat{\theta}_i^2$

$$U_{NTP}^{LQS} = \sum_{i=1}^{p}\left[\frac{t_1(\Omega_{ii}-\sigma^2)}{\left(\tau_i+d\right)^2\left(\tau_i+k\right)^2}\right] \tag{37}$$

If $(\Omega_{ii} - \sigma^2) < 0$ then, $MSE\left(\hat{\theta}_{NTP}^{LQS}\right) < MSE\left(\hat{\theta}_{NTP}\right)$.

Therefore, LQS-NTP is better than NTP estimator if $\sum_{i=1}^{p} \Omega_{ii} < \sum_{i=1}^{p} \sigma^2$.

ii. Comparison of $MSE\left(\hat{\boldsymbol{\theta}}_{NTP}^{LQS}\right)$ with $MSE\left(\hat{\boldsymbol{\theta}}_{NBB}^{M}\right)$

$MSE\left(\hat{\theta}_{NTP}^{LQS}\right) > MSE\left(\hat{\theta}_{NBB}^{M}\right)$ if and only if $\sum_{i=1}^{p} J_i < \sum_{i=1}^{p} L_i$.

$U_{NBB^M}^{NTP^{LQS}} = MSE\left(\hat{\theta}_{NTP}^{LQS}\right) - MSE\left(\hat{\theta}_{NBB}^{M}\right)$

(38)

$$U_{NBB^M}^{NTP^{LQS}} = \sum_{i=1}^{p}\left[\frac{\left(\tau_i+d_1\right)^2\tau_i\Omega_{ii}+\left((k_1+1-d_1)\tau_i+k_1\right)^2\hat{\theta}_i^2}{\left(\tau_i+d_1\right)^2\left(\tau_i+k_1\right)^2}\right] -$$

$$= \sum_{i=1}^{p}\left[\frac{\left((1-d_2)\tau_i+k_2\right)^2\hat{\theta}^2+\tau_i\left(\tau_i+(d_2+k_2)\right)^2\Omega_{ii}}{\left(\tau_i+1\right)^2(\tau_i+k_2)^2}\right] \tag{39}$$

$$= \sum_{i=1}^{p}\left[\frac{v_{22}(v_1\Omega_{ii}+v_{12})-v_{13}(v_2+v\Omega_{ii})}{v_{13}v_{22}}\right] \tag{40}$$

$$= \sum_{i=1}^{p}\left[\frac{v_{22}v_{12}-v_{13}v_2-(v_{13}v_{21}-v_{22}v_1)\Omega_{ii}}{v_{13}v_{22}}\right] \tag{41}$$

where $v_1 = \left(\tau_i + d_1\right)^2 \tau_i$, $v_{12} = \left((k_1+1-d_1)\tau_i + k_1\right)^2 \hat{\theta}_i^2$,

$v_{13} = \left(\tau_i + d_1\right)^2\left(\tau_i + k_1\right)^2$, $b_2 = \left((1-d_2)\tau_i + k_2\right)^2 \hat{\theta}^2$,

$v_{21} = \tau_i\left(\tau_i + (d_2+k_2)\right)^2$, $v_{22} = \left(\tau_i + 1\right)^2(\tau_i + k_2)^2$.

$$U_{NBB^M}^{NTP^{LQS}} = \sum_{i=1}^{p}\left[\frac{J_i-L_i}{v_{13}v_{22}}\right], \tag{42}$$

where, $J_i = v_{22}v_{12} - v_{13}v_2$, $L_i = (v_{13}v_{21} - v_{22}v_1)\Omega_{ii}$

Therefore,$J_i - L_i$<0 if $\sum_{i=1}^{p} J_i < \sum_{i=1}^{p} L_i$.such that k_1, k_2, d_1 and $d_2$are the biasing parameters of robust LQS-NTP and robust M- NBB estimators respectively.

iii. Comparison of $MSE\left(\hat{\boldsymbol{\theta}}_{NTP}^{LQS}\right)$ with $MSE\left(\hat{\boldsymbol{\theta}}_{MRT}^{M}\right)$

The MSE of robust LQS-NTP is less than MSE of Robust M- MRT if and only if and only if $\sum_{i=1}^{p} R_i < \sum_{i=1}^{p} S_i$.

$$U_{MRT^M}^{NTP^{LQS}} = MSE\left(\hat{\theta}_{NTP}^{LQS}\right) - MSE\left(\hat{\theta}_{MRT}^{M}\right) \tag{43}$$

$$U_{MRT^M}^{NTP^{LQS}} = \sum_{i=1}^{p}\left[\frac{\left(\tau_i+d_3\right)^2 \tau_i \Omega_{ii}+\left((k_3+1-d_3)\,\tau_i+k_3\right)^2\hat{\theta}_i^2}{\left(\tau_i+1\right)^2\left(\tau_i+k_3\right)^2}\right] - \sum_{i=1}^{p}\left[\frac{\tau_i \Omega_{ii}+k_4^2(1+d_4)^2\hat{\theta}_i^2}{\left(\tau_i+k_4(1+d_4)\right)^2}\right] \tag{44}$$

$$U_{MRT^M}^{NTP^{LQS}} = \sum_{i=1}^{p}\left[\frac{b_1\Omega_{ii}+b_1}{b_2}\right] - \sum_{i=1}^{p}\left[\frac{\tau_i \Omega_{ii}+h_1}{h_2}\right] \tag{45}$$

$$U_{MRT^M}^{NTP^{LQS}} = \sum_{i=1}^{p}\left[\frac{b_2h_2-b_3h_1-\left(b_3\,\tau_i-b_1h_2\right)\Omega_{ii}}{b_3h_2}\right], \tag{46}$$

where, $b_1 = \left(\tau_i + d_3\right)^2 \tau_i$, $b_2 = \left((k_3+1-d_3)\,\tau_i + k_3\right)^2 \hat{\theta}_i^2$, $b_3 = \left(\tau_i + 1\right)^2\left(\tau_i + k_3\right)^2$,
$h_1 = k_4^2(1+d_4)^2\hat{\theta}_i^2$ and $h_2 = \left(\tau_i + k_4(1+d_4)\right)^2$

$$U_{MRT^M}^{NTP^{LQS}} = \sum_{i=1}^{p}\left[\frac{R_i-S_i}{b_3h_2}\right], \tag{47}$$

where, $R_i = b_2h_2 - b_3h_1$ and $S_i = \left(b_3\,\tau_i - b_1h_2\right)\Omega_{ii}$.
Hence, $R_i - S_i < 0$, if and only if $\sum_{i=1}^{p} R_i < \sum_{i=1}^{p} S_i$, where k_3, k_4, d_3 and d_4 are the biasing parameters of robust LQS-NTP and robust M-MRT estimators.

3.2 Robust Choice of Biasing Parameters *k* and *d* for the robust LQS-NTP Estimator.

If $\hat{\theta}_{LQS} \sim N(0,1)$ with covariance matrix equals N^2H^{-1}. When $\sqrt{n}\left(\hat{\theta}_{LQS}^2 - \theta\right) \sim N(0, N^2H^{-1})$, therefore, the assumption holds for practical purpose, where $N^2 = \frac{L_0^2 E[\Psi^2(\varepsilon/l_0)]}{E[\phi'(\varepsilon/l_0)]^2}$. Such that l_0 is the scale estimate.
Also, the unbiased estimator $\theta_{iLQS} = \hat{\theta}_{iLQS}^2$ that is $E(\hat{\theta}_{iLQS}) = \theta_{\mathrm{iLQS}}$ and the unbiased estimator $\Omega_{ii} = N^2/\tau_i$ such that $N^2 = \frac{c^2(n-p)^{-1}\sum_{i=1}^{n}[(\varepsilon_i/c)]^2}{\sum_{i=1}^{n}\left[\frac{1}{n}\omega'(\varepsilon_i/c)\right]^2}$.
Biasing parameters for some robust estimators are hereby expressed as follows:
Robust Ridge Regression (RRR) estimator biasing parameter estimated by Lukman *et al*. (2014) is defined as:

$$\hat{k}_{Ridge} = \frac{pN^2}{\sum_{i=1}^{p}\hat{\theta}_{iLQS}^2}. \tag{48}$$

Biasing parameter for robust M-Liu estimator by Kibria and Lukman (2020) and defined as:

$$\hat{d}_{Liu} = 1 - N^2\left[\frac{\sum_{i=1}^{p}\frac{1}{\tau_i(\tau_i+1)}}{\sum_{i=1}^{p}\frac{\hat{\theta}_{iM}^2}{(\tau_i+1)^2}}\right]. \tag{49}$$

Biasing parameter for Robust M-Ridge-Liu estimator can be expressed as:

$$\hat{k}_{RL} = \frac{1}{p}\sum_{i=1}^{p}\frac{N^2}{\hat{\theta}_i^2 - d\left(\frac{N^2}{\tau_i+\hat{\theta}_{iM}^2}\right)} \tag{50}$$

and

$$\hat{d}_M = min\ \left\{\frac{\hat{\theta}_{iM}^2}{\frac{N^2}{\tau_i}+\hat{\theta}_{iM}^2}\right\}. \tag{51}$$

Robust M Dawoud-Kibria (2020) derived the biasing parameter as follows:

$$\hat{k}(DK^M) = \frac{1}{p}\sum_{i=1}^{p}\frac{N}{(1+d_{TP})\left(\frac{N}{\tau_i}+2\hat{\theta}_{iM}^2\right)}. \tag{52}$$

where $d_{TP} = min\left(\frac{\hat{\theta}_{iM}^2}{\left(\frac{N}{\tau_i}\right)+\hat{\theta}_{iM}^2}\right)_{i=1}^{p}$

Therefore, following the research of Yang and Chang (2010) and by minimizing (34) with respect to k and d the biasing parameters for robust LQS-NTP can be derived as:

$$\hat{k}_{NTP}^{LQS} = k = \frac{N^2(\tau_i+d)-(1-d)\tau_i\hat{\theta}_{iLQS}}{(\tau_i+1)\hat{\alpha}_{iR}}, \quad (53)$$

such that $d = 0<d<1$ and

$$\hat{d}_{NTP}^{LQS} = d = \frac{\sum_{i=1}^{p}\frac{(\theta_{iLQS}^2-N^2)}{(\tau_i+1)^2}}{\sum\frac{(N^2+\tau_i\hat{\theta}_{iLQS}^2)}{(\tau_i+1)^2\tau_i}}. \quad (54)$$

4 Application to Real-life data

Real-life datasets used to demonstrate and evaluate the performance of all the estimators considered in this study include the Boston Housing Dataset and the Cement Dataset. These datasets have been widely employed in previous studies, including Lukman and Ayinde (2018)**,** among other authors, for investigating regression estimation methods under challenging data conditions.

4.1 Application to Boston Data set

The Boston Housing Dataset is a well-known benchmark dataset in statistics and machine learning, originally compiled from housing and socioeconomic information for the Boston, Massachusetts area. It contains 14 variables, comprising the response variable MEDV (median value of owner-occupied homes, in thousands of dollars) and 13 explanatory variables: CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, B, and LSTAT. These variables represent crime rate, land use, industrial activity, proximity to the Charles River, air pollution, housing characteristics, accessibility, property taxes, pupil–teacher ratio, demographic characteristics, and socioeconomic status. The regression model for the data set can be expressed as follows:

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+...+\beta_{13}x_{13} \quad (55)$$

Where; : y =MEDV: Median value of owner-occupied homes (in \$1000s), x_1= CRIM: Per capita crime rate by town, x_2 =ZN: Proportion of residential land zoned for lots over 25,000 sq.ft., x_3= INDUS: Proportion of non-retail business acres per town, x_4= CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise), x_5 =NOX: Nitric oxides concentration (parts per 10 million), x_6= RM: Average number of rooms per dwelling, x_7 = AGE: Proportion of owner-occupied units built prior to 1940, x_8 =DIS: Weighted distances to five Boston employment centers, x_9 = RAD: Index of accessibility to radial highways, x_{10} = TAX: Full-value property tax rate per \$10,000, x_{11} =PTRATIO: Pupil-teacher ratio by town, x_{12} =B: $1000(B_k-0.63)^2$, where B_k is the proportion of Black residents by town, x_{13} =LSTATX:% lower status of the population.

The dataset was found suitable for regression analysis because it combines socioeconomic, environmental, geographical, and housing characteristics. However, the diagnostic results in Table 1 indicate substantial multicollinearity among the explanatory variables. Multicollinearity was observed for INDUS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, B, and LSTAT, with VIF values ranging from 11.10 to 85.03. The most severe cases were recorded for PTRATIO (85.03), RM (77.95), NOX (73.89), and TAX (61.23). In contrast, CRIM (2.10), ZN (2.84), and CHAS (1.15) showed no multicollinearity. Furthermore, observations 186, 361, 368, 369, 370, 371, 372, and 412 were identified as outliers. Thus, the dataset presents a challenging combination of multicollinearity and extreme observations, making it appropriate for assessing the performance of robust regression estimators. The multiple linear regression model is specified with MEDV as the dependent variable and the 13 explanatory variables as predictors. The results of the multicollinearity assessment are presented in Table 1, while Table 2 compares the MSE performance of the estimators considered. The results show that the proposed robust LQS-NTP regression estimator produced the lowest MSE (35.134) across the considered estimators. The least MSE values is highlighted in bold in Table 2.

Table 1: Multicollinearity (Variance Inflation Factor - VIF)

Variable	VIF	Interpretation
CRIM	2.10	No concern
ZN	2.84	No concern
INDUS	14.49	High multicollinearity
CHAS	1.15	No concern
NOX	73.89	Severe multicollinearity
RM	77.95	Severe multicollinearity
AGE	21.39	High multicollinearity
DIS	14.70	High multicollinearity
RAD	15.17	High multicollinearity
TAX	61.23	Severe multicollinearity
PTRATIO	85.03	Severe multicollinearity
B	20.10	High multicollinearity
LSTAT	11.10	High multicollinearity

Table 2: MSEs and coefficients of all the Estimators using Botson data

Coef.	**OLS**	**RE**	**LE**	**KL**	**MRT**	**DK**	**M**	**M-RE**	**M-LE**	**M-KL**	**M-MRT**	**M-DK**	**LQS**	**LQS-NTP**
β_0	36.4595	36.3992	11393.765	36.3389	36.3691	36.2787	18.93117	36.3992	11393.77	36.339	36.3691	33.218	4.1809	61.424
β_1	-0.108	-0.108	25687.153	-0.10796	-0.10797	-0.1079	-0.10581	-0.10798	25687.15	-0.108	-0.10797	-0.10652	-0.0756	-0.1195
β_2	0.04642	0.04643	168607.78	0.04643	0.04643	0.04644	0.035204	0.04643	168607.8	0.0464	0.04643	0.0467	-0.0414	0.0443
β_3	0.02056	0.02047	111564.08	0.02037	0.02042	0.02028	-3.66E-05	0.02047	111564.1	0.0204	0.02042	0.01562	0.1823	0.0588
β_4	2.68673	2.68676	994.07277	2.68678	2.68677	2.68681	1.609756	2.68676	994.0728	2.6868	2.68677	2.68784	2.6666	2.6767
β_5	-17.767	-17.732	6097.5719	-17.6977	-17.715	-17.663	-10.3687	-17.7322	6097.572	-17.698	-17.715	-15.918	5.3217	-32.024
β_6	3.80987	3.81291	73921.643	3.81595	3.81443	3.81899	5.05555	3.81291	73921.64	3.816	3.81443	3.97357	4.8739	2.5497
β_7	0.00069	0.00067	732581.14	0.00065	0.00066	0.00062	-0.02337	0.00067	732581.1	0.0006	0.00066	-0.00056	-0.0005	0.0103
β_8	-1.4756	-1.4747	45714.456	-1.47375	-1.47421	-1.4729	-1.10571	-1.47466	45714.46	-1.4738	-1.47421	-1.42688	0.7659	-1.8506
β_9	0.30605	0.30583	93444.985	0.30561	0.30572	0.30539	0.195714	0.30583	93444.98	0.3056	0.30572	0.29426	-0.0738	0.3968
β_{10}	-0.0123	-0.0123	4287797.6	-0.01233	-0.01233	-0.0123	-0.01119	-0.01233	4287798	-0.0123	-0.01233	-0.01217	0.0001	-0.0136
β_{11}	-0.9527	-0.9518	205317.05	-0.95089	-0.95135	-0.95	-0.77216	-0.95182	205317	-0.9509	-0.95135	-0.90276	-0.7406	-1.3377

β_{12}	0.00931	0.00932	4208049.5	0.00933	0.00932	0.00934	0.011003	0.00932	4208049	0.0093	0.00932	0.00977	0.0034	0.0058
β_{13}	-0.5248	-0.5246	119799.4	-0.52449	-0.52456	-0.5244	-0.34182	-0.52462	119799.4	-0.5245	-0.52456	-0.51749	-0.6354	-0.5807
MSE	41.6218	41.5024	5.681E+17	315.3	41.4428	41.2643	44.71361	44.5853	6.10E+17	338.72	44.5213	38.1025	63.209	**35.134**

4.1 Application to Portland Cement Data

The Cement Dataset is a classical small-sample dataset widely used in regression analysis and the study of multicollinearity. It consists of 13 observations on Portland cement, with measurements of the percentages of four chemical components and the heat evolved after 180 days of hardening. The dataset originated from Woods *et al*. (1932) and has subsequently been used extensively in statistical literature to evaluate regression estimators.

The response variable, Heat (Y), represents the heat evolved in calories per gram, while the four explanatory variables are tricalcium aluminate (C_3A), tricalcium silicate (C_3S), tetracalcium aluminoferrite (C_4AF), and dicalcium silicate (C_2S). The diagnostic results indicate severe multicollinearity among all four explanatory variables, as their VIF values substantially exceed the conventional threshold of 10. The VIF values are 54.26 for C_3A, 160.78 for C_3S, 32.05 for C_4AF, and 19.87 for C_2S, with C_3S exhibiting the most severe multicollinearity. The multiple linear regression model for the Cement Dataset is expressed in Equation (56).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + .\beta_4 x_4 \tag{56}$$

Furthermore, Table 3 presents the MSE comparison of the estimators considered in the study. The results indicate that the proposed robust LQS-NTP estimator achieved the lowest MSE (29.9634**)** among all other estimators considered in study as indicated in bolded form in the Table.

Table 3: MSEs and coefficients of all the Estimators using Portland Cement data set

Coef	OLS	RE	LE	RL	MRT	DK	M	LQS	M-RE	M-LE	M-RL	M-MRT
β_0	100.368	94.5678	388.9449	98.3574	91.9124	83.4566	102.373	114.128	94.5678	388.9449	97.526	91.912
β_1	-0.9446	-1.1247	2621.949	-1.0071	-1.2072	-1.4698	-0.8696	-0.5519	-1.1247	2621.949	-1.03287	-1.2072
β_2	-0.9952	-1.0683	15740.11	-1.0206	-1.1018	-1.2084	-0.9647	-0.8383	-1.0683	15740.11	-1.03106	-1.1018
β_3	-1.0192	-1.0349	4628.614	-1.0246	-1.0421	-1.0649	-1.0156	-1.0723	-1.0349	4628.614	-1.02687	-1.0421
β_4	-0.0356	0.07797	34732.37	0.00375	0.12998	0.29558	-0.0771	-0.2805	0.07797	34732.37	0.02003	0.13
MSE	75.8407	67.3167	3.58E+10	3651.71	63.5841	52.4076	78.1209	129.69	69.3406	3.69E+10	3658.46	65.496

Discussion

The empirical results demonstrate the effectiveness of the proposed Robust LQS-NTP estimator in addressing the simultaneous presence of multicollinearity and extreme observations in linear regression models. The motivation for the proposed method arises from the limitations of conventional estimators, particularly the sensitivity of OLS to extreme observations and the instability of regression estimates under severe multicollinearity. The proposed method combines the robustness of the Least Quantile of Squares estimator with the shrinkage capability of the New Two-Parameter estimator, thereby providing a framework for dealing with both challenges concurrently. The Boston Housing Dataset provides strong evidence of the difficulties associated with the data structure. The diagnostic results revealed substantial multicollinearity, particularly for PTRATIO, RM, NOX, and TAX, with VIF values of 85.03, 77.95, 73.89, and 61.23, respectively. Several other explanatory variables also exhibited high multicollinearity. In addition, observations 186, 361, 368, 369, 370, 371, 372, and 412 were identified as outliers. Thus, the dataset represents a suitable empirical case for evaluating estimators designed to simultaneously accommodate multicollinearity and extreme observations.

The MSE results for the Boston Housing data indicate that the proposed LQS-NTP estimator recorded the lowest MSE of 35.134 among all the estimators considered. This performance is notable because the competing estimators include both conventional biased estimators and robust variants. The result suggests that incorporating the high-breakdown property of LQS into the NTP framework improves estimation efficiency when the data contain both severe multicollinearity and outlying observations. A similar pattern was observed with the Portland Cement Dataset. The four explanatory variables exhibited severe multicollinearity, with VIF values of 54.26, 160.78, 32.05, and 19.87 for C_3A, C_3S, C_4AF, and C_2S, respectively. The particularly high VIF of 160.78 for C_3S confirms the severity of the collinearity problem in this dataset.
The comparative MSE results further support the superiority of the proposed method. For the Cement Dataset, the LQS-NTP estimator achieved an MSE of 29.9634, which was the smallest among all the estimators considered. This result is consistent with the findings from the Boston Housing Dataset, where LQS-NTP also produced the minimum MSE. The consistency of the results across two different real-life datasets provides empirical evidence that the proposed estimator can effectively improve estimation performance under severe multicollinearity and data contamination.
Hence, the results indicate that the proposed LQS-NTP estimator provides a favourable bias–variance trade-off. Its performance can be attributed to the combination of the robust LQS estimation procedure, which limits the influence of extreme observations, and the two-parameter shrinkage mechanism of NTP, which helps control the adverse effects of multicollinearity. The empirical findings therefore support the theoretical motivation for developing the proposed estimator and demonstrate its potential usefulness for real-life regression applications.

5 Conclusion

This study developed a Robust Least Quantile of Squares–New Two-Parameter (LQS-NTP) estimator for linear regression models affected by multicollinearity and extreme observations. The proposed estimator integrates the high-breakdown robustness of LQS with the shrinkage properties of the NTP estimator to provide improved estimation under challenging data conditions. The empirical applications using the Boston Housing and Portland Cement datasets confirmed the existence of serious multicollinearity. The Boston Housing data also contained several identified outliers, while the Cement data exhibited extremely high VIF values across all explanatory variables. These characteristics provided suitable settings for evaluating the proposed estimator against existing conventional, biased, and robust estimators. The results showed that the LQS-NTP estimator consistently outperformed the competing estimators in terms of MSE. It recorded the lowest MSE of 35.134 for the Boston Housing Dataset and 29.9634 for the Portland Cement Dataset. These findings demonstrate the ability of the proposed estimator to provide more efficient regression estimates in the presence of severe multicollinearity and extreme observations. Therefore, the Robust LQS-NTP estimator is recommended as an effective alternative to existing estimators for regression problems involving simultaneous multicollinearity and extreme observations. Its consistent performance across the two real-life datasets indicates its potential applicability in empirical studies where conventional regression methods may produce unstable or inefficient estimates.

Conflict of interest: Authors declare that there is conflict of interest

REFERENCES

Adejumo, T. J., Ayinde, K., Akomolafe, A. A., Makinde, O. S., & Ajiboye, A. S. (2023). Robust-M new two-parameter estimator for linear regression models: Simulations and applications. *African Scientific Reports, 2*(3), 138. https://doi.org/10.46481/asr.2023.2.3.138

Altukhaes, W. B., Roozbeh, M., & Mohamed, N. A. (2024). Feasible robust Liu estimator to combat outliers and multicollinearity effects in restricted semiparametric regression models. *AIMS Mathematics, 9*(11), 31581–31606. https://doi.org/10.3934/math.20241519

Altukhaes, W. B., Roozbeh, M., & Mohamed, N. A. (2024). Feasible robust Liu estimator to combat outliers and multicollinearity effects in restricted semiparametric regression models. *AIMS Mathematics, 9*(11), 31581–31606. https://doi.org/10.3934/math.20241519

Awwad, M., & Abonazel, M. R. (2022). 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, 34*(9), e6779. https://doi.org/10.1002/cpe.6779

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Birkes, D. and Dodge, Y. D. (1993). Alternative Methods of Regression, John Wiley & Sons, New York.

Dawoud, I., & Abonazel, M. R. (2021). Robust Dawoud–Kibria estimator for handling multicollinearity and outliers in the linear regression model. *Journal of Statistical Computation and Simulation, 91*(17), 3678–3692. https://doi.org/10.1080/00949655.2021.1945063

Dawoud, I., & Kibria, B. M. G. (2020). A new biased estimator to combat the multicollinearity of the Gaussian linear regression model. *Stats, 3*(4), 526–541. https://doi.org/10.3390/stats3040033

Erisoğlu, M., Karakoca, A., & Yurtaslan, A. (2024). Novel robust estimators for the linear regression model with multicollinearity and outlier problems. *REVSTAT—Statistical Journal, 22*(4). https://doi.org/10.57805/revstat.v22i4.686

Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*(346), 383–393. https://doi.org/10.1080/01621459.1974.10482962

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. John Wiley & Sons.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics, 35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Huber, P. J. (1973). Robust regression: Asymptotically normal estimators. *The Annals of Statistics, 1*(5), 799–821. https://doi.org/10.1214/aos/1176342503

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons.

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics—*

Simulation and Computation, 32(2), 419–435. https://doi.org/10.1081/SAC-120017499

Kibria, B. M. G., & Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: Simulations and applications. *Scientifica, 2020*, Article 9758378, 16 pages. https://doi.org/10.1155/2020/9758378

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.

Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics—Theory and Methods, 22*(2), 393–402. https://doi.org/10.1080/03610929308831027

Lukman, A. F., Arowolo, O., & Ayinde, K. (2014). Some robust ridge regression for handling multicollinearity and outliers. *International Journal of Sciences: Basic and Applied Research, 16*(2), 192–202.

Lukman, A. F., Ayinde, K., Binuomote, S., & Clement, O. A. (2019). Modified ridge-type estimator to combat multicollinearity. *Journal of Statistical Computation and Simulation, 89*(18), 3496–3512. https://doi.org/10.1080/00949655.2019.1666800

Lukman, A. F., & Ayinde, K. (2018). Detecting influential observations in two-parameter Liu-Ridge estimator. *Journal of Data Science, 16*(2), 201–218.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). John Wiley & Sons.

Owolabi, A. T., Ayinde, K., & Alabi, O. O. (2022). A new ridge-type estimator for the linear regression model with correlated regressors. *Concurrency and Computation: Practice and Experience*. https://doi.org/10.1002/cpe.6933

Oyeleke, K. T., Olatayo, T. O., & Efuwape, B. T. (2025). Handling multicollinearity and outliers: A comparative study of one- and two-parameter estimators using real-life data. *International Journal of Development Mathematics*. https://doi.org/10.62054/ijdm/0104.14

Özkale, M. R., & Kaçiranlar, S. (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics—Theory and Methods, 36*(15), 2707–2725. https://doi.org/10.1080/03610920701386877

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*(388), 871–880. https://doi.org/10.1080/01621459.1984.10477105

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley & Sons. https://doi.org/10.1002/0471725382

Susanti, Y., Pratiwi, H., Sulistijowati, S. H., & Liana, T. (2014). M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics, 91*(3), 349–360.

https://doi.org/10.12732/ijpam.v91i3.7

Woods, H., Steinour, H. H., & Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. *Industrial & Engineering Chemistry, 24*(11), 1207–1214. https://doi.org/10.1021/ie50275a002

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.

Yang, H., & Chang, X. (2010). A new two-parameter estimator in linear regression. *Communications in Statistics—Theory and Methods, 39*(6), 923–934. https://doi.org/10.1080/03610920902807911

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics, 15*(2), 642–656. https://doi.org/10.1214/aos/1176350366