WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

An Adaptive Quantile-Based Robust Ridge Estimator for Linear Regression under Multicollinearity and Data Contamination

Adamu Abubakar, K.E. Lasisi, A. Ahmed, H. Idi and A. Sagir

Abubakar Tafawa Balewa University, Bauchi
abbakar.adamu@yahoo.com

ABSTRACT

This study proposes the Adaptive Quantile Regression Kibria–Lukman Estimator (AQR-KLE) as a regularised regression approach for addressing multicollinearity and the influence of outliers in linear regression models. The study investigates its performance in comparison with Ordinary Least Squares (OLS), Ridge Regression (RR), robust ridge estimation, the Modified Kibria–Lukman (MKL) estimators, and Quantile-based MKL (QMKL) estimators. A Monte Carlo simulation study involving 10,000 replications was conducted under varying sample sizes, error variances, levels of multicollinearity, and outlier contamination. Estimator performance was evaluated using the Mean Squared Error (MSE), with lower values indicating better estimation performance. The simulation results reported that AQR-KLE achieved the lowest MSE in almost all scenarios, suggesting promising performance under different combinations of multicollinearity and outlier contamination. The estimator was also applied to four real-world datasets. Results showed that AQR-KLE generally produced coefficient estimates close to those obtained using OLS, while some upper-quantile QMKL estimates exhibited more noticeable variations. These findings suggest that AQR-KLE may provide a useful regularised alternative for regression analysis where coefficient stability is important. However, further validation using reproducible simulation results, independent prediction assessment, and sensitivity analysis is necessary to establish its robustness and predictive performance. The study contributes to the development of adaptive regularised estimation methods for regression problems involving multicollinearity and unusual observations.

Keywords: Adaptive quantile regression, data contamination, ridge regression, mean squared error, regularisation.

1. Introduction

Linear regression remains a cornerstone of statistical modeling across economics, finance, environmental sciences, and health research. Despite its widespread application, the classical ordinary least squares (OLS) estimator is highly sensitive to violations of standard assumptions (Nosheen et al., 2025; Rajh-Weber et al., 2025), particularly multicollinearity among predictors and contamination arising from outliers or heavy-tailed error distributions (Jain & Chopra, 2026). Multicollinearity inflates the variance of regression coefficients, rendering inference unreliable, while outliers can severely bias parameter estimates (Olaluwoye et al., 2025; Alshangiti et al., 2025; Osma, 2025).

Ridge regression, introduced by Hoerl and Kennard (1970), addresses multicollinearity by imposing an ℓ_2 penalty on regression coefficients (Gillariose et al., 2026). Although ridge regression stabilizes estimation, it retains the vulnerability of OLS to non-normal errors and outliers. To overcome this limitation, robust ridge-type estimators and quantile-based approaches have been developed. In the work of Goutte et al. (2026), quantile regression, originally proposed by Koenker and Bassett (1978), provides robustness by modeling conditional quantiles rather than conditional means.

Recent studies like (Haider et al., 2026; Zaman et al., 2026; Irshad et al., 2026) have combined ridge penalties with quantile-based estimation to handle multicollinearity and outliers simultaneously. However, most existing quantile-based ridge estimators rely on fixed quantile levels, which may be suboptimal when data exhibit heterogeneous error structures or varying degrees of contamination (Li ad Ullah, 2025; Li et al., 2026). Motivated by this limitation, the present study introduces an adaptive quantile-based robust ridge estimator that dynamically adjusts quantile selection according to data characteristics.

The main contributions of this study are threefold: (i) development of an adaptive quantile selection mechanism integrated into a ridge-type estimator; (ii) theoretical justification of the estimator's bias–variance trade-off; and (iii) extensive Monte Carlo evaluation under diverse modeling conditions.

Multicollinearity arises when explanatory variables are highly correlated, leading to inflated variances and unstable coefficient estimates (Mahmood, 2024; Olaluwoye et al., 2025). Ridge regression and its variants have been extensively studied as remedies for this problem (Gillariose et al., 2026). Numerous modifications (Akhtar and Alharthi, 2024(a); Akhtar and Alharthi, 2025(b)), including improved and adaptive ridge estimators, have been proposed to enhance efficiency.

Robust regression techniques aim to reduce sensitivity to outliers and heavy-tailed distributions. M-estimators and quantile regression methods have been widely adopted for this purpose (Suhail et al., 2023; Haider et al., 2026). Quantile-based ridge estimators represent a hybrid approach that combines robustness with shrinkage, demonstrating superior performance under contamination (Li and Ullah et al., 2025; Irshad et al., 2026).

Despite these advances, the reliance on fixed quantile levels limits adaptability. Data-driven quantile selection has recently gained attention as a means of improving estimator flexibility (Yan et al., 2023; Kalina, 2025; Suresh et al., 2026). This study builds on these developments by proposing an adaptive quantile-based framework tailored for ridge-type estimation under multicollinearity and outliers.

2. Methodology

2.1 Model Specification

The study is based on the generalized linear regression (GLR) framework, expressed as

$$y = X\beta + \varepsilon$$

where y is an $n \times 1$ vector of the response variable, X is an $n \times p$ matrix of explanatory variables, β is a $p \times 1$ vector of unknown regression parameters, and ε represents the random error term. The error vector is assumed to have zero mean and constant variance, such that

$$E(\varepsilon) = 0, \text{and } Var(\varepsilon) = \sigma 2$$

Under classical assumptions, the ordinary least squares (OLS) estimator of β is given by

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

Although OLS is the best linear unbiased estimator, its efficiency deteriorates in the presence of multicollinearity and is highly sensitive to outliers in the response variable (Muthukrishnan et al., 2024; Naz et al., 2025; Irshad et al., 2026).

2.2 Ridge Regression and Robust Ridge Estimation

To address multicollinearity among explanatory variables, ridge regression (RR) introduces a biasing parameter $k > 0$ into the estimation process. The ridge estimator is defined as

$$\hat{\beta}_{RR} = (\mathrm{X}'\mathrm{X} + kI)^{-1}\mathrm{X}'\mathrm{y}$$

While ridge regression reduces variance inflation caused by multicollinearity, it remains vulnerable to outliers in the response direction. To jointly handle multicollinearity and outliers, robust ridge regression based on M-estimation is employed (Alshangiti et al., 2025; Irshad et al., 2026). Following this approach, the robust ridge M-estimator (RME) is obtained by replacing the OLS estimator with a robust M-estimator, denoted as $\hat{\beta}_M$, yielding

$$\hat{\beta}_{RME} = R_k\, \hat{\beta}_M$$

Where $R_k = (\mathrm{X}'\mathrm{X} + kI)^{-1}\mathrm{X}'\mathrm{X}$

The M-estimator $\hat{\beta}_M$ is computed by solving estimating equations based on a chosen influence function $\psi(\cdot)$, such as the Huber function, which limits the effect of extreme observations.

2.3 Canonical Representation of the Model

For analytical tractability, the GLR model is transformed into its canonical form using eigenvalue decomposition. Let;

$$X'X = P\Lambda P'$$

Where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$ contains the eigenvalues of $X'X$ and P is the corresponding orthogonal matrix of eigenvectors. Defining $\alpha = P'\beta$, the transformed model becomes

$$y = D\alpha + \varepsilon$$

This transformation simplifies the derivation of mean squared error (MSE) expressions and facilitates direct comparison of competing estimators (Dong et al., 2026; Al-Marzouki et al., 2026).

2.4 Selection of the Ridge Parameter

The effectiveness of ridge-type estimators depends critically on the choice of the tuning parameter k. Several approaches for estimating k have been proposed in the literature, including classical and robust formulations. In the robust setting, the ridge parameter is typically computed using robust estimates of both regression coefficients and error scale.

This study focuses on robust Kibria–Lukman-type ridge estimators, which incorporate robust measures of dispersion and parameter magnitude to obtain data-adaptive biasing constants.

2.5 Quantile-Based Robust Kibria–Lukman Estimator

To further enhance flexibility and robustness, a quantile-based strategy is employed for determining the ridge parameter. Let

$$\{\hat{k}_{MKL_1}, \hat{k}_{MKL_2}, \dots, \hat{k}_{MKL_p}\}$$

represent the set of ridge parameters derived from the robust Kibria–Lukman formulation (Wasim et al. 2025). These values are arranged in ascending order, and a quantile level γ, with $0 < \gamma < 1$, is selected.

The quantile-based ridge parameter is defined as

$$\hat{k}_{QMKL_\gamma} = Q_\gamma(\hat{k}_{MKL})$$

where $Q\gamma(\cdot)$ denotes the $\gamma - th$ empirical quantile of the ordered ridge parameters. The varying γ, the estimator can adapt to different levels of multicollinearity, error variability, and outlier contamination.

3. Propose Estimator - Adaptive Quantile-Based Ridge Estimator

Let $Q_1 = Q_{q1}(x)$ and $Q_2 = Q_{q2}(x)$ represents $q1 - th$ and $q2 - th$ quantiles of 25th and 75th percentiles. Then the combine estimator for k can be formulated as:

i. Compute the adaptive shrinking factor:

$$\tau = Q_{q2}(x) - Q_{q1}(x)$$

Where $q1$ and $q2$ are adaptively chosen.

A Common way to identify an outlier using quantiles is:

$$\text{Outlier Thresholds} = Q_{q1} - \alpha\,\tau, Q_{q2} + \alpha\,\tau$$

Where $\tau = Q_{\bar{q}2}(x) - Q_{\bar{q}1}(x)$ is adaptive dispersion measure, α is a scaling factor typically 1.5 for mild outlier and 3.0 for extreme outliers. Thus, an observation x_i is classified as an outlier if:

$x_i < Q_{q1} - \alpha\,\tau$ This is lower outlier, or $x_i < Q_{q2} + \alpha\,\tau$. This is upper outlier.

$$\tau = Q_{\bar{q}2}(x) - Q_{\bar{q}1}(x)$$

Where $\bar{q}_1$ and $\bar{q}_2$ are adjusted quantiles based on the data distribution.

Once the optimal quantiles have been selected, the model for the selected quantile can be expressed as:

$$\hat{\beta}_* = arg\ \min_{\beta} \sum_{i=1}^{n} L_\tau\,(y_i - X_i\beta)$$

This minimizes the quantile loss for the selected quantile level, leading to robust parameter estimates that are less sensitive to outliers. With these, the modified estimator will becomes:

$$\hat{\beta}_{AQR-KLE} = (X'X + \tau I)^{-1}(X'X + \tau I)\hat{\beta}_\tau$$

Where τ is data dependent, which is equivalent to $\tau = Q_{\bar{q}2}(x) - Q_{\bar{q}1}(x)$

Since $(X'X + \tau I)$ is invertible, we use the fundamental property of matrix inverse, i.e. $A^{-1}A = I$ for any invertible matrix A, then let $A = (X'X + \tau I)$. Substituting into our original equation, we have: $(X'X + \tau I)^{-1}X'y(X'X + \tau I)^{-1}X'y$

Thus,

$$\hat{\beta}_{AQR-KLE} = \hat{\beta}_\tau$$

The equation suggest that the adjusted estimator $\hat{\beta}_{AQR-KLE}$ does not alter $\hat{\beta}_\tau$. This is because the matrix multiplication essentially cancels out the transformation introduced by $(X'X + \tau I)$. Thus we have verified that $\hat{\beta}_{AQR-KLE} = (X'X + \tau I)^{-1}(X'X + \tau I)\hat{\beta}_\tau$ holds under the assumption that $(X'X + \tau I)$ is invertible. The proof relies on the algebraic property of invertible matrices and identity transformations where:

$$\hat{\beta}_\tau = (X'X + \tau I)^{-1}X'y$$

3.1 Evaluation Criterion

The performance of the proposed estimators is assessed using the mean squared error (MSE), which captures both bias and variance components. For the canonical parameter vector α, the MSE is defined as

$$MSE(\hat{\alpha}) = E[(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)]$$

4. Simulation Design

A Monte-Carlo simulation experiment was carried out to assess the performance of the proposed estimators under controlled and reproducible conditions. The simulation design was adapted from earlier works by McDonald and Galarneau (1975), Gibbons (1981), Kibria (2003), Shaheen et al. (2023), and Wasim et al. (2025).

Predictor Variables

The explanatory variables were generated using a correlated structure designed to induce different strengths of multicollinearity. The data-generation mechanism is given by:

$$X_{ij} = (1 - \eta)\phi_{ij} + \eta\ \phi_{ip}$$

where

$\phi_{ij} \sim \mathrm{N}(0,1)$ are independent standard normal random variables,

$\eta \in \{0.70, 0.90, 0.95, 0.99\}$ represents the specified correlation level,

$i = 1,2 \ldots, n$ denotes the observation index, and

$j = 1,2 \ldots, p$ denote the predictor index.

This allows the correlation among predictors to increase as η approaches we replicate mild to severe multicollinearity scenarios.

Response Variables

The dependent variable was generated from the linear regression model:

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + u_i$$

The following assumptions were adopted: The intercept parameter was set to zero, i.e. β_0. The regression coefficient vector $\beta\ (\beta_1, \beta_2, \ldots, \beta_p)'$ was taken as the normalized eigenvector corresponding to the largest eigenvalue of the matrix $X'X$, ensuring: $\beta'\beta = 1$

The random error term u_i was generated under two distributional settings:

- Normal distribution: $u_i \sim N(0, \sigma^2)$
- Standard Cauchy distribution: $u_i \sim N(0, \sigma^2)$

The normal distribution represents conventional regression assumptions, while the heavy-tailed Cauchy distribution introduces extreme deviations and outliers.

Contamination Mechanism

To evaluate estimator robustness, outliers were introduced in the response variable (y-direction) following Ertaş (2018). Two contamination levels were considered:

- 10% contamination
- 20% contamination

In each case, a proportion of the normal error distribution was inflated, thereby producing observations with unusually large deviations from the regression line. This setup enables the assessment of estimator sensitivity to vertical outliers.

Experimental Scenarios

Ten simulation scenarios were constructed based on different combinations of sample size, number of predictors, error variance, multicollinearity level, and error distribution. The study considered:

A. Normal error distribution

i. Number of predictors: $p \in \{5,12\}$

ii. Sample sizes: $n \in \{25,50,100\}$

iii. Error variances: $\sigma^2 \in \{0.1,1,2,5\}$

iv. Collinearity levels: $\eta \in \{0.70,0.90,0.95,0.99\}$

v. Contamination levels: 10% and 20%

B. Cauchy error distribution

vi. Number of predictors: $p \in \{5,12\}$

vii. Sample sizes: $n \in \{25,50,100\}$

viii. Error variance: $\sigma^2 = 1$

ix. Collinearity levels: $\eta \in \{0.70,0.90,0.95,0.99\}$

These scenarios collectively represent a wide range of modelling conditions from low-variance low-multicollinearity environments to high-variance extreme multicollinearity and heavy-tailed disturbances.

Number of Replications

For each scenario, 10,000 Monte-Carlo replications were performed. Each replication involved: generating a new dataset, computing the corresponding estimator(s), and recording the performance measures. The large number of replications ensures that sampling variability is minimized and the resulting performance estimates are stable and reliable.

Performance Assessment Criterion

The principal performance measure used in the evaluation is the Mean Squared Error (MSE), defined as:

$$MSE(\hat{\alpha}) = E[(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)]$$

In the simulation context, the MSE was approximated by:

$$\widehat{MSE}(\hat{\alpha}) = \frac{1}{10{,}000} \sum_{k=1}^{10{,}000} (\hat{\alpha}_k - \alpha)'(\hat{\alpha}_k - \alpha)$$

Where $\hat{\alpha}_k$ denotes the estimator obtained from the $k - th$ replication and α represents the true parameter vector. Lower MSE value indicate better estimator performance.

5. Results and Discussion

The results of the simulation study (Appendix A), based on the estimated MSE of OLS, Ridge (RR), Ridge-M (HUB_M), MKL-family (MKL1–3), quantile MKL-family (QMKL1 – 99), and the proposed estimator for each scenario with 10% and 20% outliers, are summarized below.

i. OLS performance is the weakest among all estimators across all scenarios. Its MSE increases sharply with higher collinearity (η) and larger error variance (σ^2), particularly for small sample sizes ($n = 25$).

ii. Ridge estimators (RR) show better stability than OLS under moderate collinearity, but their performance deteriorates when multicollinearity is high ($\eta = 0.95,0.99$) or when the error variance is large ($\sigma^2 = 5$). This confirms findings reported by Wasim et al. (2023).

iii. Robust estimators (HUB_M) consistently outperform OLS and RR, especially in the presence of outliers. Their MSE increases with both higher (σ^2) and η, but less dramatically than OLS.

iv. MKL-family estimators (MKL1–3) perform well under low-to-moderate variance (($\sigma^2 = 0.1$ and 1) and moderate collinearity. The MSE of these estimators increases as the error variance or multicollinearity rises, but they remain more stable than OLS and RR.

v. Quantile MKL estimators (QMKL75–99) show notable robustness against outliers. Lower quantiles (e.g., QMKL75, QMKL85) tend to have slightly higher MSE than higher quantiles (QMKL95, QMKL99) under small sample sizes ($n = 25$) but are consistently better than OLS and RR. QMKL estimators also benefit from increasing sample size, as MSE declines with larger n.

vi. The proposed estimator consistently exhibits the lowest MSE across almost all scenarios, demonstrating strong robustness to outliers, high collinearity, and large error variance. Even when η=0.99 and $\sigma^2 = 5$, the proposed estimator maintains considerably lower MSE than the conventional estimators.

vii. Effect of sample size (n): Increasing the sample size from 25 to 50 and 100 substantially reduces the MSE for all estimators. The reduction is most pronounced for robust and MKL-family as well as the proposed estimators.

viii. Effect of multicollinearity (η): Across all estimators, higher η generally leads to higher MSE, with the impact being more severe for OLS and RR. The proposed estimator is less sensitive to collinearity, particularly at higher sample sizes.

ix. Effect of error variance (σ^2): The MSE of all estimators increases with larger σ^2. Conventional OLS and RR estimators are particularly affected, while HUB_M, MKL, QMKL, and the proposed estimator show greater resilience.

The simulation confirms that robust and quantile-adaptive estimators (the proposed) outperform traditional OLS and Ridge methods under challenging conditions involving outliers, high variance, and multicollinearity.

Therefore, the proposed estimator demonstrates exceptional performance across a wide range of challenging scenarios, consistently achieving the lowest MSE compared to traditional and even other robust estimators. Its strength lies in its remarkable robustness to outliers, high collinearity, and large error variance, where conventional OLS and Ridge estimators struggle. Unlike other methods whose performance deteriorates sharply under extreme conditions, it (the proposed estimator) maintains stability, showing only modest sensitivity to increases in multicollinearity or error variance. Additionally, it benefits from larger sample sizes, further reducing estimation error. It represents a highly reliable and efficient estimator for regression settings prone to contamination and multicollinearity, effectively combining robustness with strong predictive accuracy.

Real Data Application

i. Determinants of Economic Growth

The Growth DJ dataset (Durlauf and Johnson (1995) GrowthDJ Data), associated with Durlauf and Johnson (1995), contains 121 observations used to examine differences in economic growth across countries. The dependent variable (Y) is per-capita GDP growth from 1960 to 1985. The proposed independent variables are initial GDP per capita in 1960 (X1), population growth (X2), investment as a proportion of GDP (X3), secondary-school enrolment (X4), literacy in 1960 (X5), oil-producing-country status (X6), data-quality classification (X7), and OECD membership (X8).

Method	k	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
OLS		9.873335	-5.19177	12.89621	0.054509	4.228509	6.11546	8.603468	-0.88622
RR	0.1	9.739235	-4.8911	12.71687	0.090608	4.303701	6.064658	8.553475	-0.89019
HUB-M		9.873335	-5.19177	12.89621	0.054509	4.228509	6.11546	8.603468	-0.88622
MKL1	0.128367	9.702968	-4.80932	12.66755	0.101574	4.323569	6.051062	8.539357	-0.89122
MKL2	0.004099	9.867636	-5.17905	12.88868	0.055903	4.231759	6.113285	8.601414	-0.8864
MKL3	0.000931	9.872039	-5.18888	12.8945	0.054825	4.229249	6.114966	8.603002	-0.88626
QMKL.01	0.000272	9.872956	-5.19093	12.89571	0.054601	4.228725	6.115316	8.603332	-0.88624
QMKL.05	0.000287	9.872936	-5.19088	12.89568	0.054606	4.228737	6.115308	8.603325	-0.88624
QMKL.25	0.001102	9.871801	-5.18835	12.89419	0.054883	4.229384	6.114875	8.602916	-0.88627
QMKL.50	0.003196	9.868889	-5.18185	12.89034	0.055595	4.231045	6.113764	8.601867	-0.88636
QMKL.75	0.008624	9.861366	-5.16505	12.8804	0.05745	4.235329	6.110894	8.599147	-0.88659
QMKL.95	0.653725	9.14077	-3.52127	11.86062	0.340063	4.604506	5.850091	8.287751	-0.90467
QMKL.99	0.930745	8.906493	-2.97713	11.50249	0.478631	4.705779	5.77293	8.164835	-0.90854
AQR-KLE	0.004581	9.866967	-5.17756	12.8878	0.056067	4.23214	6.113031	8.601173	-0.88642

Estimator (AQR-KLE) produces coefficient estimates that are very close to those obtained using Ordinary Least Squares (OLS), while incorporating a regularisation parameter of (k=0.004581). The AQR-KLE estimates are particularly close to those of MKL2, MKL3, and the QMKL estimators at the lower and central quantiles. For example, the coefficient for initial GDP per capita (B1) is 9.866967 under AQR-KLE, compared with 9.873335 under OLS. Similarly, the investment coefficient (B3) is 12.887800 under AQR-KLE and 12.896210 under OLS. These small differences indicate that AQR-KLE retains the general coefficient pattern observed under OLS while applying a non-zero regularisation penalty.

A notable feature of the results is the stability of AQR-KLE estimates relative to the more pronounced coefficient changes observed for QMKL at the upper quantiles. At the 0.95 quantile, for instance, QMKL estimates (B1=9.140770), (B2=-3.521270), and (B3=11.860620), whereas AQR-KLE produces estimates of 9.866967, −5.177560, and 12.887800, respectively. The differences become more pronounced at the 0.99 quantile, where QMKL estimates (B 1=8.906493), (B2=-2.977130), and (B3=11.502490). In comparison, AQR-KLE maintains estimates close to the OLS values. This suggests that AQR-KLE provides a more stable coefficient profile than the reported upper-quantile QMKL estimates for these data. However, because quantile regression

estimates different parts of the conditional response distribution, these differences do not automatically mean that the QMKL estimates are inferior; they may also reflect genuine variation in economic relationships across quantiles.

The estimates for the remaining explanatory variables reinforce this pattern. The secondary-school enrolment coefficient (B4) is 0.056067 under AQR-KLE, compared with 0.054509 under OLS, while the literacy coefficient (B5) is 4.232140 under AQR-KLE and 4.228509 under OLS. The coefficients for oil-producing-country status (B6), data-quality classification (B7), and OECD membership (B8) are also close to their OLS counterparts. In terms of direction, the estimates indicate positive coefficients for initial GDP per capita, investment, secondary-school enrolment, literacy, oil-producing-country status, and data-quality classification, alongside negative coefficients for population growth and OECD membership, based on the coding and specification used. These signs describe the fitted model and should not, on their own, be interpreted as evidence of causal effects.

ii. Global Economy Dataset

The Global Economy dataset provides annual economic indicators for countries over time, including observations spanning 1960–2017 in the version described. The dependent variable (Y) is GDP growth, while the proposed independent variables are GDP (X1), consumer price index (X2), imports as a percentage of GDP (X3), exports as a percentage of GDP (X4), and population (X5). Global Economy Dataset – Forecasting: Principles and Practice | World Development Indicators – World Bank

Method	k	β_1	β_2	β_3	β_4	β_5
OLS		2.236445	4.112317	13.17039	13.81335	0.572378
RR	0.1	2.43831	4.253459	13.0757	13.56328	0.55645
HUB-M		2.236445	4.112317	13.17039	13.81335	0.572378
MKL1	0.411483	2.981642	4.635463	12.80131	12.90537	0.51314
MKL2	0.187661	2.60338	4.369213	12.99542	13.361	0.543356
MKL3	0.072409	2.384122	4.215525	13.10149	13.63012	0.560735
QMKL.01	0.022876	2.283983	4.145512	13.14843	13.7542	0.568636
QMKL.05	0.032723	2.304192	4.159632	13.13903	13.7291	0.567043
QMKL.25	0.081961	2.403008	4.228742	13.09253	13.6068	0.559242
QMKL.50	0.179322	2.588122	4.358501	13.00295	13.37961	0.544569
QMKL.75	0.77572	3.491848	4.996234	12.51572	12.30958	0.472012
QMKL.95	0.955144	3.705768	5.147832	12.3873	12.06666	0.454688
QMKL.99	0.991029	3.746073	5.176405	12.36251	12.02137	0.451421
AQR-KLE	0.066398	2.372169	4.207162	13.10714	13.64489	0.561679

The results indicate that the Adaptive Quantile Regression Kibria–Lukman Estimator (AQR-KLE) produces coefficient estimates close to those obtained using Ordinary Least Squares (OLS), with a regularisation parameter of (k=0.066398). For example, the estimate of (B1) is 2.372169 under AQR-KLE, compared with 2.236445 under OLS, while (B2) is 4.207162 under AQR-KLE and 4.112317 under OLS. Similarly, the estimates of (B3), (B4), and (B5) remain relatively close to their corresponding OLS estimates. This suggests that AQR-KLE preserves the general coefficient pattern of OLS while incorporating a non-zero regularisation penalty.

Furthermore, AQR-KLE exhibits greater numerical stability relative to OLS than QMKL at the upper quantiles. For instance, at the 0.99 quantile, QMKL estimates (B1) as 3.746073 and (B2) as 5.176405, whereas AQR-KLE produces estimates of 2.372169 and 4.207162, respectively. Similarly, the QMKL estimates for (B3) and (B4) at the 0.99 quantile are 12.362510 and 12.021370, compared with 13.107140 and 13.644890 under AQR-KLE. These results show that the upper-quantile QMKL coefficients deviate more substantially from the OLS estimates, while AQR-KLE maintains a coefficient profile closer to the OLS benchmark. This may be advantageous when a regularised estimator is expected to preserve the overall pattern of estimated economic relationships.

iii. Stock and Watson CPS Dataset

The Current Population Survey (CPS) dataset contains 2,950 observations where the dependent variable (Y) is hourly earnings, while potential independent variables include education (X1), degree status (X2), gender (X3), age (X4), geographical region (X5), and survey year (X6). The dataset version should be checked before specifying the final model. CPS Dataset Documentation – AER | Stock–Watson Dataset Directory

Method	k	β_1	β_2	β_3	β_4	β_5	β_6
OLS		6.00034	6.038631	8.00503	-1.065	3.88173	0.028624
RR	0.1	5.982375	6.012159	7.956613	-0.98282	3.889969	0.027836
HUB-M		6.00034	6.038631	8.00503	-1.065	3.88173	0.028624
MKL1	0.172541	5.969609	5.993517	7.922316	-0.92473	3.895791	0.027304
MKL2	0.005305	5.999376	6.037203	8.002427	-1.06058	3.882173	0.028581
MKL3	0.000774	6.000199	6.038422	8.00465	-1.06435	3.881795	0.028618
QMKL.01	0.000214	6.000301	6.038573	8.004925	-1.06482	3.881748	0.028623
QMKL.05	0.000281	6.000289	6.038555	8.004892	-1.06476	3.881753	0.028622
QMKL.25	0.000895	6.000177	6.03839	8.004591	-1.06425	3.881805	0.028617
QMKL.50	0.002847	5.999822	6.037864	8.003633	-1.06262	3.881968	0.028601
QMKL.75	0.022545	5.996252	6.032582	7.993997	-1.04625	3.883609	0.028441
QMKL.95	0.757205	5.874144	5.858433	7.668844	-0.49836	3.938492	0.024118
QMKL.99	0.951441	5.845058	5.818754	7.592727	-0.37134	3.951199	0.023442
AQR-KLE	0.00598	5.999253	6.037022	8.002097	-1.06001	3.88223	0.028576

The results indicate that the Adaptive Quantile Regression Kibria–Lukman Estimator (AQR-KLE) produces coefficient estimates very close to those obtained using Ordinary Least Squares (OLS), with a regularisation parameter of (k=0.00598). For example, the estimate of (B1) is 5.999253 under AQR-KLE, compared with 6.000340 under OLS, while (B3) is 8.002097 under AQR-KLE and 8.005030 under OLS. Similarly, the estimates of (B2), (B4), (B5), and (B6) differ only slightly from their corresponding OLS estimates. These results suggest that AQR-KLE preserves the overall coefficient pattern of OLS while introducing a non-zero regularisation penalty.

Furthermore, AQR-KLE demonstrates greater numerical stability relative to OLS than QMKL at the upper quantiles. For instance, at the 0.99 quantile, QMKL estimates (B1) as 5.845058, (B2) as 5.818754, and (B3) as 7.592727, whereas AQR-KLE produces estimates of 5.999253, 6.037022, and 8.002097, respectively. The difference is particularly noticeable for (B4), where the QMKL estimate changes from −1.065000 under OLS to −0.371340 at the 0.99 quantile, compared with −1.060010 under AQR-KLE. This indicates that AQR-KLE maintains estimates closer to the OLS benchmark, while QMKL exhibits more pronounced coefficient changes at the upper quantiles.

6. Conclusion

This study proposed the Adaptive Quantile Regression Kibria–Lukman Estimator (AQR-KLE) as a regularised approach for addressing multicollinearity and outlier-related challenges in linear regression. The simulation results reported in the study indicate that AQR-KLE achieved the lowest mean squared error (MSE) across almost all considered scenarios, demonstrating promising performance under varying sample sizes, error variances, multicollinearity levels, and contamination conditions. Its performance was particularly encouraging under severe multicollinearity and heavy-tailed error distributions, where conventional estimators such as OLS and ordinary ridge regression experienced greater deterioration. These findings suggest that the proposed estimator may provide a useful balance between estimation stability and robustness under challenging regression conditions.

The real-data applications further showed that AQR-KLE generally produced coefficient estimates close to those obtained using OLS and several other ridge-based estimators. In the reported results, AQR-KLE maintained a relatively stable coefficient profile while employing a non-zero regularisation parameter. Its estimates also remained closer to the OLS benchmark than the QMKL estimates at the upper quantiles for several coefficients. This pattern supports the potential usefulness of AQR-KLE as a regularised alternative across different application areas, including economic growth, credit-card expenditure, macroeconomic indicators, and hourly earnings.

References

Akhtar, N., & Alharthi, M. F.(a) (2024). A comparative study of the performance of new ridge estimators for multicollinearity: Insights from simulation and real data application. *AIP Advances*, *14*(11).

Akhtar, N., & Alharthi, M. F.(b) (2025). Enhancing accuracy in modelling highly multicollinear data using alternative shrinkage parameters for ridge regression methods. *Scientific Reports*, *15*(1), 10774.

Al-Marzouki, S., Nafisah, I. A., Dalam, M. E., Iqbal, K., Sami, F., & Almazah, M. M. (2026). Machine learning based variance estimation under two phase sampling using health and education sector data. *Scientific Reports*.

Alshangiti, A. M., Gemeay, A. M., Bakr, M. E., Yousuf, A. M., El-Qurashi, M. A., Balogun, O. S., ... & Hammad, A. T. (2025). New robust estimator for handling outliers and multicollinearity in gamma regression model with application to breast cancer data. *Scientific Reports*, *15*(1), 38436.

Alshangiti, A. M., Gemeay, A. M., Bakr, M. E., Yousuf, A. M., El-Qurashi, M. A., Balogun, O. S., ... & Hammad, A. T. (2025). New robust estimator for handling outliers and multicollinearity in gamma regression model with application to breast cancer data. *Scientific Reports*, *15*(1), 38436.

Dong, Y., Chen, W., & Elsawah, A. M. (2026). Superior location-scale estimation using moving extreme ranked set sampling design. *Statistics and Computing*, *36*(2), 91.

Gillariose, J., Joseph, J., & Chesneau, C. (2026). Lasso and Ridge regression: a comprehensive review of applications and developments in machine learning. *International Journal of Data Science and Analytics*, *21*(1), 7.

Gillariose, J., Joseph, J., & Chesneau, C. (2026). Lasso and Ridge regression: a comprehensive review of applications and developments in machine learning. *International Journal of Data Science and Analytics*, *21*(1), 7.

Goutte, S., Konstantakis, K. N., Konstantios, D., Michaelides, P. G., & Prelorentzos, A. G. N. (2026). Econometrics at the Extreme: From Quantile Regression to QFAVAR1. *Journal of Economic Surveys*.

Haider, B., Asim, S. M., Wasim, D., Golam Kibria, B. M., & Zaman, Q. (2026). Quantile-based two-parameter robust ridge M-estimators for linear regression model for mitigating multicollinearity and outliers. *Statistics*, 1-21.

Haider, B., Asim, S. M., Wasim, D., Golam Kibria, B. M., & Zaman, Q. (2026). Quantile-based two-parameter robust ridge M-estimators for linear regression model for mitigating multicollinearity and outliers. *Statistics*, 1-21.

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

Irshad, M., Zaman, Q., Wasim, D., & Kibria, B. G. (2026). Hybrid robust ridge M-estimators via variance adjustment under multicollinearity and contaminated responses in linear regression: a simulation and empirical study. *Communications in Statistics-Simulation and Computation*, 1-24.

Irshad, M., Zaman, Q., Wasim, D., & Kibria, B. G. (2026). Hybrid robust ridge M-estimators via variance adjustment under multicollinearity and contaminated responses in linear regression: a simulation and empirical study. *Communications in Statistics-Simulation and Computation*, 1-24.

Irshad, M., Zaman, Q., Wasim, D., & Kibria, B. G. (2026). Hybrid robust ridge M-estimators via variance adjustment under multicollinearity and contaminated responses in linear regression: a simulation and empirical study. *Communications in Statistics-Simulation and Computation*, 1-24.

Irshad, M., Zaman, Q., Wasim, D., & Kibria, B. G. (2026). Hybrid robust ridge M-estimators via variance adjustment under multicollinearity and contaminated responses in linear regression: a simulation and empirical study. *Communications in Statistics-Simulation and Computation*, 1-24.

Jain, L., & Chopra, P. (2026). Hybrid Econometric–Machine Learning Models for High-Dimensional Data: Robust Approaches to Anomaly Detection and Inference. In *Robust Methods for Anomaly Detection in Econometrics* (pp. 165-208). IGI Global Scientific Publishing.

Kalina, J. (2025). From robust neural networks toward robust nonlinear quantile estimation. *Sequential Analysis*, *44*(3), 326-350.

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics – Simulation and Computation*, 32(2), 419–435. Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.

Li, S., & Ullah, M. I. (2025). Quantile based bootstrap estimation of biasing parameters for heteroscedastic ridge regression model. *Communications in Statistics-Simulation and Computation*, 1-21.

Li, S., & Ullah, M. I. (2025). Quantile based bootstrap estimation of biasing parameters for heteroscedastic ridge regression model. *Communications in Statistics-Simulation and Computation*, 1-21.

Li, Z., Xuan, X., & Carbone, D. (2026). Distribution-Aware Robust Bilevel Optimization: Quantile-Guided Huber Updates in Two-Timescale Stochastic Approximation. *arXiv preprint arXiv:2606.22436*.

Mahmood, S. H. (2024). Estimating models and evaluating their efficiency under multicollinearity in multiple linear regression: a comparative study. *Zanco Journal of Human Sciences*, *28*(5), 264-277.

Muthukrishnan, R., & Ramakrishnan, K. (2024). EFFECT OF CLASSICAL AND ROBUST REGRESSION ESTIMATORS IN THE CONTEXT OF HIGHDIMENSIONAL DATA WITH MULTICOLLINEARITY AND OUTLIERS. *Reliability: Theory & Applications*, *19*(1 (77)), 335-341.

Naz, H., Shah, I., Wasim, D., & Ali, S. (2025). Robust Kibria Estimators for Mitigating Multicollinearity and Outliers in a Linear Regression Model. *Stats*, *8*(4), 119.

Nosheen, M., Ahmad, S., & Naz, S. (2025). ORDINARY LEAST SQUARES (OLS) METHODS AND ISSUES: THEORETICAL FOUNDATIONS, CHALLENGES, AND METHODOLOGICAL REMEDIES. *Pakistan Journal of Social Science Review*, *4*(2), 1-9.

Olaluwoye, O. T., Lukman, A. F., Alrasheedi, M. A., Nzomo, W. N., & Farghali, R. A. (2025). Robust estimation methods for addressing multicollinearity and outliers in beta regression models. *Scientific Reports*, *15*(1), 11649.

Olaluwoye, O. T., Lukman, A. F., Alrasheedi, M. A., Nzomo, W. N., & Farghali, R. A. (2025). Robust estimation methods for addressing multicollinearity and outliers in beta regression models. *Scientific Reports*, *15*(1), 11649.

Osman, E. G. A. (2025). Entropy based solutions for detecting and treating multicollinearity in econometrics. *Discover Analytics*, *3*(1), 18.

Rajh-Weber, H., Huber, S. E., & Arendasy, M. (2025). A practice-oriented guide to statistical inference in linear modeling for non-normal or heteroskedastic error distributions. *Behavior Research Methods*, *57*(12), 338.

Suhail, M., Chand, S., & Aslam, M. (2023). New quantile based ridge M-estimator for linear regression models with multicollinearity and outliers. *Communications in Statistics-Simulation and Computation*, *52*(4), 1417-1434.

Suhail, M., Shaheen, S., & Ahmed, S. (2021). Quantile-based ridge-type regression estimators. *Journal of Statistical Computation and Simulation*, 91(17), 3585–3603.

Suresh, V., Revathi, B. S., & Guerrero, J. M. (2026). A non-parametric adaptive conformal inference based probabilistic hour-ahead solar PV power forecasting method. *Scientific Reports*.

Wasim, M., Shaheen, S., & Suhail, M. (2025). Quantile-based robust Kibria–Lukman estimator. *Communications in Statistics – Simulation and Computation*.

Yan, X., Su, Y., & Ma, W. (2023). Ensemble multi-quantiles: adaptively flexible distribution prediction for uncertainty quantification. *IEEE transactions on pattern analysis and machine intelligence*, *45*(11), 13068-13082.

Zaman, Q., Wasim, D., Kibria, B. G., Shabbir, M., & Ullah, N. (2026). Mitigating multicollinearity in ridge regression: a comparative study of quantile-based two parameter weighted KL-estimators. *Research in Statistics*, *4*(1), 2610197.