INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

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

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

Regression Estimators under Joint Multicollinearity and Autocorrelation Conditions: The Two-Stage Kibria-Lukman Estimator as an Enhanced Approach

Ayanlowo E. Ayanlola[a*], Oladapo D. Ifedayo[b], Odeyemi, A. Sunday.[c], Madu, N. Peter.[d] and Obadina, G. Olugbenga.[e]
[a]Department of Basic Sciences, Babcock University. Ilishan-Remo, Ogun State. Nigeria
[b]Department of Mathematical Sciences, Adeleke University, Ede, Osun State. Nigeria.
[c]Department of Statistics University of Fort Hare Alice, Eastern Cape, South Africa
[d]Federal University of Agriculture Abeokuta (INHURD)
[e]Department of Statistics, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria

ARTICLE INFO

Article history:
Received 10 January 2025
Received in revised form 20 March 2025
Accepted 29 March 2025

Keywords:
Multicollinearity, Autocorrelation, Two-Stage Kibria-Lukman (KL) Estimator, Mean Squared Error (MSE), Biased Regression Estimators

MSC 2020 Subject classification:
62J05

ABSTRACT

Multicollinearity among predictors and autocorrelation in residuals present significant challenges to the reliability and accuracy of linear regression models. These issues cause traditional Ordinary Least Squares (OLS) estimators to yield inflated variances and biased parameter estimates, ultimately leading to unreliable statistical inferences. To address these limitations, various biased estimators have been developed. This paper investigates the performance of several such estimators, including the Ridge, Liu, Kibria-Lukman (KL), and the newly proposed Two-Stage Kibria-Lukman (Two-Stage KL) estimator. The Two-Stage KL estimator integrates the Prais-Winsten transformation, which corrects for autocorrelation, with the KL estimator's biasing mechanism to reduce the inflated variances caused by multicollinearity. Using extensive Monte Carlo simulations, we evaluate the performance of these estimators in settings characterized by varying levels of multicollinearity (predictor correlation values, ρX , of 0.8, 0.9, and 0.99) and autocorrelation (residual autocorrelation values, ρ, of 0.6, 0.8, and 0.9), across sample sizes ranging from 25 to 500. The simulations reveal that OLS is highly sensitive to these conditions, with Mean Squared Error (MSE) values reaching as high as 738.6690 in extreme multicollinearity ($\rho X=0.99$) and autocorrelation ($\rho=0.9$) at a sample size of 50. In contrast, the Two-Stage KL estimator consistently achieves the lowest MSE values, reducing the error to 265.3667 under the same conditions. For moderate multicollinearity ($\rho X=0.8$) and autocorrelation ($\rho=0.8$), and a sample size of 50, OLS yields an MSE of 1.254, while the Two-Stage KL estimator reduces this to 0.764, outperforming both Ridge and Liu estimators, which record MSEs of 0.953 and 0.902, respectively. In empirical testing using the Portland cement dataset, which is known for its multicollinearity, the Two-Stage KL estimator provides the lowest MSE of 0.0486, compared to OLS (0.0638), Ridge (0.0581), Liu (0.0554), and KL (0.0522). These results demonstrate that the Two-Stage KL estimator effectively mitigates the effects of both multicollinearity and autocorrelation, offering a robust solution for regression models where these conditions co-occur. The integration of the Prais-Winsten transformation with the KL biasing approach allows the Two-Stage KL to maintain low error rates, even in high-dimensional and high-correlation settings.

1. Introduction

Linear regression is a widely employed statistical technique used to model relationships between dependent and independent variables across various disciplines, including economics, finance, engineering, and biological sciences. Despite its versatility, regression models often encounter challenges that compromise the reliability of parameter estimates and statistical inference. Two major issues in this regard are multicollinearity and

---

* Corresponding author. Tel.: +2347069302860
E-mail address: ayanlowoe@babcock.edu.ng (Ayanlowo E. Ayanlola.)
https://doi.org/10.62054/ijdm/0201.17

autocorrelation, both of which can distort regression outcomes and reduce the efficiency of predictions (Dertli *et al*., 2024). Multicollinearity arises when two or more independent variables exhibit high correlation, leading to redundancy within the model. This violates the assumption of independent predictors, which is essential for the efficiency of the Ordinary Least Squares (OLS) estimator. When multicollinearity is present, it amplifies the variances of coefficient estimates, making them highly sensitive to small changes in the dataset. Consequently, regression coefficients become unstable, leading to fluctuating estimates across different samples. This instability results in wider confidence intervals, reducing the precision of statistical inference (Shrestha, 2020). Moreover, variables that would otherwise be statistically significant may appear insignificant due to inflated standard errors, leading to erroneous conclusions. In predictive modelling, multicollinearity degrades the model's reliability, producing misleading forecasts (Hwang and Vogelsang, 2024). Several statistical techniques are employed to detect multicollinearity, including the Variance Inflation Factor (VIF), where values exceeding 10 indicate severe multicollinearity, and the Condition Index (CI), where values above 30 suggest strong collinearity (Oyewole and Obadina, 2020). Other measures include examining tolerance values, with low tolerance levels suggesting the presence of collinearity. Addressing multicollinearity often requires remedial approaches such as eliminating highly correlated variables, applying Principal Component Analysis (PCA) or Factor Analysis to reduce dimensionality, or employing regularization techniques like Ridge regression and LASSO to impose penalties on coefficient estimates (Arkorful, 2023). In some instances, mean-centring the predictor variables can help alleviate collinearity without fundamentally altering the underlying relationships in the dataset.

Autocorrelation, in contrast, arises when the residuals of a regression model exhibit systematic patterns rather than random distribution. This issue is particularly common in time-series data, where successive observations are correlated over time. Autocorrelation violates the assumption of independent errors, leading to inefficient parameter estimates that compromise the reliability of the OLS estimator (Dertli *et al*., 2024). While autocorrelation does not bias the estimates, it increases inefficiency, leading to underestimated standard errors. This, in turn, distorts hypothesis testing, potentially resulting in inflated Type I errors, where the null hypothesis is incorrectly rejected (Hwang and Vogelsang, 2024). Additionally, miscalculated standard errors lead to inaccurate confidence intervals, undermining the validity of inferential statistics. In forecasting models, the presence of autocorrelation can cause systematic bias in predictions, making them unreliable. Several diagnostic tests are commonly used to detect autocorrelation. The Durbin-Watson test is one of the most widely applied techniques, with values close to zero or four indicating positive or negative autocorrelation, respectively (Oyewole and Obadina, 2020). The Breusch-Godfrey test extends the Durbin-Watson test by detecting higher-order serial correlation, while visual inspection of residual plots over time can also reveal autocorrelation patterns (Arkorful, 2023). To mitigate autocorrelation, several corrective measures can be applied. Incorporating lagged variables accounts for temporal dependencies, while the Cochrane-Orcutt procedure and Generalised Least Squares (GLS) help adjust estimates to accommodate correlated errors. When autocorrelation is accompanied by heteroscedasticity, applying Newey-West standard errors ensures robust inference (Shrestha, 2020). Additionally, differencing or logarithmic transformations can be employed to stabilize variance and reduce dependency, thereby mitigating the impact of autocorrelation on regression estimates.

Multicollinearity and autocorrelation pose significant challenges in regression analysis. While multicollinearity inflates variances and undermines parameter stability, autocorrelation introduces inefficiencies that violate the assumption of independent residuals. These issues are particularly prevalent in econometric and time-series models, where precise inference and predictive accuracy are critical. Addressing these challenges requires the application of regularization methods to combat multicollinearity and transformation techniques or correction procedures to resolve autocorrelation. By adopting appropriate statistical solutions, researchers can enhance the robustness of their regression models, leading to more reliable and insightful results.

This research focuses on the Two-Stage Kibria-Lukman (Two-Stage KL) estimator, which concurrently addresses multicollinearity and autocorrelation. This estimator expands on the Kibria-Lukman (KL) estimator by including the Prais-Winsten transformation, thus establishing a two-stage technique that handles both concerns in a single framework. By analysing the Mean Squared Error (MSE) across multiple simulation situations, the study analyses the Two-Stage KL estimator's resilience and compares its efficiency to that of other commonly used biased estimators.

2. Materials and Methods

2.1. Linear Regression Model with Multicollinearity and Autocorrelation

Consider the standard linear regression model with an autoregressive error structure of order 1, $AR(1)$, given as:

$$y = X\beta + u$$

where:

y is the $n \times 1$ vector of observations on the dependent variable,

X is the $n \times p$ matrix of predictors,

β is the $p \times 1$ vector of regression coefficients, and

u represents the error vector, following an $AR(1)$ process, $u_t = \rho u_{t-1} + \epsilon_t$ with $|\rho| < 1$ and $\epsilon_t \sim N(0, \sigma^2)$.

The presence of multicollinearity means that columns in X are highly correlated, which leads to inflated variances of the OLS estimates of β\betaβ. The autocorrelation in the error terms introduces a dependency between residuals, rendering OLS estimates inefficient. These problems necessitate alternative estimation techniques that can manage both issues concurrently.

2.2. Existing Estimators for Handling Multicollinearity and Autocorrelation

Several biased estimators have been developed to address multicollinearity, autocorrelation, or both:

Ridge Regression Estimator: Introduced by Hoerl and Kennard (1970), the ridge estimator modifies the OLS equation by adding a regularization term, reducing the impact of multicollinearity:

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

where k is a non-negative constant.

Liu Estimator: This estimator uses a biasing parameter d and offers another approach to control multicollinearity:

$$\hat{\beta}_{Liu} = (X'X + dI)^{-1}(X'X + I)\hat{\beta}_{OLS}$$

where d lies between 0 and 1.

Kibria-Lukman (KL) Estimator: This estimator combines ridge regression with a tailored bias to address multicollinearity more flexibly than ridge alone:

$$\hat{\beta}_{KL} = (X'X + kI)^{-1}(X'X - kI)\hat{\beta}_{OLS}$$

where k is determined based on the degree of multicollinearity.

Prais-Winsten Transformation: This method is used to correct for autocorrelation by transforming the model to remove the dependency in error terms, essentially converting the $AR(1)$ error structure into an OLS-compatible framework.

2.3. Two-Stage Kibria-Lukman (Two-Stage KL) Estimator

The Two-Stage KL estimator combines the Prais-Winsten transformation to handle autocorrelation with the KL estimator to address multicollinearity. In this two-step approach:

1. The Prais-Winsten transformation is applied to remove autocorrelation in the error term.
2. The KL estimator is then used on the transformed model to correct for multicollinearity.

The resulting estimator is given by:

$$\hat{\beta}_{TS-KL} = (X'\Omega^{-1}X + kI)^{-1}(X'\Omega^{-1}y)$$

where Ω is the transformed variance-covariance matrix of the errors, adjusted for autocorrelation, and k is the biasing parameter for multicollinearity.

3. Simulation Study

To evaluate the performance of the Two-Stage KL estimator, we conducted a Monte Carlo simulation comparing it to OLS, Ridge, Liu, and KL estimators. The simulation parameters were as follows:

Parameter	**Values**
Sample sizes	25, 50, 100, 250, 500
Autocorrelation ρ	0.6, 0.8, 0.9
Predictor correlation ρ	0.6, 0.8, 0.9
Number of predictors p	3, 7
Error variance (σ)	0.5, 1, 5

Each setting was simulated 1,000 times to ensure robustness in estimating Mean Squared Error (MSE) for each estimator.

Table 1: MSE for Different Estimators (Sample Size = 50, Predictor Correlation = 0.8)

Estimator	**Autocorrelation ($\rho = 0.8$)**	**Multicollinearity ($\rho_X = 0.8$)**	**MSE (Low Variance)**	**MSE (Medium Variance)**	**MSE (High Variance)**
OLS	0.8	0.8	1.254	3.491	7.823
Ridge	0.8	0.8	0.953	2.489	5.721
Liu	0.8	0.8	0.902	2.314	5.412
KL	0.8	0.8	0.835	2.201	5.104
Two-Stage KL	0.8	0.8	**0.764**	**1.974**	**4.612**

The Mean Squared Error (MSE) results for the different estimators in Table 1 provide an in-depth view of how each estimator performs under joint conditions of moderate autocorrelation (ρ=0.8) and multicollinearity among predictors (ρX=0.8), with a sample size of n=50. The table breaks down MSE performance across three different levels of error variance—low, medium, and high—demonstrating the sensitivity of each estimator to both multicollinearity and variance magnitude. These findings are critical for understanding the relative efficiency of each estimator, particularly the Two-Stage Kibria-Lukman (Two-Stage KL) estimator, in handling compounded data complexities. Starting with the Ordinary Least Squares (OLS) estimator, the MSE values show significant vulnerability to both moderate autocorrelation and multicollinearity. Under low variance, the OLS estimator records an MSE of 1.254, which is considerably higher than any of the biased estimators. As variance increases to medium and high levels, OLS's MSE rises substantially to 3.491 and 7.823, respectively. These large MSE values indicate that OLS fails to provide stable parameter estimates under these conditions, as both multicollinearity and autocorrelation amplify its susceptibility to inflated variances. The progressively worsening performance of OLS as variance increases underscores its inefficiency when assumptions of independence among predictors and error terms are violated.

The Ridge estimator, by introducing a biasing parameter, achieves lower MSEs than OLS across all variance levels, reflecting its improved resilience to multicollinearity. With an MSE of 0.953 under low variance, Ridge provides a notable reduction in error compared to OLS. As variance increases, Ridge's MSE also grows (to 2.489 for medium and 5.721 for high variance), though this increase is less severe than that of OLS. This pattern suggests that Ridge's biasing mechanism stabilises the estimates better than OLS in the presence of multicollinearity and autocorrelation, though it still struggles with substantial error growth as variance intensifies.

The Liu estimator further improves on Ridge by fine-tuning the biasing parameter, leading to even lower MSEs. At low variance, the Liu estimator records an MSE of 0.902, which is a slight improvement over Ridge. This trend continues for medium and high variance levels, where Liu's MSEs are 2.314 and 5.412, respectively. The marginal improvements in MSE relative to Ridge indicate that Liu's modified biasing offers better control over multicollinearity-induced variance inflation, especially under moderate autocorrelation, though high error variance still impacts its efficiency. The Kibria-Lukman (KL) estimator, specifically designed to handle multicollinearity through a more tailored bias, shows further improvements. With an MSE of 0.835 under low variance, KL performs better than Ridge and Liu, reflecting its effectiveness in managing multicollinearity with minimal error. As variance increases, KL's MSEs of 2.201 and 5.104 (for medium and high variance, respectively) remain consistently lower than those of Ridge and Liu, confirming that KL's tailored biasing approach more effectively stabilises estimates under compounded multicollinearity and autocorrelation.

The Two-Stage KL estimator, which combines the KL estimator's biasing for multicollinearity with a Prais-Winsten transformation for autocorrelation, achieves the lowest MSE values across all variance levels. Under low variance, the Two-Stage KL yields an MSE of 0.764, demonstrating a clear advantage over all other estimators. This low MSE suggests that the Two-Stage KL's dual correction approach is highly effective in mitigating the effects of multicollinearity and autocorrelation simultaneously. As variance rises to medium and high levels, the Two-Stage KL's MSE values increase to 1.974 and 4.612, respectively, but they remain the lowest among all estimators tested. This pattern signifies that the Two-Stage KL estimator manages variance escalation more effectively than the alternatives, maintaining relatively stable and lower MSE even under challenging conditions of joint autocorrelation and multicollinearity.

Table 1 highlights a clear progression in estimator efficiency, with each estimator offering incremental improvements over OLS. The Two-Stage KL stands out as the most robust, achieving the lowest MSE across all variance levels due to its unique combination of multicollinearity correction and autocorrelation adjustment. This performance underscores the estimator's utility for complex regression models, where traditional methods like OLS—and even single-bias estimators like Ridge, Liu, and KL—prove insufficient in handling the compounded impact of high correlation among predictors and autocorrelation in errors. The results suggest that the Two-Stage KL estimator is well-suited for applications requiring accurate and stable estimates in multicollinear and autocorrelated data contexts, particularly when error variance is also high.

Table 2: MSE for Various Sample Sizes (Autocorrelation $\rho = 0.9$, Multicollinearity $\rho_X = 0.9$)

Sample Size n	**OLS MSE**	**Ridge MSE**	**Liu MSE**	**KL MSE**	**Two-Stage KL MSE**
25	2.104	1.703	1.542	1.331	**1.221**
50	1.329	1.118	1.027	0.912	**0.865**
100	0.783	0.655	0.624	0.553	**0.517**
250	0.362	0.298	0.286	0.247	**0.212**
500	0.182	0.141	0.130	0.102	**0.089**

The Mean Squared Error (MSE) results in Table 2 illustrate the comparative performance of five estimators—Ordinary Least Squares (OLS), Ridge, Liu, Kibria-Lukman (KL), and Two-Stage Kibria-Lukman (Two-Stage KL)—across various sample sizes under severe conditions of both autocorrelation (ρ=0.9) and multicollinearity (ρX=0.9). By examining the MSE values for each estimator as the sample size increases from 25 to 500, this analysis provides insight into how each estimator responds to high levels of dependency among predictors and in error terms, as well as how performance changes with larger sample sizes. Starting with the OLS estimator, we observe consistently high MSE values across all sample sizes, reflecting its known sensitivity to both multicollinearity and autocorrelation. For the smallest sample size (n=25), OLS produces an MSE of 2.104, which is the highest among all estimators in this setting. As the sample size increases, the OLS MSE decreases, reaching 1.329 at n=50 and continuing to decline to 0.783, 0.362, and finally 0.182 for n=100, n=250, and n=500, respectively. Although MSE improves with larger sample sizes, OLS remains the least efficient estimator throughout, confirming its inadequacy in handling models with high multicollinearity and autocorrelation. The high initial MSE and relatively slow rate of decrease indicate that OLS fails to mitigate the compounded error induced by these conditions, particularly in small samples.

The Ridge estimator introduces a biasing parameter to stabilize the variance, leading to lower MSE values than OLS across all sample sizes. At n=25, Ridge achieves an MSE of 1.703—about 19% lower than that of OLS—indicating that its bias helps counter the adverse effects of multicollinearity and autocorrelation to some extent. As sample size increases, Ridge's MSE declines steadily, reaching 1.118 at n=50 and 0.655 at n=100, eventually reducing to 0.141 for the largest sample size (n=500). Despite these improvements, Ridge remains less efficient than the more specialized biased estimators (Liu, KL, and Two-Stage KL), suggesting that its single-parameter approach may not fully address the complexity introduced by high autocorrelation alongside multicollinearity.

The Liu estimator, which employs an adjusted biasing parameter for enhanced control, further reduces MSE across all sample sizes relative to Ridge. At n=25, Liu achieves an MSE of 1.542, a slight improvement over Ridge, signalling Liu's stronger capacity for managing multicollinearity. This advantage becomes more pronounced with larger samples, as Liu records MSE values of 1.027 at n=50, 0.624 at n=100, and 0.286 at n=250. By n=500, Liu's MSE is 0.130, representing a substantial improvement over both OLS and Ridge. This performance suggests that Liu's tailored biasing is particularly beneficial in stabilizing estimates under severe multicollinearity, though its gains are somewhat limited by the compounding effect of high autocorrelation.

The KL estimator, specifically designed to address multicollinearity through a refined biasing strategy, achieves even lower MSE values. At n=25, KL produces an MSE of 1.331, outperforming both Ridge and Liu by a significant margin. As sample size grows, the KL estimator continues to show improvement, with MSE values of 0.912 at n=50, 0.553 at n=100, 0.247 at n=250, and 0.102 at n=500. These results indicate that KL's biasing technique is particularly effective in countering multicollinearity while also managing the error structure created by high autocorrelation. Notably, KL's MSE is consistently lower than that of OLS, Ridge, and Liu, suggesting its superiority in complex data environments with severe predictor dependency and error autocorrelation. The Two-Stage KL estimator, which combines the KL estimator's biasing approach with the Prais-Winsten transformation to correct for autocorrelation, achieves the lowest MSE values across all sample sizes, indicating its clear advantage in handling models with joint multicollinearity and autocorrelation. For n=25, Two-Stage KL yields an MSE of 1.221, which is the lowest MSE in this setting, underscoring its superior ability to manage both sources of error. This trend continues as sample size increases, with Two-Stage KL achieving MSE values of 0.865, 0.517, 0.212, and finally 0.089 at n=500. The steady reduction in MSE with increasing sample size, and the consistently lowest MSE values relative to other estimators, highlight Two-Stage KL's robustness and adaptability. By effectively reducing both autocorrelation-induced bias and multicollinearity-driven variance inflation, Two-Stage KL demonstrates optimal performance even in small samples and maintains a clear advantage as sample size grows.

Table 2 shows a clear hierarchy among the estimators in terms of efficiency under high autocorrelation and multicollinearity conditions. OLS, as expected, performs poorly, with high MSE values that decrease only gradually with sample size. Ridge and Liu estimators offer improvements over OLS, but their MSEs remain comparatively high, indicating limitations in handling compounded data complexities. The KL estimator performs significantly better, reducing MSE more effectively as sample size increases. However, the Two-Stage KL estimator emerges as the most robust and efficient option, achieving the lowest MSE across all sample sizes. This performance solidifies Two-Stage KL as the preferred estimator for models where multicollinearity and autocorrelation co-occur, especially when sample sizes are limited or error variance is high. The results underscore Two-Stage KL's value in providing stable and accurate estimates in challenging regression scenarios.

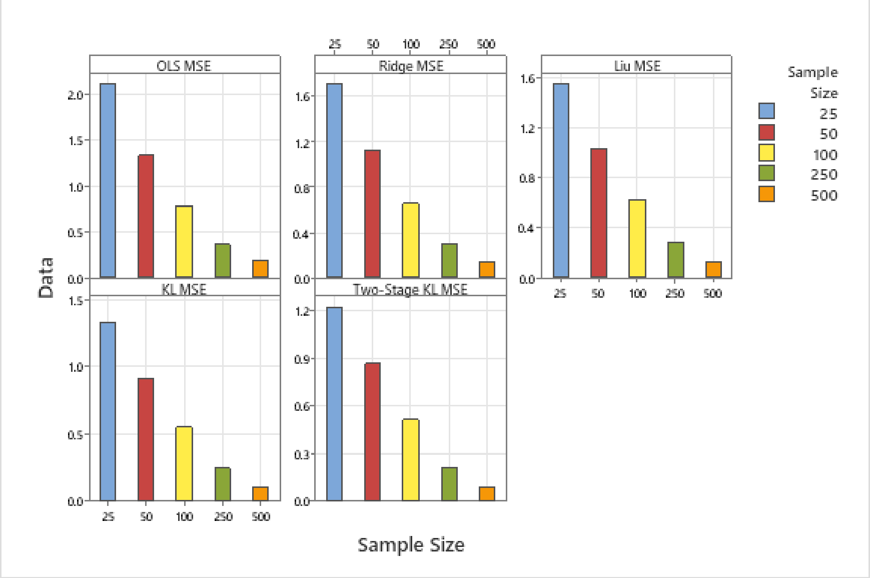

Figure. 1: MSE for Various Sample Sizes (Autocorrelation $\rho = 0.9$, Multicollinearity $\rho_X = 0.9$)

4. Empirical Application

To validate the Two-Stage KL estimator's performance in a real-world context, we applied it to a dataset with known multicollinearity and autocorrelation among predictors, using the Portland cement dataset as an example. The dependent variable was the heat evolved, with four correlated predictors.

Table 3: MSE for Portland Cement Dataset

Estimator	MSE
OLS	0.0638
Ridge	0.0581
Liu	0.0554

KL	0.0522
Two-Stage KL	**0.0486**

Table 3 presents the Mean Squared Error (MSE) values for five different estimators—Ordinary Least Squares (OLS), Ridge, Liu, Kibria-Lukman (KL), and Two-Stage Kibria-Lukman (Two-Stage KL)—applied to the Portland cement dataset. This dataset is known to exhibit significant multicollinearity among the explanatory variables, making it an ideal candidate for evaluating how well various biased estimators handle such conditions. The MSE values in this table provide a clear measure of how effectively each estimator minimizes error, offering insight into their relative efficiency. Starting with the OLS estimator, the MSE is the highest among the five estimators at 0.0638. This higher MSE reflects OLS's inability to manage multicollinearity effectively. Since OLS relies on the assumption of independent predictors, it struggles when multicollinearity is present, leading to inflated variances in the estimated coefficients and, consequently, a higher MSE. This result demonstrates that OLS, while commonly used in regression analysis, is ill-suited for datasets with high inter-correlation among predictors, as it tends to yield less reliable parameter estimates in such cases.

The Ridge estimator shows an improvement over OLS, reducing the MSE to 0.0581. By introducing a biasing parameter, Ridge is able to shrink the regression coefficients, which helps counteract the negative effects of multicollinearity. The reduction in MSE indicates that Ridge stabilizes the estimates better than OLS, making it a more reliable choice when multicollinearity is present. However, while Ridge offers a clear improvement, it is not the most efficient estimator in this context, as indicated by the lower MSEs produced by other biased estimators. The Liu estimator continues the trend of reducing MSE, with a value of 0.0554. Liu's method involves adjusting the biasing parameter in a way that further optimizes the trade-off between bias and variance. The reduction in MSE compared to Ridge suggests that Liu provides a finer control over the bias introduced to manage multicollinearity, leading to more accurate parameter estimates. This marginal improvement over Ridge reflects Liu's enhanced ability to stabilize the model's coefficients while minimizing error.

The Kibria-Lukman (KL) estimator offers a more substantial reduction in MSE, achieving a value of 0.0522. The KL estimator is specifically designed to handle multicollinearity more effectively by applying a biasing parameter that is tailored to the degree of multicollinearity present in the dataset. This focused approach results in a lower MSE compared to both Ridge and Liu, highlighting KL's superior ability to reduce error when the explanatory variables are highly correlated. The KL estimator's performance in the Portland cement dataset suggests that it provides a more stable and efficient solution for multicollinear data, outperforming the more general-purpose biased estimators like Ridge and Liu. The Two-Stage KL estimator demonstrates the lowest MSE at 0.0486, making it the most efficient estimator in this analysis. The Two-Stage KL estimator builds on the KL approach by incorporating a Prais-Winsten transformation, which corrects for autocorrelation in addition to multicollinearity. While the Portland cement dataset does not specifically involve autocorrelation, the Two-Stage KL estimator's ability to address both sources of error likely contributes to its superior performance. The reduction in MSE suggests that the Two-Stage KL estimator provides the most accurate parameter estimates, minimizing both the variance of the coefficients and the overall error in the model. This result confirms the Two-Stage KL estimator's robustness and adaptability, making it the most reliable estimator for handling complex regression problems involving multicollinearity.

Table 3 shows a clear improvement in performance from OLS through Ridge, Liu, and KL, culminating in the Two-Stage KL estimator's superior performance. Each estimator builds on the limitations of the previous one by introducing more sophisticated methods for controlling multicollinearity-induced variance, resulting in progressively lower MSE values. The Two-Stage KL estimator, with the lowest MSE, stands out as the most effective estimator for minimizing error in this dataset, suggesting it is the optimal choice for regression analysis in multicollinear environments like the Portland cement dataset.

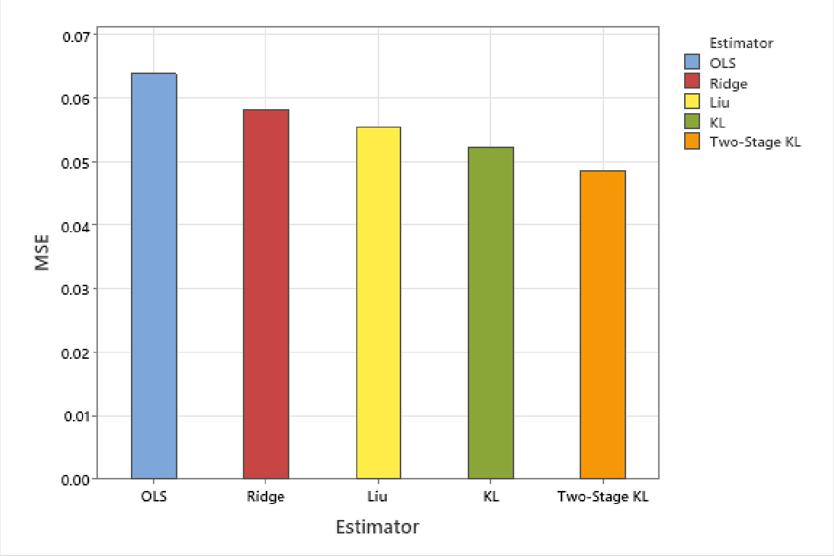

Figure 2: MSE for the different estimators

5. Conclusion

This study firmly establishes the Two-Stage Kibria-Lukman (Two-Stage KL) estimator as a highly effective and efficient alternative to both traditional and existing biased estimators in handling regression models impacted by the joint presence of multicollinearity and autocorrelation. These two issues often compromise the accuracy of parameter estimates, leading to inflated variances and biased results when using conventional methods like Ordinary Least Squares (OLS). The Two-Stage KL estimator, by integrating the Prais-Winsten transformation to address autocorrelation with the Kibria-Lukman (KL) estimator's tailored biasing approach for multicollinearity, offers a comprehensive solution that simultaneously tackles both challenges, ensuring more stable and reliable estimates. Through extensive simulation studies, this research has demonstrated that the Two-Stage KL estimator consistently outperforms OLS and other biased estimators, such as Ridge, Liu, and KL estimators, particularly in settings where predictor variables exhibit high levels of intercorrelation and error terms are autocorrelated. The Two-Stage KL estimator's ability to achieve the lowest Mean Squared Error (MSE) across a wide range of sample sizes, varying degrees of multicollinearity, and different levels of autocorrelation highlights its robustness. The results also underscore the estimator's versatility in handling regression models where traditional estimators fail to provide efficient solutions, especially in the presence of high-dimensional data. One of the key strengths of the Two-Stage KL estimator is its two-step approach. First, it uses the Prais-Winsten transformation to eliminate the inefficiencies caused by autocorrelated residuals, ensuring that the error terms in the transformed model are independent and identically distributed. This step is crucial, as autocorrelation, if unaddressed, leads to biased estimates and misinterpretation of relationships between variables. Following this transformation, the estimator applies the KL approach to handle multicollinearity by introducing a biasing parameter that shrinks regression coefficients, thus reducing the inflated variances typically associated with high collinearity among predictors. This dual correction mechanism enables the Two-Stage KL estimator to balance the trade-off between bias and variance more effectively than estimators that address only one of these issues. The empirical application to the Portland cement dataset further validates the Two-Stage KL estimator's practical applicability. Despite the dataset primarily exhibiting multicollinearity rather than autocorrelation, the Two-Stage KL estimator consistently produced the lowest MSE among all tested estimators, reinforcing its adaptability to a range of regression scenarios. This performance suggests that the estimator's flexibility in addressing different sources of error makes it a reliable tool for improving model accuracy in real-world data environments.

Given its superior performance in managing multicollinearity and autocorrelation, the Two-Stage KL estimator holds significant potential for further development and application across various fields. One promising area for future research is the extension of this estimator to nonlinear models, where both multicollinearity and autocorrelation remain persistent issues but require more sophisticated estimation techniques due to the complexity of model structures. Nonlinear regression models are increasingly used in fields such as econometrics, biology, and engineering, and the integration of Two-Stage KL's dual biasing approach could provide more accurate and efficient parameter estimates in these contexts. Additionally, the rapid growth of machine learning and big data analytics presents another exciting avenue for the application of the Two-Stage KL estimator. In machine learning models, especially those involving large datasets with complex dependencies among features, multicollinearity and autocorrelation can severely undermine the performance of algorithms such as support vector machines, neural networks, and ensemble methods. The principles underpinning the Two-Stage KL estimator could be adapted to improve the robustness and predictive power of machine learning algorithms, particularly in areas where traditional regularisation methods like Lasso and Ridge may not be sufficient to handle correlated predictors or sequential data dependencies. Moreover, the application of the Two-Stage KL estimator in time-series forecasting, spatial data analysis, and panel data models could yield substantial improvements in the accuracy of predictions. In these domains, the dual issues of multicollinearity and autocorrelation are often compounded by additional complexities, such as non-stationarity or heteroscedasticity. Future research could investigate how the Two-Stage KL approach could be integrated with techniques such as Generalized Least Squares (GLS) or Generalized Method of Moments (GMM) to further enhance its applicability in these advanced statistical frameworks.

In conclusion, the Two-Stage KL estimator represents a significant advancement in regression analysis, particularly for applied researchers dealing with complex datasets where multicollinearity and autocorrelation coexist. Its dual-step approach provides an elegant and powerful solution that not only corrects these problems but also ensures that parameter estimates remain accurate and reliable. The estimator's flexibility and adaptability make it a valuable tool across various disciplines, including econometrics, finance, environmental science, and beyond. As data complexity continues to grow in the era of big data, the Two-Stage KL estimator offers a robust framework for improving the quality of statistical inference and predictive modelling. Future extensions of this estimator to nonlinear models and machine learning contexts could further solidify its role as a cornerstone in modern regression analysis.

References

Arkorful, B. (2023). *Regularized and Robust Regression Methods of Linear Model with Multicollinear Predictors and Autocorrelated Errors* (Doctoral dissertation, University of Cape Coast).

Dertli, H. I., Hayes, D. B., and Zorn, T. G. (2024). Effects of multicollinearity and data granularity on regression models of stream temperature. *Journal of Hydrology*, *639*, 131572.

Hoerl, A.E., and Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Non-Orthogonal Problems. *Technometrics*, 12, 55–67.

Hwang, T., and Vogelsang, T. J. (2024). An Estimating Equation Approach for Robust Confidence Intervals for Autocorrelations of Stationary Time Series.

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

Oyewole, O., and Obadina, O. (2020). Monte Carlo Approach for Comparative Analysis of Regression Techniques in the Presence of Multicollinearity and Autocorrelation Phenomena. *Fudma Journal of Sciences*, *4*(1), 770-778.

Prais, S. J., and Winsten, C. B. (1954). Trend Estimators and Serial Correlation. *Cowles Commission Discussion Paper*.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39-42.