Pakistan Journal of Social Sciences Review (PJSSR)
ISSN (e) 2959-8052 (p) 2959-8044

Volume 4, Issue 2, 2025

ORDINARY LEAST SQUARES (OLS) METHODS AND ISSUES: THEORETICAL FOUNDATIONS, CHALLENGES, AND METHODOLOGICAL REMEDIES

[1]Maryam, [2]Dr Misbah Nosheen, [3]Shehzad Ahmad, [4]Dr Sadia Naz

[1]MPhil scholar Department of Economics Hazara University Mansehra

[2]Chairperson Department of Economics Hazara University Mansehra

[3]PhD Scholar Electrical and Computer Engineering COMSATS University Islamabad

[4]Hazara University Mansehra

DOI: https://doi.org/

Keywords *(Ordinary Least Squares, Heteroscedasticity, Multicollinearity, Endogeneity, Autocorrelation, Econometrics, Instrumental Variables, Regression Diagnostics).*

Article History
Received on 1 May2025
Accepted on 21 May 2025
Published on 20 May 2025

Corresponding Author: * Misbah Nosheen

Abstract

Ordinary Least Squares (OLS) is a foundational estimation technique widely used in econometrics, statistics, and data analysis for examining linear relationships between variables. This paper reviews the theoretical underpinnings of OLS, its classical assumptions, and the conditions under which it provides Best Linear Unbiased Estimators (BLUE). It further explores the key practical challenges associated with real-world applications, such as heteroscedasticity, autocorrelation, multicollinearity, and endogeneity. Drawing upon empirical literature, the study evaluates the limitations of OLS and presents solutions including robust standard errors, generalized least squares, instrumental variable approaches, and regularization techniques. The integration of modern econometric tools with machine learning methodologies is also discussed as a promising direction for improving OLS-based inference in high-dimensional and complex data environments. The paper concludes that while OLS remains a central tool for empirical researchers, its effectiveness depends on rigorous diagnostics, adherence to assumptions, and methodological adaptability.

INTRODUCTION

Ordinary Least Squares (OLS) is a foundational estimation technique widely used in econometrics, statistics, and data analysis for examining linear relationships between variables. This paper reviews the theoretical underpinnings of OLS, its classical assumptions, and the conditions under which it provides Best Linear Unbiased Estimators (BLUE). It further explores the key practical challenges associated with real-world applications, such as heteroscedasticity, autocorrelation, multicollinearity, and endogeneity. Drawing

upon empirical literature, the study evaluates the limitations of OLS and presents solutions including robust standard errors, generalized least squares, instrumental variable approaches, and regularization techniques. The integration of modern econometric tools with machine learning methodologies is also discussed as a promising direction for improving OLS-based inference in high-dimensional and complex data environments. The paper concludes that while OLS remains a central tool for empirical researchers, its effectiveness depends on rigorous diagnostics, adherence to assumptions, and methodological adaptability.
Keywords: Ordinary Least Squares, Heteroscedasticity, Multicollinearity, Endogeneity, Autocorrelation, Econometrics, Instrumental Variables, Regression Diagnostics

1.Introduction

Ordinary Least Squares (OLS) is one of the most fundamental and widely used estimation techniques in statistics and econometrics. It provides a relatively simple yet powerful method to estimate relationships among variables by minimizing the sum of squared residuals—the differences between observed and predicted values of the dependent variable (Ludbrook, 2010). OLS has remained a cornerstone of empirical research across disciplines such as economics, finance, psychology, and health sciences, due to its ease of implementation, interpretability, and desirable statistical properties under classical assumptions.

The classical linear regression model under OLS assumes linearity in parameters, random sampling, no perfect multicollinearity, exogeneity of regressors, and homoscedasticity (Hansen, 2022). When these assumptions hold, the Gauss-Markov theorem assures that OLS estimators are the Best Linear Unbiased Estimators (BLUE). These characteristics make OLS attractive for tasks like estimating coefficients, conducting hypothesis tests, making predictions, and drawing causal inferences (Sokolov, Ray & Corman, 2021).

However, in practical applications, data often deviate from these ideal conditions. Violations of OLS assumptions can lead to inefficient, biased, or inconsistent estimates, thus undermining inference. For instance, heteroscedasticity, where the variance of the error term varies across observations, results in inefficient estimators and biased standard errors, which can distort statistical tests (White, 1980). Similarly, autocorrelation, especially common in time-series and panel data, violates the assumption of uncorrelated errors and inflates the risk of Type I errors (Sladekova & Field, 2024).

Another critical issue is multicollinearity, where independent variables are highly linearly correlated. This does not bias the OLS estimates but inflates their variance, leading to instability and difficulty in interpreting the individual effect of predictors (Chaudhary et al., 2022). Additionally, endogeneity, perhaps the most severe violation, occurs when regressors are correlated with the error term, resulting in biased and inconsistent estimates. This can arise from omitted variables, measurement errors, or simultaneity (Abdallah, Goergen & O'Sullivan, 2015; Khatib, 2025).

Recognizing these limitations, econometricians have developed a rich array of diagnostic tests and alternative estimation strategies. For instance, White's heteroscedasticity-consistent standard errors adjust inference without affecting coefficient estimates (White, 1980). For autocorrelation, techniques like the Newey-West standard errors offer robustness in time-series contexts where serial correlation is present (Kumar, 2023). In addressing multicollinearity, tools like Variance Inflation Factors (VIFs) are used for detection, while ridge regression and principal component analysis (PCA) are suggested as remedies, albeit with some trade-offs in interpretability (Ayinde, Alabi & Nwosu, 2021).

To tackle endogeneity, instrumental variables (IV) and two-stage least squares (2SLS) are commonly employed. However, the effectiveness of these techniques is heavily dependent on the validity and strength of the instruments used (Otse, Obunadike & Abubakar, 2025). When instruments are weak or invalid, the results can be as misleading as those from naive OLS estimation. More recently, innovations in quasi-experimental designs and machine learning-enhanced IV strategies have been proposed to strengthen identification in high-dimensional and complex data settings (Cinelli & Hazlett, 2025).

Another set of concerns arises from model specification errors, such as omitted variable bias, incorrect functional forms, or measurement errors. These issues can severely bias OLS estimates and lead to flawed conclusions (Bach, 2021). Tools such as the Ramsey RESET test help detect functional form misspecifications, while the Hausman test and Durbin-Wu-Hausman test are used to assess the presence of endogeneity.

Given the increasing complexity of data—larger samples, more variables, and dynamic relationships—OLS is no longer sufficient as a standalone technique in many modern research contexts. The integration of regularization techniques such as LASSO, machine learning algorithms for variable selection, and double machine learning frameworks for handling endogeneity and high-dimensionality, are becoming essential to modern applied econometrics (Abdallah et al., 2015; Bach, 2021).

Despite these challenges, OLS continues to be the most widely taught and used estimation technique in empirical research. Its popularity is not without justification: OLS provides a clear framework for estimation and hypothesis testing that is both intuitive and grounded in robust statistical theory. However, the responsibility falls on the researcher to ensure that the assumptions are not blindly taken for granted. Failing to test for heteroscedasticity, autocorrelation, multicollinearity, or endogeneity can result in misleading results, undermining the credibility of the analysis and any policy or theoretical implications drawn from it.

Furthermore, the application of OLS in sensitive areas such as public health, poverty alleviation, environmental economics, or gender equity must be particularly rigorous. Poor model specification or misinterpretation of regression results can lead to policies that are not only ineffective but potentially harmful. Therefore, a sound understanding of when and how OLS assumptions break down—and what can be done to address these violations—is essential for responsible empirical work.

This paper aims to provide a comprehensive analysis of the OLS method by:

1. Explaining its theoretical underpinnings,
2. Identifying major violations of assumptions in empirical practice,
3. Reviewing scholarly contributions that offer solutions, and
4. Discussing the implications of these issues for policy, theory, and further research.

In what follows, the literature review synthesizes contributions from both classical and recent empirical studies on OLS challenges. The discussion section then analyzes these findings and compares the effectiveness of proposed solutions across different empirical contexts. The conclusion reflects on the enduring relevance of OLS and offers recommendations for strengthening its application in modern statistical analysis.

2. Literature Review:

3. The literature surrounding Ordinary Least Squares (OLS) estimation is both foundational and expansive, reflecting its centrality in the history and practice of statistical modeling and econometrics. OLS was formally introduced by Legendre and Gauss in the early 19th century and has since become the most widely applied method for estimating the parameters of linear regression models. At its core, OLS aims to minimize the sum of squared residuals between the observed values and those predicted by the model. The theoretical justification for its use lies in the Gauss-Markov theorem, which establishes that, under a set of classical assumptions, the OLS estimators are the Best Linear Unbiased Estimators (BLUE) (Hansen, 2022). These assumptions include linearity in parameters, zero mean of error terms, homoscedasticity (constant variance of errors), no autocorrelation (independence of errors), and the absence of perfect multicollinearity among explanatory variables.

4. However, real-world data rarely conform perfectly to these assumptions. One of the most commonly violated conditions is homoscedasticity. White (1980) was the first to formally address this issue by proposing heteroscedasticity-consistent standard errors, which allow for valid inference even

when error variances differ across observations. His contribution significantly altered empirical practice by allowing researchers to retain coefficient estimates while adjusting standard errors to reflect the true variance structure. Sladekova and Field (2024) further expanded this line of research by recommending diagnostic tools such as quantile LOWESS plots and the Breusch-Pagan test to detect heteroscedasticity, particularly in cross-sectional datasets. They noted that while heteroscedasticity does not bias OLS coefficients, it makes the model inefficient and undermines statistical inference if not properly corrected.

5. Another major concern with OLS arises in the analysis of time-series or panel data, where the assumption of independent error terms is often violated due to autocorrelation. Autocorrelation inflates the likelihood of Type I errors and leads to inefficient estimators. Kumar (2023) emphasized the widespread use of the Durbin-Watson test for detecting first-order autocorrelation and highlighted Newey-West standard errors as a robust alternative for adjusting standard errors under arbitrary autocorrelation patterns. However, these adjustments assume a correctly specified error structure, which is not always attainable. Consequently, while Newey-West estimators offer more reliable inference, they also require careful model specification and diagnostics.

6. Multicollinearity, or the presence of high correlation between independent variables, is another critical issue in OLS regression. Although multicollinearity does not bias the estimates, it inflates their variances, making them highly sensitive to small changes in data and complicating interpretation. Chaudhary et al. (2022) outlined the use of Variance Inflation Factors (VIFs) to identify problematic variables and recommended remedies such as removing redundant predictors, combining variables, or applying dimensionality-reduction techniques like Principal Component Analysis (PCA) and Ridge Regression. Ayinde, Alabi, and Nwosu (2021) proposed a novel partitioning and extraction strategy that reduces the severity of multicollinearity while maintaining model interpretability. Despite the utility of these approaches, they often involve trade-offs between precision and transparency, highlighting the need for context-sensitive solutions.

7. Endogeneity is arguably the most serious violation of OLS assumptions, as it directly biases and invalidates coefficient estimates. Endogeneity occurs when an explanatory variable is correlated with the error term, typically due to omitted variables, measurement error, or simultaneity. Instrumental Variables (IV) and Two-Stage Least Squares (2SLS) are standard techniques used to address this issue. Otse, Obunadike, and Abubakar (2025) reviewed various applications of IV methods and emphasized the importance of strong and valid instruments, which are often difficult to find and validate in empirical research. Abdallah, Goergen, and O'Sullivan (2015) highlighted how failure to properly address endogeneity can lead to incorrect inferences and proposed a combined approach using econometrics and machine learning for better control of complex error structures.

8. Recent methodological developments have introduced new perspectives on dealing with endogeneity. Cinelli and Hazlett (2025) presented a sensitivity analysis framework for instrumental variable models, allowing researchers to quantify how robust their estimates are to potential omitted variable bias. Their framework aids in transparent reporting of assumptions and the limitations of causal claims. This is particularly useful in social sciences and public policy research, where unobserved heterogeneity is common.

9. Model specification errors, such as the omission of relevant variables or incorrect functional form, also pose a significant threat to the validity of OLS estimates. Bach (2021) emphasized the use of specification tests like the Ramsey RESET test to detect functional form misspecification and the Hausman test to compare the consistency of OLS with other estimators like Generalized Least Squares or IV. Sokolov et al. (2021) argued for combining statistical diagnostics with substantive theory to avoid overfitting and misinterpretation, especially in models with high-dimensional data or complex interactions.

10. Moreover, the intersection of econometrics and machine learning has introduced innovative approaches to improving the performance of OLS under non-ideal conditions. Abdallah et al. (2015) suggested that techniques such as LASSO and Elastic Net can be used for variable selection prior to running an OLS or IV regression, particularly in settings with many potential confounders. These methods have led to the emergence of double machine learning approaches, where algorithms estimate nuisance parameters while preserving the ability to conduct valid statistical inference on parameters of interest.

11. Altogether, the literature reveals that while OLS is a powerful and elegant estimation technique, its application in empirical research is fraught with challenges that require rigorous diagnostics and thoughtful methodological adjustments. From early theoretical proofs of unbiasedness to modern innovations integrating artificial intelligence, the evolution of OLS reflects the dynamic tension between simplicity and complexity in empirical modeling. The studies reviewed collectively underscore that responsible application of OLS depends on careful attention to underlying assumptions and the use of appropriate corrective techniques tailored to specific data characteristics.

3. Main issues in Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) estimation relies on a series of classical assumptions that, when violated, can lead to significant issues in estimation and inference. One of the most commonly encountered problems is heteroscedasticity, which occurs when the variance of the error term is not constant across observations. This violates one of the key Gauss-Markov assumptions and results in inefficient estimates and unreliable standard errors, making statistical tests invalid (White, 1980). While the coefficient estimates remain unbiased, heteroscedasticity affects the precision of these estimates and leads to misinterpretation. To address this, researchers often use heteroscedasticity-consistent standard errors, also known as White's robust standard errors, which adjust inference without altering the coefficients themselves (Kumar, 2023).

Another frequent issue, particularly in time-series and panel data, is autocorrelation or serial correlation. When the residuals are correlated across time, the OLS assumption of error independence is violated, leading to inefficient estimates and underestimated standard errors (Chaudhary et al., 2022). This causes overstatement of the statistical significance of coefficients. Autocorrelation is especially problematic in macroeconomic models, where shocks are likely to persist over time. Corrective techniques such as Newey-West standard errors or Cochrane-Orcutt iterative procedures are often applied to mitigate the impact of autocorrelation (Kumar, 2023).

Multicollinearity represents another substantial issue in OLS, especially in models involving many explanatory variables. It refers to a situation in which two or more independent variables are highly linearly correlated, making it difficult to isolate the individual effect of each predictor (Ayinde & Nwosu, 2021). While multicollinearity does not bias the OLS estimates, it inflates the standard errors, making it more likely that important variables will be deemed statistically insignificant. Tools such as Variance Inflation Factors (VIF) help detect multicollinearity, and remedies include removing or combining variables, or applying ridge regression (Otse et al., 2025).

The problem of endogeneity is perhaps the most serious issue undermining the credibility of OLS estimates. Endogeneity arises when one or more independent variables are correlated with the error term, leading to biased and inconsistent parameter estimates (Abdallah et al., 2015). This can result from omitted variables, simultaneity, or measurement error. For instance, if a variable that affects both the dependent and independent variables is omitted from the model, the included regressors may pick up its effect, distorting causal interpretations. Instrumental Variable (IV) estimation and Two-Stage Least Squares (2SLS) are widely used to address endogeneity, though these methods require strong assumptions about the validity and relevance of instruments (Cinelli & Hazlett, 2025).

Closely related to endogeneity is the omitted variable bias, which occurs when a relevant variable is left out of the regression model. This omission biases the estimates of included variables if the omitted factor is correlated with them and the outcome variable. The magnitude and direction of the bias depend on the relationship between the omitted and included variables (Bach, 2021). Researchers can apply specification tests such as the Ramsey RESET test to detect model misspecification and ensure a more comprehensive inclusion of relevant predictors.

Another often overlooked but critical problem is measurement error, particularly in the independent variables. When the values of predictors are measured with error, the classical OLS assumptions are violated, leading to attenuation bias—where estimated coefficients are biased toward zero (Abdallah et al., 2015). This is especially concerning in social science and development research where data collection challenges are frequent. The IV method can also help address measurement error when a valid instrument is available.

The sensitivity of OLS to outliers and influential data points is also well-documented. Since OLS minimizes the sum of squared residuals, large deviations exert disproportionate influence on the regression line. Outliers can skew results and may lead to misleading conclusions. Analysts often use diagnostics such as leverage values, Cook's distance, and DFBETAS to identify and mitigate the influence of such observations (Sokolov et al., 2021). In more robust approaches, alternative estimators like M-estimators or quantile regression may be considered.

OLS also suffers when non-linearity in the data structure is ignored. The linearity assumption implies a straight-line relationship between predictors and the outcome, but many real-world relationships are non-linear. In such cases, a linear model becomes misspecified, leading to biased estimates. Researchers can resolve this by applying transformations (e.g., log, square, interaction terms), or by using non-linear regression models and machine learning methods when appropriate (Bach, 2021).

Finally, small sample sizes and overfitting present challenges in many empirical studies. With too few observations relative to the number of predictors, OLS estimates become highly unstable, and overfitting becomes likely. This results in high variance and poor out-of-sample prediction. The adjusted R^2 statistic or cross-validation techniques are useful in addressing these concerns, and regularization methods like Lasso and Ridge regression are increasingly being integrated to penalize overly complex models (Cinelli & Hazlett, 2025).

In conclusion, while OLS is a foundational tool in regression analysis, its practical application often deviates from theoretical assumptions. Violations such as heteroscedasticity, autocorrelation, multicollinearity, endogeneity, and others can compromise the reliability of estimates. However, numerous diagnostic and corrective techniques have been developed to address these issues, ensuring that OLS continues to serve as a robust and flexible tool for empirical research when applied thoughtfully and with methodological rigor.

4.Discussion

The discussion of Ordinary Least Squares (OLS) methods revolves around their strengths, limitations, and the evolving techniques developed to address assumption violations. Despite its simplicity and interpretability, OLS requires a set of strict assumptions to provide valid results, making its application both foundational and vulnerable to misuse if diagnostic checks are ignored.

OLS continues to be widely applied because it provides Best Linear Unbiased Estimators (BLUE) under the Gauss-Markov assumptions: linearity, independence, homoscedasticity, and no multicollinearity (Hansen, 2022). However, in real-world data, these assumptions are frequently violated. For instance, heteroscedasticity—non-constant variance of residuals—is often encountered in cross-sectional data. This distorts standard errors and makes hypothesis tests unreliable. White (1980) proposed a heteroscedasticity-consistent covariance matrix estimator, allowing for robust inference without altering point estimates. Such innovations demonstrate the adaptability of OLS under relaxed assumptions.

Similarly, autocorrelation poses a significant challenge in time series and panel data models. When residuals are correlated over time, as often happens in economic or financial datasets, OLS estimators remain unbiased but become inefficient. Furthermore, the estimated standard errors are biased, which inflates Type I error probabilities. Techniques like Newey-West standard errors, which correct for autocorrelation and heteroscedasticity, offer a pragmatic solution (Kumar, 2023). However, their validity depends on proper lag length selection and underlying assumptions that may still be difficult to verify.

Multicollinearity, another frequent violation, inflates the variances of OLS estimates, making coefficient estimates highly sensitive to minor changes in data. This reduces the reliability of statistical inferences and complicates the interpretation of results. Tools like Variance Inflation Factors (VIF) help detect multicollinearity, while methods like ridge regression offer solutions by imposing a penalty on large coefficients (Ayinde & Nwosu, 2021). Though ridge regression sacrifices some interpretability, it stabilizes estimates, particularly in high-dimensional data environments.

Endogeneity is arguably the most damaging problem for causal inference using OLS. When regressors are correlated with the error term due to omitted variables, measurement errors, or simultaneity, OLS estimators become biased and inconsistent (Abdallah et al., 2015). Instrumental Variable (IV) estimation, including Two-Stage Least Squares (2SLS), is a traditional remedy. However, the credibility of IV estimates depends on the strength and validity of the instruments used. Weak instruments lead to biased estimates and large standard errors, making inference difficult (Cinelli & Hazlett, 2025). The identification of valid instruments remains one of the most contentious and complex tasks in empirical research.

The rise of quasi-experimental methods and machine learning offers new avenues for addressing OLS limitations. Techniques like regression discontinuity design, difference-in-differences, and synthetic control methods provide robust alternatives for causal inference under weaker assumptions than traditional OLS. Moreover, combining machine learning tools with econometric models, such as using LASSO for variable selection or double machine learning for treatment effect estimation, enhances the robustness and flexibility of analysis (Bach, 2021).

Model specification remains a foundational concern. Misspecification, including omitted variable bias and functional form errors, undermines the integrity of OLS results. Tests such as the Ramsey RESET test help diagnose such errors, but the correction requires theoretical guidance and domain knowledge (Sokolov et al., 2021). Additionally, measurement error in explanatory variables biases coefficients toward zero—a phenomenon known as attenuation bias. IV methods or structural modeling are often employed to mitigate such issues, although they come with their own assumptions and identification challenges.

OLS also demonstrates sensitivity to outliers and influential observations, which can disproportionately affect the regression line. Diagnostic tools like leverage plots, Cook's distance, and DFBETAS are instrumental in identifying such cases. When outliers dominate the estimation, robust regression techniques such as M-estimators or quantile regression may be more appropriate (Sladekova & Field, 2024).

A key advancement in addressing OLS shortcomings is the use of robust standard errors, including cluster-robust and bootstrap methods. These alternatives provide valid inference even when traditional assumptions about the error structure are violated. Cluster-robust errors are particularly useful in panel data where observations are grouped, and within-group correlations are likely (Khatib, 2025).

Moreover, the integration of Bayesian methods into linear regression modeling has gained attention. Bayesian regression incorporates prior beliefs and updates them with data, offering more flexible inference, especially in small samples or high-dimensional spaces. While not a replacement for OLS, Bayesian methods provide a complementary perspective that acknowledges uncertainty more explicitly (Bach, 2021).

The choice between OLS and alternative methods should be context-driven. For example, in well-controlled experiments with randomized treatment assignment, OLS can provide unbiased estimates of treatment effects even in small samples. Conversely, in observational data prone to confounding, methods correcting for endogeneity and omitted variable bias are necessary.

In conclusion, OLS remains an indispensable tool in empirical analysis, valued for its simplicity and theoretical elegance. However, real-world data rarely meet the classical assumptions underlying OLS, necessitating the use of diagnostic tests, robust standard errors, and alternative estimation techniques. Advances in econometrics and computational tools have enriched the analyst's toolkit, allowing for more robust, transparent, and credible empirical research. Nonetheless, the effectiveness of these methods hinges on the researcher's understanding of the data, the assumptions of each technique, and the empirical context in which they are applied. As such, the OLS framework must be applied with caution, awareness, and methodological rigor to ensure that its results remain both valid and informative.

5.Conclusion

Ordinary Least Squares (OLS) is the foundation of statistical and econometric analysis that is valued for its simplicity, ease of interpretation, and efficiency under classical assumptions. The current paper encapsulated the key ideas of OLS estimation and critically reviewed the main problems which can arise with real-world applications such as heteroscedasticity, autocorrelation, multicollinearity, and endogeneity.

While OLS estimators are unbiased and efficient under ideal conditions, it most often occurs with real data that there is a breakdown of these assumptions, potentially tainting the validity of the estimates. These have been addressed through methodological advancements like heteroscedasticity-consistent standard errors, generalized least squares, instrumental variable methods, and regularization techniques.

The empirical evidence emphasizes that no one-size-fits-all solution exists; instead, researchers should precisely diagnose the type of assumption violation and select suitable remedies according to the context and data features. In addition, the increasing use of machine learning algorithms together with standard econometric techniques has great potential to increase OLS-based inference robustness and reliability.

Lastly, OLS continues to be an anchor utility in the empirical analyst's arsenal, but effectiveness is a function of strict adherence to its assumptions and careful management of violations. Substantive research and methodological developments will be essential in expanding the power and domain of OLS to ever more complex data environments.

References

Abdallah, W., Goergen, M., & O'Sullivan, N. (2015). Endogeneity: How failure to correct for it can cause wrong inferences and some remedies. British Journal of Management, 26(4), 791–804. https://doi.org/10.1111/1467-8551.12113

Ayinde, K., Alabi, O. O., & Nwosu, U. I. (2021). Solving multicollinearity problem in linear regression model: The review suggests new idea of partitioning and extraction of the explanatory variables. Journal of Mathematics and Statistics Studies, 2(1), 12–20.

Bach, P. S. (2021). Applications in high-dimensional econometrics (Doctoral dissertation, Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky). https://doi.org/10.25592/uhhfdm.9383

Chaudhary, A. A., Nazir, N., Riaz, A., Sadiq, N., & Riaz, N. (2022). Autocorrelation: What happens if the error or disturbance terms are correlated in time-series data. Journal of Quantitative Methods, 6(2), 45–59.

Cinelli, C., & Hazlett, C. (2025). An omitted variable bias framework for sensitivity analysis of instrumental variables. Biometrika. Advance online publication. https://doi.org/10.1093/biomet/asaf004

Hansen, B. E. (2022). A modern Gauss–Markov theorem. Econometrica, 90(3), 1283–1294. https://doi.org/10.3982/ECTA18358

Khatib, S. F. (2025). An assessment of methods to deal with endogeneity in corporate governance and reporting research. Corporate Governance: The International Journal of Business in Society, 25(3), 606–630. https://doi.org/10.1108/CG-10-2023-0451

Kumar, N. K. (2023). Autocorrelation and heteroscedasticity in regression analysis. Journal of Business and Social Sciences, 5(1), 9–20.

Ludbrook, J. (2010). Linear regression analysis for comparing two measurers or methods of measurement: But which regression? Clinical and Experimental Pharmacology and Physiology, 37(7), 692–699. https://doi.org/10.1111/j.1440-1681.2010.05376.x

Otse, E. J., Obunadike, G. N., & Abubakar, A. (2025). Linear regression approach to solving multicollinearity and overfitting in predictive analysis. Journal of Science Research and Reviews, 2(1), 108–117.

Sladekova, M., & Field, A. P. (2024, June). Quantifying heteroscedasticity in linear models using quantile LOWESS intervals. Proceedings of the International Conference on Statistical Modeling, 112–125.

Abdallah, W., Goergen, M., & O'Sullivan, N. (2015). Endogeneity: How failure to correct for it can cause wrong inferences and some remedies. British Journal of Management, 26(4), 791-804.

Sokolov, D., Ray, N., & Corman, E. (2021). Least squares for programmers. arXiv preprint arXiv:2105.01295. https://arxiv.org/abs/2105.01295

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838. https://doi.org/10.2307/1912934