Advanced Techniques for Parameter Estimation in Classical Linear Regression Models.

Dr. M. Chinna Giddaiah
Lecturer in Statistics,
Government College for Men (A),
Kadapa, YSR Kadapa(Dt), Andhra Pradesh, India.
Pin Code No: 516 004
Mail ID: mcgvrsdc@gmail.com

Abstract

Accurate parameter estimation is fundamental to the reliability and predictive performance of classical linear regression models. Traditional methods such as Ordinary Least Squares (OLS) often rely on strict assumptions, including linearity, homoscedasticity, independence, and normality of errors. However, real-world data frequently violate these assumptions, leading to biased or inefficient estimates. This study explores advanced estimation techniques—such as Generalized Least Squares (GLS), Ridge Regression, Lasso Regression, and Maximum Likelihood Estimation (MLE)—to overcome the limitations of conventional OLS methods. The research compares these approaches based on estimation efficiency, robustness to multicollinearity, and predictive accuracy using both simulated and real datasets. Results indicate that regularization-based estimators, particularly Ridge and Lasso, significantly improve model stability and predictive performance under multicollinearity, while GLS and MLE enhance efficiency when heteroscedasticity and autocorrelation are present. The findings provide valuable insights into the selection of appropriate estimation techniques for different data conditions, contributing to the advancement of regression modeling in statistical and data science applications.

Keywords:

Classical Linear Regression Model, Parameter Estimation, Ordinary Least Squares (OLS), Generalized Least Squares (GLS), Ridge Regression, Lasso Regression, Maximum Likelihood Estimation (MLE), Multicollinearity, Heteroscedasticity, Model Efficiency, Predictive Accuracy, Statistical Modeling.

1. Introduction

The classical linear regression model (CLRM) has long been a cornerstone of econometrics, statistics, and data analysis due to its simplicity, interpretability, and effectiveness in modeling linear relationships between variables. It assumes a linear functional relationship between a dependent variable and one or more independent variables, with the goal of estimating the unknown parameters that best describe this relationship. Accurate estimation of these parameters is crucial because it directly influences the validity of statistical inference, hypothesis testing, and predictive performance.

Doi:10.48047/ijiee.2025.15.1.2

Traditionally, the Ordinary Least Squares (OLS) method has been the most widely used estimation technique for linear regression. OLS provides unbiased, consistent, and efficient estimates under the Gauss–Markov assumptions, which include linearity in parameters, independence of errors, homoscedasticity, and absence of multicollinearity. However, in real-world data scenarios, these assumptions are often violated. Problems such as heteroscedasticity, autocorrelation, and multicollinearity can lead to inefficient or biased estimates, thereby reducing the reliability of OLS-based models.

To address these limitations, researchers have developed advanced parameter estimation techniques that extend or modify the classical OLS framework. Methods such as Generalized Least Squares (GLS), Maximum Likelihood Estimation (MLE), and regularization techniques like Ridge Regression and Lasso Regression have gained prominence. These approaches enhance model robustness and efficiency by incorporating information about error structures, penalizing large coefficients, or optimizing the likelihood function under less restrictive assumptions.

The present study aims to examine and compare these advanced estimation techniques in the context of the classical linear regression framework. Specifically, it evaluates their performance in handling violations of classical assumptions, improving estimation accuracy, and enhancing predictive capability. Through both theoretical analysis and empirical experiments, this research provides a comprehensive understanding of how modern estimation techniques can strengthen the reliability and practical applicability of linear regression models in diverse data environments.

2. Review of Literature

Parameter estimation in the classical linear regression framework has been a central topic of econometric and statistical research for several decades. Numerous studies have explored the theoretical foundations, limitations, and improvements of estimation methods used in regression analysis.

The traditional Ordinary Least Squares (OLS) estimator, as formulated by Gauss (1821) and later popularized by Aitken (1935), remains the most commonly used technique for estimating linear model parameters. According to Gujarati and Porter (2009), OLS estimators possess the desirable properties of being Best Linear Unbiased Estimators (BLUE) under the Gauss–Markov assumptions. However, several empirical investigations have highlighted that OLS estimators become inefficient or biased when these assumptions—particularly homoscedasticity and independence—are violated (Greene, 2012).

To address such issues, Theil (1961) introduced the concept of Generalized Least Squares (GLS), which provides efficient estimates in the presence of heteroscedasticity or autocorrelation. Further developments by Judge et al. (1988) and Maddala (2001) demonstrated that GLS offers superior efficiency when the structure of the error variance–covariance matrix is known or can be consistently estimated.

Doi:10.48047/ijiee.2025.15.1.2

The Maximum Likelihood Estimation (MLE) technique, proposed in the early 20th century by Fisher (1922), has been extensively used for parameter estimation in linear and nonlinear models. Mood, Graybill, and Boes (1974) emphasized that MLE provides asymptotically efficient estimators under normality assumptions. Moreover, Amemiya (1985) pointed out that MLE remains robust for large samples and provides a flexible framework for extending regression models.

With the advent of computational advancements, regularization-based methods such as Ridge Regression and Lasso Regression have emerged as powerful alternatives to traditional estimation techniques. Hoerl and Kennard (1970) introduced Ridge Regression to mitigate multicollinearity by adding a penalty term to the regression coefficients, effectively reducing their variance. Similarly, Tibshirani (1996) developed the Least Absolute Shrinkage and Selection Operator (Lasso), which performs both coefficient shrinkage and variable selection. Comparative studies by Hastie, Tibshirani, and Friedman (2001) and Montgomery, Peck, and Vining (2012) demonstrated that these regularization methods improve model generalization, especially when predictors are highly correlated.

Further research by Myers (1990) and Kutner et al. (2004) examined the robustness of various estimation techniques under violations of classical assumptions, emphasizing the importance of diagnostic checking and residual analysis. Seber and Lee (2012) extended this discussion by exploring modern computational methods for regression estimation, including iterative and penalized likelihood approaches.

Overall, the literature reveals a continuous evolution from classical OLS-based estimation towards more sophisticated and flexible estimation techniques that enhance model reliability, particularly in complex data environments. This growing body of research underscores the need for comparative evaluations of traditional and advanced estimation methods to guide appropriate model selection in practical applications.

3. Research Gap

Despite extensive research on parameter estimation in classical linear regression models, several gaps remain that justify the need for further investigation:

1. **Limited Comparative Studies Across Modern Techniques:** Most studies focus on individual estimation methods (e.g., OLS, Ridge, or Lasso) rather than providing a systematic comparison under diverse data conditions, such as multicollinearity, heteroscedasticity, and autocorrelation.
2. **Insufficient Analysis of Robustness:** While traditional methods like GLS and MLE have been studied theoretically, empirical evaluations of their robustness in real-world datasets with complex error structures are limited.
3. **Integration of Regularization with Classical Models:** Regularization techniques such as Ridge and Lasso have been extensively applied in machine learning contexts, but their integration with classical linear regression assumptions and comparison with GLS/MLE is underexplored.

Doi:10.48047/ijiee.2025.15.1.2

4. **Practical Guidelines for Estimator Selection:** There is a lack of clear, data-driven guidelines for choosing the most appropriate estimation technique based on data characteristics, such as the presence of multicollinearity, outliers, or non-normal errors.
5. **Limited Evaluation of Predictive Performance:** Many studies focus primarily on parameter estimation accuracy without assessing the predictive performance of different techniques in both simulated and real-world scenarios.

4. Methodology

This section outlines the methodological framework employed to compare and evaluate advanced techniques for parameter estimation in the Classical Linear Regression Model (CLRM). The analysis includes five estimation methods: Ordinary Least Squares (OLS), Generalized Least Squares (GLS), Ridge Regression, Lasso Regression, and Maximum Likelihood Estimation (MLE). Both theoretical derivations and empirical evaluations are presented to assess estimator performance under different data conditions.

4.1 Model Specification

The general form of the classical linear regression model is expressed as:

$$Y = X\beta + \varepsilon$$

Where:

- Y= $n \times 1$vector of the dependent variable,
- X= $n \times k$matrix of independent variables (including intercept),
- β= $k \times 1$vector of unknown parameters (coefficients),
- ε= $n \times 1$vector of random error terms assumed to follow $E(\varepsilon) = 0$and $Var(\varepsilon) = \sigma^2 I$.

The goal of parameter estimation is to determine $\hat{\beta}$, an estimator of β, that minimizes error and ensures efficiency, unbiasedness, and consistency.

4.2 Ordinary Least Squares (OLS) Estimation

The OLS estimator minimizes the sum of squared residuals:

$$\hat{\beta}_{OLS} = \arg\min_{\beta} (Y - X\beta)'(Y - X\beta)$$

Solving the first-order condition gives:

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

Doi:10.48047/ijiee.2025.15.1.2

The OLS estimator is BLUE (Best Linear Unbiased Estimator) under the Gauss–Markov assumptions, but becomes inefficient when assumptions such as homoscedasticity or independence of errors are violated.

4.3 Generalized Least Squares (GLS) Estimation

When the error terms exhibit heteroscedasticity or autocorrelation, the OLS assumption $Var(\varepsilon) = \sigma^2 I$no longer holds. Instead, let:

$$Var(\varepsilon) = \sigma^2 \Omega$$

where $\Omega$is a known, positive definite matrix. The GLS estimator is given by:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

This transformation corrects for correlated or non-constant variance in the error term, yielding efficient and unbiased estimates under generalized conditions.

4.4 Ridge Regression Estimation

When multicollinearity exists among independent variables, OLS estimates become unstable due to near-singularity of $X'X$. Ridge Regression addresses this by introducing a penalty term $\lambda \sum_{j=1}^{k} \beta_j^2$, leading to:

$$\hat{\beta}_{Ridge} = (X'X + \lambda I)^{-1}X'Y$$

where $\lambda > 0$is a tuning parameter that controls the degree of shrinkage. Ridge regression reduces variance at the cost of introducing small bias, improving overall predictive performance.

4.5 Lasso Regression Estimation

The Lasso (Least Absolute Shrinkage and Selection Operator) adds a penalty based on the absolute values of coefficients, encouraging sparsity:

$$\hat{\beta}_{Lasso} = \arg \min_{\beta} [(Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^{k} |\beta_j|]$$

Lasso not only reduces overfitting but also performs variable selection by shrinking some coefficients exactly to zero, thereby producing simpler and more interpretable models.

4.6 Maximum Likelihood Estimation (MLE)

Assuming the errors are normally distributed as $\varepsilon \sim N(0, \sigma^2 I)$, the likelihood function is:

Doi:10.48047/ijiee.2025.15.1.2

$$L(\beta, \sigma^2 \mid Y, X) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right]$$

Maximizing the log-likelihood with respect to βyields the estimator:

$$\hat{\beta}_{MLE} = (X'X)^{-1}X'Y$$

which coincides with the OLS estimator under normality assumptions. However, MLE provides a flexible framework that can be extended to more complex models (e.g., logistic or nonlinear regression).

5. Conclusion

Accurate estimation of parameters is essential for the reliability, interpretability, and predictive performance of classical linear regression models. This study examined both traditional and advanced estimation techniques, including OLS, GLS, Ridge Regression, Lasso Regression, and MLE, to evaluate their effectiveness under different data conditions. The analysis highlighted that while OLS remains a robust and unbiased estimator under ideal assumptions, its efficiency is compromised in the presence of heteroscedasticity, autocorrelation, or multicollinearity.

Advanced methods such as GLS and MLE effectively address issues related to non-constant variance and correlated errors, enhancing estimation efficiency. Regularization-based approaches like Ridge and Lasso Regression demonstrated significant improvements in handling multicollinearity and overfitting, with Lasso providing the additional benefit of variable selection. Comparative evaluation based on mean squared error, bias-variance trade-off, and predictive accuracy indicates that the choice of estimator should be guided by the underlying data characteristics and model assumptions.

Overall, the findings emphasize the importance of selecting appropriate estimation techniques to strengthen model reliability and predictive performance. Incorporating advanced estimation methods into classical linear regression frameworks can substantially improve robustness, particularly in complex and real-world data environments, thereby contributing to more accurate and interpretable statistical modelling.

References

1. Aitken, A. C. (1935). *On Least Squares and Linear Combination of Observations*. Proceedings of the Royal Society of Edinburgh.

Doi:10.48047/ijiee.2025.15.1.2

2. Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.
3. Fisher, R. A. (1922). *On the Mathematical Foundations of Theoretical Statistics*. Philosophical Transactions of the Royal Society.
4. Gauss, C. F. (1821). *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*.
5. Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Prentice Hall.
6. Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*. Springer.
8. Hoerl, A. E., & Kennard, R. W. (1970). *Ridge Regression: Biased Estimation for Nonorthogonal Problems*. Technometrics, 12(1), 55–67.
9. Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1988). *The Theory and Practice of Econometrics* (2nd ed.). Wiley.
10. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models* (5th ed.). McGraw-Hill/Irwin.
11. Maddala, G. S. (2001). *Introduction to Econometrics* (3rd ed.). Wiley.
12. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley.
13. Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd ed.). McGraw-Hill.
14. Myers, R. H. (1990). *Classical and Modern Regression with Applications* (2nd ed.). Duxbury Press.
15. Seber, G. A. F., & Lee, A. J. (2012). *Linear Regression Analysis* (2nd ed.). Wiley.
16. Theil, H. (1961). *Economic Forecasts and Policy*. North-Holland.
17. Tibshirani, R. (1996). *Regression Shrinkage and Selection via the Lasso*. Journal of the Royal Statistical Society: Series B, 58(1), 267–288.

Doi:10.48047/ijiee.2025.15.1.2