Proceedings of CONF-MPCS 2025 Symposium: Leveraging EVs and Machine Learning for Sustainable Energy Demand Management
DOI: 10.54254/2753-8818/2025.GL24049

Mathematical Foundations and Comparative Analysis of the Least Squares Method

Yanxi Li

Beijing 101 High School, Beijing, China
li_yx_1118@outlook.com

Abstract: The least squares method is a fundamental statistical approach used to determine the best-fitting function for a given set of data by minimizing the sum of squared residuals. Originating from the work of Carl Friedrich Gauss and Adrien-Marie Legendre in the early 19th century, it has since become a cornerstone in regression analysis, data fitting, and various scientific applications. This paper aims to explore the mathematical principles underlying the least squares method, including its derivation, geometric interpretation, and theoretical comparisons between Ordinary Least Squares and Weighted Least Squares. We compare Ordinary Least Squares with Weighted Least Squares, highlighting their respective assumptions and use cases. Our findings demonstrate that while Ordinary Least Squares provide unbiased and efficient estimations under homoskedasticity, Weighted Least Squares are preferable when dealing with heteroskedastic data. By presenting a comprehensive analysis of the least squares method, we offer insights into its theoretical foundations, providing valuable guidance for researchers and practitioners in data analysis.

Keywords: Least Squares Method, Regression Analysis, Ordinary Least Squares, Weighted Least Squares

1. Introduction

The least squares method is a foundational mathematical technique in data analysis and regression modeling. Initially proposed by Legendre in 1805 and later refined by Gauss, it has become indispensable across scientific and engineering disciplines [1]. The core idea is to minimize the sum of squared differences between observed and predicted values, thereby yielding the most accurate parameter estimates in a given model.

This paper explores the mathematical foundations, derivation, and geometric interpretation of the least squares method. It addresses two key questions: How does the method provide optimal parameter estimation in linear regression models? What are the differences between Ordinary Least Squares and Weighted Least Squares, and under what conditions should each method be applied?

A deep understanding of the least squares method is critical, as regression analysis serves as a powerful tool for predicting trends, analyzing relationships, and making informed decisions across diverse fields, including economics, finance, and machine learning. By analyzing the mathematical principles behind the least squares method, this paper aims to provide a deeper comprehension of its theoretical significance and practical utility.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.GL24049

2. Mathematical principles of the Least Squares method

2.1. Basic concept

When analyzing data, two variables often exhibit a certain degree of association, and it is crucial to describe their relationship mathematically. For instance, consider a study examining the relationship between children's height and their sleep duration. By collecting data on a group of children, recording their average sleep hours per night and their corresponding heights, a scatter plot can be created to visualize the distribution. Although the data points may not perfectly align, an underlying trend can often be observed. Instead of relying on intuition to estimate this relationship, a mathematical approach is required to determine the most suitable function that best represents the trend in the data. The challenge lies in identifying a function that effectively captures the relationship between the two variables, enabling predictions and deeper insights into how one variable influences the other.

To address this challenge, the least squares method is introduced. This is a fundamental statistical technique used to determine the best-fitting function by minimizing the overall discrepancy between observed and predicted values. By systematically reducing the total squared differences between actual data points and their estimated values, the least squares method ensures that the resulting model provides an optimal representation of the data. This approach is particularly essential in regression analysis, where it allows for accurate parameter estimation and reliable inference. Due to its mathematical rigor and efficiency, the least squares method has become a cornerstone in data modeling, with widespread applications in economics, engineering, and various scientific disciplines [2].

2.2. Formula derivation

In the least squares method, we estimate the parameters pf a simple linear regression model by minimizing the sum of squared residuals (SSR) [3].

We define the residuals as:

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i) \tag{1}$$

Where y_i the observed value, x_i is the independent variable, $\beta_0$is the intercept, $\beta_1$is the slope, $\varepsilon_i$is the residual error.

Given the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{2}$$

that best fits the given data in the sense of minimizing the sum of squared residuals:

$$SSR = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{3}$$

To find the optimal values of β_0 and β_1, we take the partial derivatives of SSR with respect to β_0 and β_1 and set them to zero:

$$\frac{\partial SSR}{\partial \beta_0} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-1) = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{4}$$

$$\frac{\partial SSR}{\partial \beta_1} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-x_i) = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) = 0 \tag{5}$$

Slove β_0:

$$\sum_{i=1}^{n} y_i = n\,\beta_0 + \beta_1 \sum_{i-1}^{n} x_i$$

DOI: 10.54254/2753-8818/2025.GL24049

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \tag{6}$$

Slove β_1:

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - (\bar{y} - \beta_1 \bar{x}) \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \beta_1 \bar{x} \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i = \beta_1 (\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i)$$

$$\beta_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{7}$$

Where $\bar{x}$ and $\bar{y}$ are the means of x_i and y_i, respectively.
Thus, we find the least squares estimates for β_0 and β_1.

3. Geometric interpretation of the Least Squares method

3.1. Visual representation of line fitting

In a two-dimensional linear regression problem, the goal of the Least Squares method is to find a line that minimizes the sum of squared vertical distances from all data points to the line (Figure 1). From a geometric perspective, this process is equivalent to projecting the data points onto the fitted line and minimizing the squared deviations of these projection [4].

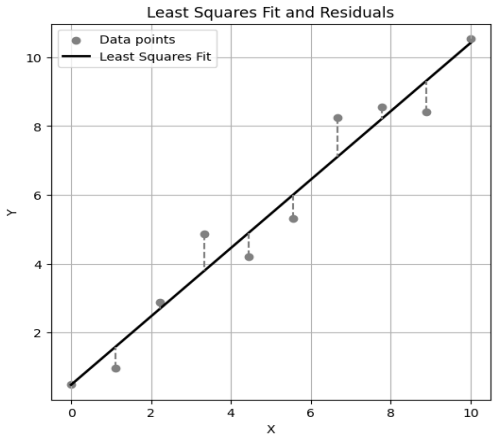

Figure 1: Vertical distances from data points to the fitted line

By minimizing the sum of squared vertical residuals, the fitted line provides the best possible approximation of the data trend. This ensures that the fitted line represents the overall pattern of the data with minimal total error.

3.2. Vector space perspective

The least square method can also be understood as a projection problem in vector spaces—especially, how to find the vector in the column space of a matrix that is closest to the observed data.

DOI: 10.54254/2753-8818/2025.GL24049

Given a linear system [5]:

$$Ax = b \tag{8}$$

where A is an $m \times n$ matrix, x is the $n \times 1$ unknow vector, and b is the $m \times 1$ observation vector. Since the system is overdetermined, an exact solution $Ax = b$ is usually impossible. Instead, we seek an x that minimizes:

$$\hat{x} = \arg\min_{x} \|Ax - b\|^2 \tag{9}$$

The columns of A span a subspace known as the column space of A. Since b may not lie in this space, we aim to find the projection of b onto it, denoted as Ax, to minimize the residual $r = b - Ax$.

The projection condition requires r to be orthogonal to the column space of A:

$$A^T(b - Ax) = 0$$

$$A^TA\hat{x} = A^Tb \tag{10}$$

This is the normal equation. If A^TA is invertible:

$$\hat{x} = (A^TA)^{-1}A^Tb \tag{11}$$

This provides the best x minimizing $\|Ax - b\|^2$.

4. Ordinary Least Squares vs. Weighted Least Squares

4.1. Basic assumptions of OLS

Ordinary Least Squares (OLS) is one of the most fundamental linear regression techniques. Its primary goal is to find the best-fitting line for a dataset by minimizing the sum of squared residuals between observed and predicted values.

$$\sum_{i=1}^{n}(y_i - X_i\beta)^2 \tag{12}$$

To ensure that OLS estimates are optimal—meaning they are unbiased, efficient, and have the minimum variance—the method relies on several key assumptions [6, 7]:

Linearity: OLS assumes that there is a linear relationship between the independent variables (X) and the dependent variable (Y), meaning the data can be adequately described by a linear model. If the true relationship is nonlinear and OLS is still used, the model may suffer from systematic bias, leading to inaccurate predictions.

Constant Variance of Errors: OLS requires that the error terms (ε) have constant variance across all observations. This means that the magnitude of errors should not change as the independent variable varies. If the error variance changes across observations, a condition known as heteroscedasticity, OLS estimates remain unbiased but are no longer optimal, as the standard errors may be miscalculated, affecting statistical inference such as confidence intervals and hypothesis testing.

Independence of Errors: The error terms (ε) should be independent of each other, meaning the error in one observation should not influence the error in another. This assumption is particularly important in time series data, where errors often exhibit autocorrelation—a situation where past errors are correlated with future ones. If errors are correlated, OLS estimates remain unbiased but may underestimate standard errors, leading to misleading hypothesis test results.

Normality of Errors: OLS assumes that the error terms follow a normal distribution, especially when conducting hypothesis tests such as t-tests or F-tests. While the Central Limit Theorem (CLT)

DOI: 10.54254/2753-8818/2025.GL24049

ensures that OLS estimates approximate normality in large samples, normality remains important in small-sample settings. If the error terms deviate significantly from normality, transformations or robust regression techniques may be needed to improve estimation reliability.

4.2. Weighted Least Squares

4.2.1. The foundation of WLS

Weighted Least Squares (WLS) extends OLS to account for heteroskedasticity, where the variance of errors is not constant across observations [8]. OLS assumes that all observations have equal error variance (homoskedasticity), but in real-world applications, this assumption is often violated, so using OLS directly can lead to inefficient parameter estimates and unreliable statistical inferences [9].

WLS resolves this issue by assigning each observation a weight that reflects its confidence level or importance in the regression. The fundamental idea is to minimize the weighted sum of squared residuals:

$$\sum_{i=1}^{n} w_i \, (y_i - X_i \beta)^2 \tag{13}$$

where w_i is the weight assigned to observation i.

By solving the weighted least squares estimation, the regression coefficients:

$$\hat{x} = (A^T W A)^{-1} A^T W b \tag{14}$$

Compared to OLS, WLS adjusts for heteroskedasticity, providing more efficient estimates when error variance is not constant.

4.2.2. Weighting methods in WLS

In WLS, choosing appropriate weights is crucial for addressing the problem of heteroscedasticity. The core idea is to assign greater weight to more reliable data points while reducing the influence of less reliable ones, thereby improving the stability and interpretability of the regression model.

If the error variance σ_i^2 is known, the optimal weight is given by:

$$w_i = \frac{1}{\sigma_i^2} \tag{15}$$

In this case, observations with smaller errors contribute more, while those with larger errors have less impact, making the model more robust.

If the error variance varies with the independent variable, one can assume:

$$w_i = \frac{1}{X_i^2} \text{ or } \sigma_i^2 = X_i^{\gamma} \tag{16}$$

to estimate the relationship between variance and the predictor variable. If the error variance is unknown, it can be estimated using OLS residuals:

$$e_i = y_i - X_i \hat{\beta}, \quad w_i = \frac{1}{e_i^2} \tag{17}$$

This reduces the influence of observations with high error, ensuring that the model maintains a good fit even in the presence of heteroscedasticity.

DOI: 10.54254/2753-8818/2025.GL24049

4.3. Comparison of OLS and WLS

OLS is widely used due to its simplicity and efficiency. When the assumptions of linearity, constant variance (homoscedasticity), and uncorrelated errors hold, OLS provides unbiased and minimum variance estimates. It performs well in most cases and is easy to interpret. [10] However, if the error variance is not constant, OLS estimates become inefficient, meaning they no longer have the smallest possible variance, and hypothesis testing results may be misleading.

WLS improves upon OLS when heteroscedasticity is present by assigning different weights to observations. This adjustment ensures that data points with more reliable measurements (lower variance) have greater influence on the estimated parameters, leading to more precise estimates. However, WLS requires knowing or estimating the error variance structure, which is not always straightforward. If the estimated weights are inaccurate, WLS can introduce new biases rather than solving the problem. Additionally, WLS is computationally more complex than OLS, making it less practical for small datasets with minimal heteroscedasticity.

In practice, analysts often begin with OLS and then test for heteroscedasticity using statistical methods such as the Breusch-Pagan test [11]. If significant heteroscedasticity is detected, WLS or other robust regression techniques may be applied to improve estimation accuracy.

5. Conclusion

This paper have explored the theoretical foundations and applications of the least squares method, emphasizing its role in regression analysis and statistical modeling. We began by deriving the mathematical formulation of the least squares method, demonstrating how it minimizes the sum of squared residuals to achieve optimal parameter estimation. Additionally, we provided a geometric interpretation, illustrating the method's significance in data fitting.

A key focus of our discussion was the comparison between OLS and WLS. While OLS remains a widely used and effective method under the assumption of homoskedasticity, it becomes inefficient when dealing with heteroskedastic data. WLS, by assigning different weights to observations, addresses this limitation and improves estimation accuracy when error variances are not constant. However, WLS requires prior knowledge or estimation of variance structures, which can introduce complexities and potential biases.

Despite its effectiveness, the least squares method has certain limitations. For example, it assumes that the underlying relationship between variables is linear, which may not always be the case. Future research could focus on extending the least squares method to non-linear models, improving robustness against outliers, and developing advanced weighting techniques for WLS to enhance its applicability.

Overall, the least squares method remains an indispensable tool in statistical analysis, with extensive applications in economics, engineering, and social sciences. By providing a clear understanding of its principles and methodologies, this study contributes to the broader field of data modeling, offering valuable insights for both theoretical research and practical implementation.

References

[1] *Stigler, S. M. (1981). Gauss and the Invention of Least Squares. The Annals of Statistics, 9(3), 465–474.*
[2] *Wooldridge, J. M. (2019). Introductory Econometrics: A Modern Approach (7th ed.). Cengage Learning.*
[3] *Soch, J. (2021, October 27). Ordinary least squares for simple linear regression. Retrieved from https://statproofbook.github.io/P/slr-ols.html*
[4] *Seber, G. A. F., & Lee, A. J. (2012). Linear Regression Analysis (2nd ed.). Wiley.*
[5] *Margalit, D. Rabinoff, J. The method of least squares. Retrieved from https://textbooks.math.gatech.edu/ila/least-squares.html*

DOI: 10.54254/2753-8818/2025.GL24049

[6] *Burton, A. L. (2021). OLS (Linear) Regression. The Encyclopedia of Research Methods in Criminology and Criminal Justice, 509–514. doi:10.1002/9781119111931.ch10*

[7] *Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. Journal of the American Statistical Association, 57(298), 369–375.*

[8] *White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838.*

[9] *Oldenburg, R. (2025). Geometric Weighted Least Squares Estimation. Statistics, Optimization and Information Computing, 13, 611-615.*

[10] *Davidson, R., & MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. Oxford University Press.*

[11] *Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47(5), 1287–1294.*