

LEAST SQUARES CRITERION?

Authored by
mohammad looti

October 31, 2025

RECOMMENDED CITATION

mohammad looti (2025). *LEAST SQUARES CRITERION?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=63895>

LEAST SQUARES CRITERION

Primary Disciplinary Field(s): Statistics, Econometrics, Applied Mathematics, Data Science

1. Core Definition

The **Least Squares Criterion**, often simply referred to as Least Squares Estimation, is a foundational optimization principle used across numerous quantitative disciplines, particularly in statistical modeling and regression analysis. Its primary function is to provide a method for estimating the unknown parameters in a linear model by minimizing the discrepancy between the observed data and the values predicted by the model. This discrepancy, known as the error or residual, is quantified by squaring it; thus, the criterion seeks to minimize the **Sum of Squared Residuals (SSR)**.

In practical terms, the criterion posits that among all possible lines or curves that could fit a set of observed data points, the 'best' fit is the one that results in the smallest possible value for the sum of the vertical distances (residuals) between each data point and the fitted line, where these distances are first squared. The act of squaring the errors serves two crucial purposes: first, it ensures that all calculated errors are positive, meaning positive and negative residuals do not cancel each other out, thereby providing an accurate measure of overall model inaccuracy; second, it heavily penalizes large errors. By disproportionately weighting larger deviations, the Least Squares Criterion forces the estimated model parameters to be highly sensitive to outliers and to focus on achieving a fit that minimizes significant predictive failures, resulting in a single, unique line of best fit known as the Regression Line.

The initial source content accurately highlights the underlying assumption driving this criterion: "Predictions from models are not always correct." Since models are inherently approximations of reality and rarely perfectly predict outcomes, the Least Squares Criterion provides a systematic, mathematically tractable mechanism for calculating the most plausible parameter values given the inherent inaccuracies. It stands in contrast to other fitting methods, such as the Least Absolute Deviations (LAD) method, which minimizes the sum of the absolute values of the residuals. The mathematical elegance and resulting desirable statistical properties of the squared error formulation have cemented its role as the dominant method for fitting linear models, especially the technique known as **Ordinary Least Squares (OLS)**.

2. Mathematical Foundation and Objective Function

The mathematical foundation of the Least Squares Criterion involves defining an objective function, which is the quantity that the model seeks to minimize. For a standard linear regression model where the dependent variable Y_i is related to k independent variables $X_{\{i,j\}}$ and a set of

unknown parameters β_j , the error term (residual) for the i -th observation is defined as the difference between the actual observed value and the predicted value (\hat{Y}_i). Mathematically, the residual $e_i = Y_i - \hat{Y}_i$. The objective of the Least Squares Criterion is to select parameter estimates ($\hat{\beta}$) that minimize the total error $S(\hat{\beta})$, where $S(\hat{\beta})$ is the sum of the squares of these residuals:

$$S(\hat{\beta}) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

To find the parameters ($\hat{\beta}$) that achieve this minimum, the method employs calculus. Specifically, the partial derivative of the objective function $S(\hat{\beta})$ with respect to each unknown parameter β_j is calculated and set equal to zero. This process yields a system of linear equations known as the **Normal Equations**. Solving the Normal Equations simultaneously provides the unique set of parameter estimates that satisfy the criterion, ensuring that the estimated regression line is positioned precisely where the squared prediction error across all data points is at its lowest possible value. This reliance on differentiable mathematics is a key reason for the computational popularity and relative ease of implementation of the Least Squares method compared to non-differentiable minimization techniques.

The simplicity and computational efficiency derived from using the squared error penalty are critical. Because the objective function is quadratic, it is convex, guaranteeing that the solution found by setting the derivatives to zero represents a global minimum, rather than merely a local minimum. This mathematical guarantee of finding the most efficient fit under specified assumptions provides immense confidence in the resulting parameter estimates. Furthermore, the estimated parameters $\hat{\beta}$ are linear functions of the observed Y values, simplifying their statistical properties and allowing for straightforward calculation of confidence intervals and hypothesis tests, forming the backbone of standard statistical inference.

3. Etymology and Historical Development

The development of the Least Squares Criterion marks a pivotal moment in the history of statistics, shifting statistical inquiry from descriptive enumeration toward predictive modeling based on probabilistic error minimization. The principle was independently discovered and published by two of the most celebrated mathematicians of the era: Carl Friedrich Gauss and Adrien-Marie Legendre. It is widely acknowledged that Gauss developed the method earlier, utilizing it in astronomical calculations as early as 1795 to predict the orbit of the asteroid Ceres. However, he did not publish his findings until 1809 in his work, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* (Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections).

The first published description, and the coining of the term "**Méthode des moindres carrés**" (Method of Least Squares), belongs to the French mathematician Adrien-Marie Legendre, who

introduced the method in 1805 in his treatise *Nouvelles méthodes pour la détermination des orbites des comètes* (New Methods for the Determination of the Orbits of Comets). Legendre provided a clear, accessible formulation of the criterion as a robust method for parameter estimation in observational sciences, particularly geodesy and astronomy, where measurement errors are unavoidable and systematic methods for error mitigation are essential. This method provided the first rigorous, general-purpose approach to dealing with overdetermined systems--systems where there are more observations than unknown parameters, making an exact solution impossible.

The subsequent history saw a period of debate regarding priority, but Gauss solidified the method's theoretical underpinning by proving, in 1823, that the Least Squares estimators are the most efficient among all linear, unbiased estimators when the errors are normally distributed--a concept later formalized into the **Gauss-Markov Theorem**. The criterion thus transitioned from a practical tool for astronomers to a mathematically proven, optimal statistical inference technique. Its migration from physical sciences into social sciences, economics (where it became foundational for econometrics), and ultimately, modern data science, established it as one of the most successful algorithms in statistical history.

4. Key Characteristics: Assumptions of Ordinary Least Squares (OLS)

While the Least Squares Criterion itself is a minimization objective, its practical implementation, Ordinary Least Squares (OLS), relies on a specific set of assumptions to ensure that the resulting estimators possess desirable statistical properties--namely, unbiasedness, consistency, and efficiency. These assumptions, often referred to as the **Gauss-Markov Assumptions**, dictate the necessary conditions under which the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

The critical OLS assumptions include: 1) **Linearity in Parameters**, meaning the model must be linear in the coefficients, although the independent variables themselves can be non-linear transformations; 2) **Random Sampling**, ensuring the data used is a representative sample from the population; 3) **Zero Conditional Mean of Errors (Exogeneity)**, which is arguably the most crucial assumption, stating that the error term must be uncorrelated with the independent variables (i.e., $E(u|X)=0$), preventing issues like omitted variable bias or endogeneity; and 4) **No Perfect Multicollinearity**, meaning none of the independent variables are perfectly correlated with one another, ensuring the design matrix is full rank and the parameters can be uniquely estimated.

Two further critical assumptions govern the structure of the residuals themselves: 5) **Homoscedasticity**, which requires that the variance of the error term is constant across all levels of the independent variables ($Var(u|X) = \sigma^2$). Violation of this assumption (heteroscedasticity) leads to inefficient estimates, though they remain unbiased. 6) **No**

Autocorrelation (or serial correlation), requiring that the errors across different observations are independent of each other ($\text{Cov}(u_i, u_j) = 0$ for $i \neq j$). This assumption is particularly relevant in time-series data, where dependency between consecutive errors is common. If these six assumptions hold, the Least Squares estimators are proven to be the most efficient available among the class of linear, unbiased estimators.

5. Applications Across Disciplines

Due to its robustness and proven statistical properties, the Least Squares Criterion forms the foundation for empirical analysis across virtually every quantitative field. In **Econometrics and Finance**, OLS regression is the default tool for modeling relationships between macroeconomic variables (e.g., inflation and unemployment), forecasting stock returns, estimating demand elasticities, and evaluating policy impacts. Its ability to disentangle the effect of one variable while controlling for others is indispensable for causal inference in non-experimental settings.

In **Psychology and Sociology**, the criterion is used extensively in multivariate analysis, such as predicting behavioral outcomes, determining the influence of environmental factors on cognitive development, and constructing complex structural equation models. The original context of the source content, rooted in psychology, demonstrates its use for estimating psychological parameters within complex models where predictive errors are expected. Furthermore, in **Engineering and Physics**, Least Squares methods are vital for curve fitting, signal processing, control theory, and calibration of sensors, where minimizing the cumulative deviation from the theoretical expectation is paramount for system reliability.

Most recently, the criterion has become foundational to **Machine Learning and Data Science**. Simple linear regression, built directly on the Least Squares principle, remains a highly interpretable benchmark model. More complex techniques, such as Ridge Regression and Lasso Regression, are extensions of the Least Squares objective function, incorporating penalty terms to manage overfitting and multicollinearity. These modern applications confirm that the core principle of minimizing squared prediction errors remains central to building predictive models, despite the evolution of computational methods.

6. Advantages and Significance

The primary significance of the Least Squares Criterion stems from the statistical properties it bestows upon its estimators, formalized by the **Gauss-Markov Theorem**. This theorem asserts that under the standard OLS assumptions, the OLS estimator is the **Best Linear Unbiased Estimator (BLUE)**. 'Best' implies that among all linear and unbiased estimators, the OLS estimator has the minimum variance, making it the most efficient estimator possible under these ideal conditions. This guarantee of statistical optimality is the chief reason for the dominance of OLS in

empirical research.

A second major advantage is the computational tractability of the method. The analytic solution derived from the Normal Equations requires only matrix algebra and avoids complex iterative optimization schemes required by many other statistical methods (such as Maximum Likelihood Estimation in certain contexts). This makes the OLS estimates straightforward to calculate, replicate, and interpret. The resulting regression coefficients ($\hat{\beta}$) have clear interpretations: they represent the estimated change in the dependent variable for a one-unit change in the independent variable, holding all other variables constant.

Finally, the output of the Least Squares procedure provides a rich set of diagnostic tools necessary for inference. Standard errors, t -statistics, F -statistics, and the **Coefficient of Determination (R^2)**--which quantifies the proportion of the variance in the dependent variable explained by the model--are direct byproducts of the squared error minimization process. These diagnostics allow researchers not only to estimate parameters but also to rigorously test hypotheses about the relationships between variables and assess the overall fit and predictive power of the model.

7. Debates and Criticisms

Despite its widespread adoption and mathematical elegance, the Least Squares Criterion is subject to several significant criticisms and limitations, primarily arising from its strict reliance on the Gauss-Markov assumptions and its sensitivity to data structure. The most cited critique concerns the penalty mechanism: because the criterion squares the residuals, the resulting estimation is highly susceptible to **outliers**.

An observation far removed from the general trend (an outlier) generates a massive squared residual, disproportionately pulling the fitted regression line toward it to minimize this large penalty. This lack of resistance means that a few extreme data points can drastically alter the parameter estimates, making the resulting model a poor representation of the central tendency of the majority of the data. This has driven the development of **Robust Regression** techniques, such as Least Absolute Deviations (LAD or L1 regression), which minimize the absolute value of the residuals, providing greater resistance to outliers by imposing a less severe penalty.

Furthermore, violations of the core assumptions severely compromise the desirable properties of the OLS estimator. If **Heteroscedasticity** (non-constant variance of errors) or **Autocorrelation** (dependent errors) is present, the OLS estimates remain unbiased, but the standard errors calculated are incorrect, rendering hypothesis tests invalid. Researchers must then resort to corrected inference procedures, such as using White's standard errors or Generalized Least Squares (GLS). If the assumption of **Exogeneity** is violated (e.g., due to measurement error or endogeneity), the OLS estimates become both biased and inconsistent, necessitating more sophisticated methods like Instrumental Variables (IV) or two-stage least squares, which are

designed to address correlation between the regressors and the error term, thereby moving beyond the basic OLS implementation of the Least Squares Criterion.

Further Reading

[Least squares - Wikipedia](#)

[Ordinary least squares \(OLS\) - Wikipedia](#)

[Least Squares Method Definition and Formula - Investopedia](#)

[Why squared error is preferred over absolute error - Stack Exchange Statistics](#)

ARABPSYCHOLOGY.COM