Research Article

# Statistically Significant Linear Regression Coefficients Solely Driven by Outliers in Finite-Sample Inference

**Felix Reichel**[1]

1. Department of Economics, Johannes Kepler University Linz, Linz, Austria

**In this paper, we investigate the impact of outliers on the statistical significance of coefficients in linear regression. We demonstrate, through numerical simulation using `R`, that a single outlier can cause an otherwise insignificant coefficient to appear statistically significant. We compare this with robust Huber regression, which reduces the effects of outliers. Afterwards, we approximate the influence of a single outlier on estimated regression coefficients and discuss common diagnostic statistics to detect influential observations in regression (e.g., studentized residuals). Furthermore, we relate this issue to the optional normality assumption in simple linear regression[1], required for exact finite-sample inference but asymptotically justified for large $n$ by the Central Limit Theorem (CLT). We also address the general dangers of relying solely on p-values without performing adequate regression diagnostics. Finally, we provide a brief overview of regression methods and discuss how they relate to the assumptions of the Gauss-Markov theorem.**

**Corresponding author:** Felix Reichel, felix.reichel@jku.at

## 1. Introduction

Linear regression is a foundational method widely used for modeling due to its simplicity, interpretability, and strong theoretical underpinnings. The classical simple linear regression (SLR) model in scalar notation is given by:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \tag{1}$$

where $\beta_0$ (the intercept) and $\beta_1$ (the slope) are the coefficients capturing the linear relationship between the predictor $X_i$ and the dependent/outcome/response variable $Y_i$, for $i = 1, \ldots, n$. Estimation via OLS

 doi.org/10.32388/KSZXTM 

yields interpretable, closed-form solutions and enables simple statistical inference using t-tests and confidence intervals, assuming that the classical linear model conditions hold. As shown in[2], under assumptions **SLR.1** through **SLR.4** for simple linear regression, the OLS estimates remain unbiased; **SLR.5**, which assumes normally distributed errors, is only required for valid finite-sample inference using the t-distribution, but can in general be relaxed in large samples due to the Central Limit Theorem. (CLT)

One of the major advantages of linear regression is the interpretability of the estimated coefficient $\hat{\beta}_1$, representing the expected change in the response $Y$ for a one-unit (assuming no re-scaling e.g. standard-units) increase in the predictor $X$. Additionally, hypothesis testing on coefficients is conducted using Student's t-statistic:

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \sim t_{n-2}, \quad H_0 : \beta_1 = 0, \tag{2}$$

which provides measures of statistical significance with the associated p-values, assuming the normality of the error term or a (sufficiently) large sample size.

Despite its upsides, a critical limitation of OLS is its strong sensitivity to outliers. The use of the squared loss function $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ for model fittings causes observations with large residuals to have a disproportionately high influence on the estimated coefficients. Although the sensitivity can in general be informative, it can also lead to highly misleading subsequent inferences and effect calculations. Even a single outlier can distort the estimated coefficient $\hat{\beta}_1$, resulting in underestimated standard errors and thus false statistical significance of the predictor variable[3].

This paper investigates how outliers can lead to misleading conclusions in regression analysis. Through simulation using `R`, we demonstrate the fragility of OLS-based inference when a finite sample inference is manipulated by insertion of a single outlier. To mitigate this issue, we additionally fit a robust Huber regression[4].

In doing so, we aim to demonstrate the limitations of classical OLS regression based on the quadratic loss function for statistical inference and advocate the use of additional robust alternatives and model diagnostic tools.

These include regression diagnostics (for example, residual analysis, leverage, studentized residuals, Cook's distance[5]), model fit statistics (for example, $R^2$, `F` tests), and formal outlier detection methods such as single outlier statistical tests[6][7][8] using residuals, to ensure reliable and transparent communication of statistically significant results.

## 2. Linear Regression and Statistical Inference

We begin by recalling the standard form of the linear regression model in matrix notation:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I), \tag{3}$$

where $Y \in \mathbb{R}^n$ is the dependent/outcome/response variable, $X \in \mathbb{R}^{n \times p}$ is the covariate matrix of predictors (assumed to have full column rank), $\beta \in \mathbb{R}^p$ is the vector of coefficients to be estimated, and $\varepsilon \in \mathbb{R}^n$ is its error term, assumed to be independent and identically distributed with zero mean and constant variance $\sigma^2$[2]. These assumptions are known as the classical linear model assumptions denoted as **MLR.1**–**MLR.5** in[2] in the context of multiple regression for example.

### 2.1. OLS Estimation

The coefficients $\beta$ are estimated using OLS, which minimizes the sum of squared residuals:

$$\hat{\beta} = \arg\min_{\beta} \|Y - X\beta\|^2. \tag{4}$$

The solution to this optimization problem is given by:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y. \tag{5}$$

This formula is valid under assumption **MLR.3** (no perfect multicollinearity), which ensures that $X^\top X$ is a invertible matrix product[2].

### 2.2. Distribution of the OLS Estimator

Under optional assumption **MLR.5**, which states that the error term $\varepsilon$ is normally distributed, the OLS estimator $\hat{\beta}$ also follows a multivariate normal distribution:

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 (X^\top X)^{-1}\right). \tag{6}$$

This result forms the basis for inference procedures such as hypothesis testing and confidence interval construction[2].

### 2.3. Estimating Variance

Because $\sigma^2$ is unknown in practice, it is estimated using the residuals from the fitted model. Under assumption **MLR.4** (homoskedasticity), an unbiased estimator of the variance is given by:

$$\hat{\sigma}^2 = \frac{1}{n-p} \|Y - X\hat{\beta}\|^2 = \frac{1}{n-p} \sum_{i=1}^{n} \hat{\varepsilon}_i^2, \tag{7}$$


doi.org/10.32388/KSZXTM

where $\hat{\varepsilon} = Y - X\hat{\beta}$ is the vector of residuals[2].

### *2.4. Standard Error and Hypothesis Testing*

The standard error of the estimated coefficient $\hat{\beta}_j$ is computed as:

$$SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 \left[(X^\top X)^{-1}\right]_{jj}}. \tag{8}$$

To test the null hypothesis $H_0 : \beta_j = 0$, we compute the corresponding $t$-statistic as:

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}. \tag{9}$$

Under the assumptions of the classical linear model—and particularly **MLR.5**—this statistic follows a Student's $t$-distribution with $n - p$ or written as $n - p - 1$ (depending on intercept inclusion) degrees of freedom:

$$t_j \sim t_{n-p}. \tag{10}$$

The two-sided p-value can be calculated as:

$$\text{p-val.} = 2 \cdot P(T > |t_j|), \quad T \sim t_{n-p}, \tag{11}$$

which tests whether the coefficient $\beta_j$ is statistically different from zero in the population regression model[2].

## 3. The Effect of a Single Outlier on Regression Coefficients

OLS is sensitive to outliers. One unusual point can strongly affect the estimated slope, especially if the point has high leverage—that means, if it lies far from the center of the predictor distribution. This can make a non-significant result appear statistically significant[3][5].

We can show this with a simple example. We generate 100 observations $(x, y)$ in `R` using the seed: 123 with no true relationship. Model 1, based on this clean data, shows no significant coefficient. In Model 2, we add one extreme outlier. This one point changes the slope to 1.62 and creates a highly significant result. Model 3 uses robust regression (Huber's M-estimator), which reduces the outlier's impact and gives a smaller slope again (see Table 1).


doi.org/10.32388/KSZXTM

| Variable | No Outlier | | With Outlier | | Robust Regression | |
|---|---|---|---|---|---|---|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| Intercept | -0.103 | 0.098 | -0.115 | 0.230 | -0.162 | 0.097 |
| Coefficient on $x$ | -0.052 | 0.107 | **1.620*** ** | **0.171** | **0.184** ** | **0.072** |
| Residual Std. Error | 0.971 (df = 98) | | 2.289 (df = 99) | | 0.966 (df = 99) | |
| $R^2$ / Adj. $R^2$ | 0.002 / -0.008 | | 0.476 / 0.471 | | - | |
| F Statistic | 0.241 (1, 98) | | **89.99*** (1, 99)** | | - | |
| Observations | 100 | | 101 | | 101 | |

**Table 1.** Impact of a Single Outlier on the Statistical Significance of Linear Regression Estimates

*Notes: Model 1 is estimated on clean data. Model 2 adds one outlier, which changes the slope and makes the result statistically significant. Model 3 uses robust regression (Huber) to reduce the influence of the outlier. Robust models often omit $R^2$ or F-statistics because they do not apply directly. See Appendix A for Residual Plots.*

*Significance levels:* $*\,p < 0.1$; $**\,p < 0.05$; $***\,p < 0.01$

We can also illustrate the effect mathematically. When a single value in the data changes, the approximate change (using the result of a first-order Taylor approximation) in the OLS estimate is approximately given by:

$$\Delta\hat{\beta} \approx (X^\top X)^{-1} x_i^\top \Delta y_i, \tag{12}$$

where $x_i$ is the row vector for the $i$-th observation. This shows that the change in the estimated slope depends on both the residual size ($\Delta y_i$) and the leverage of the (outlier) point. Leverage is defined as:

$$h_i = x_i (X^\top X)^{-1} x_i^\top, \tag{13}$$

where $h_i$ measures how far $x_i$ is from the center of the predictor space. High-leverage points can have a disproportionate effect on the fit[3].

Robust methods help reduce this sensitivity. Huber's M-estimator[4] uses a different loss function that grows quadratically near zero but linearly in the tails of the distribution, limiting the influence of large

 doi.org/10.32388/KSZXTM 

residuals. Another robust approach is Least Trimmed Squares (LTS), which fits the model using only the subset of observations with the smallest residuals[9].

Robust regression methods provide more stable and reliable estimates when datasets contain outliers.

# 4. Single Outlier Tests for Linear Regression Models

Outlier detection in linear regression models typically revolves around residuals, i.e., differences between observed and fitted values. Several test statistics target unusually large residuals to identify potential outliers.

## *4.1. Internally Studentized Residuals*

The internally studentized residual for observation $i$ is:

$$R_i := \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}, \quad \text{where} \quad e_i := y_i - \hat{y}_i, \tag{14}$$

$$h_{ii} := \mathbf{x}_i^\top (X^\top X)^{-1} \mathbf{x}_i, \quad \hat{\sigma}^2 := \frac{1}{n-p} \sum_{j=1}^{n} e_j^2. \tag{15}$$

Here, $h_{ii}$ is the leverage (diagonal of the so-called hat matrix $H := X(X^\top X)^{-1} X^\top$). The denominator rescales residuals the by the local variance.

Under the assumption of normally distributed errors (**MLR.5**), $R_i$ approximately follows a standard normal distribution for large $n$, but does not exactly follow a $t$-distribution. The **externally studentized residual**, which removes the $i$th observation when estimating variance, follows a $t_{n-p-1}$ distribution.

## *4.2. Maximum Absolute Internally Studentized Residual*

The test statistic for detecting a single outlier is the maximum studentized residual:

$$R_n := \max_{i=1,\ldots,n} |R_i|. \tag{16}$$

Under the null hypothesis $H_0$ (no outliers), an approximate critical value can be obtained using the Bonferroni correction. To control the family-wise error rate at level $\alpha$, we compare $R_n$ to the quantile of the Student's t-distribution with $n - p - 1$ degrees of freedom:

$$\mathbb{P}\left(R_n > t_{1-\alpha/(2n), n-p-1}\right) \leq \alpha. \tag{17}$$


doi.org/10.32388/KSZXTM

### 4.3. Normalized Maximum Ordinary Residual

An alternative approach avoids using the hat matrix $H$ and instead relies on the unadjusted, or raw, residuals. The corresponding test statistic is the *normalized maximum ordinary residual*:

$$R_n^* = \sqrt{n} \cdot \frac{\max_i |e_i|}{\|\mathbf{e}\|_2} = \sqrt{n} \cdot \frac{\max_i |y_i - \hat{y}_i|}{\sqrt{\sum_{j=1}^{n} e_j^2}}. \tag{18}$$

This statistic highlights large absolute residuals relative to the overall residual magnitude. Since it does not account for leverage, it can be more sensitive to large deviations in response values. However, this also means it may overlook influential outliers associated with high-leverage points.

### 4.4. Distributional Properties and Critical Values

The exact distributions of $R_n$ (maximum studentized residual) and $R_n^*$ (normalized maximum ordinary residual) are not analytically tractable. Therefore, critical values are typically estimated using one or more of the following methods:

- Bonferroni-adjusted $t$-tests applied to individual residuals,
- Monte Carlo or permutation tests under the null hypothesis $H_0$,
- Conservative bounds derived from F-distributions or inverted Student's $t$-distributions.

### 4.5. Upper Bound of Single Outlier Test Statistics

Ugah et al.[8] derive an upper bound as identical for both as:

$$R_0^* = \sqrt{\frac{(n-p)F_{\alpha/n,1,n-p-1}}{n-p-1+F_{\alpha/n,1,n-p-1}}}. \tag{19}$$


doi.org/10.32388/KSZXTM

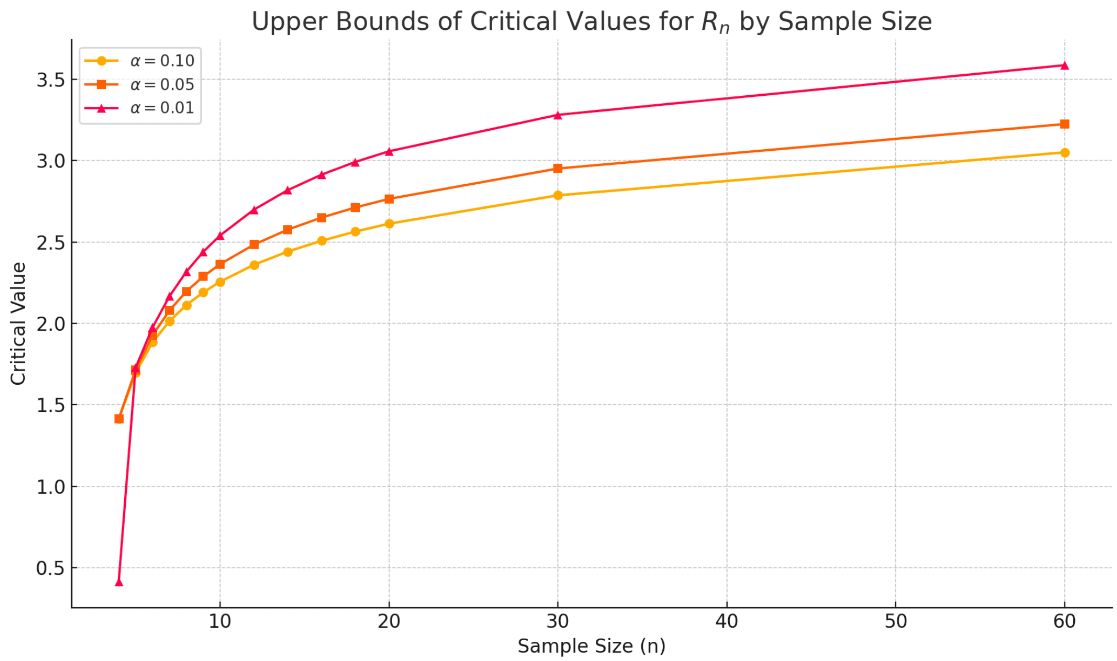


**Figure 1.** Upper bounds of critical values for $R_n$ by sample size $n$ and significance level $\alpha$.

## 5. On the Normality Assumption in Regression Analysis

Recall again **MLR.5** for exact finite-sample inference. Under this assumption, the least squares estimator

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} \tag{20}$$

is unbiased, efficient, and follows:

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1}). \tag{21}$$

This enables valid inference using standard test statistics.

The t-statistic for testing $H_0 : \beta_j = 0$ is:

$$t_j = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 (X^\top X)^{-1}_{jj}}} \sim t_{n-p}, \tag{22}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{23}$$

The residuals $\hat{\varepsilon}_i$, however, are correlated due to their dependence on the fitted values via the hat matrix $H = X(X^\top X)^{-1} X^\top$, where $\hat{\mathbf{y}} = H\mathbf{y}$ and $\hat{\boldsymbol{\varepsilon}} = (I - H)\mathbf{y}$.


doi.org/10.32388/KSZXTM

While OLS remains unbiased and consistent under weaker assumptions (e.g., finite variance and exogeneity), the exact distribution of test statistics like $t_j$ requires normality.

### *Outlier Effects on Normality Assumption in Regression Analysis*

Outliers violate this assumption in two key ways:

1. **Non-normality of residuals:** A single outlier induces skewness or kurtosis in the residual distribution, invalidating $t_j \sim t_{n-p}$. Deviations from linearity are often visible in the Q–Q plot.
   $\implies$ Inflated or deflated Type I error rates.
2. **Distortion of variance estimates:** Outliers distort $\hat{\sigma}^2$, affecting directly $SE(\hat{\beta}_j)$ and thus statistical inference. For high-leverage points $h_{ii}$ (diagonal entries of the hat matrix), even a small residual $e_i$ can disproportionately influence:
$$SE\left(\hat{\beta}_j\right) \propto \sqrt{(X^\top X)^{-1}_{jj}} \tag{24}$$
   $\implies$ Misestimated p-values, false significance, or masked effects. (Robust methods reduce the outlier's influence.)

Under outlier contamination, the OLS estimator becomes:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(0)} + \Delta(\mathbf{x}_o, y_o), \tag{25}$$

where $\Delta$ increases with leverage $h_{oo}$ and residual magnitude $|e_o|$. Thus, $\hat{\boldsymbol{\beta}}$ can become biased in finite samples and may become inconsistent, particularly if the outlier introduces dependence between the covariates and the errors.

By the **Gauss–Markov theorem,** OLS is known as the **Best Linear Unbiased Estimator (BLUE)** under classical conditions, regardless of normality. In large samples, the **Central Limit Theorem (CLT)** implies that $\hat{\boldsymbol{\beta}}$ is approximately normal, allowing for asymptotic inference even when the error distribution is not normal.

## 6. An Overview of Regression Methods

Linear regression relies on the **Gauss–Markov assumptions**, which ensure that the OLS estimator is the **BLUE**. These assumptions include linearity of the model, meaning the outcome variable $\mathbf{y}$ is expressed as a linear combination. The covariate matrix $X$ must have full column rank so that $X^\top X$ is invertible, ensuring that the parameter estimates are uniquely defined. Exogeneity is also required, meaning the regressors are uncorrelated with the error term: $\mathbb{E}[\boldsymbol{\varepsilon} \mid X] = 0$, which implies $\mathrm{Cov}(X, \boldsymbol{\varepsilon}) = 0$. The error


doi.org/10.32388/KSZXTM

terms must be homoscedastic, following a constant variance $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I$, and they must be uncorrelated across observations, i.e., $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$. While these assumptions are sufficient for OLS to be unbiased and efficient, an additional assumption of normally distributed errors, $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I)$, is required for exact finite-sample inference using t-tests and F-tests.

In practice, these conditions are often violated. To address such limitations, various extensions and robust methods have been developed. Below, we outline some of the most widely used regression methods and highlight their relation to the underlying Gauss–Markov assumption.

### *6.1. Deming Regression*

Deming regression accounts for measurement error in both $x$ and $y$, minimizing orthogonal distances:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{1 + \lambda}, \quad \lambda = \frac{\sigma_x^2}{\sigma_y^2} \tag{26}$$

Useful when both variables are noisy[10].

**Assumptions addressed:** Violates fixed $x$; assumes homoscedastic, independent errors.

### *6.2. Ridge Regression*

Ridge regression applies an $L_2$ penalty to control variance from multicollinearity:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{27}$$

[11].

**Assumptions addressed:** Mitigates multicollinearity.

### *6.3. Lasso Regression*

Lasso uses an $L_1$ penalty to induce sparsity:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{28}$$

Some coefficients may be exactly zero[12].

**Assumptions addressed:** Mitigates multicollinearity.

### *6.4. Elastic Net*

Elastic Net combines $L_1$ and $L_2$ penalties:


doi.org/10.32388/KSZXTM

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2 + \lambda_1 \sum |\beta_j| + \lambda_2 \sum \beta_j^2 \tag{29}$$

Balances sparsity and stability[13].

**Assumptions addressed:** Mitigates multicollinearity.

### *6.5. Robust Regression*

Robust regression reduces sensitivity to outliers by using a different loss function $\rho(\cdot)$ less sensitive than squared error:

$$\min_{\beta} \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta) \tag{30}$$

Huber's loss is a common default choice for example[4].

**Assumptions addressed:** Handles non-normality and heteroscedasticity.

### *6.6. Quantile Regression*

Quantile regression estimates conditional quantiles by minimizing asymmetric loss:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta), \quad \rho_\tau(u) = u(\tau - \mathbf{1}_{\{u<0\}}) \tag{31}$$

[14].

**Assumptions addressed:** Handles heteroscedasticity and non-normality.

### *6.7. Principal Component Regression (PCR)*

PCR applies PCA to $X$, then regresses $y$ on the top components. It reduces multicollinearity and variance[15].

**Assumptions addressed:** Mitigates multicollinearity.

### *6.8. Partial Least Squares (PLS)*

PLS projects $X$ onto components maximizing covariance with $y$, often outperforming PCR when predictors are correlated with the outcome[16].

**Assumptions addressed:** Mitigates multicollinearity.

### *6.9. LOESS*

LOESS fits local linear models weighted by distance:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} w_i(x)(y_i - \beta_0 - \beta_1 x_i)^2 \tag{32}$$

Flexible and nonparametric[17].

**Assumptions addressed:** Relaxes linearity; locally handles heteroscedasticity.

### *6.10. Spline Regression*

Spline regression uses piecewise polynomials with a smoothness penalty:

$$\min \sum (y_i - f(x_i))^2 + \lambda \int (f''(x))^2 dx \tag{33}$$

The smoothing parameter $\lambda$ controls complexity[18].

**Assumptions addressed:** Relaxes linearity.

### *6.11. Generalized Linear Models (GLMs)*

GLMs generalize linear models via a link function:

$$g(\mathbb{E}[y_i]) = x_i^\top \beta \tag{34}$$

Includes logistic and Poisson models[19].

**Assumptions addressed:** Handles non-normality and heteroscedasticity.

### *6.12. Generalized Additive Models (GAMs)*

GAMs allow additive nonlinear effects:

$$\mathbb{E}[y_i] = \alpha + f_1(x_{i1}) + \cdots + f_p(x_{ip}) \tag{35}$$

Each $f_j$ is estimated nonparametrically (e.g., splines)[20].

**Assumptions addressed:** Relaxes linearity; handles non-normality and mild heteroscedasticity.

## 7. Conclusion

Outliers can inflate the significance of regression coefficients, leading to incorrect inferences. The normality assumption of the residuals is critical for valid t-tests, and its violation—commonly due to outliers—necessitates careful diagnostic checks or the use of robust methods. It is advised to always conduct residual diagnostics and robust regression techniques to assess and mitigate these issues, even post outlier-removal, especially in finite-sample inference.


doi.org/10.32388/KSZXTM

Also, the use of outlier diagnostics utilizing both residuals and leverage is recommendable, such as standardized residuals, leverage plots, Cook's distance, DFBETAS, and influence plots. Normality of residuals can be assessed using tools like the Q–Q plot, Shapiro–Wilk test, or histogram of residuals.

## Appendix A. Residual Diagnostics

This appendix presents residual diagnostic plots for three models: (1) a clean OLS model without any outliers, (2) an OLS model with a single high-leverage outlier, and (3) a robust regression model fit to the contaminated data. These plots provide visual evidence of how outliers affect model assumptions and how robust methods mitigate their influence.


doi.org/10.32388/KSZXTM

## *A.1. Residual Plots: SLR OLS Model (Clean Data/No Outlier)*

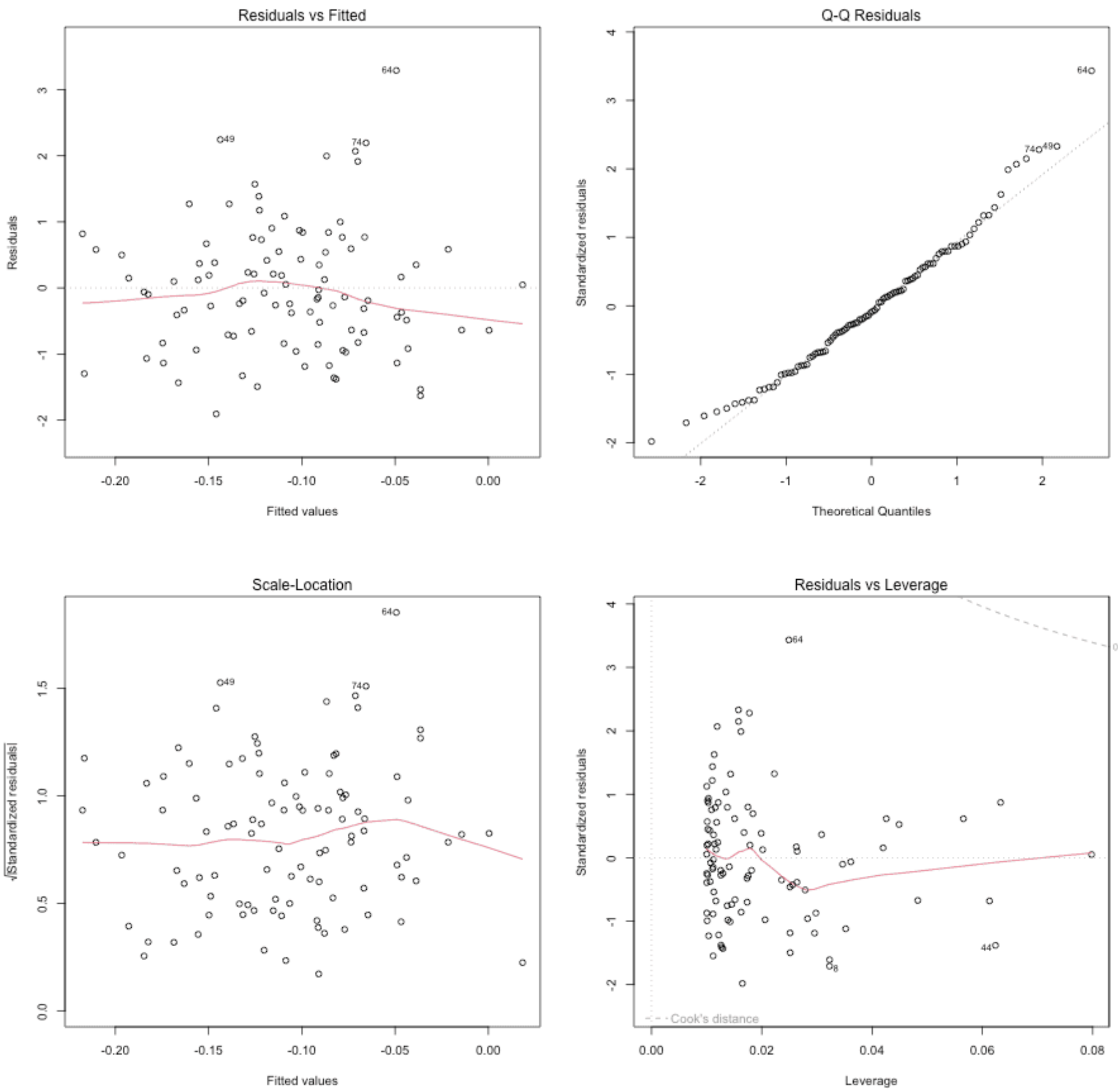


**Figure 2.** Diagnostic plots for the OLS model. The residuals appear homoscedastic (constant variance), symmetrically distributed, and independent. The Q-Q plot indicates approximate normality, validating the assumptions of OLS.


doi.org/10.32388/KSZXTM

*A.2. Residual Plots: SLR OLS Model With An Outlier*

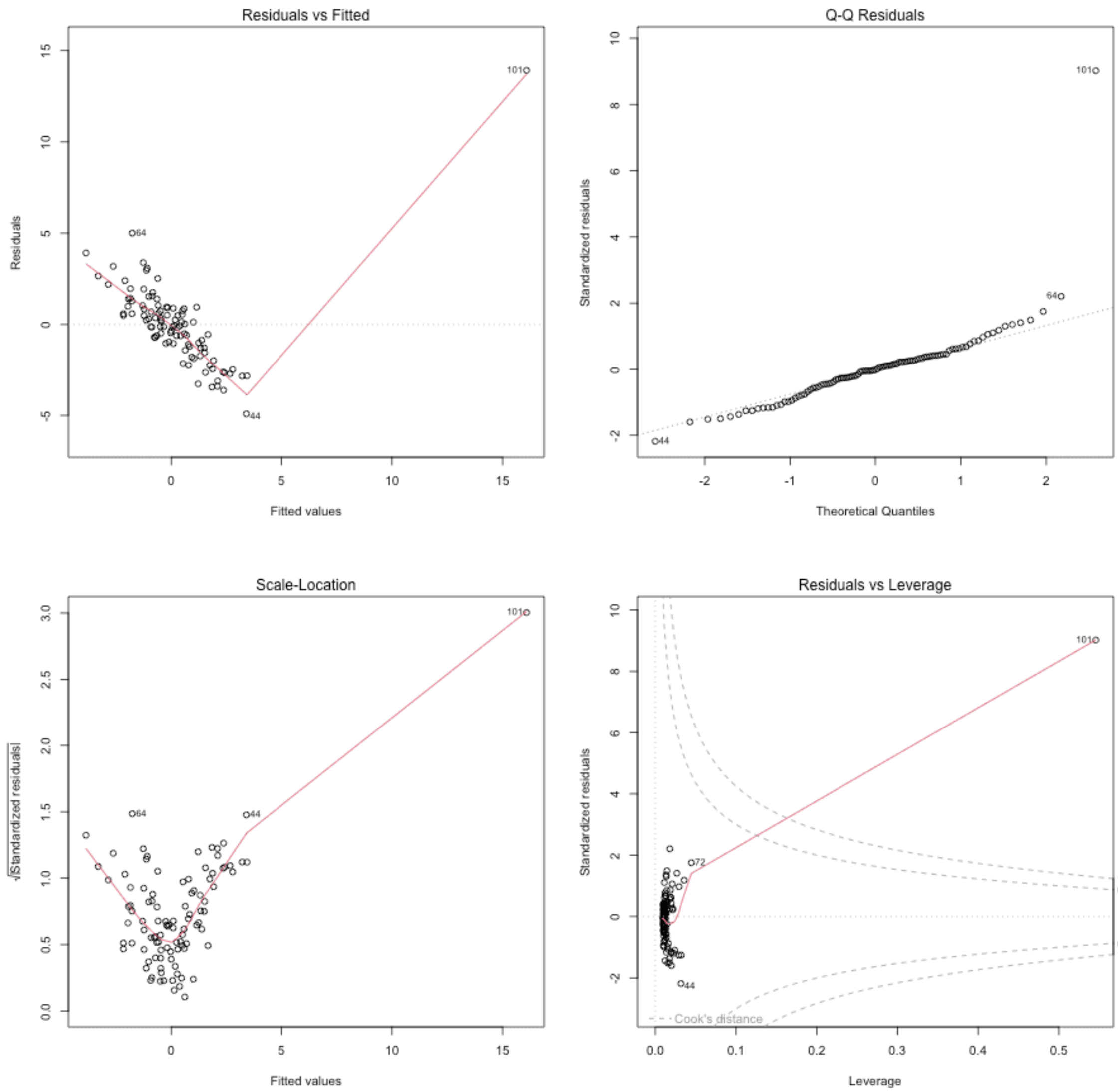


**Figure 3.** Diagnostic plots for OLS model including a high-leverage outlier. The residuals show clear distortion: heteroscedasticity, skewness, and heavy tails are evident. The Q-Q plot deviates significantly from the normal line, and the residuals vs. fitted plot shows a large residual corresponding to the outlier. This illustrates the breakdown of classical OLS assumptions.


doi.org/10.32388/KSZXTM

*A.3. Residual Plots: SLR Huber Robust With An Outlier*

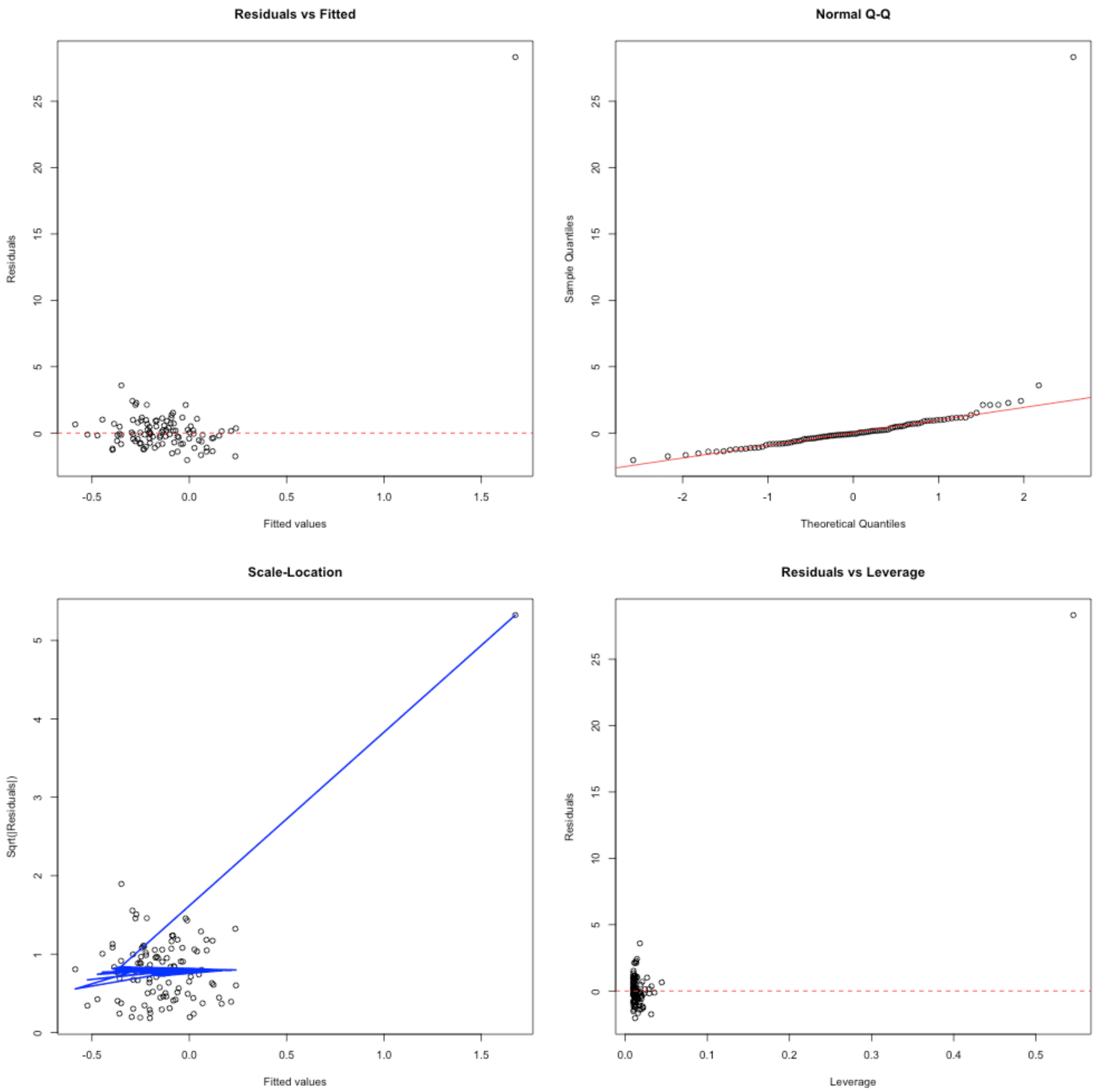


**Figure 4.** Residuals vs. fitted values for the robust regression model applied to the data with an outlier. Compared to the standard OLS fit (A.2), the residuals are more evenly spread and the extreme influence of the outlier is visibly diminished. This confirms that the robust method effectively downweights the anomalous observation, preserving the integrity of the regression fit.

## Notes

**JEL Codes:** C10, C12, C13, C80, C81.

## Acknowledgements

Thanks to T.Cunningham (MSU) this version now fixes a typo in the title.

## References

1. ^*Schmidt AF, Finan C (2018). "Linear regression and the normality assumption." Journal of Clinical Epidemiology. 98:146–151.*
2. a, b, c, d, e, f, g*Wooldridge JM (2019). Introductory Econometrics: A Modern Approach. 7th ed. Cengage Learning.*
3. a, b, c*Belsley DA, Kuh E, Welsch RE (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons.*
4. a, b, c*Huber PJ (1964). "Robust estimation of a location parameter." Annals of Mathematical Statistics. 35(1):73–101.*
5. a, b*Cook RD (1977). "Detection of influential observations in linear regression." Technometrics. 19(1):15–18.*
6. ^*Barnett V, Lewis T (1978). Outliers in Statistical Data. Wiley.*
7. ^*Weisberg S (1980). "Multiple outlier detection in regression using studentized residuals." Technometrics. 22(2):139–144.*
8. a, b*Ugah TE, Mba EI, Eze MC, Arum KC, Mba IC, Oranye HE (2021). "On the upper bounds of test statistics for a single outlier test in linear regression models." Journal of Applied Mathematics. 2021:5 pages.*
9. ^*Rousseeuw PJ, Leroy AM (1987). Robust Regression and Outlier Detection. Wiley.*
10. ^*Cornbleet PJ, Gochman N (1989). "Incorrect least-squares regression coefficients in method- comparison analysis." Clinical Chemistry. 35(4):584–585.*
11. ^*Hoerl AE, Kennard RW (1970). "Ridge regression: Biased estimation for nonorthogonal problems." Technometrics. 12(1):55–67.*
12. ^*Tibshirani R (1996). "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B. 58(1):267–288.*
13. ^*Zou H, Hastie T (2005). "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society: Series B. 67(2):301–320.*
14. ^*Koenker R, Bassett G (1978). "Regression quantiles." Econometrica. 46(1):33–50.*
15. ^*Jolliffe IT (2002). Principal Component Analysis. 2nd ed. Springer.*

16. ^*Wold H, Ruhe A, Wold S, Dunn WJ III (1984). "The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses." In: Krzanowski WJ, editor. Multivariate Analysis. Academic Press. p. 167–198.*
17. ^*Cleveland WS (1979). "Robust locally weighted regression and smoothing scatterplots." Journal of the American Statistical Association. 74(368):829–836.*
18. ^*de Boor C (1978). A Practical Guide to Splines. Springer.*
19. ^*McCullagh P, Nelder JA (1989). Generalized Linear Models. 2nd ed. Chapman and Hall.*
20. ^*Hastie T, Tibshirani R (1990). Generalized Additive Models. Chapman and Hall.*

## Declarations

**Funding:** No specific funding was received for this work.

**Potential competing interests:** No potential competing interests to declare.