Huber Regression: A Robust Alternative to Least Squares

James E. Stine and Marcus Mellor
Electrical and Computer Engineering Department
Oklahoma State University
Stillwater, OK 74078, USA

Introduction

When modeling data with regression, we often assume that errors are roughly Gaussian and that extreme observations are rare. Under these conditions, Ordinary Least Squares (OLS) is attractive because it is simple, efficient, and mathematically well-understood.

Ordinary Least Squares (OLS) regression minimizes the sum of squared residuals:

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

This objective gives OLS very desirable properties when the errors ε_i are independent, identically distributed, and approximately Gaussian:

- OLS is an *unbiased estimator* of the regression parameters.
- OLS achieves the *minimum variance* among all linear unbiased estimators (the Gauss–Markov theorem).
- Under Gaussian errors, OLS is also the *maximum likelihood estimator*.

However, the squaring of residuals means that large errors are penalized quadratically. For example, if one residual is 5 times larger than another, its contribution to the objective is $5^2 = 25$ times larger. A small number of extreme observations ("outliers") can therefore dominate the regression line, pulling it away from the trend that represents the bulk of the data.

To address this sensitivity, Peter J. Huber proposed a modified loss function that retains the quadratic penalty for small residuals (where Gaussian assumptions are reasonable) but switches to a linear penalty for large residuals (where robustness is needed) [1]. This approach preserves much of the efficiency of OLS under Gaussian errors while dramatically reducing the influence of outliers. In the next section, we introduce the Huber loss function formally.

Huber Loss Function

Peter J. Huber (1964) proposed a hybrid loss function:

$$\rho_\delta(r) = \begin{cases} \frac{1}{2} \cdot r^2, & |r| \le \delta, \\ \delta \cdot \left(|r| - \frac{1}{2} \cdot \delta\right), & |r| > \delta, \end{cases}$$

where r is the residual and δ is a tuning threshold [1].

- For $|r| \le \delta$: behaves like OLS (quadratic penalty).
- For $|r| > \delta$: behaves like Least Absolute Deviations (linear penalty).

Choosing δ

- $\delta = 1.345 \cdot \sigma$ (where σ is the residual scale) yields $\approx 95\%$ efficiency if errors are Gaussian [2].
- Larger $\delta \Rightarrow$ closer to OLS (less robust).
- Smaller $\delta \Rightarrow$ closer to the Least Absolute Deviations (LAD) regression (more robust).

When to Use Huber Regression

- Data are mostly well-behaved but contain a few extreme outliers.
- Example: salaries, income, or experimental data where a few points are corrupted.

Exercises

Work through the following steps in order. Each exercise builds on the previous one to help you connect the concepts of OLS, robustness, and the role of the Huber parameter δ.

1. Model fitting: Fit both Ordinary Least Squares (OLS) and Huber regression to the salary dataset. Record the slope and intercept for each method. Use SciPy to compute the Huber loss. Try to avoid using ChatGPT or other AI and learn how to use SciPy to compute OLS with Huber loss.
2. You can get help from `pydoc` on SciPy least squares option with `python3 -m pydoc scipy.optimize.least_squares` or `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html`.
3. Comparing influence: Examine how the slopes and intercepts differ. Which method appears more influenced by the very highest salaries in the dataset? What does this tell you about the sensitivity of OLS compared to Huber regression?
4. Effect of the tuning parameter: Re-fit the Huber model with different thresholds δ (e.g., 0.5σ, 1.345σ, and 3σ). Plot all regression lines on the same graph. How does changing δ shift the balance between efficiency (OLS-like) and robustness?
5. Reflection: Think about the nature of real-world salary data. Salaries are usually *right-skewed* (a few very high earners pull the distribution upward), and they often contain *outliers* (superstars or data entry errors). Do you think the residuals from a regression model are likely to be perfectly Gaussian in this setting? Why or why not? Based on your reasoning, which regression method (OLS, Huber, or even median/quantile regression) would you prefer, and why?

Stubbed Python Code (with guided TODOs)

The following template provides the overall structure. Each `TODO` gives you a hint for how to complete the function. You are not given the final code, but the steps are outlined within the starter code in the repository under the MLB subdirectory.

References

[1] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[2] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed. Wiley, 2009.