ISOSS

Redescending M-estimators for Robust Regression: A Comparative Analysis based on efficiency factor

Atiq Ur Rehman
Supervised by Insha Ullah and Dost Muhammad

The Islamic Countries Society of Statistical Sciences

Contents

Introduction

- The Ordinary Least Square (OLS) estimation method is used to estimate the regression coefficients.
- OLS is sensitive to outliers or to the violation of certain model assumptions. For example, the normality of the residuals.
- Outliers, data points lies far from the majority, could influence the distribution of residuals. Consequently, on the statistical inference procedures involved.

A bit more about the consequences

- The regression coefficients could be biased and inefficient.
- An insignification regression coefficient could turn to significant because of the presence of outliers in the data.
- A positive relationship could turn to negative and vice verse.

Robust methods

- Robust estimation methods have been proposed: R-estimator, LMS, LTS, S-estimator, M-estimation and others.
- I will be focusing on M-estimation.

OLS Method.

Regression model:

$$Y_i = X_i^t \beta + \epsilon_i.$$

Usual OLS estimator:

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^{n} \epsilon_i^2 = \min_{\beta} \frac{1}{2} \sum_{i=1}^{n} \left(Y_i - X_i^t \beta \right)^2$$

Note that $\rho(\epsilon_i) = \epsilon_i^2$.

M-estimation.

The idea is to minimize some other function rather than the squared loss function. For example, Huber chose to minimize

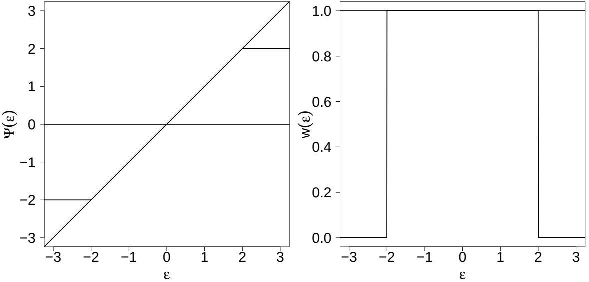
$$\rho(\epsilon_i) = \begin{cases} \frac{1}{2}\epsilon_i^2 & |\epsilon_i| \leq c \\ c|\epsilon_i| - \frac{1}{2}c^2 & |\epsilon_i| > c \end{cases},$$

$$\psi(\epsilon_i) = \begin{cases} \epsilon_i & |\epsilon_i| \leq c \\ c \text{ sign}(\epsilon_i) & |\epsilon_i| > c \end{cases},$$

where $\psi(.) = \rho'(.)$.

Huber ψ and weight functions

It gives less weight to the outlying observations and thus outliers have less effect on the estimator.



Redescending M-estimation

- Redescending m-estimator further reduce the influence of the outliers.
- The derivative of $\rho(.) = \psi(.)$ is redescending, that is, $\lim_{\epsilon_i \to \infty} \psi(\epsilon_i) = 0$.
- rather than $\lim_{\epsilon_i \to \infty} \psi(\epsilon_i) = \infty$ as in the case of OLS.
- $\lim_{\epsilon_i \to \infty} \psi(\epsilon_i) = c$ as in the case of Huber's proposal.

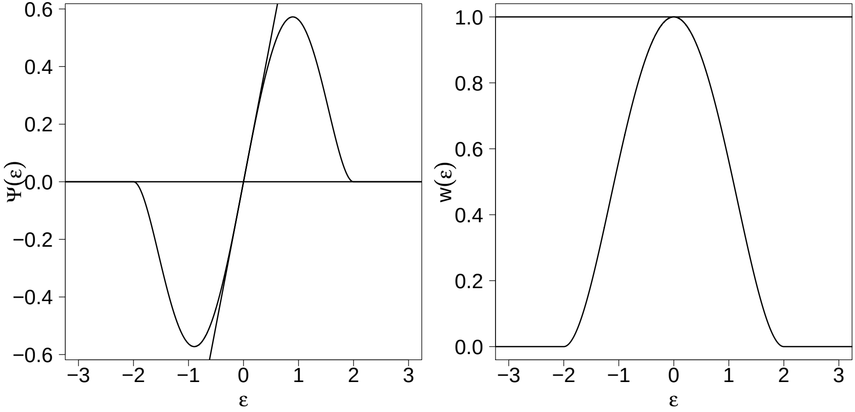
Tukey's dispersion function

$$\rho(\epsilon_i) = \begin{cases} \frac{c^2}{6}\{1 - [1 - (\frac{\epsilon_i}{c})^2]^3\} & |\epsilon_i| \le c \\ \frac{c^2}{6} & |\epsilon_i| > c \end{cases},$$

and

$$\psi(\epsilon_i) = \begin{cases} \epsilon_i[1 - (\frac{\epsilon_i}{c})^2]^2 & |\epsilon_i| \le c \\ 0 & |\epsilon_i| > c \end{cases},$$

Tukey's ψ and weight functions



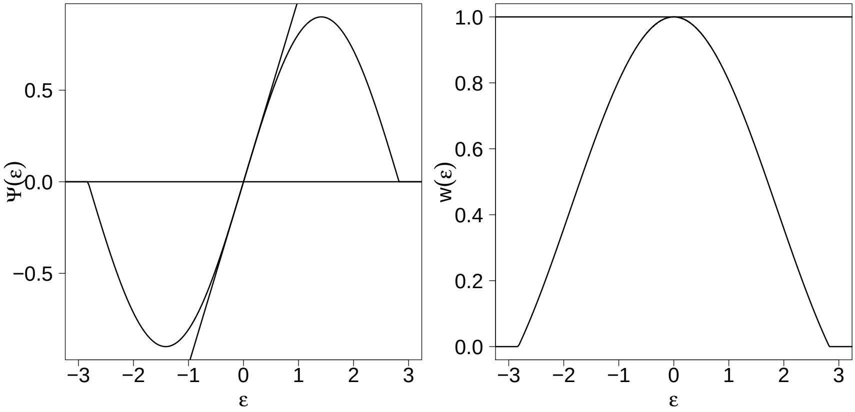
Andrew's dispersion function

$$\rho(\epsilon_i) = \begin{cases} -c * \cos(\frac{\epsilon_i}{c}) & |\epsilon_i| \leq c\pi \\ 0 & |\epsilon_i| > c\pi \end{cases},$$

and

$$\psi(\epsilon_i) = \begin{cases} \sin(\frac{\epsilon_i}{c}) & |\epsilon_i| \leq c\pi \\ 0 & |\epsilon_i| > c\pi, \end{cases}$$

Andrew's ψ and weight functions



Redescending M-estimator

For example, Andrew's Redescending M-estimator (Andrew et al. (1972)), Tukey Redescending M-estimator (Beaten and Tukey (1974)), Qadir Redescending M-estimator (Qadir (1996)), Modified form of Tukey's biweight function (Ali and Qadir (2005)), Insha's redescending M-estimator (Ullah, Qadir and Ali (2006)), Alamgir Redescending M-estimator (Alamgir et al. (2013)).

- The most recent one is Jiang et al (2018). Robust Estimation Using Modified Huber's Functions With New Tails. Technometrics.

Tuning Constant c

- All of the M-estimators mentioned above (except OLS) depend on a tuning constant $c > 0$.
- the value of c need to be chosen without sacrificing efficiency.
- Default choice in rlm function of R is 1.345 for Huber's M-estimator and 4.685 for Tukey's bisquare.
- The default choices are not always the best.

Data driven approaches

- In the era of Big data one may need to run a model multiple times—each time with a different dataset—it is difficult to choose the best value of c manually
- Automated methods to choose the value of c are required. Wang et al. (2007) has made tuning constant data-dependent for the Huber's M-estimator.

Asymptotic Results.

$$Let \quad \phi_n(\theta) = \frac{1}{nS_n}\sum_{i=1}^{n}\psi(\frac{Y_i - X_i^T\theta}{S_n})X_i$$

By Taylor's expansion

$$\begin{aligned}\phi_n(\hat{\beta}_n) =& \phi_n(\beta_0) + \phi_n'(\beta_0)(\hat{\beta}_n - \beta_0) + \frac{1}{2}(\hat{\beta}_n - \beta_0)^T\phi_n''(\beta_n')\\ &(\hat{\beta}_n - \beta_0),\end{aligned}$$

where $\phi'(.)$ and $\phi''(.)$ are the first-order and second-order derivatives $\phi(.)$.

As we have $\phi_n(\hat{\beta}_n) = 0$ and after some algebraic manipulation

$$\frac{S_n}{\sqrt{n}} \sum_{i=1}^{n} \psi(\frac{Y_i - X_i^T \beta_0}{S_n}) X_i = \sqrt{n}(\beta_n' - \beta_0) \frac{1}{n} \sum_{i=1}^{n} \psi'(\frac{Y_i - X_{\beta}^{T}{}_0}{S_n}) X_i^T X_i + o_p(1)$$

Since $S_n \to \sigma$ as $n \to \infty$, thus

$$\sqrt{n}(\hat{\beta} - \beta) \sim \mathcal{N}\left(0, \tau^{-1}\sigma^2[E(X^TX)]^{-1}\right),$$

where $\tau^{-1} = \frac{[E\psi'(\epsilon_i/\sigma)]^2}{E\psi^2(\epsilon_i/\sigma)} = \frac{b^2}{\sigma_\psi^2}$. The estimate of b can be obtained by

$$b = Pr(|\epsilon_i| \leq c)$$

and

$$\sigma_\psi^2 = \int_{-c}^{c} \psi^2(\epsilon_i) f(\epsilon_i) + c^2(1 - b)$$

$\hat{\tau}(c)$ for Huber

$$\hat{\tau}(c) = \frac{\left[\sum_{i=1}^{n} I(|\hat{\epsilon}_i| \leq c)\right]^2}{n \sum_{i=1}^{n}[I(|\epsilon_i| \leq c)\psi^2(\hat{\epsilon}_i) + c^2 I(|\hat{\epsilon}_i| > c)]},$$

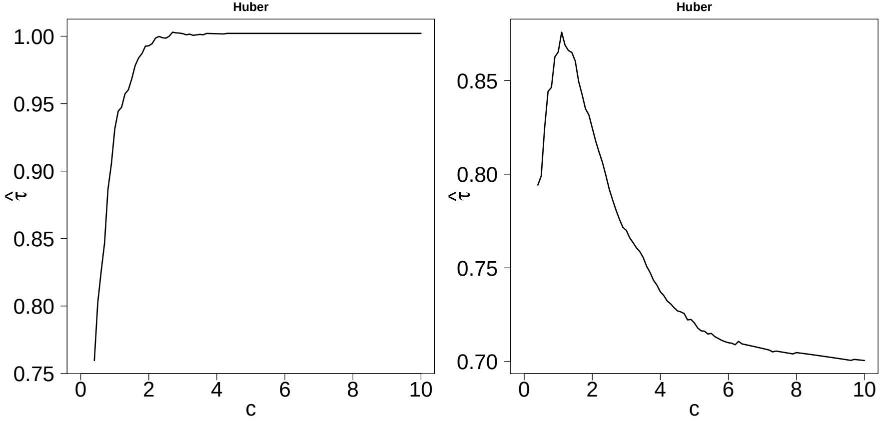
$\hat{\tau}(c)$ for Tukey's

$$\hat{\tau}(c) = \frac{\left[\sum_{i=1}^{n} I(|\hat{\epsilon}_i| \leq c)\psi'(\epsilon_i)\right]^2}{n \sum_{i=1}^{n}[I(|\epsilon_i| \leq c)\psi^2(\hat{\epsilon}_i)]},$$

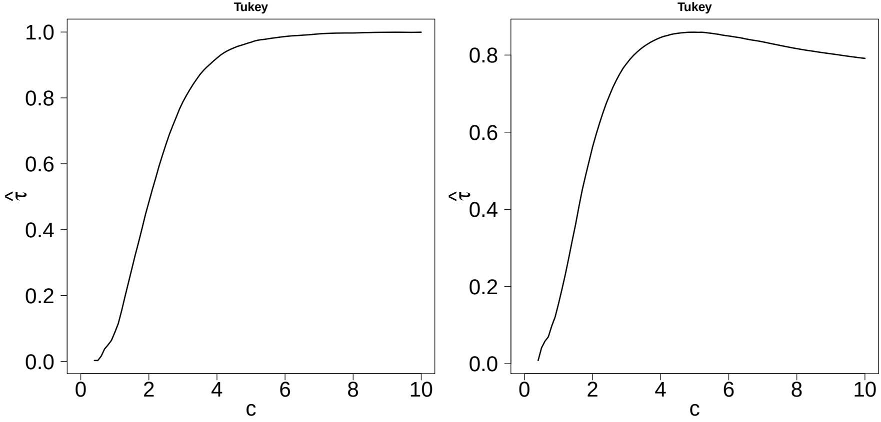
$\hat{\tau}(c)$ for Andrew

$$\hat{\tau}(c) = \frac{\left[\sum_{i=1}^{n} I(|\hat{\epsilon}_i| \leq c)\psi'(\epsilon_i)\right]^2}{n \sum_{i=1}^{n}[I(|\epsilon_i| \leq c)\psi^2(\hat{\epsilon}_i)]},$$

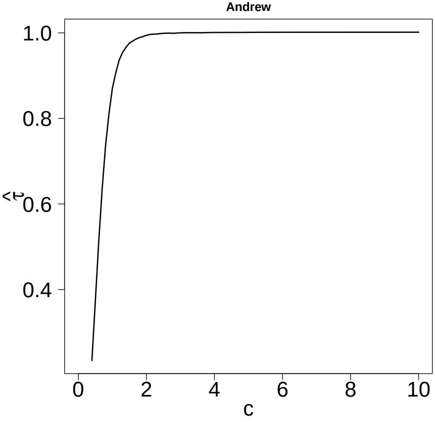
$\hat{\tau}$ of Huber for normal and non-normal residuals.

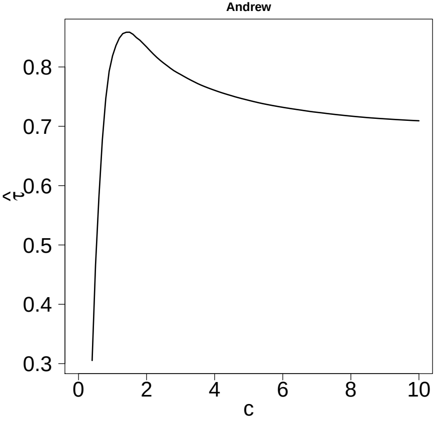


$\hat{\tau}$ of Tukey for normal and non-normal residuals.

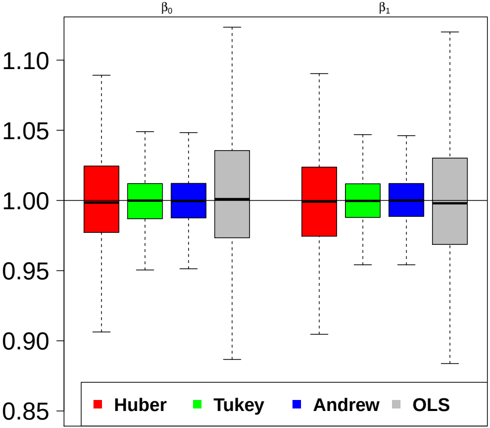


$\hat{\tau}$ of Andrew for normal and non-normal residuals.





Comparison based on $\hat{\tau}$



Summary

- Many redescending M-estimator have been proposed.
- These estimators depends on the value of c, which is chosen manually. Still no guarantee of efficiency.
- We have used efficiency factor to choose the value of c for redescending M-estimators.
- This efficiency factor helps to choose the best out of many available redescending M-estimators.