Author(s): Taskinen, Sara; Nordhausen, Klaus

Title: Least Median of Squares

Year: 2021

Version: Accepted version (Final draft)

Chapter 1

Least Median of Squares

Sara Taskinen and Klaus Nordhausen

Department of Mathematics and Statistics, University of Jyväskylä

sara.l.taskinen@jyu.fi, klaus.k.nordhausen@jyu.fi

Definition

The least median of squares (LMS) is a regression method introduced in Rousseeuw (1984) and further developed in Rousseeuw & Leroy (1987). The LMS estimator minimizes the median of the squared residuals. The method is shown to be highly robust having a breakdown point of 50% which is the highest possible. Because of this, the LMS regression method is fitted to the main bulk of the data while ignoring the rest of the data. Hence, atypical observations are easily identified. The method suffers from very low efficiency and is therefore recommended to be used primarily as an exploratory tool to identify regression outliers.

1 Introduction

Consider a data set consisting of n independent and identically distributed (iid) observations $(\boldsymbol{x}_i', y_i)$, $i = 1, \dots, n$, where $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})'$ and y_i are the observed values of predictor variables and response variables, respectively. The data are assumed to follow a linear regression model

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i,$$

where the p-vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ contains the unknown regression coefficients to be estimated based on the data and the errors ϵ_i are iid and independent

S. Taskinen & K. Nordhausen. Least Median of Squares. Submitted to: Encyclopedia of Mathematical Geosciences

of $\boldsymbol{x}_i$.
The widely used least squares (LS) estimator for β was proposed in early 1800 by Gauss and Legendre (Stigler 1981). The LS estimate of β is the $\hat{\beta}$ that minimizes

$$\sum_{i=1}^{n} r_i^2, \tag{1.1}$$

that is, the sum of squared residuals $r_i = y_i - \boldsymbol{x}_i' \hat{\beta}$. The popularity of the estimate arises mainly from its nice computational and theoretical properties. The LS estimator has a closed form solution and can be computed explicitly from data using simple matrix algebra. The estimator has also a desired property of being regression, scale and affine equivariant meaning that it is known how the estimate changes under different transformations of the data (Rousseeuw & Leroy 1987). If the error distribution is Gaussian, then the LS estimate is optimal. It is however well known that the LS estimator is extremely sensitive to outlying observations and even a single outlier can have a huge effect on the estimate. Notice that with outlier we mean here regression outlier, that is, an observation, $(\boldsymbol{x}_i', y_i)$, where the relationship between $\boldsymbol{x}_i$ and y_i differs from that of the majority of the data. For the identification of different types of regression outliers, see Rousseeuw & Leroy (1987), for example.
To improve on the LS estimator, several robust regression estimators have been proposed in the literature. Many of the proposed estimators are obtained by replacing the squared residuals by another function of the residuals. The least absolute deviation (LAD) estimator or L_1 regression estimator is obtained by minimizing the sum of the absolute values of the residuals $\sum_{i=1}^{n} |r_i|$, and a more general family of estimators, that is, the M-estimators minimize $\sum_{i=1}^{n} \rho(r_i)$, where ρ is a symmetric function ($\rho(-x) = \rho(x)$) with a unique minimum at zero (Huber 1973). Such regression estimates are robust against outliers in response variables (y_i's), however, they cannot tolerate outliers in the predictors (x_i's). Notice also that such outliers, which are also called as "leverage points", cannot always be identified using residuals based on the methods described above as the regression estimates may be strongly affected by the outliers yielding small residual values.

2 Least median of squares estimator

As mentioned in the previous section, a single outlier can strongly affect the estimates that are obtained by minimizing the sum of the squared residuals (or other function of the residuals). A regression estimator that is very robust against outliers in responses as well as outliers in predictors is obtained by replacing the sum in (1.1) with the median (Rousseeuw 1984). The least median of squares (LMS) estimator thus minimizes the median of the squared residuals

$$\operatorname*{Median}_i r_i^2. \tag{1.2}$$

In the case of a single predictor the solution can be interpreted as the midline between the two closest parallel lines which contain half of the data between them. In the case of an intercept only model, *i.e.*, when just the location of the y_i's is to be computed, the estimate is known as the Shorth estimate, the midpoint of the shortest interval to contain 50% of the data. If the amount of the points outside the parallel lines is to be reduced the method is known as least α-quantile estimator (Rousseeuw & Leroy 1987). It is shown in Rousseeuw (1984) that there always exists a solution for (1.2). However, the objective function may have many local minima (Steele & Steiger 1986). For large p the computation requires some subsampling and may be time-consuming (Souvaine & Steele 1987; Joss & Marazzi 1990). See also a cautionary note on the instability of the LMS solution in Hettmansperger & Sheather (1992).
The LMS estimate for the regression coefficient is one of the few robust estimates which does not need an initial scale estimate. The same property holds for the LS estimate. Based on the LMS estimate the scale of the residuals ϵ_i can be estimated as follows. First an initial scale s_0 is obtained as

$$s_0 = 1.4826\left(1 + \frac{5}{n-p}\right)\sqrt{\operatorname*{Median}_i r_i^2},$$

where the first constant term is for achieving consistency in case the errors follow a Gaussian distribution, and the second constant term is a finite sample correction. The initial scale is then used to define the standardized residuals $r_{i,st} = r_i/s_0$, $i = 1, \dots, n$. A general rule is that observations which have absolute standardized residuals smaller than 2.5 are considered as good data points. The actual scale based on LMS, σ_{LMS}, is then the standard deviation of the (unstandardized) residuals of the good data points.
Often in robustness studies it is of interest to evaluate what is the amount of outliers the estimator can handle. To formalize this, we need a notion of breakdown point. In the next we shortly recall the definition of finite-sample breakdown point as introduced by Donoho & Huber (1983). Let a sample of n data points be $Z = (z_1, \dots, z_n)$, where $z_i = (x_i', y_i)'$, and write $\hat{\beta}(Z)$ for an regression estimate based on the data. Now consider all possible contaminated samples Z' that are

obtained by replacing any m of the original data points by arbitrary values. Then denote the maximum bias caused by such contamination as

$$\text{bias}(m; \hat{\beta}, Z) = \sup_{Z'} ||\hat{\beta}(Z') - \hat{\beta}(Z)||.$$

Infinite bias means that m outliers can have an arbitrarily large effect on $\hat{\beta}$ and we say that the estimator "breaks down". Further, the finite-sample breakdown point is the minimum fraction of contamination that will make the estimator to break down. Mathematically, this can be expressed as

$$\epsilon^*(\hat{\beta}, Z) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} \ : \ \text{bias}(m; \hat{\beta}, Z) = \infty \right\}.$$

Rousseeuw (1984) showed that if $p > 1$ and the observations are in general position, meaning that any p observations give a unique determination of β, then the finite sample breakdown point of the LMS estimator is $([n/2] - p + 2)/n$, where $[n/2]$ denotes the largest integer less than or equal to $n/2$. When considering the limit $n \to \infty$ (with p fixed), the asymptotic breakdown point of the LMS estimator is 50%, that is, the best one can expect. The LMS thus extends the 50% breakdown property of the median to the regression setting. Another tool to study the robustness properties of an estimator is the influence function which measures the change of the estimator when an outlying observation is introduced. As the LMS estimator converges at the rate of $n^{-1/3}$ and does not have a normal limiting distribution, its influence function is not well-defined and the estimator suffers from low efficiency. Hence, Rousseeuw (1984) recommends that LMS should not be used for inferential purposes, but as an exploratory tool for diagnosing regression outliers. LMS is for example implemented in the R package `MASS` (Venables & Ripley (2002)) as the function `lmsrob`.

3 Example

Smith et al. (1984) measured the chemical contents of 53 samples of rocks in Western Australia. The data are available as the `mineral` data set in `R` (R Core Team 2020) package `RobStatTM` accompanying the book by Maronna et al. (2018). In Figure 1 we plot the chemical contents (in parts per million) of zinc vs. the chemical contents of copper. We notice that the observation 15 stands out as a clear outlier and the LS estimator is highly influence by this observation. If we reject observation 15, then the LS fit seems to model correctly the main bulk of the data. In Figure 2 the scaled residuals based on the LS fit are plotted. As the observations with absolute value of standardised residual exceeding 2.5 are identified as outliers (Rousseeuw & Leroy 1987), the horizontal lines at 2.5 and -2.5 are also added in the figure. The plot indicates that all scaled residuals are very small and although the LS method reveals two outliers, none of them is classified extremely atypical.

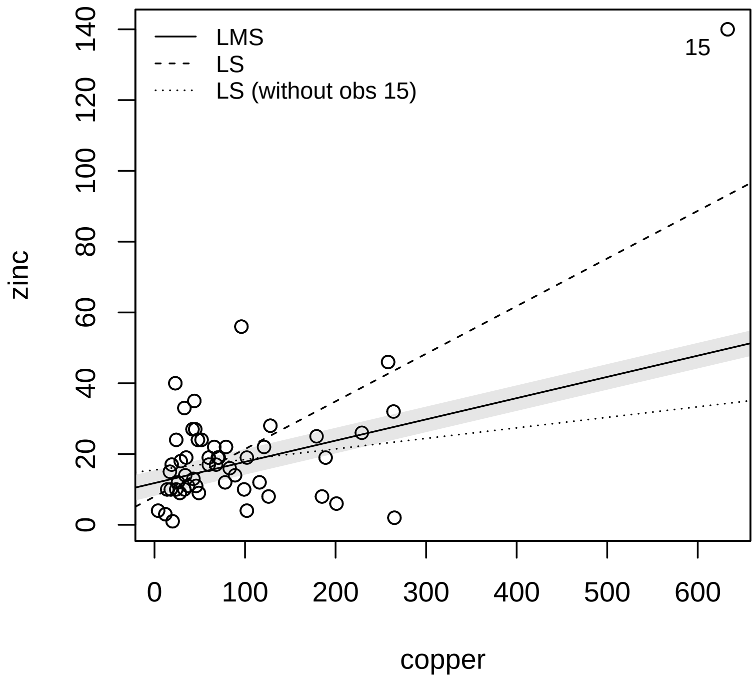

Figure 1: Zinc and copper contents of 53 samples of rocks and the regression lines based on least median of squares regression (LMS) and least squares regression (LS) applied to the original data and data with observation 15 as rejected. The grey area around the LMS regression line contains the "inner" 50% percent of the data points.

The LMS estimator is not affected by observation 15 and the resulting fit is nearly equivalent to the LS fit obtained when the observation 15 is rejected. The LMS estimates for intercept and slope are 11.79 and 0.06, respectively (as compared to the LS-estimates 7.96 and 0.13). The scaled residuals in Figure 2 indicate that the method identifies altogether eight outliers and observation 15 clearly stands out in the residual plot.

To have an idea about the model fit, Rousseeuw & Leroy (1987) define the coeffi-

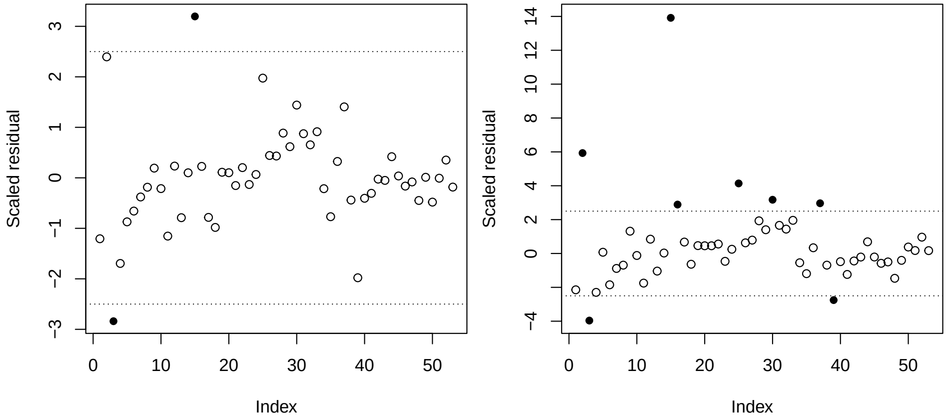

Figure 2: Scaled residuals based on LS regression (left) and LMS regression (right) applied to mineral data. An observation with an absolute value of standardised residual larger than 2.5 is identified as an outlier and marked using filled black dots.

cient of determination in the case of LMS as

$$R^2_{LMS} = 1 - \left(\frac{\text{Median}_i \, |r_i|}{\text{Mad}_i \; y_i} \right)^2,$$

where $\text{Mad}_i \; y_i$ denotes the median absolute deviation. For our example data $R^2_{LMS} = 0.59$ which is much better than the regular coefficient of determination of the LS fit of $R^2_{LS} = 0.46$.
Finally notice that in this simple example it was easy to identify an outlier just by plotting the data. However, outlier identification becomes more difficult when the number of predictors increase and cannot be performed just by plotting the data. In such a case one should proceed as guided in Rousseeuw & Leroy (1987), for example.

Summary and Conclusion

The paper introducing the LMS Rousseeuw & Leroy (1987) is considered as a seminal paper in the development of robust statistics and is, for example, appreciated in the Breakthroughs in Statistics series (Kotz & Johnson (1997)). LMS demonstrates that high breakdown regression can be performed with outliers present in response and predictors. However, due to it is low efficiency and the development of better robust regression methods such as MM-regression (Yohai

1987), LMS is nowadays mainly used for exploratory data analysis and as a initial estimate for more sophisticated methods.

Cross References

Ordinary Least Squares, Iterative Weighted Least Squares, Regression
*

Acknowledgements

References

Donoho, David L. & Peter J. Huber. 1983. The notion of breakdown point. In Peter J. Bickel, Kjell A. Doksum & Joseph L. Hodges (eds.), *A Festschrift for Erich L. Lehmann*, 157–184. Wadsworth, Belmont, CA.

Hettmansperger, Thomas P. & Simon J. Sheather. 1992. A cautionary note on the method of least median squares. *The American Statistician* 46(2). 79–83.

Huber, Peter J. 1973. Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics* 1(5). 799–821. DOI: 10.1214/aos/1176342503.

Joss, Johann & Alfio Marazzi. 1990. Probabilistic algorithms for least median of squares regression. *Computational Statistics & Data Analysis* 9(1). 123–133. DOI: https://doi.org/10.1016/0167-9473(90)90075-S.

Kotz, S. & N. L. Johnson. 1997. *Breakthroughs in Statistics Volume 3*. New York: Springer.

Maronna, Ricardo A., R. Douglas Martin, Victor J. Yohai & Matias Salibian-Barrera. 2018. *Robust Statistics: Theory and Methods (with R), 2nd Edition*. Wiley.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Rousseeuw, Peter J. 1984. Least median of squares regression. *Journal of the American Statistical Association* 79(388). 871–880. DOI: 10.1080/01621459.1984.10477105.

Rousseeuw, Peter J. & Annick Leroy. 1987. *Robust Regression and Outlier Detection*. Wiley. 1–335.

Smith, Raymond E., N A. Campbell & R Litchfield. 1984. Multivariate statistical techniques applied to pisolitic laterite geochemistry at golden grove, western australia. *Journal of Geochemical Exploration* 22(1). 193–216. DOI: https://doi.org/10.1016/0375-6742(84)90012-8.

Souvaine, Diane L. & J. Michael Steele. 1987. Time- and space-efficient algorithms for least median of squares regression. *Journal of the American Statistical Association* 82(399). 794–801.

Steele, J.M. & W.L. Steiger. 1986. Algorithms and complexity for least median of squares regression. *Discrete Applied Mathematics* 14(1). 93–100. DOI: https://doi.org/10.1016/0166-218X(86)90009-0.

Stigler, Stephen M. 1981. Gauss and the invention of least squares. *The Annals of Statistics* 9(3). 465–474. DOI: 10.1214/aos/1176345451.

Venables, W. N. & B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth. New York: Springer.

Yohai, Victor J. 1987. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics* 15(2). 642–656. DOI: 10.1214/aos/1176350366.