LEAST WEIGHTED ABSOLUTE VALUE ESTIMATOR WITH AN APPLICATION TO INVESTMENT DATA

Petra Vidnerová – Jan Kalina

Abstract

Robustness with respect to outliers in economic data belongs to one of key requirements expected from reliable statistical estimators. In the linear regression model, the least trimmed squares estimator has become a popular tool in econometric applications. Other highly robust estimators with a high breakdown point are also available, which however still far from being well known in the econometric community. These include the least trimmed absolute value estimator or least weighted squares estimator, which have been presented mainly on simulated datasets so far. In this paper, we propose a novel weighted version of the least trimmed absolute value estimator, which is denoted as the least weighted absolute value estimator. We study the performance of the novel estimator over an investment dataset. The computations include estimating the covariance matrix of the novel estimator by means of nonparametric bootstrap, or investigating computational aspects of a corresponding approximate algorithm. The novel least weighted absolute value estimator is more flexible compared to the least trimmed absolute value estimator. It is interesting to see that its performance over the dataset is similar to that of the least weighted squares, while statistical properties of the latter estimator are known to be very appealing.

Key words: robust regression, regression median, implicit weighting, computational aspects, nonparametric bootstrap

JEL Code: C21, C14

Introduction

While linear regression represents the most fundamental model in current econometrics, the least squares (LS) estimator of its parameters is notoriously known to be vulnerable to the presence of outlying measurements (outliers) in the data. The class of M-estimators, thoroughly investigated since the groundbreaking work by Huber in 1960s, belongs to the classical robust estimation methodology (Jurečková et al., 2019). M-estimators are nevertheless not robust with respect to leverage points, which are defined as values outlying on the horizontal axis (i.e.

outlying in one or more regressors). The least trimmed squares estimator seems therefore a more suitable highly robust method, i.e. with a high breakdown point (Rousseeuw & Leroy, 1987). Its version with weights implicitly assigned to individual observations, denoted as the least weighted squares estimator, was proposed and investigated in Víšek (2011). A trimmed estimator based on the L_1-norm is available as the least trimmed absolute value estimator (Hawkins & Olive, 1999), which has not however acquired attention of practical econometricians. Moreover, to the best of our knowledge, its version with weights implicitly assigned to individual observations seems to be still lacking.

Section 1 of this paper presents available implicitly weighted robust regression estimators and proposes a novel implicitly weighted estimator based on the L_1-norm, denoted as the least weighted absolute value estimator. Section 2 describes an illustration of the novel estimator on a real economic dataset of U.S. investments, together with a comparison of other (possibly) highly robust estimates.

1 Implicitly weighted robust regression

This section recalls several available highly robust estimators of parameters in linear regression and proposes a novel implicitly weighted robust estimator, denoted as the least weighted absolute value estimator. We consider the standard linear regression model

$$Y_i = \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \qquad i = 1, \ldots, n, \tag{1}$$

which may be expressed in the matrix notation as $Y = X\beta + e$. Here, $\beta = (\beta_1, \ldots, \beta_p)^T$ is the vector of parameters and the i-th row of X will be denoted as $X_i = (X_{i1}, \ldots, X_{ip})^T$ for $i = 1, \ldots, n$. We assume that there exists a (positive) common variance σ^2 of the errors $e_1, \ldots, e_n$.

The least trimmed squares (LTS) estimator of β was proposed in Rousseeuw & Leroy (1987). If we consider a fixed estimate $b = (b_1, \ldots, b_p)^T \in \mathbb{R}^p$, it will be useful to denote the residual corresponding to the i-th observation as

$$u_i(b) = Y_i - b_1 X_{i1} - \cdots - b_p X_{ip}, \quad i = 1, \ldots, n. \tag{2}$$

Ordering squared residuals as

$$u_{(1)}^2(b) \le u_{(2)}^2(b) \le \cdots \le u_{(n)}^2(b), \tag{3}$$

we arrive at the definition of the LTS estimator in the form

$$\arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^{h} u_{(i)}^2(b), \tag{4}$$

where the user must select a suitable trimming constant h fulfilling $n/2 \leq h < n$. Víšek (2006) proved that the LTS estimator may attain a high robustness (under contaminated data) but cannot achieve a high efficiency (under non-contaminated data).

The least weighted squares (LWS) regression represents a generalization of the LTS (Víšek, 2011) with the ability to combine high robustness with high efficiency. Its idea is to downweight less reliable data points by a set of continuous weights; if the outliers get zero weights, then the estimator may attain a high breakdown point. The magnitudes of non-negative weights $w_1, \dots, w_n$ must be chosen by the user, while the weights are assigned to particular observations after a permutation, which is determined automatically only during the computation based on the residuals. The LWS estimator of β is defined as the argument of minimum of

$$\arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^{n} w_i u_{(i)}^2(b). \quad (5)$$

Clearly, the least trimmed squares (LTS) estimator proposed in Rousseeuw & Leroy (1987) represents a special case of the LWS with weights equal to zero or one only.

The least trimmed absolute values (LTA) estimator is defined by means of

$$\arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^{h} |u(b)|_{(i),} \quad (6)$$

where

$$|u(b)|_{(1)} \leq |u(b)|_{(2)} \leq \cdots \leq |u(b)|_{(n)}, \quad (7)$$

and represents a trimmed version of the regression median (L_1-estimator). Rusiecki (2013) presented the LTA estimator as a novelty in the context of robust neural networks (see e.g. Kalina & Vidnerová, 2019), although it has been described before (Wilcox, 2012). This documents how the LTA estimator remains very little known. In Wilcox (2017), a comparison of the LTA and LTS is claimed to be present in a recent paper, where the topic is not mentioned at all. Still, the LTA estimator is stated again in Wilcox (2017) to have a much smaller standard error (at least in common situations) compared to the LTS, but the improvement over the LTS is to be marginal at best.

The LTA estimator is not sufficiently discussed in the most fundamental monographs on robust estimation (Jurečková et al., 2019) and we are also not aware of systematic numerical comparison of the LTA estimator with other robust estimates. Implicitly weighted estimators (Čížek, 2013; Kalina, 2014) were proposed and investigated only later than the LTA. Because there is a recent promising experience with the LWS (see e.g. Kalina & Tichavský (2020)), it is now natural to generalize the LTA estimator by means of implicit weighting.

We now propose a novel estimator denoted as the least weighted absolute value (LWA) estimator defined by

$$\arg\min_{b\in\mathbb{R}^p} \sum_{i=1}^{n} w_i |u(b)|_{(i)}. \tag{8}$$

Basically, it represents an implicitly weighted regression median. Alternatively, we may use the concept of weight functions according to Víšek (2011). Using a fixed weight function $\psi: [0,1] \to [0,1]$, which is non-increasing and continuous on $[0,1]$, we may express the LWA estimator in an equivalent way as

$$\arg\min_{b\in\mathbb{R}^p} \sum_{i=1}^{n} \psi\left(\frac{i-1/2}{n}\right) |u(b)|_{(i)}. \tag{9}$$

The computation of the robust estimators of above (LTS, LWS, LTA, LWA), which may all be perceived as implicitly weighted estimators, is intensive and an approximate algorithm must be used already for relatively small sample sizes. An approximate algorithm based on repeated randomly chosen subsets of p observations (out of all n observations) is known for the LTS. This algorithm denoted as FAST-LTS is well-established (Rousseeuw & van Driessen, 2006) and implemented in the robustbase library of R software. We point out that the function ltsReg computes a two-stage reweighted version known as the reweighted least squares (RLS) (Rousseeuw & Leroy, 1987), which has never been theoretically investigated; in our computations here, we do not use reweighted versions of the estimators. We use a straightforward modification of the FAST-LTS algorithm to compute (estimate) also the LWA estimator. This algorithm requires the user to choose the magnitudes of the weights $w_1, \ldots, w_n$ assigned to individual observations after that permutation, which minimizes (8). We use a modification of the FAST-LTS algorithm also for the LWS and LTA, while the computation of the latter estimator was investigated already in Hawkins & Olive (1999).

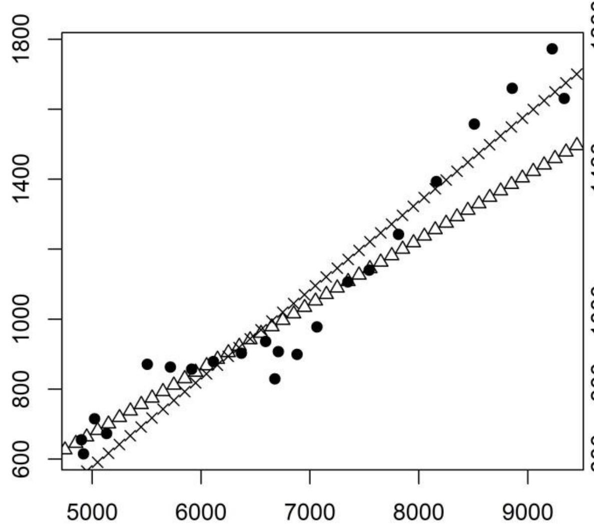
2 Numerical analysis of investment data

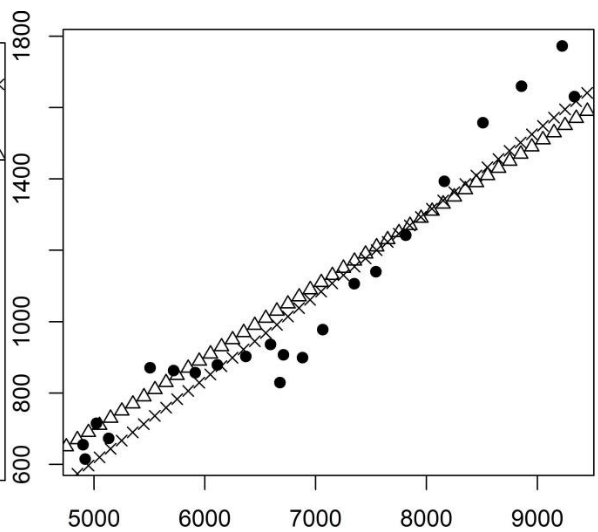
We analyze a real economic dataset of U.S. investments by means of several regression estimators to compare their performance and to study computational aspects of the novel LWA estimator. This investment dataset was analyzed in Kalina (2015) by standard regression tools. Here, we consider a regression of $n = 22$ yearly values of real gross private domestic investments in the United States in 10^9 USD against the gross domestic product. We use here R software for all the computations.

Throughout this section, we choose $h = \lfloor 3n/4 \rfloor$ for the LTS and LTA estimators, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$. For the LWS and LWA estimators, we always use trimmed linear weights (Kalina & Tichavský, 2020), allowing to combine high efficiency and high robustness. While the algorithm for each of the robust estimators, obtained by adapting the FAST-LTS algorithm, is based on repeated randomly chosen *p*-subsets of the data, we always use 10 000 of these repetitions.

Figure 1 shows the results of four robust regression estimators computed for the investment dataset, namely the LTS and LWS estimators (left image) and the LTA and LWA (right). Table 1 presents estimates of the intercept b_0 and slope b_1 for various regression estimators. Moreover, the mean square error (MSE) of each estimator evaluated in a leave-one-out cross validation study is presented in the table. Both the LTS and LTA estimators consider the observations in the years 1998, 1999 and 2000 to be the most severe outliers, which is true also for the LWS and LWA. Point estimates, as well as values of MSE, obtained by the LWA are rather similar to those of the LWS. Point estimates of the LWS and LWA are (in spite of reducing the influence of several observations) also similar to those of the least squares.

Fig. 1: Results of four robust regression estimators in the investment dataset of Section 2. Left: results of the LTS fit (triangles) and LWS fit (crosses). Right: results of the LTA fit (triangles) and LWA fit (crosses).





Source: own computation

2.1 Nonparametric bootstrap

The aim of this section is to apply bootstrap (bootstrapping, resampling with replacement) estimation techniques to estimate the variability of various implicitly weighted robust regression estimators, i.e. elements of the covariance matrix of the estimates. Let us first recall basic general principles of bootstrap estimation, which has acquired a big popularity in various statistical tasks. Incorporating the basic principles of bootstrapping, one may develop a great variety of resampling techniques that provide us with alternative possibilities of estimating such data characteristics, which would be infeasible by more standard reasoning, e.g. asymptotic results depending on the unknown distribution of regression errors in (1). The range of bootstrap methods is rather large, including residual bootstrap, nonparametric bootstrap, semiparametric bootstrap, Bayesian bootstrap etc. Also the terminology is not used in a unique way. Bootstrap is commonly used also in the task to estimate variability (i.e. the covariance matrix) of regression estimators. Theoretical foundations of bootstrap can be found e.g. in Godfrey (2009) and the references cited therein.

In this paper, we focus our attention to nonparametric bootstrap, which is (especially in our model) conceptually simple and can be computed for real data in a straightforward (although rather computationally demanding) way. We use a standard nonparametric bootstrap approach and apply it to several robust regression estimators, including the novel LWA estimator, to assess the covariance matrix of every estimator in the example with investment data. In order to have comparable results, we compute the bootstrap estimates also for the least squares, although an explicit formula $var\ b_{LS} = \sigma^2(X^TX)^{-1}$ for the least squares estimate b_{LS} is available. The results are presented again in Table 1. Particularly, the standard deviation of all point estimates of the intercept is denoted as s_0 there and for the slope as s_1; the covariance between the slope and the intercept is denoted as s_{01}.

Bootstrap results reveal the results of the LWA to be similar to those of the LWS also by means of their variability. The computations over the investment dataset reveal that the smallest variance is obtained with the least squares estimator. This is natural, as the latter is based on minimizing the residual variance, but the LWA and LWS seem to stay behind only mildly. On the other hand, the loss of efficiency of the LTS and LTA is remarkable. Just like the LWS is superior to the LTS from the point of view of efficiency for non-contaminated samples, as theoretically proven in Čížek (2013), so the LWA seems to overcome a major disadvantage of the LTA, namely the low efficiency. The nonparametric bootstrap estimation allows us to perceive the LWA as promising, based however only on this empirical result, obtained for a simplistic data set with a single regressor.

Tab. 1: Results of the analysis of the investment dataset of Section 2. The classical and robust estimates of the intercept and slope are accompanied by nonparametric bootstrap estimates of standard deviances (s_0 and s_1) and covariances (s_{01}). MSE denotes the mean square error evaluated within a leave-one-out cross validation.

Estimator	Intercept	Slope	s_0	s_1	s_{01}	MSE
LS	−582	0.239	108.9	0.016	−1.67	10 948
LTS	−375	0.207	742.0	0.106	−5.74	16 489
LWS	−601	0.242	207.2	0.031	−2.40	12 033
LTA	−312	0.204	721.6	0.112	−5.58	16 207
LWA	−551	0.232	224.8	0.030	−2.49	12 251

Source: own computation

Tab. 2: Values of five different loss functions computed for five estimators over the investment dataset. This study of Section 2.2 reveals the tightness of the algorithms for computing the individual robust regression estimators.

	Loss function				
Estimator	$\sum_{i=1}^{n} u_i^2$	$\sum_{i=1}^{h} u_{(i)}^2$	$\sum_{i=1}^{n} w_i u_{(i)}^2$	$\sum_{i=1}^{h} \lvert u\rvert_{(i)}$	$\sum_{i=1}^{h} w_i \lvert u\rvert_{(i)}$
LS	198 796	80 834	4225	995	51.7
LTS	245 484	61 298	4019	835	45.4
LWS	223 132	63 661	3914	844	45.0
LTA	247 037	62 597	4004	791	46.2
LWA	220 925	64 076	3985	826	41.3

Source: own computation

The number of bootstrap repetitions within Algorithm 1 was always chosen as 100, which seems more than sufficient for the asymptotics. We additionally performed the computations for 10 000 bootstrap samples and compared the results, which turn out to be were very close to those obtained for 100 samples. This is in accordance with our experience and a small number of bootstrap samples indeed seems to be sufficient if a single constant is estimated rather than the whole empirical distribution.

2.2 Computational aspects

We now investigate some computational aspects of the novel LWA estimator on the investment dataset. The computations of this section bring arguments in favor of using our approximate algorithm for the LWA estimator.

Table 2 presents values of five loss functions, which correspond to the LS, LTS, LWS, LTA, and LWA, respectively. The loss functions are evaluated over the investment dataset (i.e. no cross validation is desirable here) and our choices of *h* and the weights are the same as above. Here, in contrary to previous computations, it is necessary to assume $\sum_{i=1}^{n} w_i = 1$ for the LWS and LWA.

Let us now comment the results presented in Table 2. The least squares estimator minimizes $\sum_{i=1}^{n} u_i^2$ as expected and thus can be expected to yield also a rather small value of $\sum_{i=1}^{n} w_i \left|u_{(i)}\right|$. The LWA estimator has a much larger value of $\sum_{i=1}^{n} u_i^2$ compared to the LS fit. However, the algorithm used for computing the LWA has found even a much smaller value of $\sum_{i=1}^{n} w_i \left|u_{(i)}\right|$ than the LS. Analogous observations can be formulated for comparing the loss of the LWS and LWA, while each of the estimators minimizes its own specific loss function. On the whole, the results of Table 2 thus give a clear evidence in favor of the reliability of the algorithm for computing the novel LWA estimator. Such reasoning is analogous to the argument that the approximate algorithm for estimating the LWS yields a tight approximation (in a certain sense) to the precise value of the LWS; see Víšek (2011) for discussion.

Further, we discuss a suitable number of iterations within the computation of the LWA estimator. In the linear regression model (1), the minimal loss (5) obtained for the LWA regression with linear weights equals 3910.3. Figure 2 (left) shows the histogram of individual results of 1000 independent repetitions of the iterative computation; the LWA estimate is then the result corresponding to the minimal value of the loss. About 20 % of the resulting LWA estimates are highly influenced by outliers, so the loss is similar to that of the least squares, and the remaining about 80 % of the LWA estimates data have a remarkably smaller loss. Therefore, 80 % of individual repetitions are already remarkably more robust than the least squares fit.

Finally, we study the number of repetitions needed to obtain the minimal loss of the LWA estimator. The result of this study is shown in Figure 2 (right) for 1000 independent repetitions. In the mean there are 431 repetitions needed. The number of repetitions is below 1000 for 85.7 % of cases. Therefore, using 1000 repetitions is sufficiently safe to yield a reliable approximation to the true solution of the LWA estimate here.

Conclusions

While the LTS estimator belongs to popular robust regression estimators with many applications, its implicitly weighted analogue with continuous weights, i.e. the LWS estimator, has acquired much less attention in the analysis of econometric data. Nevertheless, available numerical studies reveal its ability to outperform the LTS estimator or even MM-estimators, which are currently considered to be the most successful robust estimators. Therefore, we extend the idea of implicit weighting to the context of the LTA estimator. The new estimator, denoted as the LWA estimator, is more flexible compared to the LTA, allowing also different weights than only zeros or ones. The LWA estimator can be interpreted as an implicitly weighted regression median (i.e. L_1-estimator). We use here a (simple) investment dataset to study the performance of the LWA estimator and to investigate computational aspects of an approximate algorithm for its computation. The LWA estimator performs similarly to the LWS on this dataset, and particularly gives a better fit compared to the LTS and LTA estimators.

As future work, additional computations are needed to investigate the application potential of the LWA (including data with a larger *p*, comparisons based on robust versions of MSE, or comparisons with other regression estimators). Theoretical properties of the novel estimator (i.e. asymptotic efficiency or robustness) remain unknown as well. In addition, we plan to develop a general approach to specifying suitable weights for a particular dataset, parametrizing the weight function by a small number of parameters and optimizing them.

Acknowledgments

This work was supported by the Czech Science Foundation grants GA18-23827S (P. Vidnerová) and GA19-05704S (J. Kalina).

References

Čížek, P. (2013). Reweighted least trimmed squares: An alternative to one-step estimators. *Test*, 22, 514-533.

Godfrey, L. (2009). *Bootstrap tests for regression models*. London: Palgrave Macmillan.

Hawkins, D.M. & Olive, D.J. (1999). Improved feasible solution algorithms for high breakdown estimation. *Computational Statistics and Data Analysis*, 30, 1-11.

Jurečková, J., Picek, J., & Schindler, M. (2019). *Robust statistical methods with R*. 2nd edn. Boca Raton: CRC Press.

Kalina, J. (2014). On robust information extraction from high-dimensional data. *Serbian Journal of Management*, 9(1), 131-144.

Kalina, J. (2015). Three contributions to robust regression diagnostics. *Journal of Applied Mathematics, Statistics and Informatics*, 11(2), 69-78.

Kalina, J. & Tichavský, J. (2020). On robust estimation of error variance in (highly) robust regression. *Measurement Science Review*, 20, 6-14.

Kalina, J. & Vidnerová, P. (2019). Robust training of radial basis function neural networks. *Lecture Notes in Artificial Intelligence*, 11508, 113-124.

Rousseeuw, P.J. & Leroy, A.M. (1987). *Robust regression and outlier detection.* New York: Wiley.

Rousseeuw, P.J. & van Driessen, K. (2006). Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery,* 12, 29-45.

Rusiecki, A. (2013). Robust learning algorithm based on LTA estimator. *Neurocomputing,* 120, 624-632.

Víšek, J.Á. (2006). The least trimmed squares, Part III: Asymptotic normality. *Kybernetika,* 42, 203-224.

Víšek, J.Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*, 47, 179-206.

Wilcox, R. (2012). *Modern statistics for the social and behavioral sciences. A practical introduction*. Boca Raton: CRC Press.

Wilcox, R.R. (2017). *Introduction to robust estimation and hypothesis testing*. 4th edn. London: Academic Press.

Contact

Petra Vidnerová
The Czech Academy of Sciences, Institute of Computer Science
Pod Vodárenskou věží 2, 182 07, Praha 8, Czech Republic
petra@cs.cas.cz

Jan Kalina
The Czech Academy of Sciences, Institute of Computer Science
Pod Vodárenskou věží 2, 182 07, Praha 8, Czech Republic
kalina@cs.cas.cz