High Breakdown Weighted LAD Regression Estimator

Hassan S. Uraibi
Hatem Abd A. Alwan
Al-Qadisiyah University - College of Administration and Economics
Corresponding Author: Hatem Abd A. Alwan

***Abstract* :** The least Absolute Deviation estimator has a high breakdown point probably reach 50%, so it provides a good alternative to the LS estimator when vertical outliers are present in the data set. However, It may lose this feature when the design matrix x is having at least one leverage point. Many authors have been efforts in the literature to assign a down weight to leverage point and suggested the weighted LAD regression which is denoted as WLAD for increasing the breakdown point of LAD. Most weighted functions that are discussed in the literature are based on robust Mahalanobis distance which is a familiar approach to identifying leverage points. Unfortunately, this robust distance could be affected in appearing high leverage points or when the masking and swamping phenomena have happened. Consequently, robust Mahalanobis distance may not detect all leverage points, and then WLAD would be a non-robust method, and its estimator surely will break down. In this paper, we proposed improving Mahalanobis's distance based on weighted fast and consistent high breakdown estimator location and scale matrix, which may make WLAD is more robust than the previous ones. The simulation studies have been done and the results of our proposed UWLAD is compared with AWLAD and GWLAD methods which well-known in the literature. The result shows that the performance of the UWLAD method is more efficient and reliable than others.
Keywords: Weighted LAD, masking and swamping, robust location and scale estimators

INTRODUCTION: The Least Squares (LS) estimators are considered the best linear unbiased estimators when its assumptions are met. When the distribution of random errors is heavy-tailed and the data set is having some outliers, the regression model does not comply with the normality assumption which assumed that the distribution parameters are zero mean and constant variance. Due to that, the estimated regression coefficients would radically be influenced and we will get misleading results. Therefore, robust regression methods are recommended to overcome such problems, such as M-estimate (Huber and Ronchetti, 1981), LMS (Rousseeuw, 1984), LTS (Rousseeuw and Leroy, 1987) and Least Absolute Deviation (Huber, 1987).

For a long time, the Least Absolute of Deviation (LAD) method is well-studied in the statistical literature as a robust method that relies on the sample median when the distribution of errors is heavy-tailed. Unfortunately, LAD regression has a breakdown point of (1/n) when the leverage points are present in the data set (Croux et al., 2003). The classical Mahalanobis Distance (MD) measure was the most familiar choice at that time to identify the leverage points. Rousseeuw and Zomeren(1990) said that MD is suffering from masking phenomena effects, so it non robust measure. Some practitioners of statistics were believed the best alternative choice is hat matrix (Hoagline and Welsch, 1978) while they were not aware it having proportional relation with the classical squared Mahalanobis distance (Chatterjee and Hadi, 2015), so it is non robust measure too.

Ellis and Morgenthaler (1992) pointed out that down-weighting the leverage point can improve the conditional breakdown point of the LAD regression. Hubert and Rousseeuw (1997),Giloni et al. (2006a), Giloni et al. (2006b), and Arslan (2012) suggested weighted LAD regression based on the values of robust Mahalanobis distances to increase the value of the breakdown point. The crucial point here is that the leverage points in multiple regression data is very hard to identify when the number of covariate variables exceeds 2. However, respectful efforts have done in the last decade of the twentieth century to remedy the leverage points problem. Most of these efforts concentrated on improving the performance of Mahalanobis distance by replacing classical mean and variance with robust estimators, see (Stahel,1981; Donoho 1982; Rousseeuw and Zomeren,1990; and Yohai and Maronna, 1976). With regret, Alhussieny (2021) and Midi et al (2020) showed that Robust MD is sensitive to the presence of high leverage points that may be a source to masking and swamping cases, so cannot be used to down weights of leverage points. Uraibi and Midi (2015, 2017, 2019), Midi et al (2020), and Alhussieny (2021) incorporated fast and consistent location and scatter matrices with RMD such as RFCH and RMVN that introduced by Olive and Hawkins (2010) to reduce the effect of masking and swamping phenomena.

2. WLAD based on Robust Location and Scale Estimators

Consider the linear regression model

$$y_i = \mathbf{x}_i^{\mathrm{T}} \beta + \varepsilon_i, \qquad i = 1,2,\dots,n \qquad (1)$$

Where, y_i is a response, $\mathbf{x}_i$ is the p dimensional of independent variables which may include intercept, β is a p −vector of unknown regression coefficients, ε_i is the vector of iid random errors with median equals to zero. The $\hat{\beta}^{LAD}$ estimator minimizes the sum of the absolute values of the residuals as follows,

$$\hat{\beta}^{LAD} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left| y_i - \mathbf{x}_i^{\mathrm{T}} \beta \right| \tag{2}$$

The $\hat{\beta}^{LAD}$ in linear regression setting is not sensitive to the presence of outliers or heavy-tailed errors, but it is non-robust when the independent variables $\mathbf{x}$ having leverage point. To relieve the effects of leverage points on $\hat{\beta}^{LAD}$; Giloni *et al.* (2006a) proposed a $\hat{\beta}^{GWLAD}$ estimator,

$$\hat{\beta}^{GWLAD} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \left| y_i - \mathbf{x}_i^{\mathrm{T}} \beta \right| \tag{3}$$

where the weight w_i is inversely proportional to the MD_i^2 from the clean subset such that

$$w_{G,i} = \min\left\{1, \frac{\chi^2_{(0.05,p)}}{MD_i^2}\right\} \tag{4}$$

However, a large MD_i^2 value is not necessarily to be a leverage point (Masking case) and the contrary is correct (Swamping case) due to the influence of outliers on classical mean and variance (Hadi, 1992). Arslan (2012) replaced p instead of $\chi^2_{p,0.05}$ and used Robust measure of MD_i^2 which denoted as (RMD_i^2) to assign down weights to leverage points,

$$w_{A,i} = \min\left\{1, \frac{p}{\sqrt{MD_i^2}}\right\} \tag{5}$$

where, RMD_i^2 are computed based on the robust estimators of location and scale matrix of Fast MCD (Rousseeuw and Driessen,1999) which the frequently used in the literature. In spite of that the $\chi^2_{p,0.975}$ critical value is the most familiar threshold that has been used to recognize the leverage points with RMD_i^2, but it has some shortcomings. For instance, the distribution of RMD_i^2 may not a chi-square, thus defining the exact threshold becomes a difficult matter. Moreover, the performance of it would be not well behaved with moderate sample sizes or the number of dimensions increases, see (Becker and Gather ;1999, Hardin and Rocke ;2005, Cerioli et al.; 2009 and Cerioli et al.; 2008). Alhussieny (2021) has been shown that the RMD_i^2 failed to identify all leverage points when single high leverage point is present in the data and it may be a source to masking and swamping cases. Consequently, RMD_i^2 is not a sufficient measure to detect and assigning low weights for all leverage points. In other words, unless improving the performance of RMD_i^2, would not suitable to use it as a weight function for LAD regression. Uraibi (2019) incorporated the Reweighted fast consistent and high breakdown RFCH estimators (Olive and Hawkins;2010) to compute RMD_i^2 vector as follows,

$$RMD^2_{RFCH} = (x - \hat{\mu}_{RFCH})^T\, \hat{\Sigma}^{-1}_{RFCH}\, (x - \hat{\mu}_{RFCH}) \tag{6}$$

Assume that the joint distribution of every vector of x is normal, and then (0.975) tolerance ellipsoid can be constructed from $S = \sum_i RMD^2_{RFCH}(x_i) : RMD^2_{RFCH}(x_i) \leq \chi^2_{(0.975,p)}\}$ and let $\delta = \frac{(0.975\times n)}{(2\times S)}$ the upper tail of chai-square critical value can be computed through $q = \min(\delta, 0.995)$ with p degrees of freedom. Finally, the weight function can be written as follows,

$$\mathrm{w}_{U,i} = \min\left\{1, \frac{\chi^2_{(q,p)}}{RMD^2_{RFCH,i}}\right\} \tag{7}$$

we adopt (4,5,6) weight functions with LAD,

$$\hat{\beta}^{UWLAD} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathrm{w}_{U,i} \left| y_i - \mathbf{x}_i^{\mathrm{T}} \beta \right| \tag{8}$$

and we consider

$$\hat{\beta}^{GWLAD} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathrm{w}_{G,i} \left| y_i - \mathbf{x}_i^{\mathrm{T}} \beta \right| \tag{9}$$

$$\hat{\beta}^{AWLAD} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathrm{w}_{A,i} \left| y_i - \mathbf{x}_i^{\mathrm{T}} \beta \right| \tag{10}$$

where LAD is weighted with (4,5) and (7) functions respectively.

SIMULATION

The simulation studies have been carried out to know the performance robustness of UWLAD algorithm compared with LAD, AWLAD, GWLAD. The design matrix $X_{n\times p}$ of predictors is generated randomly from multivariate normal distribution with zero means and $\rho^{|i-j|}$ variance and covariance matrix, $\rho = 0.20$. The maximum value of X_1 times by 100 to create HLP and $m = \alpha \times n$ observations of X_4 times by 100 to create another leverage points, where α the percentage of outlying observation. The first m of random errors $\varepsilon_{m\times 1}$ vector are generated from t distribution with 2 degree of freedom and the remaining $\varepsilon_{(n-m)\times 1}$ are generated from random normal distribution $N(0{,}2)$. We assume that there is a relationship between X_3 and $\varepsilon_{n\times 1}$ such that $h(X_3) = 1 + 0.4\, X_3$ to create the heteroscedasticity problem. When the population of regression parameters which denoted as $\beta_{p\times 1}$ is known, the response variable $y_{n\times 1}$ can be computed as follows,

$$y = X_{n\times p} \quad \beta_{p\times 1} + \varepsilon_{n\times 1} \tag{11}$$

This simulation scenario has been considered two times when α equals to 0.05 with different size of $n = \{50{,}100{,}300\}$ respectively. The average of coefficients, bias, mean square error, $\{\hat{\beta}, Bias, SD[\hat{\beta}], MSE[\hat{\beta}]\}$ are computed from 3000 generated datasets for each scenario where the dimension $p = 7$ (without intercept). The $\sum \beta_{7\times 1} = 7$ would be the criterion for sum of estimated coefficients of each algorithm, the closet one is the best. The of $\sum Bias$ surely would be monotonic with the difference between $\left|\sum \hat{\beta}_{7\times 1} - \sum \beta_{7\times 1}\right|$ of that algorithm with minimum $SD[\hat{\beta}]$ and average of $MSE[\hat{\beta}]$. The algorithm which possesses the lowest values of these criteria should the

Table 1: Coefficient, Bias, SD, and Ave.*MSE* estimates of LAD,AWLAD,GWLAD and UWLAD of (3000) simulation data sets when $n = 50$ and $\alpha = 0.05$ outlier and LP.

Estimate	LAD	AWLAD	GWLAD	UWLAD
$\sum \hat{\beta}_{7\times 1}$	88.955	7.556	7.557	7.563
$\left\lvert\sum \hat{\beta}_{7\times 1} - \sum \beta_{7\times 1}\right\rvert$	81.955	0.556	0.557	0.563
$SD[\hat{\beta}]$	13.489	0.302	0.303	0.285
Ave. $MSE[\hat{\beta}]$	14334.958	0.333	0.336	0.274

Table 2: Coefficient, Bias, SD, and Ave.*MSE* estimates of LAD, AWLAD,GWLAD and UWLAD of (3000) simulation data sets when $n = 100$ and $\alpha = 0.05$ outlier and LP.

Estimate	LAD	AWLAD	GWLAD	UWLAD
$\sum \hat{\beta}_{7\times 1}$	88.482	7.444	7.444	7.460
$\left\lvert\sum \hat{\beta}_{7\times 1} - \sum \beta_{7\times 1}\right\rvert$	81.482	0.444	0.444	0.460
$SD[\hat{\beta}]$	9.156	0.165	0.166	0.149
Ave. $MSE[\hat{\beta}]$	12154.954	0.144	0.145	0.114

Table 1,2 and 3 present the simulation estimates of LAD, AWLAD, GWLAD and UWLAD when the data are contaminated by 0.05 outliers in the residuals and 0.05 LP in the design matrix with single HLP, where $n = \{50{,}100{,}300\}$, respectively. It is obvious that the $\sum \beta_{7\times 1}$ of LAD is much affected by LP and HLP, as we note it values in three tables $(88.955{,}88.482{,}85.119)$ are far from the true value (7). Note the absolute values of $\left|\sum\hat{\beta}_{7\times1}-\sum\beta_{7\times1}\right|$ equivalent to $(81.955,81.482,78.119)$ to confirm that LAD is non-robust method when LP and HLP are present in the data. Consequently, this performance of LAD impacts on $SD[\hat{\beta}]$ and Ave. $MSE[\hat{\beta}]$ Which are the highest values than others.

Table 3: Coefficient, Bias, SD, and Ave.*MSE* estimates of LAD,AWLAD,GWLAD and UWLAD of (3000) simulation data sets when $n=300$ and $\alpha=0.05$ outlier and LP.

Estimate	LAD	AWLAD	GWLAD	UWLAD
$\sum\hat{\beta}_{7\times1}$	85.119	7.423	7.421	7.422
$\left\lvert\sum\hat{\beta}_{7\times1}-\sum\beta_{7\times1}\right\rvert$	78.119	0.423	0.421	0.422
$SD[\hat{\beta}]$	5.177	0.070	0.069	0.062
Ave. $MSE[\hat{\beta}]$	10725.837	0.056	0.056	0.047

The performances of AWLAD and GWLAD methods are very close to each other, but both are more robust than LAD method. Tables 1,2,3 are displayed that the $\sum\hat{\beta}_{7\times1}$ of AWLAD and GWLAD methods are better than UWLAD and that is very clear if we see their bias in the three tables. On the other hand, it is notable that the values $SD[\hat{\beta}]$ and Ave. $MSE[\hat{\beta}]$ of UWLAD are lowest than AWLAD and GWLAD values. Therefore, UWLAD method is more efficient and robust than AWLAD and GWLAD.

Conclusion

From the research that has been conducted, it is possible to conclude that LAD is highly sensitive to the presence of LP. Both weighted functions which were used AWLAD, GWLAD were not sufficient to deal with high leverage points. That is due to their cut-off points were not optimal to identify all leverage points that were masking and swamping cases existence in the dataset. So, the WLAD method that is based on these weighted functions is perhaps unreliable, particularly with high leverage points. Consequently. This paper suggests UWLAD as a robust bootstrap method to mitigate the effect of LP and HLP. In spite of AWLAD, GWLAD is robust against outliers and provides $\sum\hat{\beta}_{7\times1}$ very close to the criterion but both are failed to deal with the problems of heterogeneous and HLP. In contrast to RRAWLAD, RRGWLAD, the RRUWLAD method has assigned more precise weights for all leverage points and it is not much affected the heterogeneous problem. Finally, from the findings of the simulation, it can be concluded that UWLAD is quite convincing and encouraged so that it is recommended to use in scientific applications.

Alhussieny, SawsanA.(2021). The Algorithms of Grouped Diagnostics for Multivariate Outliers by Using Location and Scale Matrix with Application. Msc Dissertation (in Arabic). Dept. of Statistics, University of Al Qadisiyah, IRAQ.

Amado, C,& Pires,A.M.(2004) Robust bootstrap with non random weights based on the influence function, Communications in Statistics, Simulation and Computation, 33, 377-396.

Arslan, O. (2012). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. Computational Statistics & Data Analysis, 56(6), 1952-1965.

Atkionson, A. C. (2008). Fitting mixtures of regression lines with the forward search. Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security, 19, 271.

Becker, C., & Gather, U. (1999). The masking breakdown point of multivariate outlier identification rules. Journal of the American Statistical Association, 94(447), 947-955.

Cerioli, A., Riani, M., & Atkinson, A. C. (2009). Controlling the size of multivariate outlier tests with the MCD estimator of scatter. Statistics and Computing, 19(3), 341-353.

Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons

Croux, C., Filzmoser, P., Pison, G., & Rousseeuw, P. J. (2003). Fitting multiplicative models by robust alternating regressions. Statistics and Computing, 13(1), 23-36.

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Technical report, Harvard University, Boston. URL http://www-stat. stanford. edu/~ donoho/Reports/Oldies/BPMLE. pdf.

Ellis, S. P., & Morgenthaler, S. (1992). Leverage and breakdown in L 1 regression. Journal of the American Statistical Association, 87(417), 143-148.
Giloni, A., Sengupta, B., & Simonoff, J. S. (2006). A mathematical programming approach for improving the robustness of least sum of absolute deviations regression. Naval Research Logistics (NRL), 53(4), 261-271.
Giloni, A., Simonoff, J. S., & Sengupta, B. (2006). Robust weighted LAD regression. Computational Statistics & Data Analysis, 50(11), 3124-3140.
Hadi, A. S. (1992). A new measure of overall potential influence in linear regression.Computational and Statistical Data Analysis. 14:1-27.
Hardin, J., & Rocke, D. M. (2005). The distribution of robust distances. Journal of Computational and Graphical Statistics, 14(4), 928-946.
Hoaglin, D. C and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. American Statistician. 32:17-22.
Huber, P. (1987). The place of the L1-norm in robust estimation. Statistical Data Analysis Based on the L1-Norm and Related Methods (Y. Dodge, ed.).
Huber, P. J., & Ronchetti, E. M. (1981). Robust statistics, ser. Wiley Series in Probability and Mathematical Statistics. New York, NY, USA, Wiley-IEEE, 52, 54.
Hubert, M., & Rousseeuw, P. J. (1997). Robust regression with both continuous and binary regressors. Journal of Statistical Planning and Inference, 57(1), 153-163.
J. Yohai, V., & A. Maronna, R. (1976). Location estimators based on linear combinations of modified order statistics. Communications in Statistics-Theory and Methods, 5(5), 481-486.
Koenker, R. (2005). Quantile regression.,(Cambridge University Press: Cambridge, UK). Quantile Regression. Cambridge University Press. Cambridge, UK.
Olive, D. J., & Hawkins, D. M. (2010). Robust multivariate location and dispersion. Preprint, see (www. math. siu. edu/olive/preprints. htm).
Rousseeuw P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics. 41:212–223
Rousseeuw, P. and Van Zomeren, B. (1990).Unmasking multivariate outliers and leverage points. Journal of American Statistical Associations. 85: 633-639
Rousseeuw, P. J. (1984). Least median of squares regression. Journal of theAmerican Statistical Association. 79: 871–880.
Rousseeuw, P.J. and Leroy, A.M. (1987). Robust Regression and Outlier Detection. New York: Wiley
Stahel, W. A. (1981). Robust Estimation: Infinitesimal Optimality and Covariance Matrix Estimators. PhD. thesis (in German), ETH, Zurich
Uraibi, H. S. (2019). Weighted Lasso Subsampling for HighDimensional Regression. Electronic Journal of Applied Statistical Analysis, 12(1), 69-84.
Uraibi, H. S., & Midi, H. (2019). ON ROBUST BIVARIATE AND MULTIVARIATE CORRELATION COEFFICIENT. Economic Computation & Economic Cybernetics Studies & Research, 53(2).
Uraibi, H. S., Midi, H., & Rana, S. (2015). Robust stability best subset selection for autocorrelated data based on robust location and dispersion estimator. Journal of Probability and Statistics, 2015.
Uraibi, H. S., Midi, H., & Rana, S. (2017). Robust multivariate least angle regression. J. Scienceasia, 43(1), 56-60.