International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319 – 6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 9; September 2017; Page No. 6150-6155
DOI: http://dx.doi.org/10.24327/ijcar.2017.6155.0883

CASE REPORT

BREAKDOWN PLOT OF SYMMETRY DISTANCE BASED ROBUST ESTIMATORS

Muthukrishnan R and Reka R*

Department of Statistics, Bharathiar University, Coimbatore-641046, Tamil Nadu, India

A R T I C L E I N F O

Article History:

Received 19th June, 2017
Received in revised form 3rd July, 2017 Accepted 18th August, 2017
Published online 28th September, 2017

Key words:

LS - Symmetric Distance - LTSD – MSD- Outliers.

A B S T R A C T

The Least Median Square and Least Trimmed Square are the most popular methods that have a high breakdown point (50%), but when the outliers are clustered, these methods can breakdown at lower percentages of outliers. In this paper, the breakdown property of LMS, LTS and LTSD are investigated with the presence of large percentage of clustered outliers in the data. The concept of symmetry distance (SD) based method is proposed, called the M estimator based symmetry distance (MSD). The superiority of the proposed method has been demonstrated by considering break down property over the LMS, LTS and LTSD methods, when large percentage of clustered outliers and/or a large deviation in the inliers population under real and simulating environment.

INTRODUCTION

Regression analysis is a vital statistical tool usually employed in computer vision for a huge variety of tasks. The least squares method is the traditional and commonly used method of computation in regression analysis. It achieves most favourable results when the error distribution is Gaussian. However, it becomes unreliable if the noise has non-zero mean components and/or if there are outliers in the data. The outliers may be clusters, large measurement errors, or impulse noise corrupting the data. At a transition between two homogeneous regions of the image, samples belong to one region may turn into outliers for fits to the other region.

The breakdown concept is frequently employed to evaluate a regression method. The definition of robustness in this perspective often is focused on the view of the breakdown point. The breakdown point of regression method is the smallest amount of outlier contamination that may strength the value of the estimate outside an arbitrary range. Breakdown point is one of the most significant qualities of the robust estimators. An estimator is said to be more robust if it has larger breakdown point.

In this paper the LMS, LTS and LTSD methods are briefly discussed in their usual form within the robust regression literature. Then the concept of symmetry distance is introduced into model fitting and named as the M estimator based Symmetry Distance (MSD). The left over part of the paper consists of the experimental results which show that the performance of MSD in the context of breakdown property is much better than LMS, LTS and LTSD methods; particularly when there is a large percentage of clustered outlier and standard deviation in the data.

Corresponding author: **Reka R**
Department of Statistics, Bharathiar University, Coimbatore-641046, Tamil Nadu, India

Robust Regression Estimators

Regression analysis is a statistical procedure for studying and modelling the relationship between variables. The most familiar form of regression analysis is the least square method, which achieves optimum results when the data includes normally distributed random errors. The general form of the linear model is as follows:

$$y_i = x_{i1}\theta_1 + ... + x_{ip}\theta_p + e_i\,(i = 1,......n) \quad (1)$$

where, y_i is the response variable and $x_{i1}........., x_{ip}$ are the explanatory variables. The error term e_i is usually assumed to be normally distributed with mean zero and standard deviation σ. The ordinary least square (LS) method estimates $\hat{\theta}$ by

$$\min \sum_{i=1}^{n} r_i^2 \quad (2)$$

where, residual sss. Even though the least square method has a low computational cost and high efficiency, it is tremendously sensitive to outliers. In fact, even one single outlier can influence the result to a large degree. The Least Square method has a breakdown point of 0, because only one single extreme outlier is sufficient to compel the Least Square method to produce arbitrarily large values. In order to reduce the influence of outliers, a number of robust estimators have been developed and the most popular methods are briefly furnished as follows.

Rousseeuw (1984) proposed the least median of squares estimator (LMS), a simple idea of replacing the sum by a median in the least squares. It finds the parameters to be estimated by minimizing the median of squared residuals corresponding to the data points. The least median of squares estimator can be written as

$$\min_{\hat{\theta}} \underset{i}{med}\, r_i^2 \tag{3}$$

The LMS method has excellent global robustness and high breakdown point (i.e., 50%). However, the relative efficiency of the LMS method is poor when Gaussian noise present in the data.

Rousseeuw (1984) proposed the Least Trimmed Squares (LTS) estimator. It mates a complement way for robust estimators of minimizing the objective function is

$$\hat{\theta} = \min_{\theta} \sum_{j=1}^{h} (r^2)_{j:n} \tag{4}$$

where h is the number of data points, residuals are included in the sum. This will find a robust estimate by identifying the (n-h) points which hold large residuals as outliers and discarding trimmers of data set. The trimmed data set results in least square estimates which can be visualized as h as close as to the number of good points in the data set. It is less sensitive to local effects than LMS, it has more local stability and it has better statistical efficiency than LMS.

Symmetry Distance

Symmetry is considered a pre-attentive characteristic that enhances recognition and reconstruction of shapes and objects. Symmetry exists in many man-made and natural objects. Symmetry Distance (Zabrodsky *et al.* (1995)) as a quantifier of the minimum effort required turning a given shape into a symmetric shape. This effort is measured by the mean of the square distances each point is moved from its location in the original shape to its location in the symmetric shape. Symmetry distance is to find the orientation of symmetric objects from their images and to find locally symmetric regions in images. Given n points $x_i, i = 1,...N$ and a reference vector C, the point symmetry distance (Su and Chou (2001)) between a point x_j and C is defined as follows:

$$d_s(x_j, C) = \min_{\substack{i=1,...N \\ and\, i \neq j}} \frac{\|(x_j - C) + (x_i - C)\|}{\|(x_j - C)\| + \|(x_i - C)\|} \tag{5}$$

when $x_i = (2C - x_j)$ exists, $d_s(x_j, C) = 0$.

However, a point could be exploited several times as the balancing point with regard to the centre. Thus the notion of symmetry is not properly entranced. In order to forefend a point being exploited as a symmetric point extent one time by other points, the refinement of the point symmetry distance between a point x_j and C is defined as follows:

$$D_s(x_j, C) = \min_{\substack{i=1,...N \\ and\, i \neq j \\ and\, i \notin \Re}} \frac{\|(x_j - C) + (x_i - C)\|}{\|(x_j - C)\| + \|(x_i - C)\|} \tag{6}$$

where $\Re$ is a set of points that have been used as symmetric point. The symmetry distance is written as

$$SD_n(x, C) = \frac{1}{n} \sum_{i=1}^{n} D_s(x_i, C) \tag{7}$$

If the SD of a perceptual structure is equal to zero, it is perfectly symmetric; if the SD of a perceptual structure is very big, it has little symmetry.

Least Trimmed Symmetry Distance (LTSD)

Wang and Sutter (2002) proposed Least Trimmed Symmetry Distance by combining the LTS procedure with symmetric distance measure. Mathematically, LTSD estimate can be written as

$$\hat{\theta} = \arg \min_{\theta, C} SD_h(x, C) \tag{8}$$

Only *h* data points with the smallest sorted residuals are used to calculate the symmetry distance. The estimated parameters correspond to the least symmetry distance. In LTSD the concept of trimming induces loss of information.

M- Estimator based Symmetry Distance (MSD)

The concept of symmetry is applied with M-estimator to overcome the limitation of LTSD. The proposed method is namely M-estimator based Symmetry Distance (MSD) and mathematically defined by

$$\hat{\theta} = \arg \min_{\theta, C} \rho(SD_n(x, C)) \tag{9}$$

The computational steps of MSD are as follows:

Step 1: Randomly select a subset which contains p+1 data points.
Step 2: Compute point symmetry.
Step 3: Compute symmetry distance SD_i based on point symmetry.
Step 4: Calculate the weight function w_i by M estimator.
Step 5: Calculate the $\hat{\theta}$ based on *n* observations.
Step 6: Repeat the above procedure for all possible subsets and compute $\hat{\theta}$.
Step 7: Finally, output $\hat{\theta}$ with the lowest SD

Experimental Results

This section provides the experimental results in the context of fitting models with symmetrical structures using the proposed method MSD along with the other methods which were discussed in the previous sections. Fitting a model is one of the major tasks of computer vision. Circle fitting is very popular topics that have symmetry characteristic. First we considered a simulating environment and then a real image. The first experiment provides the results of (i) LMS and LTS and (ii) MSD and LTSD fits in the context of circle fitting along with breakdown plots. The superiority of the proposed MSD method over LMS, LTS and LTSD has been studied and the results are provided in the second experiment. The breakdown plots of MSD with other methods are also presented under varying standard deviations. The experiment three provided the results based on real image.

Experiment 1

In circle fitting, the centre point at (x_0, y_0) with a radius r in (x, y) co-ordinates has the following form

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \quad (10)$$

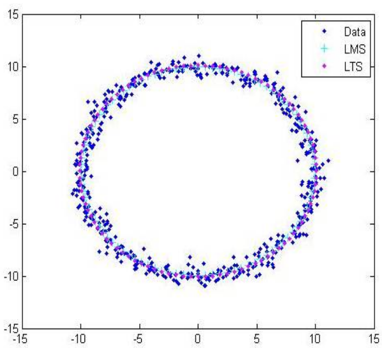
It can be written as

$$\left(x_1 + \frac{b_1}{2a}\right)^2 + \left(x_2 + \frac{b_2}{2a}\right)^2 = \frac{\|b\|^2}{4a^2} - \frac{c}{a} \quad (11)$$

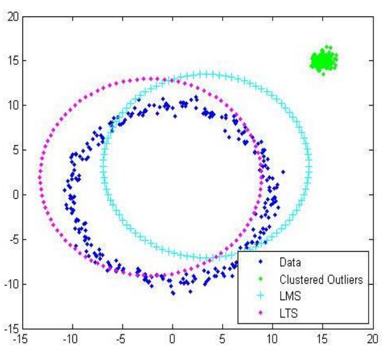
The centre and radius is given by

$$z = (z_{1,} z_2) = \left(\frac{-b_1}{2a}, \frac{b_2}{2a}\right); r = \sqrt{\frac{\|b\|^2}{4a^2} - \frac{c}{a}} \quad (12)$$

The 500 data points have been generated with radius10 and (0,0) as the centre point. Considering the experiment without outliers, the obtained results under the methods LMS and LTS are is shown in figure 1(a). In the case of contamination, specifically, in the presence of clustered outlier, the data were contaminated within the region of radius 15 and the fitted models are shown in figure 1(b). It is noted that, the LMS method gets affected even, less than 20% contamination whereas the LTS method can tolerate more than 20% of clustered outliers, hence the fitted model with 30% contamination is displayed.



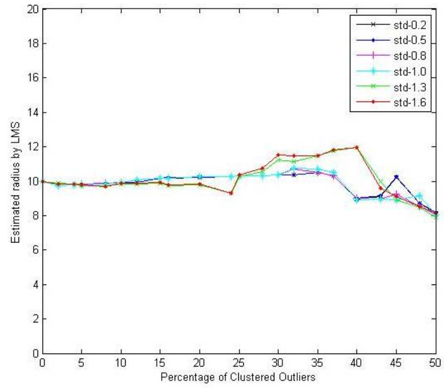
(a)



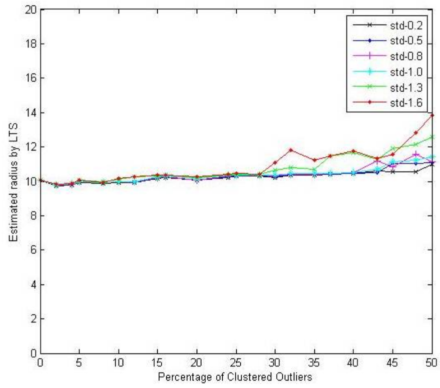
(b)

Figure 1 Model Fitting with LMS and LTS (a) without outliers (b) with outliers

Now, the robustness property, breakdown is considered under various levels of contaminations. The estimated radius by varying the standard deviation of the inliers under LMS and LTS and thus obtained breakdown plots are displayed in the figure 2.



(a)



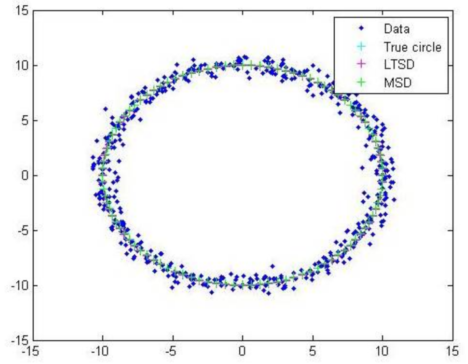
(b)

Figure 2 Breakdown Plots (a) LMS (b) LTS

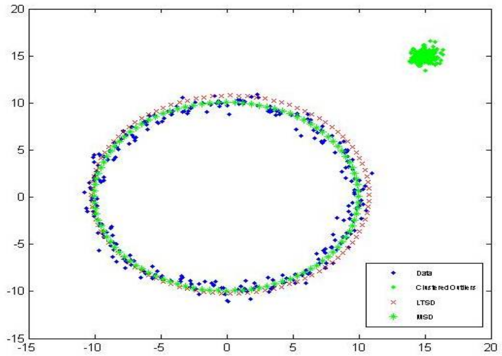
From figure 2(a), it can be seen that when the standard deviation of inliers is not more than 1.0, LMS can estimate the right result under high percentage of outliers (less than 30%). However, when the standard deviation of inliers is more than 1.0, LMS does not produce the right result even when the percentage of outliers is less than 20%. From figure 2(b), it noted that when the standard variance of inliers is 0.5 to 1.0, the LTS can estimate the results even under less than 43% clustered outliers; while when the standard variance of inliers is more than 1.0, LTS does not give the right results even when less than 30% of the data are outliers.

From the discussions above, it can be observed that there are several conditions under which LMS and LTS failed to be robust. A crucial point is that these methods measure only one single statistic; the least median of residuals or the least sum of trimmed squared of the residuals, omitting other characteristics of the data. If the failures are considered, the results lose the most basic and common feature of the inliers with respect to the fitted circle-symmetry.

Figure 3(a) shows that the LTSD and MSD are the right models, where the true centre is (0,0) and radius is 10. It is noted that the LTSD method estimated the centre (0.28,0.27) and radius 10.52 and MSD method produces the centre (-0.03,0.02) and radius 10.01. It is noted that, the LTSD can tolerate upto 40% contamination, whereas the MSD can tolerate more than 40% of clustered outliers, hence the fitted model with 45% contamination is displayed in the figure 3(b).



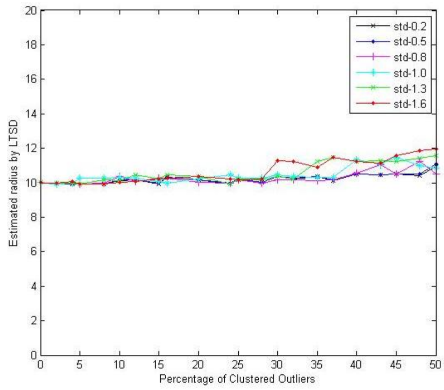
(a)



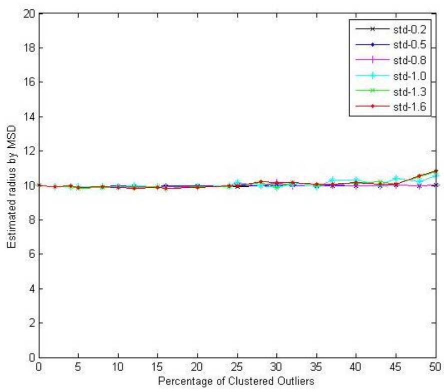
(b)

Figure 3 Model Fitting with LTSD and MSD (a) without outliers (b) with outliers

The breakdown plot of LTSD and MSD is shown in the figure 4. It is noted that when the standard deviation of inliers is not more than 1.0; LTSD can produce the right results under high percentage of outliers (less than 40%).



a



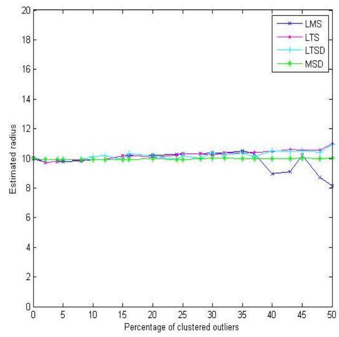
b

Figure 4 Breakdown Plots (a) LTSD (b) MSD

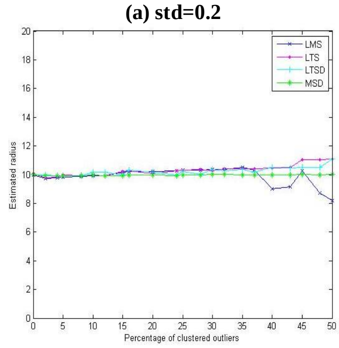
However, when the standard deviation of inliers is more than 1.0, LTSD does not produce the right result even when the percentage of outliers is less than 30%. Further, it is observed that when standard deviation is above 1.0, the method MSD can tolerate upto 47% of outliers, while the standard deviation below 1.0, the breakdown point is about 50%.

Experiment 2

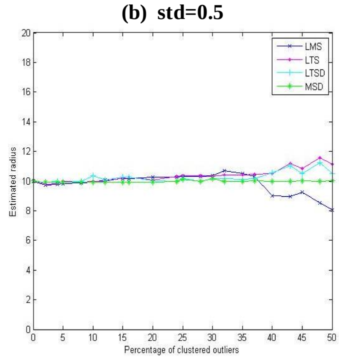
The performance of MSD over the other methods has been carried out when varying the standard deviation (0.2, 0.5, 0.8, 1.0, 1.3, and 1.6) and various levels of contaminations for the generated data which is used in the previous experiment. The breakdown plots of MSD along with LMS, LTS and LTSD under varying standard deviation are displayed in the figure 5.



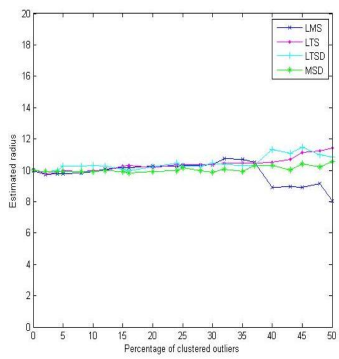
(a) std=0.2



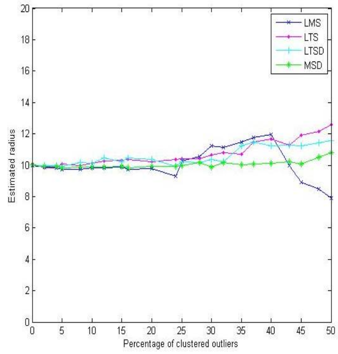
(b) std=0.5



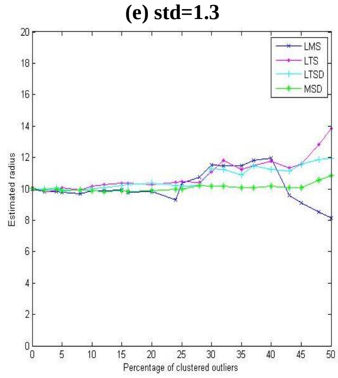
(c) std=0.8



(d) std=1.0



(e) std=1.3



(f) std=1.6

Figure 5 Breakdown Plots under various standard deviations

It is observed that when the level of contamination and standard deviations increases LMS and LTS does not produce the reliable estimates, because these methods are based on least median of squared residuals and Least trimmed residuals. The symmetry distance based robust estimator like LTSD and the proposed MSD produces the reliable estimates. When the contamination level and the standard deviation increases the LTSD can tolerate upto 30% but MSD can perform upto 47%. It is noted that the proposed MSD has higher breakdown point than the LTSD.

Experiment 3

To study the performance of the MSD in the context of model fitting over the other procedures a real image with outliers (LED ball with stick) was considered and is given in figure 6(a). The detected edge of the image was obtained by using canny operator (Canny (1986)) with threshold 0.09 which is given in the figure 6(b). The detected edge (circle) under the methods LMS, LTS, LTSD and MSD are displayed in the figure 6(c). Since the stick is considered as the clustered outliers in the image, the robust estimators LMS and LTS methods fails to detect the edge of the LED ball. The LTSD and MSD correctly detected the edge of the LED ball in the image. It is noted that the LTSD has small deviation in the model fitting, but the proposed MSD method more correctly fitted than LTSD. It is noted that the estimated centre and radius under the methods LMS, LTS, LTSD and MSD is [(322.31, 229.61), 151.45], [(215.16, 228.31), 160.32], [(182.26, 177.09), 162.13] and [(181.88, 174.02), 158.02] respectively.

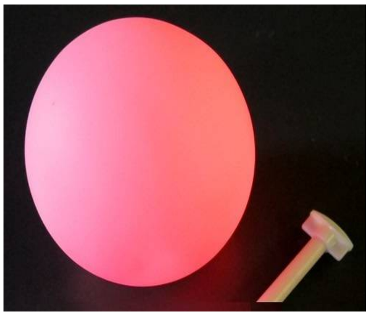
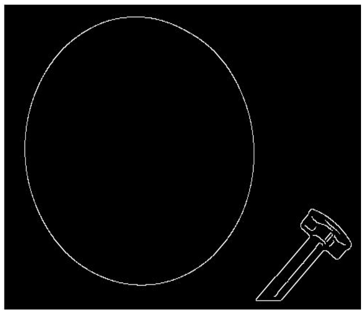
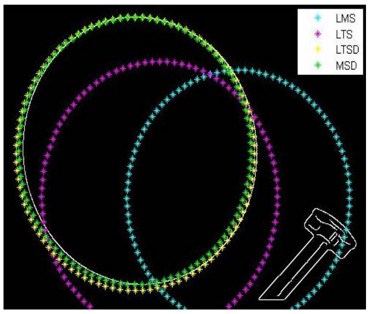


Figure 6 (a) Real Image



(b) Edge Detection



(c) Model fitting by MSD with other methods

Summary

In general, the LS method is very sensitive to outliers and thus it doesn't fit well by considering all the data points. The most popular robust regression estimators like LMS and LTS breakdown at lower percentage of outliers when the outliers are clustered. Since, concept of trimming induces loss of information, the symmetry based LTS method also fails to tolerate certain amount of clustered outliers. To overcome these limitations, the concept of symmetry distance is applied in M estimator and proposed a method namely MSD. The superiority of the proposed method MSD has been studied by compared with LMS, LTS and LTSD under real and simulating environment in the context of circle fitting and also by considering the breakdown plots. It is concluded that if the data contains clustered outliers the MSD estimator is the most suitable one instead of LMS, LTS and LTSD. It is very much applicable in performing computer vision tasks, since it deals with multiple structures; one structure is outlier to the other ones.

References

1. D.Umbach, K.N.Jones, A Few Methods for Fitting Circles to Data, *IEEE Transactions on Instrumentation and Measurement*, 52(6), (2000) 1881-1885.

2. N.R. Draper, H.Smith, *Applied Regression Analysis*, Third Edition, Wiley Interscience Publication, United States (1998).
3. W. Gander, H. Golub, R. Strebel, Least-Squares fitting of circles and ellipses, *BIT Numerical Mathematics*, 34, (1994) 558-578.
4. F.R.Hampel, P.J.Rousseeuw, E.Ronchetti, W.A.Stahel, *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, New York (1986)
5. P.J.Huber, Robust estimation of a location parameter, *Annals of Mathematical Statistics,* 35, (1964) 73-101.
6. P.Meer, D.Mintz, A.Rosenfeld, D.Y.Kim, Robust Regression Methods for Computer Vision: A Review, *International Journal of Computer Vision* 6(1), (1991) 59-70.
7. R.Muthukrishnan, M.Radha, M Estimators in Regression Models, *Journal of Mathematics Research,* 2(4), (2010) 23-27.
8. R.Muthukrishnan, R.Reka, E.D.Boobalan, Model Fitting Based on Robust Regression Procedures, *International Journal of Statistics and Analysis,* 5(2), (2015) 151-160.
9. P.J.Rousseeuw, Least Median of Squares Regression, *Journal of the American Statistical Association,* 79, (1984) 871-880.
10. M.C.Su, C.H.Chou, A Modified Version of the K-Means Algorithm with a distance based on Cluster Symmetry, *IEEE, Transactions on Pattern Analysis and Machine Intelligence,* 23(6), (2001) 674-680.
11. J.Canny, A Computational Approach to Edge Detection, *IEEE, Transactions on Pattern Analysis and Machine Intelligence,* 6, (1986) 679-698.
12. J.W.Tukey, *The Future of Data Analysis,* The Annals of Mathematical Statistics, 33, (1962) 1-67.
13. H.Wang and D.Suter, *LTSD: A Highly Efficient Symmetry-based Robust Estimator.* In 7th International Conference on Automation, Robotics and Computer Vision (ICARCV02), (2002) 332-337.
14. H.Wang and D.Suter, Using Symmetry in Robust Model Fitting. *Pattern Recognition Letters,* 24(16), (2003) 2953-2966.
15. Y.Susanti, H. Pratiwi, H. Sri Sulistijowati, T.Liana., M Estimation, S Estimation, and MM Estimation in Robust Regression. *International Journal of Pure and Applied Mathematics,* 91(3), (2014) 349-360.

How to cite this article:

Muthukrishnan R and Reka R (2017) 'Breakdown Plot of Symmetry Distance Based Robust Estimators', *International Journal of Current Advanced Research,* 06(09), pp. 6150-6155. DOI: http://dx.doi.org/10.24327/ijcar.2017.6155.0883

****