Comparative Performance of MM and Least Trimmed Squares (LTS) Robust Regression Methods

Carolina B. Baguio

Abstract

This study compares the performance of MM and Least Trimmed Squares (LTS) Robust Regression methods with the Ordinary and Modified Least Squares. The data of Hawkins-Bradu-Kass (1984) were used for the investigation. This is a much referenced data of 75 observations with one response variable and three independent variables. These data are known to be quite troublesome in terms of masking and swamping. Masking refers to bad data points being camouflaged because they are clustered; while swamping refers to good data points which appear to be outliers. In the study, it was shown empirically that LTS performs better than the other mentioned methods in terms of finite efficiency, goodness - of - fit and breakdown point.

Keywords: Robust regression, M-estimates, Least Trimmed Squares, Efficiency, Breakdown Point

✍ **CAROLINA B. BAGUIO** is a Professor of Mathematics in the Department of Mathematics, College of Science and Mathematics, MSU-IIT, Iligan City, Philippines. She holds a Ph.D. in Applied Mathematics from the Mindanao Polytechnic State College, Cagayan de Oro City.

1 Introduction

The least squares method of estimation has maintained its popularity until today. It was realized that outliers in the data, which do not appear to come from the normal distribution but may have arisen from a distribution with long (heavy) tails or from gross errors, have unusually large influence on the least squares estimates. Consequently, robust methods of estimation have been developed to reduce the influence of outliers in the data, on the estimates.

Robust regression is an alternative procedure to ordinary least squares which can be used in the presence of outliers and the distribution of the error terms is non-normal. There are quite a number of these methods, however, this paper will only attempt to focus on the MM and Least Trimmed Squares Robust Regression Methods and compare the respective efficiencies in terms of the residual errors, goodness-of-fit employing the multiple coefficient of determination and the breakdown point. The Least Median Squares(LMS) is not included in the study since it was found in the literature that this method perform poorly with respect to Ordinary Least Squares (OLS).

This article differs from the work of Schumacker R.E., et al (2002) wherein they did comparison on OLS versus LTS and MM robust regression using moderately large set of data which is 500. The criteria for comparison also differ from this paper.

Moreover, this study specifically deals with robustness study on the effect on the estimates of the coefficients of linear regression if data in Hawkins-Bradu-Kass (1984) is used. These data can be easily shown to violate the assumption of normality and even linearity because of the masking and swamping of the observations. It is quite difficult to look for some realistic data which are masked and swamped.

Obviously, this paper is different from a related unpublished dissertation by Anderson, C. of University of North Texas which compares five robust regression methods with OLS in the criteria for comparison since efficiency, biasedness and test of null hypothesis were used in this dissertation.

2 Objectives of the Paper

1. To present the concept of robust regression estimator and the procedures of the MM and LTS Robust Regression Methods.

2. To compare the efficiency of the two methods with the Ordinary Least Squares (OLS) in terms of finite efficiency or residual error, goodness-of-fit and breakdown point.

3 Properties of Robust Regression Estimator

Ryan(1997) stressed that a robust regression estimator should possess the following properties:

1. To perform almost as well as the Ordinary Least Squares(OLS) when the latter is the appropriate choice. This means when the errors are normally distributed with data being free of mistakes and influential data points.

2. To perform much better than OLS when the conditions in (1) are not satisfied.

3. Not be overly difficult to compute or understand.

4 Criteria for Assessing Regression Models

1. Efficiency refers to the variance of the sampling distribution for the estimator. High efficiency estimators have small variance in the sampling distribution for the estimator. The efficiency of the selected robust technique is defined as the ratio of the mean-squared error of this technique divided by the OLS mean square error. There are two types of efficiency such as finite and asymptotic efficiency. *Finite sample efficiency* refers to the variance of the sampling distribution for the estimator as it is applied in small sample settings while *Asymptotic efficiency* refers to the way an estimator performs as the sample size gets larger. In this paper, the former is to be utilized in comparing two techniques.

2. The Multiple Coefficient of Determination is usually employed to measure the goodness-of-fit of the model. The closer it is to 1 or 100% then the better is the fit of the model.

3. High breakdown point. This refers to the smallest fraction of the anomalous data which render the estimator useless. The literature is recommending a breakdown point greater than 10%. High breakdown point is the largest percentage of data points that can be arbitrarily changed and not unduly influence the estimator. For example the median has 50% breakdown point while the mean has a breakdown point of $1/n \times 100\%$.

5 The Trimmed Mean

One problem with the median, however, is that its value is determined by only 1 or 2 values in the data set-information is lost. The trimmed mean represents a compromise between the mean and the median (Huber, 1981). The trimmed mean is computed by putting the observations in order. Next, trim the numbers by removing the d largest and d smallest observations, and then compute the average of the remaining numbers. d can be between 0 and $n/2$. Trimming enough data gives the sample median. Rules of thumb are that 20%-25% ($d = .2 \cdot n$) trimming works well in a wide range of settings(Wilcox, 1997). Another approach to selecting the trimming amount is to calculate the mean for 0, .10, .20 and then use the trimming value that corresponds to the smallest standard error (Leger and Romano, 1990). While Baguio(1999) proposed an ad hoc procedure to trim the data adaptively using the length of the tails in the distribution. The breakdown point for the trimmed mean is the trimming percentage.

6 M-Estimators

Huber [(1973),(1981)] introduced a class of estimators known as M-estimators, with objective function of minimizing the sum of the symmetric function of the residuals r represented by $\sum \rho(r^2)$ with a unique minimum at zero. The computation of the regression weights are done iteratively until a convergence criterion is met. There were several proposed function of the residuals discussed in the literature. The breakdown point of M-estimator is $1/n \times 100\%$.

7 Least Trimmed Squares (LTS)

The estimator of LTS is obtained by minimizing the sum of the ordered squared residuals ranging from 1 to h, from smallest to largest, and the value of h is determined by $h = [n/2] + [(p+1)/2]$. The symbol $[\cdot]$ means the "integer portion of". In terms of the trimming proportion α, the value of h can also be computed as $h = [n(1-\alpha)+1]$ as proposed by Rousseeuw and Leroy (1987, p.134).The LTS is appealing since the objective function is not based on the fit at any particular point but on the residuals. If the exact number of bad data points are trimmed excluding the good data points then this result to optimal estimator OLS applied to the good data points.

8 MM Estimation and Regression

MM estimation is a special type of M-estimator developed by Yohai (1987) where the function of the residuals is a bounded loss function scaled and fine tuned. MM- estimator involves three estimations. The first stage involves calculation of an estimator (S-estimation) with high breakdown point. Then a robust M-estimation is done on the second stage using the S-estimate for the initial values. The third stage is computing for the final M-estimates of the regression parameters. MM estimates are computed using the facility of S-Plus with several statistics for inference and diagnostics.

Robust regression models are useful for fitting linear relationships when the random variation in the data is not Gaussian (normal) or when the data contain significant outliers. In such situations, standard linear regression may return inaccurate estimates. The robust MM regression method returns a model that is almost identical in structure to a standard linear regression model. This allows the production of familiar plots and summaries with a robust model.

In Robust MM Regression, robust initial regression coefficients are used as starting values. The robust regression coefficients are found by minimizing a scale parameter, S, while χ may be one of several bounded loss functions that serves the purpose of minimizing the empirical influence of troublesome residuals. χ is an integral of $\chi(u)$ in the formula

$$\sum_{i=1}^{n} \chi(y_i - x_i b)/c_0 S = (n-p)\beta$$

where $\chi(u) = u^6 - 3u^4 + 3u^2$, $u \leq 1$, c = tuning constant = 1.548 and $\beta = 0.50$.

9 Empirical Results

In order to compare the efficiency of the MM and LTS Robust Regression methods, the data of Hawkins-Bradu-Kass (1984) is utilized. This is a much referenced data consisting of 75 observations which is known to be quite troublesome in terms of masking and swamping. Masking refers to bad data points being camouflaged because they are clustered while swamping refers to good data points which appear to be outliers.

The facilities of Statistical Package for Social Science (SPSS) version 10 and S-PLUS 2000 for the construction of figures 1 to 6 and for the computations of the different estimates of the Robust Regression Models, respectively. These two robust methods were also compared to OLS and modified OLS which are inappropriate methods in order to find out the disadvantage in using these methods:

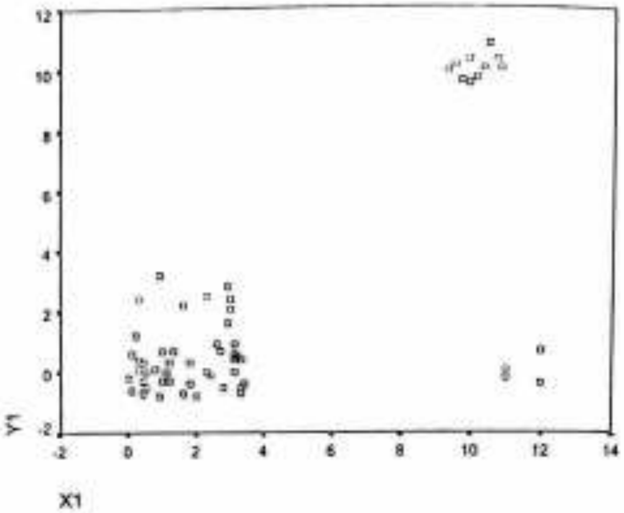

Figure 1: Scatter plot of $Y1$ versus $X1$

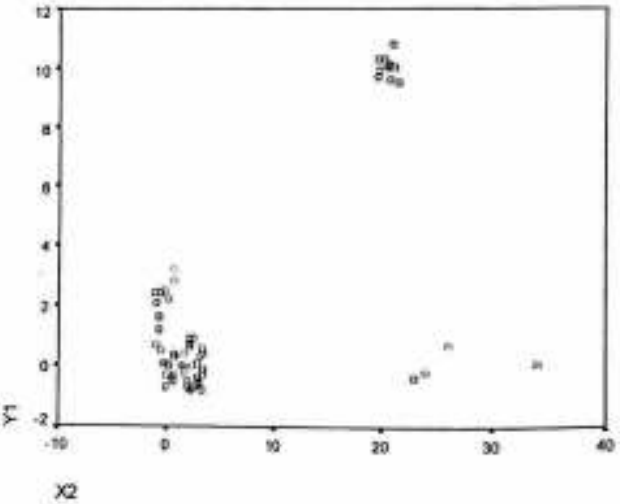

Figure 2: Scatter plot of $Y1$ versus $X2$

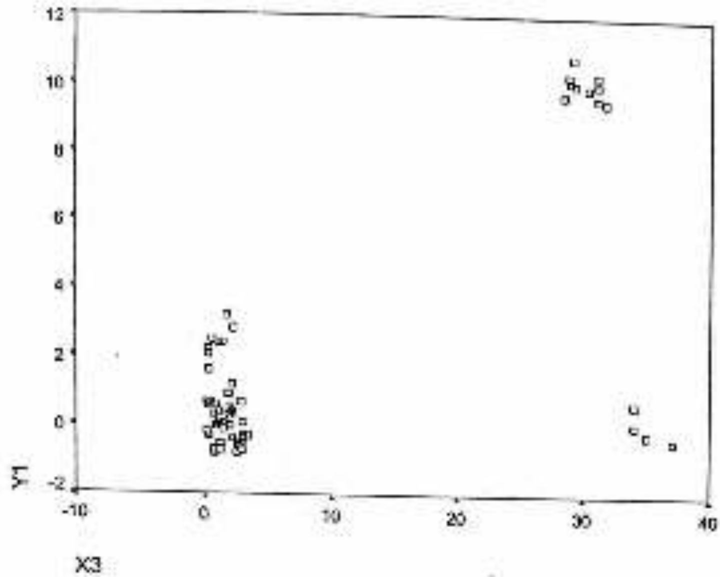

Figure 3: Scatter plot of $Y1$ versus $X3$

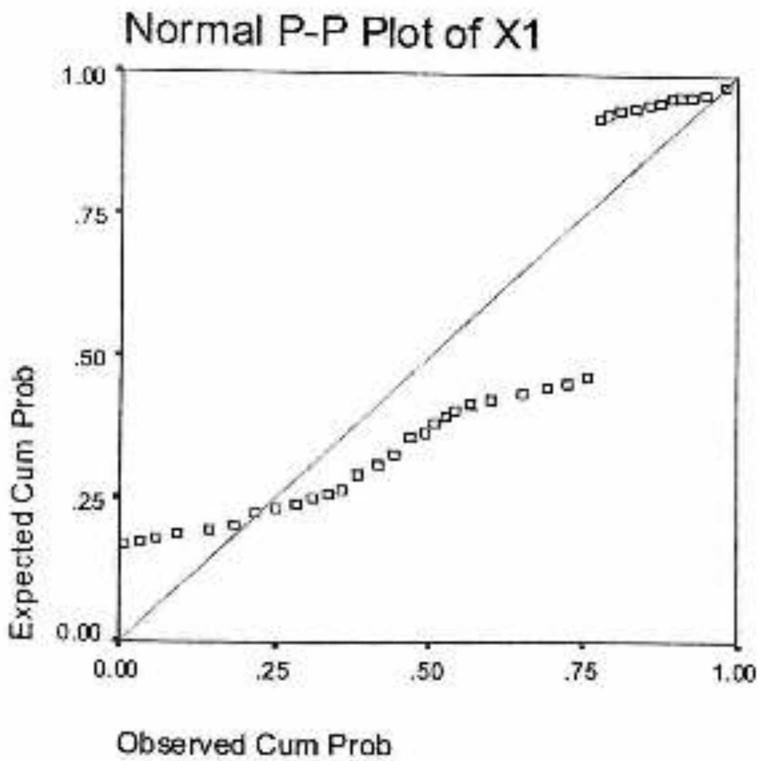

Figure 4: Normal PP plot of $Y1$ vs. $X1$

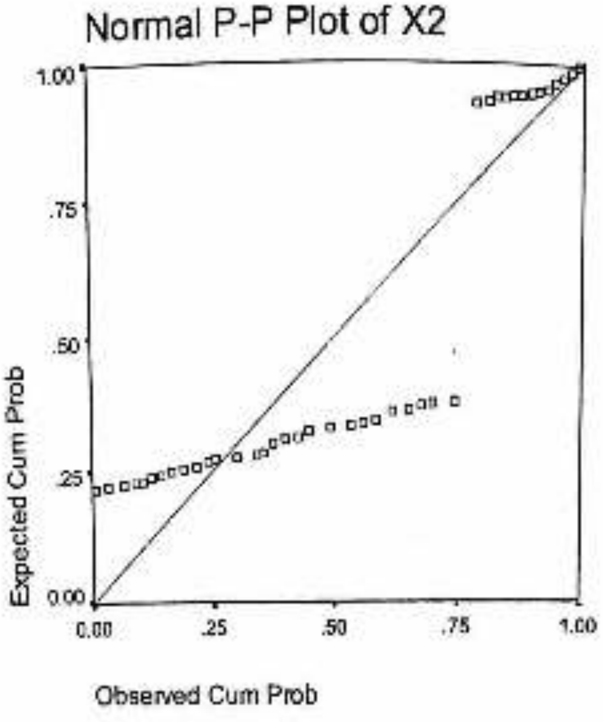

Figure 5: Normal PP plot of $Y1$ vs. $X2$

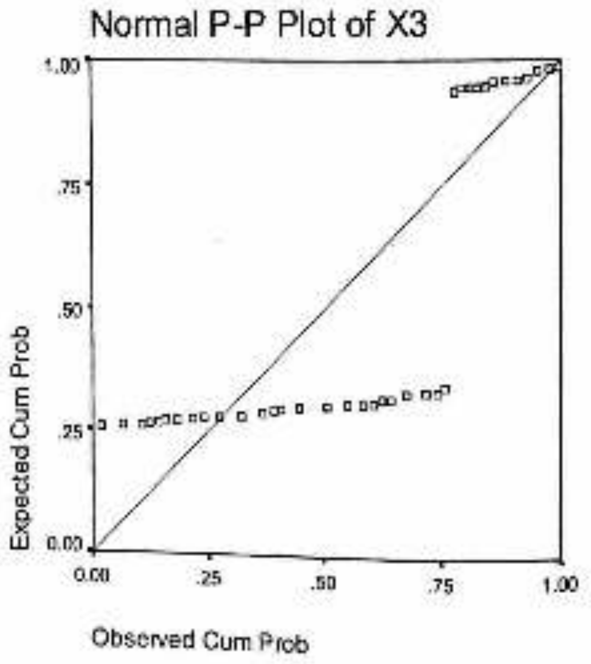

Figure 6: Normal PP plot of $Y1$ vs. $X3$

Figures 1 to 6 give the graphical display of the Hawkins-Baradu-Kass data. It is evident that the there is the clumping and masking of data. The normal PP plots in figures 3 to 6 indicate the data deviate from normality due to the presence of outliers. The OLS method is inappropriate for this kind of data. Table 1 gives the comparative models and the various estimates inherent to the regression models.

Method	Inter-cepts	$X1$	$X2$	$X3$	Resi-dual Error	Break-down Point	Effi-ciency	Coeff. of Determi-nation
MM	−0.18	.08	.04	−.052	.7894	1.33%	35.10%	2%
LTS	−1.10	.196	.21	.16	.6831	13.33%	30.4%	88.29%
OLS	0.39	.24	−.33	.38	2.25	1.33%		60.20%
Modified OLS (1st 10 outliners discarded)	−0.18	.081	.04	−.05	.56	13.33%	24.9%	4.3%

Table 1: Estimates of Regression Models using the MM,LTS, OLS and Modified OLS Methods

The basis for comparing the performance of the mentioned Robust Regression are small residual error, coefficient of multiple determinations close to 1, efficiency close to 100% with a breakdown point of greater than 10%.

It can be observed from the table that the Regression model using LTS gives smaller residual error of .6831 and relatively higher coefficient of determination which is 88.29% with a breakdown point of 13.33% and efficiency of 30.4% compared to the MM method. However, comparing the four methods, the LTS performs better in terms of residual error, breakdown point, efficiency and goodness-of-fit. It is apparent from the result that MM Regression method yields high efficiency but poor breakdown point and goodness-of-fit.

On the other hand, the model obtained using the OLS method has a relatively higher residual error of 2.25 although the value of the coefficient of determination $R^2 = 60.20\%$ is quite higher than the MM method however, both have the same breakdown point of 1.33%

It can also be observed from the table that if the ten outliers which can be determined from the scatter plot are discarded, the resulting residual error is small but the fit is not quite good since the coefficient of variation explained by the model is only 4.3% although the breakdown point is 13.33% which is greater than 10%.

10 Conclusion

For data which are masked and swamped thereby violating the assumptions of normality, the LTS Robust method performs better than the OLS and MM Robust Regression methods. The basis of comparison is on the magnitudes of the residual error, the goodness of fit measured by the coefficient of multiple determination and the breakdown point greater than 10%.

11 Recommendation

There are still other Robust methods to be assessed like the S-estimators mentioned in the literature. Hence, for a future study, it is recommended that the efficiency of these methods in terms of its efficiency, goodness-of-fit and breakdown point be evaluated.

References

[1] Baguio, Carolina B. *Adaptive Robust Estimator of a Location Parameter*, A doctoral dissertation submitted to Mindanao Polytechnic State College, Cagayan de Oro City.

[2] Baguio, C. and Jamboy, N. *Comparative Study of Least Median Square (LMS) and Least Trimmed Squares (LTS)*, National Convention in Statistics Proceedings, 2002.

[3] Hawkins-Bradu-Kass. *Location of Several Outliers in Multiple Regression using elemental sets.* Technometrics,26(1984) 197-208.

[4] Huber, P.J. *Robust Regression: Asymptotics, Conjectures and Monte Carlo.* The Annals of Statistics, **1** (1973) 799-892.

[5] Huber, P. J. *Robust statistics*, Wiley, New York, 1981.

[6] Leger, C. & Romano, J.P. *Bootstrap adaptive estimation: the trimmed-mean example.* The Canadian Journal of Statistics, **18** (1990) 297-314.

[7] Rousseeuw, P.J. and Leroy, A.M., *Robust Regression and Outlier Detection*, John Wiley, New York, 1987.

[8] Ryan, Thomas P., *Modern Regression Methods*, Wiley Series in Probability and Statistics, John Wiley, New York, 1997.

[9] Schumacker, R.E. et al. *A Comparison of OLS and LTS and MM Robust Regression in S-PLUS.* Paper presented at the Soutwest Education Research Association 25th Annual meeting, Austin, Texas, 2002.

[10] Wilcox, Rand R. *Introduction to Robust Estimation and Hypothesis Testing.* Academic Press, New York, 1997.

[11] Yohai, Victor J. *High Breakdown Point and High Efficiency Robust Estimates for Regression.* Annals of Statistics, **15**(1987) 642-656.