International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Comparative Study of Robust Regression Methods in Modeling the Currency in Circulation in Malaysia

Azme bin Khamis[1], Nur Azreen binti Abdul Razak[2]

Faculty of Science, Technology and Human Development, University of Tun Hussein Onn Malaysia,
86400 Parit Raja, Batu Pahat, Johor, Malaysia

Abstract: ***This study presents a comparative study on robust regression method via M-estimator and Least Trimmed Square (LTS) estimator against Ordinary Least Square (OLS) estimator in modelling economic indicators in Malaysia. The aim is to estimate the parameters of multiple linear regression model in the presence of outliers and to evaluate the estimator performance of OLS, LTS and M estimation. The coefficient of determination known as R-squared values were used as criteria to evaluate the estimator performance. Research findings showed that LTS estimator outperform M estimator and OLS estimator in terms of maximum of R-squared values.***

Keywords: robust, LTS, M-estimator, OLS, R-squared

1. Introduction

Regression is one of the most commonly used statistical method. Regression analysis is conducted to determine the relationship between two or more variables having cause-effect relations and also to make a prediction as stated by [1]. Out of many possible regression techniques, the ordinary least squares (OLS) method has been generally adopted because in light of the fact that simplicity of computation.

However, OLS estimation can be influenced by the presence of outliers, observations which deviate far from the linear relation of the response variable and the exploratory variables. Despite the fact that though outliers usually bias the OLS predictions towards outliers, they are often implemented in empirical analysis as stated by [2]. Thus, it can be assumed that in the presence of outlier, Least squares estimation is inefficient and can be biased [3].

To remedy this problem, new statistical tools have been developed that are not so easily affected by outliers. An alternative is to utilize robust regression methods, [4]. Robust regression is an essential method for analysing data that are contaminated with outliers. It can be habituated to detect outliers and to provide resistant results in the presence of outliers [5]. These are robust methods, such as Least Median of Squares (LMS), Least Trimmed Squares (LTS), Huber M Estimation, MM Estimation, Least Absolute Value Method (LAV) and S Estimation [6,7,8,9].

Circulation is the process of repeating the use of individual units of a currency for transactions in terms of monetary economics. Thus, currency in circulation (CIC) is the total value of currency either in coins and paper currency that has ever been issued minus the amount that has been removed from the economy by the central bank. As stated by [10], CIC is one of the main factors that push money market liquidity and are vital indicators for monetisation and demonetization of the economy. The share of the CIC in money supply and its ratio to nominal Gross Domestic Product reveals its relative importance in any economy, see [11].

Several researches on Currency in Circulation have been done. [10] modelled and predicted the monthly CIC in Ghana using ARIMAX and SARIMA models. [12] modelled and predicted the CIC for Liquidity Management in Nigeria. In another study, [13] studied on Currency and Coinage Circulation in India. [14] discussed and modelled the CIC in Nigeria. See also [15,16].

Consequently, researchers should consider a comparison of OLS parameter estimates with robust regression parameter estimates. Therefore, the purpose of this study was to compare robust regression methods; LTS and *M* estimation against OLS regression estimation method in terms of the determination of coefficient by modelling the currency in circulation in Malaysia.

2. Methodology

2.1 Materials

In this study, we are going to model the Currency in Circulation using Multiple Linear Regression method and estimate the parameter by using different estimator that will be discussed in details on the following subsection. Data that were used in this study are the economic indicators consist of Currency in Circulation (CIC) as a dependent variable and the independent variables consist of Exchange Rate (EXC), External Reserve (EXT) and Reserve Money (RM) that were monthly basis starting from January 1998 to January 2016. The data were gained from Ministry of Finance Malaysia and Department of Statistics Malaysia. All statistical analyses and OLS, LTS and M-estimation results were performed and compared by using *R* language.

2.2 Multiple Linear Regression (MLR)

Regression analysis consists of an accumulation of techniques that are acclimated to explore relationships between variables. A main objective of regression analysis is to estimate the unknown parameters in the regression

model and also called fitting the model to data, see [17]. Regression model can be either linear or non-linear when each term is either the predictor variable or a constant. While a linear equation has one basic form, nonlinear equations can take many different forms. A linear regression model can be expressed with the following equation:

$$Y_i = \beta_0 + \beta_i + \varepsilon_i \quad (1)$$

where, β_0 is the slope, β_i and i = 1, …, p is the slope of unknown parameter and ε_i is a random error component usually assumed to be normally distributed with mean zero, Y_i is the dependent variable or response and X_i is independent variable or the predictor.

Multiple linear regression analysis is an extension of simple linear regression analysis, used to determine the association between two or more independent variables and a single continuous dependent variable. The multiple linear regression equation is as follows:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \quad (2)$$

where, Y_i is the predicted value of the dependent variable, X_1 through X_p are p distinct independent variables, β_0 is the value of Y when all of the independent variables (X_1 through X_p) are equal to zero and β_1 through β_p are the estimated regression coefficients.

Generally, ordinary least squares (OLS) estimation is used to estimate the parameters in multiple regression. [1] defined the assumptions of OLS are that residual errors should be normally distributed, have equal variance at all levels of the independent variable known as homoscedasticity and be uncorrelated with both the independent variables and with each other. However, OLS estimation in the multiple regression are affected by the occurrence of outliers and missing data [18]. If the data contains missing data or outliers, then the sample estimates and results can be misleading.

2.3 Outliers data in Multiple Linear Regression

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. Outliers should be investigated carefully as they often contain valuable information about the process under the data gathering and recording process. In general, outliers may arise and be classified into two types such as man-made one and random one. Man made one may arise due to miss-reporting, instrument error and sampling while random one may exist due to random chance for drawing sample from a population. Each different reason may require different treatments, see [19].

OLS estimator is extremely sensitive to multiple outliers in linear regression analysis and can disturb the assumption of normality, one of the most important components of statistical studies [20] . It can even be easily biased by just a single outlier because of its low breakdown point.

Unlike OLS estimator, robust regression provides robust regression estimators even in the presence of multiple outliers [21]. The primary purpose of robust regression techniques is to provide resistant results in the presence of outliers and fit a model that describes the information in the majority of the data. As defined by [22], robust method habituated that these techniques should perform well on both with outliers and on without outliers. The impact of outliers when using robust regression is minimized by giving smaller weight for outliers in the estimation procedure.

2.4 Robust Regression

Robust least squares refers to a variety of regression methods designed to be robust or less sensitive to outliers. Robust regression is an alternative to least squares regression when data exist with outliers and it can also be used for the purpose of detecting influential observations. [23] Verbally expressed that the properties of efficiency, breakdown point and bounded influence are habituated to define the quantification of robust technique performance in a theoretical sense. The simplest robust approach of robust regression is M-estimation, see [24, 25, 26, 27]. Least trimmed squares (LTS) estimation is a robust method with high breakdown point, which can withstand high proportion of outliers and still maintains its robustness. Here, the methods used for estimation parameters such as OLS, LTS and M-estimation could be introduced as follows, see [4].

2.4.1 Ordinary Least Square (OLS) estimation:

The OLS is used to find the best estimate of β's with the least squares criterion which minimizes the sum of squared distances of all of the points from the actual observation to the regression surface. OLS estimators are sensitive to the presence of observations that lie outside the norm for the regression model of interest. The OLS estimation method in residual form can be represented as:

$$Min \sum_{i=1}^{n} e_i^2 \quad (3)$$

where, e_i^2 or $\sum (Y - \hat{Y})^2$ denoted as the sum of the squared residual between the actual and predicted values. This method consists of the minimization of the sum of the squared residuals. However, in spite of the computational simplicity of LS method, this estimator is now being criticized more and more for its dramatic lack of robustness. Additionally, even there is a single outlier, it can have a large influence on the results of regression equation, see [28].

2.4.2 M estimation

The most common general method is M-estimation in the context of robust regression was first introduced by Huber (1973) as a result of making the least squares approach robust. The class of M-estimator models contains all models that are derived to be maximum likelihood models. Rather than minimize the sum of squared errors as the objective, the M-estimate minimizes a function ρ f the errors. They are based on the idea of replacing the squared residuals, e_i^2 , with another function of the residuals, given by:

$$Min \sum_{i=p}^{h} \rho(e_i^2) \quad (4)$$

Where ρ is a symmetric function with a unique minimum at zero. *M*-estimates are calculated using iteratively reweighted least squares (IRLS). In IRLS, the initial fit is

calculated and then a new set of weights is calculated based on the results of the initial fit. The iterations are continued until a specified number of iterations are finished or a convergence criterion is met.

2.4.3 Least Trimmed Square (LTS) estimation:

Rousseeuw (1984) developed the least trimmed squares (LTS) estimation method. LTS offers a more efficient way to find robust estimates and this method is given by,

$$Min\sum_{i=1}^{h}\left(e_i^2\right) \quad (5)$$

where $\left(e_i^2\right)$ are the ordered squared residuals, from smallest to largest and the value of *h* must be determined based on trimming the data values where $h=\left[\frac{n}{2}\right]+\left[\left(\frac{p+1}{2}\right)\right]$ with *n* and *p* being sample size and number of parameters, respectively. The largest squared residuals are excluded from the summation in this method, which allows those outlier data points to be excluded completely. Depending on the value of *h* and the outlier data configuration, LTS can be very efficient. In fact, if the exact same numbers of outlying data points are trimmed, this method is computationally equivalent to OLS.

After computing the parameters estimation of each model using OLS, LTS and M-estimator, we evaluate the performance of all estimators by comparing the value of the parameter estimation, *p*-value and the adjusted *R*-squared value of each model. The *p*-value for each term tests the null hypothesis that the coefficient is equal to zero or no effect. A low *p*-value (< 0.05) indicates that the null hypothesis can be rejected. In other words, a predictor that has a low *p*-value is likely to be a meaningful addition to model because changes in the predictor's value are related to changes in the response variable. Conversely, a larger or insignificant *p*-value suggests that changes in the predictor are not associated with changes in the response.

R-squared or R^2 is a statistical measure of how close the data are to the fitted regression line and also known as the coefficient of determination, see [29]. In general, the higher the *R*-squared, the better the model fits your data and is calculated as shown:

$$R^2=\frac{SSR}{SST}=\frac{\sum_{j=1}^{N}\left(\hat{y}_j-\bar{y}\right)^2}{\sum_{j=1}^{N}\left(y_j-\bar{y}\right)^2} \quad (6)$$

Where, SST is the total sum of squares, SSR is the regression sum of squares, *N* is the number of observations, $\bar{y}$ is the mean value, $\hat{y}$ is the predicted value and y is the response value.

3. Results and Discussion

Outliers can have a big influence on estimating the parameters. Therefore, scatter plot of each model is implemented by plotting the residuals against the fitted values to prove that the outliers do exist in the data. Refer Figure 1.

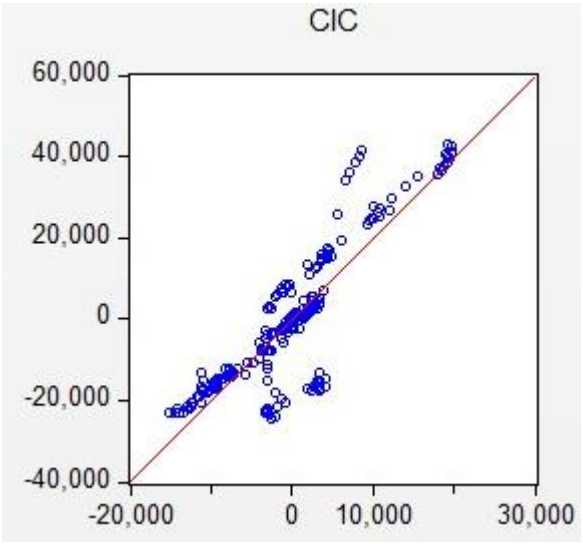

Figure 1: Scatter plot for CIC data

The plots of each model in Figure 1 seem to indicate that the residuals and the fitted values are correlated respectively, as they are in a homoscedastic linear model with normally distributed errors. Besides that, the plots indicate dependency between the residuals and the fitted values. Based on the plots, it can be summarized that outliers do exist in CIC.

There are two robust regression methods in this study were comparatively evaluated against OLS regression method. These robust methods consist of M-estimation and LTS estimator that are most commonly used regression method when data contains outliers. One of the effective performance statistics is the coefficient of determination (*R*-square). The results of the parameter estimations based on different estimators and *R*-squared including *p*-values of each models are given at Table 1. Is it clearly seen that the parameters estimation of each model are slightly different to each other when it is being compared. For example, The M-estimator produces a much larger negative impact of CIC than LTS and OLS (-10063.81 versus -5732.80 versus 519.34).

Table 1: Estimation Parameter based on different estimator

Model	Equation	R^2	*F*-test/ (*p*-value)
OLS	CIC = 519.38 – 187.17*EXC + 0.05*EXT + 0.38*RM	0.9461	1265/ (0.0000)
M-Est	CIC = -10063.81 + 2450.32*EXC + 0.02*EXT + 0.51*RM	0.9970	1283/ (0.0000)
LTS	CIC = -5732.80 + 51.82*EXC + 0.223*EXT + 1.86*RM	0.9978	146.54/ (0.0000)

Besides that, the corresponding *p*-value of 0.0000 based on *F*-test for each model using different estimator indicate a strong rejection of the null hypothesis that all non-intercept coefficients are equal to zero. In other words, most of model are statistically significant. Meanwhile, the *R*-squared value of each model in Table 1 shows a strong positive correlation

respectively. By comparing each estimator for each model, CIC is more fitted when using LTS estimator followed by M and OLS estimator. In conclusion, LTS estimator outperform M and OLS estimator in terms of maximum of *R*-squared values.

4. Conclusion

Multiple regression is a popular statistical tool used by researchers to explain or determine the relationships between a dependent variable and multiple independent variables.

In this study, two robust regression methods, LTS and M-estimator were comparatively evaluated against OLS regression method. Table 1 shows a comparison of all results, respectively. As seen from these table, therefore, it can be concluded that, LTS was the best method that fits data well, the second better method was the M method and OLS was the last efficient method in this study. Mean square error (MSE) in a simulation study can be used as criteria to evaluate the estimator as a recommendation for a future study.

5. Acknowledgment

The authors would like to express their appreciation for the support of the sponsors with Project Vote: U556 by Research and Innovation Centre (R&D), Research, Innovation, Commercialization & Consultancy Office (ORICC) Universiti Tun Hussein Onn Malaysia. This paper was partly sponsored by the Centre for Graduate Studies UTHM. Both supports is gratefully acknowledged.

References

[1] Uyanik, G. K., & Guler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240.
[2] Wen, Y. W., Tsai, Y. W., Wu, D. B. C., & Chen, P. F. (2013). The Impact of Outliers on Net-Benefit Regression Model in Cost- Effectiveness Analysis. *PLoS ONE*, *8*(6), 1–9.
[3] Cetin, M., & Toka, O. (2011). The Comparison of S-estimator and M-estimators in Linear Regression. *Journal of Science*, *24*(4), 1–6.
[4] Schumacker, R. E., Monahan, M. P., & Mount, R. E. (2002). A Comparison of OLS and Robust Regression using S-Plus. *Multiple Linear Regression Viewpoints*, *28*(2), 10–13.
[5] Abd-Almonem, M. (2015). Robust Methods in Regression Analysis: Comparison and Improvement. *Journal of Science*, 38–53.
[6] Berk, R. A. (1990) A Primer on Robust Regression. M*odern Methods of Data Analysis Newbury Park*, CA: Sage Publications, Inc, 292-323.
[7] Birkes, D. & Dodge,Y. (1993). *Alternative Methods of Regression*, John Wiley & Sons, Canada.
[8] Staudte, R. G. & Sheatherm, S. J. (1990). *Robust Estimation and Testing*, John Wiley & Sons, New York.
[9] Wilcox, R. R. (1997). *Introduction to Robust Estimation and Hypothesis Testing*, CA:Academic Press, San Diego.
[10] Nasiru, S., Luguterah, A., & Anzagra, L. (2013). The Efficacy of ARIMAX and SARIMA Models in Predicting Monthly Currency in Circulation in Ghana, *3*(5), 73–81.
[11] Luguterah A., Suleman N., Anzagra L., (2013). A Predictive Model for Monthly Currency in Circulation in Ghana. Mathematical Theory and Modeling, 3(4): 43-52.
[12] Ikoku, A. (2014). Modeling and Forecasting Currency in Circulation for Liquidity Management in Nigeria, *5*(1), 79–104.
[13] Chinnammai, S. (2013). A Study on Currency and Coinage Circulation in India. *International Journal of Trade, Economics and Finance*, *4*(1), 43–47.
[14] Okereke, O. E., Ire, K. I., & Irokwe, O. (2015). Time Series Analysis of Currency in Circulation in Nigeria. *Journal of Natural Sciences Research*, *5*(19), 51–56.
[15] Cassino, V., Misich, P., & Barry, J. (1994). Forecasting the demand for currency. *Reserve Bank Bulletin*, *601*(1), 27–33.
[16] Obinska-Wajda, E. (2016). The New Institutional Economic Main Theories. *"E-Finanse" Financial Internet Quarterly*, *12*(1), 78–85.
[17] Alma, O. G. (2011). Comparison of Robust Regression Methods in Linear Regression. *International Journal Contemp. Math. Sciences*, *6*(9), 409–421.
[18] Naugher, H. J. K. (2000). Outliers lie: An illustrative example of identifying outliers and applying robust models. M*ultiple Linear Regression Viewpoints*, 26(2), 2-6.
[19] Anscombe, F. J. (1960) Rejection of outliers. *Technometrics*, 123–147.
[20] Cankaya, S., & Abacı, S. H. (2015). A Comparative Study of Some Estimation Methods in Simple Linear Regression Model for Different Sample Sizes in Presence of Outliers. *Journal of Agriculture-Food Science and Technology*, *3*(6), 380–386.
[21] Donoho, D. L. & Huber P. J. (1983). The notion of breakdown point. *Wadsworth*, 157–184.
[22] Hampel, F. R., Ronchetti, E. M., Rousseeuw P. J., and Stahel, W.A. (1986) *Robust Statistics. The Approach based on Influence Functions*, John Wiley and Sons, New York.
[23] Bhatia, K. K. S. & Bha, A. K. (1975). Analysis of rainfall distribution in monsoon season. *Indian forester*,238-248.
[24] Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73–101.
[25] Hampel, F. R. (1974) . The influence curve and its role in robust estimation. J*ournal of the American Statistical Association*, 69, 383–393.
[26] Tukey, J. W. (1977). *Exploratory data analysis*. New York: Addison-Wesley.
[27] Rousseeuw, P. J. (1984) Least median of squares regression. *Journal of the American Statistical Association*, 871–880.
[28] Rousseeuw, P.J. and Leroy, A.M. (1987). Robust Regression and Outlier Detection. *Series in Applied Probability and Statistics*, 1-329.
[29] Williams, R. (2015). Review of Multiple Regression. *Journal of Science*, 1–12.