IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 3 Ver. III (May - June 2017), PP 33-37
www.iosrjournals.org

Comparison of Adaptive and M Estimation in Linear Regression

Chikhla Jun Gogoi[1], Bipin Gogoi[2]
[1]*Research Scholar Department of Statistics Dibrugarh University*
[2]*Professor Department of Statistics Dibrugarh University*

***Abstract**: In the presence of outliers least square estimation is insufficient and can be biased. Adaptive estimation is robust estimation that can be used to improve the accuracy of the estimate by reducing the influence of outliers. Adaptive estimators can be effective in achieving low mean squared error for a variety of non normal distributions of errors. M estimation is the extension of the maximum likelihood estimation and is also a robust estimation. In this paper, comparison is made between OLS Estimation, Adaptive estimation and M estimation under a particular situation (An example).*
***Keyword**: OLS estimation, Adaptive estimation, M estimation, Robust regression,.*

I. Introduction

Adaptive estimation is a robust estimation. The objective of adaptive estimation is to improve the accuracy of the estimate by reducing the influence of outliers. When some observations are adaptively downweighted, the influence of outliers is greatly reduced and then adaptive estimation can be considered. M estimation is the extension of the maximum likelihood estimation. It is also a robust estimation.

1.1 Objectives of Adaptive Estimation:

Our objective in Adaptive regression is to develop an estimator that
1. uses information from all the data points if the error distribution appears to be non normal.
2. effectively downweights outliers so that their influence is limited .
3. is robust in the sense that small changes in the data will not greatly change the estimates.
4. can be computed easily.

1.2 The multiple Linear Regression Model:

In this section for the estimation of parameters in a linear model a WLS approach will be used.
The multiple regression model is
$y_i = \beta_0 x_{i,0} + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + \varepsilon_i$ for i= 1,2,...,n
In matrix form
Y=Xβ+ε
Where Y is the n$\times$ 1 vector containing the dependent variable, X is an n$\times$(p+1) matrix containing the independent variables, β is the (p+1)$\times$ 1 vector of parameters to be estimated and ε is the n$\times$ 1vector of errors.

1.3 Adaptive Estimation:

We begin the adaptive estimation by comparing the standardised deleted residuals
$d_i = e_i[\frac{n-p-2}{SSE(1-h_{ii})-e_i^2}]^{1/2}$ for i= 1,2,...,n
Where $e_i$is the ordinary residual, h_{ii} is the ith diagonal element of the hat matrix $X_R(X_R^{/}X_R)X_R^{/}$ and SSE is the usual sum of squared residuals from the regression based on the n observations in the reduced model.
We will use all p+1 independent variables in this model. We will weight the observations so that the c.d.f of the studentised deleted residuals, after weighting, will approximate the c.d.f of the t distribution with v= n-(p+1)-1 =n-p-2 df , which will be denoted by $T_{n-p-2}(.)$.
We then smooth the c.d.f of these standardized deleted residual*s by*using a normal kernel with a bandwith of h= 1.587 $\hat{\sigma}\ n^{-1/3}$, as suggested by Polansky (1998). Since the observations are studentised the variance should not depart too much from $\sigma^2 = 1,$ so a value of h=1.587 $n^{-1/3}$is used to obtain the smoothed distribution function.
Let D={$d_1, d_{2,} ..., d_n$} be the set of studentised deleted residuals. The smoothed c.d.f. at point d over the set of all studentised deleted residuals (D) is computed as

$$\widehat{F_h}(d;D) = \frac{1}{n}\sum_{i=1}^{n} \emptyset\left(\frac{d-d_i}{h}\right)$$

Where $\emptyset(.)$ is the c.d.f. of the standard normal distribution. After the smoothed c.d.f. of these studentized deleted residuals is obtained, we centre the studentised deleted residuals by subtracting the estimated median $\tilde{d}$, which is determined by a search process so that
$\widehat{F_h}(\tilde{d}; D) = 0.5$. The centered studentised deleted residuals are calculate as
$d_{c,i} = d_i - \tilde{d}$ for i= 1,...,n and will be called the residuals. The set of residuals will be denoted by $D_c = \{d_{c,1}, \dots, d_{c,n}\}$ and we will let $t_i = T_{n-p-2}^{-1}\left(\widehat{F_h}(d_{c,i}; D_c)\right)$. to weight the observations we use
$w_i = \frac{t_i}{d_{c,i}}$ for i= 1,...,n. If the error terms are normally distributed, then smoothed c.d.f. of the centered studentised deleted residuals should approximate the c.d.f. of the t distribution with v=n-p-2 degrees of freedom and the weights should approximate one. If the ith observation is an outlier, then $d_{c,i}$ will be large relative to t_i so that the ith observation will be given a small weight. After we have computed the weights wi for i= 1,2,...,n, they can be used as the diagonal elements in the weighing matrix W with zero off- diagonal elements. We perform the WLS regression by premultiplying both sides of the model by W to obtain
WY=WXβ+Wε
This can be written as the transformed model
$Y^* = X^*\beta + \varepsilon^*$
Where $Y^* = WY, X^* = WX \; and \; \varepsilon^* = W\varepsilon$. then we will use the OLS method to compute the parameter estimates.

1.5 M Estimation
M estimation is a robust estimation and is a estimation of maximum likelihood type.
If estimator at M estimation is $\hat{\beta} = \beta_n(x_1, x_2, \dots, x_n)$then
E[$\beta_n(x_1, x_2, \dots, x_n)$] = β ... (1)
Equation (1) shows that the estimator $\hat{\beta} = \beta_n(x_1, x_2, \dots, x_n)$ unbiased and has minimum variance, so M-estimator has the smallest variance estimator compared to other estimators of variance:

$$\text{var}(\hat{\hat{\beta}}) \geq \frac{[\widetilde{\beta'}]^2}{nE(\frac{d}{d\beta} lnf(x_i;\beta))^2}$$

where $\hat{\hat{\beta}}$ is the other linear and unbiased estimator of β. M estimation is an extension of the maximum likelihood estimate method and a robust estimation. In this method it is possible to eliminate some of the data , which in some cases is not always appropriate to do especially if it is eliminated is an important data or seed, whose case often encountered in agriculture. M estimation principle is to minimize the residual function ρ:

$$\widehat{\beta_M} = \min \rho(y_i - \textstyle\sum_{j=0}^{k} x_{ij}\beta_j) \qquad (2)$$

We have to solve

$$min \sum_{i=1}^{n} \rho(u_i) = min \sum_{i=1}^{n} \rho(\frac{e_i}{\sigma}) = min \sum_{i=1}^{n} \rho(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{\sigma})$$

To obtain (2) and we set estimator for σ:

$$\hat{\sigma} = \frac{MAD}{.6745} = \frac{median|e_i - median(e_i)|}{.6745}$$

For ρ function we use Tukey's bisquare objective function:

$$\rho(u_i) = \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^4}, \quad |u_i| \leq c$$
$$= \frac{c^2}{6} \quad |u_i| > c$$

Furthermore we look for first partial derivative $\widehat{\beta_M}$to β so that

$$\sum_{i=1}^{n} x_{ij}\varphi(\frac{y_i - \sum_{j=0}^{k} x_{i,j}\beta}{\hat{\sigma}}) = 0, j = 1,2, \dots, k \qquad (3)$$

Where $\varphi = \rho^{/}, x_{ij}$ is the ith observation on the jth independent variable and $x_{i0} = 1$
Draper and Smith give a solution for equation (3)by defining a weighted function

$$\text{W}(e_i) = \frac{\varphi(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}})}{\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta}{\hat{\sigma}}} \qquad (4)$$

Because $u_i = \frac{e_i}{\hat{\sigma}}$, we can rewrite equation (4) with

$$w_i = [1 - (\frac{u_i}{c})^2]^2 \quad ,|u_i| \leq c$$
$$=0 \quad |u_i| > c$$

We take c= 4.685 for Tukey's bisquare weighted function. So equation (3) becomes

$\sum_{i=1}^{n} x_{ij} w_i (y_i - \sum_{j=1}^{k} x_{ij} \beta) = 0$ j=0,1,2,...,k (5)

Equation (5) can be solved by iteratively reweighted least squares method. In this method we assume that there is an initial estimate $\widehat{\beta^0}$ and $\widehat{\sigma_i}$ is a scale estimate. If j is the number of parameters then

$\sum_{i=1}^{n} x_{ij} w_i^0 (y_i - \sum_{j=0}^{k} x_{ij} \beta^0) = 0, \quad j = 0,1,2,\ldots,k$ (6)

In matrix notation, equation (6)can be written as

$$X^{/} W_i X \beta = X^{/} W_i Y$$

Where W_i is a n× n matrix with its diagonal elements are weighted. Equation (7) is known as weighted least square equation. Solution for this equation gives an estimator for β, i.e. $\hat{\beta} = \left(X^{/} W_i X\right)^{-1} \left(X^{/} W_i Y\right)$.

Algorithm:

1. Estimate regression coefficients on the data using OLS.
2. Test assumptions of the regression model.
3. Detect the presence of outliers in the data.
4. Calculate estimated parameter $\widehat{\beta^0}$ with OLS.
5. Calculate residual value $e_i = y_i - \hat{y}_i$
6. Calculate value $\hat{\sigma}_i = 1.4826MAD$
7. Calculate value $u_i = \frac{e_i}{\hat{\sigma}_i}$
8. Calculate the weighted value

$$w_i = \left[1 - \left(\frac{u_i}{4.685}\right)^2\right]^2, \quad |u_i| \leq 4.685$$
$$= 0 \qquad , \quad |u_i| > 4.685$$

9. Calculate $\hat{\beta}_M$ using weighted least squares (WLS) method with weighted w_i.
10. Repeat steps 5-8 to obtain a convergent value of $\hat{\beta}_M$.
11. Test to determine whether independent variables have significant effect on the dependent variable.

1.5 Example:

The data shown in the table below are from a study by Coleman et al. (1966) has concerned the relationship of several factors to the mean verbal test scores of sixth graders(y). The data, which were also analysed by Mosteller and Tukey(1977) and by Rousseeuw and Leroy (1987), include staff salaries per pupil (x1), percent of white collar fathers (x2), socioeconomic status composite deviation (x3), mean of teachers' verbal test scores (x4), and mean of mothers' educational level (x5).

Table: Data on the mean verbal scores of sixth graders in 20 schools.

School	Y	x1	x2	x3	x4	x5
1	37.01	3.83	28.87	7.2	26.6	6.19
2	26.51	2.89	20.1	-11.71	24.4	5.17
3	36.51	2.86	69.05	12.32	25.7	7.04
4	40.7	2.92	65.4	14.28	25.7	7.1
5	37.1	3.06	29.59	6.31	25.4	6.15
6	33.9	2.07	44.82	6.16	21.6	6.41
7	41.8	2.52	77.37	12.7	24.9	6.86
8	33.4	2.45	24.67	-0.17	25.01	5.78
9	41.01	3.13	65.01	9.85	26.6	6.51
10	37.2	2.44	9.99	-0.05	28.01	5.57
11	23.3	2.09	12.2	-12.86	23.51	5.62
12	35.2	2.52	22.55	0.92	23.6	5.34
13	34.9	2.22	14.3	4.77	24.51	5.8
14	33.1	2.67	31.79	-0.96	25.8	6.19
15	22.7	2.71	11.6	-16.04	25.2	5.62
16	39.7	3.14	68.47	10.62	25.01	6.94
17	31.8	3.54	42.64	2.66	25.01	6.33
18	31.7	2.52	16.7	-10.99	24.8	6.01
19	43.1	2.68	86.27	15.03	25.51	7.51
20	41.01	2.37	76.73	12.77	24.51	6.96

The estimated parameters are:

Parameters	OLS estimation	M Estimation	Adaptive Estimation
β_0	19.89	19.60	19.90
β_1	-1.791	-1.793	-1.790
β_2	.04362	0.04650	0.04360
β_3	.5561	.5570	.5556
β_4	1.12	1.12	1.11
β_5	-1.812	-1.79	-1.81

The estimated regression line under OLS estimation is
$Y=19.89-1.791x_1+.04362x_2+.5561x_3+1.12x_4-1.812x_5$
The estimated regression line under M estimation is
$Y=19.60-1.793x_1+.04658x_2+.5570x_3+1.12x_4-1.79x_5$
The estimated regression line under Adaptive estimation is
$Y=19.90-1.790x_1+.04360x_2+.5556x_3+1.11x_4-1.81x_5$

The MSE under OLS estimation is 4.354214
The MSE under M estimation is 4.351321
The MSE under Adaptive estimation is 4.30

1.6 Example:

The data given below was collected by Haith (1976) on water quality and land use in 20 river basins in New York State. These data, which are listed in the yable below , have been analysed by several authors, including Simpson et al. (1992) and Ryan(1997). The data include the nitrogen concentration (mg/L) of river water and several land use variables. The nitrogen concentration will be used as the dependent variable. The independent variables are the percentages of commercial, agricultural, forest, and residential land in river basin.

Table: the New York rivers data set.

	River basin	Y(Nitrogen)	x1(Commercial)	x2(Agric.)	x3(Forest)	x4(Res.)
1	Olean	1.10	0.29	26	63	1.2
2	Cassadaga	1.01	0.09	29	57	0.7
3	Oatka	1.90	0.58	54	26	1.8
4	Neversink	1.00	1.98	2	84	1.9
5	Hackensack	1.99	3.11	3	27	29.4
6	Wappinger	1.42	0.56	19	61	3.4
7	Fishkill	2.04	1.11	16	60	5.6
8	Honeoye	1.65	0.24	40	43	1.3
9	Susquehanna	1.01	0.15	28	62	1.1
10	Chenango	1.21	0.23	26	60	0.9
11	Tioughnioga	1.33	0.18	26	53	0.9
12	West Canada	0.75	0.16	15	75	0.7
13	East Canada	0.73	0.12	6	84	0.5
14	Saranac	0.80	0.35	3	81	0.8
15	Ausable	0.76	0.35	2	89	0.7
16	Black	0.87	0.15	6	82	0.5
17	Schohari	0.80	0.22	22	70	0.9
18	Raquette	0.87	18.00	4	75	0.4
19	Oswegatchie	0.66	13.00	21	56	0.5
20	Cohocton	1.25	0.13	40	49	1.1

Table: The estimated parameters are:

Parameters	OLS estimation	M Estimation	Adaptive Estimation
β_0	4.01	4.012	3.950
β_1	-.0323	-0.0324	-0.0328
β_2	-0.02310	-0.02309	-0.02401
β_3	-0.03610	-0.03607	-0.03680
β_4	-0.0245	-0.0239	-0.0247

The estimated regression line under *OLS* estimation is
$Y=4.01-.0323\ X_1-.02310\ X_2-.03610\ X_3\ -.0245\ X4$
The estimated regression line under M estimation is
$Y=4.012-.0324\ X_1\ -.02309\ X_2-.03607\ X_3-.0239\ X_4$
The estimated regression line under Adaptive estimation is
$Y=3.950-.0328X_1-.02401X_2-.03680X_3-.0247X_4$

The MSE under OLS estimation is 0.07277
The MSE under M estimation is 0.072877
The MSE under Adaptive estimation is 0.092212

II. Conclusion

In the first example it can be observed that MSE of the regression model under Adaptive estimation is smaller than the OLS estimation and M estimation. So according to this result the Adaptive estimation is suitable for the data. But in the second example the MSE under Adaptive estimation is greater than the OLS and M estimation, so either OLS or M estimation is applicable to the data.

References

[1] Badan Pusat Statistik, Production of Paddy Maize and Soybeans, www.bps.go.id/release/Production of Paddy Maize and Soybeans, 2012.

[2] Birkes D. and Dodge, Y. ,(1993) Alternative Methods of Regression, John Wiley Sons Inc., New York, .

[3] Chen,C., Robust Regression and *Outlier* Detection with the ROBUSTREG Procedure, Statistics and Data Analysis, paper 265-27, SAS Institute Inc., Cary, NC.

[4] Draper,N.R. and Smith,H.(1998). Applied Regression Analysis, Third Edition, Wiley Interscience Publication, United States,.

[5] Haith, D.A.(1976). Land use and water quality in New York rivers. Journal of the Environmental Engineering Division, Proceedings of the American Society of Civil Engineers, 102, 1-15

[6] Montgomery D.C. and Peck,E.A.(2006) An Introduction to Linear Regression Analysis, John Wiley Sons Inc., New York, .

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

[8] Rousseeuw P.J.and Yohai,V.J.(1984). Robust Regression by Mean of S Estimators, Robust and Nonlinear Time Series Analysis, New York, , 256-274, doi: 10.1007/978-1-4615-7821-5-15.

[9] Ryan,T.p.(1997). Modern Regression Methods, Wiley , New York.

[10] Salibian,M., Yohai,V.J.(2006). A Fast Algoritm for S-Regression Estimates, Journal of Computational and Graphical Statistics, 15, No. 2 , 414-427,doi: 10.1198/106186006X113629. 360 Y. Susanti, H. Pratiwi, S. Sulistijowati H., T. Liana

[11] Simpson,D.G.,Ruppert ,D.,and Caroll,R.J.(1992). On one step GM estimates and stability of inferences in linear regression. Journal of the American Statistical Association, 87,439-450

[12] Susanti,Y. and Pratiwi,H.(2011). Robust Regression Model for Predicting the Soybean Production in Indonesia, Canadian Journal on Scientific and Industrial Research, 2, No. 9 , 318-328.

[13] Susanti,Y., Pratiwi,H., and Liana, T.(2009).Application of M-estimation to Predict Paddy Production in Indonesia, presented at IndoMS International Conference on Mathematics and Its Applications (IICMA), Yogyakarta, .

[14] Yohai,V.J.,(1987). High Breakdown Point and High Efficiency Robust Estimates for Regression, The Annals of Statistics, 15, No. 20 , 642-656, doi: 10.1214/aos/1176350366.