New M-Estimator Objective Function in Simultaneous Equations Model (A Comparative Study)

Ahmed H. Youssef

Professor of Applied Statistics and Econometrics
Institute of Statistical Studies and Research, Cairo University, Egypt

Ahmed Amin El-Sheikh

Associated Professor of Applied Statistics and Econometrics
Institute of Statistical Studies and Research, Cairo University, Egypt

Eman Taha Hassan Mohammed

Master student of Applied Statistics and Econometrics,
Institute of Statistical Studies and Research, Cairo University, Egypt

Copyright © 2013 Ahmed H. Youssef et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this paper, a suggested M-estimator (S-M) objective function will be introduced. In addition, A comparative study among different two-stage M-estimation functions including S-M and two stage least squares (2SLS) has been introduced using Monte Carlo Simulations to investigate the properties of these functions, when the errors term of the reduced form follows heavy tail distributions.

Keywords: Two-stage M- estimation, simultaneous equations model, two stage least squares (2SLS), robust estimation, heavy tail distribution, Huber M-estimator, Tukey M-estimator

1-Introduction

Least squares (LS) estimates when the errors are heavy-tailed distributed has certain disadvantages that LS has low breakdown point, which tends to zero as the sample size increases. Also, LS estimators may not be efficient or asymptotically efficient which mean that LS estimates will not be BLUE. One treatment is to remove influential observations from the least squares, which is called diagnostic technique.

Robust regression analysis provides an alternative approach to the LS regression model when the fundamental assumptions are unfulfilled by the nature of the data. M-estimation method is one of the most common method of robust regression. Huber (1964) has introduced it as a new approach towards a theory of robust estimation; He treated the asymptotic theory of estimating a location parameter for contaminated normal distributions. Also, in (1973) he introduced M- estimator of regression.

Kim and Muller (2007) have introduced two stage Huber estimator (2SH) to get a simple, a robust and an efficient estimator in the case of random regressors and asymmetric errors stage. They concluded that 2SH estimator has smaller standard errors than (2SLS) and two stage deviations (2SLAD).

Simultaneous equations model can be considered as a tool for investigating the interdependence between variables that is difficult to explain by single-equation models. The method of two – stage least squares (2SLS) can be considered the most common method used for estimating simultaneous equations models which based on the least squares regression estimation. Greene (2008)

This paper aims to introduced a suggested M- estimator objective function, apply it to estimate simultaneous two equations model, and study the properties of this function compared to different M- estimator objective functions 2SH and Tukey M-estimator. The effective of vertical outliers in endogenous variable (heavy tail errors) on properties of estimators will be investigated using Monte Carlo simulations which are executed to evaluate the performance of different robust M-estimators.

2- M- Estimator

M-estimator is considered as the next step in the direction of the robust estimation, where Huber (1964) has introduced it as a new approach towards a theory of robust estimation. In (1973), he extended the idea of M-estimator for regression by minimizing a symmetric smooth function of the residuals over the parameter estimates, which can presented as follows:

$$\sum_{i=1}^n \rho(r_i) = \sum_{i=1}^n \rho(y_i - x_i' \hat{\beta}) \quad \text{or} \quad \sum_{i=1}^n \rho(r_i) = \sum_{i=1}^n \rho\left(\frac{y_i - x_i' \hat{\beta}}{\hat{\sigma}}\right), \quad (1)$$

Where, $\hat{\sigma}$ the estimated scale of residuals, ρ is a symmetric, positive definite function with a unique minimum at zero, and is chosen to be less increasing than square.

Let, $\beta = (\beta_1, \dots, \beta_p)$ is a vector of regression model parameters including the intercept term, The M-estimator of β based on the function $\rho(r_i)$ is the vector $\hat{\beta}$ which is the solution of the following p equations;

$$\sum_{i=1}^n \psi(r_i) \frac{\partial r_i}{\partial \beta_j} = 0, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p, \quad (2)$$

Where, $\psi(r_i)$ is called Psi function. Also, it is called the influence function, which measures the influence of a datum i on the value of the parameter estimate.

The M- estimator for β , based on the function ρ and the data is the value $\hat{\beta}$ which minimizes $\sum_{i=1}^n \psi(y_i - x_i \beta)$, $\hat{\beta}$ is determined by solving the set of p simultaneous equations as

$$\sum_{i=1}^n \psi(y_i - x_i \beta) x_i' = 0, \quad (3)$$

By defining the weighted function $w(r) = \frac{\psi(r)}{r}$, and let, $w_i = w(r_i)$, hence the p simultaneous equations in (3) becomes:

$$\sum_{i=1}^n (y_i - x_i \beta) w_i x_i' = 0, \quad (4)$$

Equation (4) can be combined into the following single matrix equation

$$X'WX\beta = X'Wy$$

And hence

$$\hat{\beta} = (X'WX)^{-1} X'Wy, \quad (5)$$

Equation (5) can be solved as weighed least squares (WLS). Bhar (2010).

It is noted that weights depend upon the residuals, the residuals depend upon the estimated coefficients, and at the same time the estimated coefficients depend upon the weights, so an iterative solution, which is called iteratively reweighed least-squares (IRLS) is required in this case as:

1. Select initial estimates $\hat{\beta}_0$ such as the least-squares estimates.
2. At each iteration t , calculate residuals $r_i^{(t-1)}$ and associated weights

$$w_i^{(t-1)} = w[r_i^{(t-1)}] \text{ from the previous iteration.}$$

3. Solve WLS estimates as:

$$\hat{\beta}^{(t)} = [X'W^{(t-1)}X]^{-1} X'W^{(t-1)}y,$$

where X is the matrix of the model, and $W^{(t-1)} = \text{diag} \{w_i^{(t-1)}\}$ is the current weighted matrix.

4. Steps 2. and 3 are repeated until the estimated coefficients converge. Fox (2002) continued until consecutive estimates $\hat{\beta}^{(t-1)}$ and $\hat{\beta}^{(t)}$ are sufficiently close to one another.

3- The 2SLS Estimation

Consider the structural equation model as

$$y_t = Y_t' \gamma + x_{1t}' \beta + u_t, \quad t = 1, 2, \dots, T, \quad (6)$$

where y_t is the dependent variable, Y_t is a vector ($G \times 1$) of endogenous variables, x_1 is a vector ($K_1 \times 1$) of exogenous random variables and u_t is the error term. Kim and Muller (2007) have interested in estimating the structural parameters $\alpha_0 = (\gamma', \beta')'$, they denoted x_{2t} the vector ($K_2 \times 1$) of exogenous random variables that are absent from equation (6). Assuming that, $Y_t' = (Y_{1t}, Y_{2t}, \dots, Y_{Gt})$, then the reduced form can be represented for each Y_{jt} as

$$Y_{jt} = x_t' \Pi_j + V_{jt}, \quad \begin{matrix} j = 1, \dots, G, \\ t = 1, 2, \dots, T \end{matrix} \quad (7)$$

Where $x_t' = (x_{1t}', x_{2t}')$, Π_j is a vector ($K \times 1$) of unknown parameters with $K = K_1 + K_2$, and V_{jt} is the error term of the reduced form.

The 2SLS estimator is obtained basing on LS regression of y_t on Y_t and x_1 . Thus, the 2SLS name arises from the two regressions in the procedure:

1. Stage 1. Obtain the least squares predictions from regression of Y_t on x_t .

2. Stage 2. Estimate α by least squares regression of y_t on $\hat{Y}_t$ and x_1 . Green (2008)

4-Two-Stage Huber Estimation

The first stage of M- estimation that yields $\hat{\Pi}$ as estimate of Π_j in (7) would be obtained as the solution of

$$\sum_{t=1}^T \rho(Y_{jt} - x_t' \Pi_j) \quad (8)$$

The second stage of Huber estimation that yields $\hat{\alpha}$ as estimate of $\alpha_0 = (\gamma', \beta')'$ in (6) would be obtained as the solution of

$$\min \sum_{t=1}^T \rho(y_t - \hat{z}_t' \alpha) \quad (9)$$

where $\hat{z}_t = [\hat{Y}_{jt}, x_{1t}]$. Huber function has to be modified using some scale

estimator of the error term so, Flavin (1999), has been used a standardized MAD to obtain the required scale estimators as follows:

1- $\hat{\sigma}_j = 1.483MAD$, where $MAD = \text{median}\{|r_{jt} - \text{median}\{r_{jt}\}|\}$ for equation (8)

2- $\hat{\sigma}_t = 1.483MAD$, where $MAD = \text{median}\{|r_t - \text{median}\{r_t\}|\}$ for equation (9)

where r_{jt} are the residuals obtained from the LS regression of Y_{jt} on x_t and r_t are the residuals from the LS regression of y_t on $\hat{z}_t$.

a- Huber Objective Function

Huber M- estimator that we considered is represented as

$$\rho_H(r_i/\hat{\sigma}) = \begin{cases} \frac{1}{2} \left(\frac{r_i}{\hat{\sigma}} \right)^2 & \rightarrow \text{for } \left| \frac{r_i}{\hat{\sigma}} \right| \leq c \\ c \left| \frac{r_i}{\hat{\sigma}} \right| - \frac{1}{2} c^2 & \rightarrow \text{for } \left| \frac{r_i}{\hat{\sigma}} \right| > c \end{cases} \quad i = 1, \dots, n \quad (10)$$

where ρ_H a smooth and symmetric function, c is a tuning value, we use $c=2$

b- Huber Weight Function

The Huber weight function of the residuals for equation (10) according to Flavin (1999) is defined as

$$w_H(r_i) = \begin{cases} 1 & \text{for } \left| \frac{r_i}{\hat{\sigma}} \right| \leq c \\ \frac{c \hat{\sigma}}{|r_i|} & \text{for } \left| \frac{r_i}{\hat{\sigma}} \right| > c \end{cases} \quad i = 1, 2, \dots, n \quad (11)$$

where r_i is a general concept for residuals whether in the first stage or in the second stage in 2SLS.

5- Tukey M- Estimator

In addition to Huber M- estimator, Tukey M- estimator is considered as one of the most famous M- estimator. Tukey objective function is defined as

$$\rho_T(r_i) = \begin{cases} \frac{k^2}{6} \left\{ 1 - \left[1 - \left(\frac{r_i}{k} \right)^2 \right]^3 \right\} & \text{for } |r_i| \leq k, \\ \frac{k^2}{6} & \text{for } |r_i| > k, \end{cases} \quad i = 1, \dots, n \quad (12)$$

where k is a tuning value, $k = 4.685$, and the Tukey weight function $w_T(r_i)$ is defined as:

$$w_T(r_i) = \begin{cases} \left[1 - \left(\frac{r_i}{k} \right)^2 \right]^2 & \text{for } |r_i| \leq k \\ 0 & \text{for } |r_i| > k \end{cases}, \quad i = 1, \dots, n \quad (13)$$

where $w_T(r_i)$ denotes Tukey residuals weight function. Fox (2002)

6- Suggested M - Estimator (S-M)

In this section, suggested M- objective function (S-M) has been introduced, which combined between LS objective function and another logistic function that approaches the idea of Cauchy function that has the form,

$$\rho_{ch} = \frac{c^2}{2} \log \left(1 + \left(\frac{r_i}{c} \right)^2 \right), \quad i = 1, \dots, n \quad (14)$$

Where c is a tuning value, the 95% asymptotic efficiency on the standard normal distribution is obtained with the tuning constant $c=2.3849$. Bhar (2010).

According to LS, and Cauchy function in (14), the suggested function can be defined as

$$\rho_{S-M}(r_i) = \begin{cases} \frac{r_i^2}{2} & \text{if } |r_i| < k \\ \ln(r_i^2 + 1) & \text{if } |r_i| \geq k \end{cases} \quad i = 1, 2, \dots, n \quad (15)$$

where k is a tuning value and $\rho_{S-M}(r_i)$ satisfies the same properties of ρ above.

The first derivative of (15), which is called influence function,

$$\psi_{S-M}(r_i) = \begin{cases} r_i & \text{if } |r_i| < k \\ \frac{2r_i}{(r_i^2 + 1)} & \text{if } |r_i| \geq k \end{cases} \quad i = 1, 2, \dots, n \quad (16)$$

Since, weighted function w_{S-M} is defined as $\frac{\psi_{S-M}(r_i)}{r_i}$ so,

$$w_{S-M}(r_i) = \begin{cases} 1 & \text{if } |r_i| < k \\ \frac{2}{r_i^2 + 1} & \text{if } |r_i| \geq k \end{cases} \quad i = 1, 2, \dots, n \quad (17)$$

We graph the ρ -function, Ψ -function and the weight function in Fig. 1.

Figure 1: Graph of S-M function: (A) Objective function (B) Ψ -function (C) Weight function.

To determine the value of k which makes the estimator more efficient, a simulation study has been executed by considering the weight function in equation (17) at different values of k (1, 1.5, 2, 2.5, and 3) for different samples sizes (15, 30, 50, 100) when the errors follow Chi-square distribution with 2 degrees of freedom, ($\text{Chisq}(n,2)$) as a heavy tailed distribution.

The results of the simulation are shown in tables (1) and (2), which display the properties of the S-M function at different values of k . These tables show the estimates of the three parameters (γ , β_0 , and β_1), bias and mean square errors \

(MSE) compared with 2SLS. It can be seen from tables (1) and (2) when $k = 3$ the function has least MSE for most different sample sized except for $n = 30$ but at the same time it has the less bias. So, we will use S-M with $k = 3$ (S-M (3)) in our comparison to investigate its performance compared with other M- estimator functions based on Bias and MSE.

Table (1): Estimated parameters, Bias and MSE: (n = 15, 30)

Criteria		2SLS	S-M- Estimator				
			$k=1$	$k=1.5$	$k=2$	$k=2.5$	$k=3$
$n=15$	$\hat{\gamma}$	0.024	0.492	0.49	0.499	0.497	0.506
	Bias	0.524	0.008	0.01	0.001	0.003	0.006
	MSE	0.361	1.149	1.147	1.139	1.111	1.113
	$\hat{\beta}_0$	7.109	2.581	2.647	2.634	2.644	2.563
	Bias	6.109	1.581	1.647	1.634	1.644	1.563
	MSE	44.118	64.367	64.621	64.63	63.329	63.336
	$\hat{\beta}_1$	0.496	0.188	0.202	0.177	0.174	0.17
	Bias	0.296	0.012	0.002	0.023	0.026	0.03
	MSE	0.929	1.352	1.301	1.367	1.34	1.325
$n=30$	$\hat{\gamma}$	0.299	0.468	0.467	0.449	0.466	0.482
	Bias	0.201	0.032	0.033	0.051	0.034	0.018
	MSE	0.241	0.66	0.681	0.687	0.689	0.646
	$\hat{\beta}_0$	4.531	2.659	2.683	2.838	2.714	2.573
	Bias	3.531	1.659	1.683	1.838	1.714	1.573
	MSE	26.801	35.73	37.628	38.934	38.489	35.695
	$\hat{\beta}_1$	0.421	0.256	0.259	0.263	0.245	0.255
	Bias	0.221	0.056	0.059	0.063	0.045	0.055
	MSE	0.622	0.764	0.794	0.788	0.821	0.736

Table (2): Estimated parameters, Bias and MSE: (n = 50,100)

Criteria		2SLS	S-M- Estimator				
			<i>k=1</i>	<i>k=1.5</i>	<i>k=2</i>	<i>k=2.5</i>	<i>k=3</i>
<i>n=50</i>	$\hat{\gamma}$	1.8	0.484	0.484	0.473	0.474	0.467
	Bias	1.3	0.016	0.016	0.027	0.026	0.033
	MSE	4.629	0.283	0.29	0.316	0.312	0.278
	$\hat{\beta}_0$	-7.178	2.571	2.576	2.655	2.621	2.679
	Bias	8.178	1.571	1.576	1.655	1.621	1.679
	MSE	254.153	17.823	18.403	20.45	20.093	18.427
	$\hat{\beta}_1$	0.649	0.146	0.14	0.132	0.132	0.142
	Bias	0.449	0.054	0.06	0.068	0.068	0.058
	MSE	0.706	0.4	0.394	0.419	0.408	0.369
<i>n=100</i>	$\hat{\gamma}$	0.505	0.502	0.499	0.492	0.507	0.504
	Bias	0.005	0.002	0.001	0.008	0.007	-0.004
	MSE	0.178	0.194	0.207	0.219	0.213	0.173
	$\hat{\beta}_0$	3.173	2.397	2.413	2.424	2.301	2.349
	Bias	2.173	1.397	1.413	1.424	1.301	1.349
	MSE	15.813	11.644	12.435	13.173	12.302	10.235
	$\hat{\beta}_1$	0.13	0.208	0.206	0.202	0.197	0.204
	Bias	0.07	0.008	0.006	0.002	0.003	0.004
	MSE	0.157	0.155	0.165	0.178	0.169	0.141

7. Simulation Study for comparing Different M- estimators

The Simulation Study was designed to investigate the properties of some M-estimators compared with 2SLS, when there are outliers in exogenous variables (vertical outliers). Outliers in exogenous variables were reflected as a heavy tailed errors distribution.

a. Simulations Model

The Simulation Study was executed considering two simultaneous equations:

$$y_t = Y_t \gamma + (1, x_{1t}) \beta + u_t = 0.5Y_t + 1 + 0.2x_{1t} + u_t \quad (18)$$

where $\gamma = -B_{12} = 0.5$, $\beta' = (\beta_0, \beta_1) = (-\Gamma_{11}, -\Gamma_{12}) = (1, 0.2)$, x_{1t} is the second element of x_t , and u_t is the first element of U_t and the second equation is,

$$Y_t = 0.7y_t + 1 + 0.4x_{2t} + 0.5x_{3t} + \varepsilon_t \quad (19)$$

where x_{2t} and x_{3t} are, respectively, the third and the fourth elements of x_t , and ε_t is the second element of U_t . The compact form for the model of equations (18) and (19) can be represented as:

$$B \begin{bmatrix} y_t \\ Y_t \end{bmatrix} + \Gamma x_t = U_t \quad (20)$$

By considering the model (20), the reduced form of equations (18) and (19) was conducted as:

$$\begin{bmatrix} y & Y \end{bmatrix} = X \begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} + \begin{bmatrix} \nu & V \end{bmatrix} \quad (21)$$

Where, $\begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} = -\Gamma'(B')^{-1}$ and $\begin{bmatrix} \nu & V \end{bmatrix} = U(B')^{-1}$.

b. Simulation Frame Work

Simulations for two simultaneous equations model as Kim and Muller (2007) have been executed by considering the equations (18) to (21) as follows:

1- $\begin{bmatrix} \nu & V \end{bmatrix}$ in equation (21) were drawn from the standard normal $N(0, 1)$, the Lognormal with log-mean = 0 and log-standard deviation = 1, denoted by LN(0, 1), the Student-t with 4 degrees of freedom, denoted by t(4), and Chisq(2).

2- The second to the fourth columns in X are drawn from the normal distribution with mean (0.5, 1, -0.1), variances equal to 1. X is fixed at each replication.

3- $\begin{bmatrix} y & Y \end{bmatrix}$ were generated using the reduced form parameters basing on the generated exogenous variables X , and the generated errors as equation (21).

4- 1000 replications were used for different sample sizes 15, 30, 50, and 100. For each replication, the value of the parameters $(\gamma, \beta_0, \beta_1)$ has been estimated.

5- The following methods have been compared:

a- Two Stage Least Square (2SLS)

b- Two stage Huber M- estimator with a tuning value =2, which is denoted as 2SH(2).

c- Two stage Tukey M-estimators with a tuning value =4.685, 2ST(4.685).

d- Suggested M-estimator with tuning value =3, S-M(3).

6- The comparison among these M- estimators based on Bias, MSE, and relative efficiency (RE) that is defined as $\frac{\text{Var of 2SLS}}{\text{Var of the robust M - estimator}}$

7- Estimating the parameter using M- estimators based on iterative reweighted least square (IRLS), which depends on weight functions as in equations (11), (13), and (17).

c. The Simulation Results

1- Normal Errors:

In table (3), when $n=15$, it is noted that 2SH (2) and S-M(3) gave the best results respectively while at $n=30$, 2SLS results are better than the other methods of estimation. In table (4), we can conclude generally, that 2SLS is the most efficient estimator at large samples sizes, and its bias decreases as sample size increases, where it has the least MES at large sample sizes. The estimated values of robust methods converges each other.

2- Log Normal Errors

In table (3), when $n=15$, 2SLS and 2SH (2) gave the best results compared with the other two methods. When $n=30$, it can be seen generally that 2ST (4.685) has a better results among the other methods of estimation. In Table (4), we can see that the behavior of S-M(3) and 2ST(4.685) is the best compared with the other two methods.

3-Student-t Errors

In table (5), when $n=15$ or 30, it can be seen generally that 2SH (2) has a better results among the other method of estimation. In table (6), when $n=50$, it can be noted that 2SH (2) and S-M(3) gave the best results respectively while at $n=100$, 2SLS and 2SH (2) gave the best results respectively compared with the other two methods.

4- Chi- Square Errors

In table (5), when $n=15$, 2SLS is the most efficient method compared to other estimators while at $n=30$, 2SLS and 2SH (2) gave the best results respectively compared with the other two methods. In table (6), when $n=50$, 2SH (2) and S-M(3) gave the best results respectively while at $n=100$, it is noted that 2SH (2) gave the best results compared with the other methods.

**Table (3) Estimated parameters, Bias, MSE, and RE: N (0, 1), and LN (0, 1):
(n = 15, 30)**

<div style="display: flex; align-items: center; justify-content: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg);">n</div> <div>Criteria</div> </div>		N (0, 1)				LN (0, 1)			
		2SLS	2SH(2)	2ST (4.685)	S-M (3)	2SLS	2SH (2)	2ST (4.685)	S-M (3)
15	$\hat{\gamma}$	0.684	0.49	0.475	0.485	0.123	0.511	0.496	0.483
	Bias	0.184	0.01	0.025	0.015	0.377	0.011	0.004	0.017
	MSE	0.758	0.249	0.355	0.286	0.263	0.442	0.629	0.579
	RE	1.000	2.909	2.043	2.533	1.000	0.274	0.192	0.209
	$\hat{\beta}_0$	0.528	1.023	1.064	1.039	5.277	2.279	2.191	2.34
	Bias	0.472	0.023	0.064	0.039	4.277	1.279	1.191	1.34
	MSE	5.488	2.064	2.911	2.387	24.759	20.088	25.551	24.82
	RE	1.000	2.552	1.811	2.207	1.000	0.350	0.268	0.281
	$\hat{\beta}_1$	0.436	0.212	0.218	0.222	0.366	0.175	0.147	0.175
	Bias	0.236	0.012	0.018	0.022	0.166	0.025	0.053	0.025
	MSE	0.267	0.306	0.48	0.352	1.019	0.52	0.703	0.671
	RE	1.000	0.690	0.440	0.600	1.000	1.908	1.415	1.478
30	$\hat{\gamma}$	0.401	0.493	0.503	0.5	0.192	0.51	0.508	0.498
	Bias	0.099	0.0067	0.0028	0	0.308	0.01	0.008	0.002
	MSE	0.09	0.123	0.189	0.155	0.173	0.234	0.244	0.249
	RE	1.000	0.651	0.423	0.516	1.000	0.333	0.320	0.313
	$\hat{\beta}_0$	1.56	1.02	0.996	1.002	4.538	2.373	2.041	2.154
	Bias	0.56	0.02	0.004	0.002	3.538	1.373	1.041	1.154
	MSE	0.797	1.099	1.687	1.409	19.213	12.491	9.587	10.68
	RE	1.000	0.440	0.286	0.343	1.000	0.631	0.787	0.717
	$\hat{\beta}_1$	0.025	0.208	0.211	0.213	0.56	0.194	0.214	0.236
	Bias	0.175	0.0078	0.011	0.013	0.36	0.006	0.014	0.036
	MSE	0.217	0.151	0.228	0.19	0.909	0.299	0.297	0.303
	RE	1.000	1.232	0.816	0.980	1.000	2.606	2.625	2.582

**Table (4) Estimated parameters, Bias, MSE, and RE: N (0, 1), and LN (0, 1):
(n = 50, 100)**

n \ Criteria		N (0, 1)				LN (0, 1)			
		2SLS	2SH(2)	2ST (4.685)	S-M (3)	2SLS	2SH (2)	2ST (4.685)	S-M (3)
50	$\hat{\gamma}$	0.469	0.506	0.502	0.507	0.143	0.5	0.495	0.497
	Bias	0.031	0.006	0.002	0.007	0.357	0	0.005	0.003
	MSE	0.05	0.051	0.078	0.068	0.153	0.102	0.073	0.074
	RE	1.000	0.961	0.628	0.721	1.000	0.255	0.356	0.351
	$\hat{\beta}_0$	1.231	0.99	1.006	0.986	5.224	2.436	2.143	2.193
	Bias	0.231	0.01	0.006	0.014	4.224	1.436	1.143	1.93
	MSE	0.501	0.544	0.847	0.747	19.721	7.141	4.071	4.369
	RE	1.000	0.824	0.529	0.600	1.000	0.370	0.680	2.917
	$\hat{\beta}_1$	0.184	0.196	0.198	0.2	0.269	0.199	0.191	0.188
	Bias	0.016	0.004	0.002	0	0.469	.001	0.009	0.012
	MSE	0.066	0.067	0.105	0.087	0.762	0.139	0.118	0.117
	RE	1.000	0.985	0.629	0.759	1.000	3.899	4.596	4.638
100	$\hat{\gamma}$	0.437	0.506	0.513	0.507	0.728	0.504	0.513	0.509
	Bias	0.063	0.006	0.013	0.007	0.228	0.004	0.013	0.009
	MSE	0.037	0.044	0.063	0.057	0.586	0.073	0.049	0.047
	RE	1.000	0.751	0.525	0.579	1.000	7.317	10.936	11.381
	$\hat{\beta}_0$	1.23	0.983	0.958	0.976	1.466	2.399	2.032	2.118
	Bias	0.23	0.017	0.042	0.024	0.466	1.399	1.032	1.118
	MSE	0.381	0.451	0.646	0.577	24.141	5.426	2.751	2.99
	RE	1.000	0.728	0.509	0.569	1.000	6.897	14.190	13.749
	$\hat{\beta}_1$	0.164	0.203	0.205	0.207	0.203	0.197	0.199	0.199
	Bias	0.036	0.003	0.005	0.007	0.003	0.003	0.001	0.001
	MSE	0.034	0.032	0.048	0.042	0.154	0.064	0.045	0.044
	RE	1.000	1.032	0.688	0.787	1.000	2.406	3.422	3.500

Table (5) Estimated parameters, Bias, MSE, and RE: t (4), and Chisq (2):
(n = 15, 30)

Criteria <i>n</i>		t (4)				Chisq (2)			
		2SLS	2SH(2)	2ST (4.685)	S-M (3)	2SLS	2SH (2)	2ST (4.685)	S-M (3)
15	$\hat{\gamma}$	0.072	0.477	0.454	0.461	-0.024	0.492	0.52	0.506
	Bias	0.428	0.023	0.06	0.039	0.524	0.008	0.02	0.006
	MSE	0.418	0.368	0.596	0.444	0.361	0.801	1.443	1.113
	RE	1.000	0.640	0.397	0.531	1.000	0.107	0.060	0.077
	$\hat{\beta}_0$	2.098	1.052	1.108	1.078	7.109	2.788	2.283	2.563
	Bias	1.098	0.052	0.108	0.078	6.109	1.788	1.283	1.563
	MSE	4.184	3.032	4.785	3.653	44.118	48.592	77.745	63.336
	RE	1.000	0.983	0.624	0.817	1.000	0.150	0.089	0.112
	$\hat{\beta}_1$	0.472	0.222	0.24	0.236	0.496	0.188	0.161	0.17
	Bias	0.272	0.022	0.04	0.036	0.296	0.012	0.039	0.03
	MSE	0.536	0.479	0.772	0.574	0.929	0.925	1.635	1.325
	RE	1.000	0.965	0.600	0.807	1.000	0.909	0.515	0.635
30	$\hat{\gamma}$	0.34	0.507	0.495	0.511	0.299	0.487	0.478	0.482
	Bias	0.16	0.007	0.005	0.11	0.201	0.013	0.022	0.018
	MSE	0.877	0.186	0.299	0.228	0.241	0.384	0.787	0.646
	RE	1.000	4.576	2.846	3.942	1.000	0.547	0.267	0.325
	$\hat{\beta}_0$	1.581	0.983	1.005	0.971	4.531	2.971	2.484	2.573
	Bias	0.581	0.017	0.005	0.029	3.531	1.971	1.484	1.573
	MSE	5.55	1.695	2.689	2.045	26.801	28	39.595	35.7
	RE	1.000	3.075	1.938	2.550	1.000	0.594	0.383	0.431
	$\hat{\beta}_1$	0.168	0.2	0.202	0.2	0.421	0.215	0.254	0.255
	Bias	0.032	0	0.002	0	0.221	0.015	0.054	0.055
	MSE	0.389	0.248	0.382	0.289	0.622	0.468	0.895	0.736
	RE	1.000	1.565	1.016	1.343	1.000	1.225	0.642	0.782

Table (6) estimated parameters, Bias, MSE, and RE: t (4), and Chisq (2):
(n = 50, 100)

n \ Criteria		t (4)				Chisq (2)			
		2SLS	2SH(2)	2ST (4.685)	S-M (3)	2SLS	2SH (2)	2ST (4.685)	S-M (3)
50	$\hat{\gamma}$	0.513	0.499	0.499	0.5	1.8	0.495	0.495	0.467
	Bias	0.013	0.001	0.001	0	1.3	0.005	0.005	0.033
	MSE	0.124	0.097	0.127	0.105	4.629	0.168	0.292	0.278
	RE	1.000	1.278	0.976	1.181	1.000	17.497	10.066	10.614
	$\hat{\beta}_0$	0.891	1	1.003	0.998	-7.178	2.924	2.362	2.679
	Bias	0.109	0	0.003	0.002	8.178	1.924	1.362	1.679
	MSE	1.415	1.048	1.41	1.171	254.15	15.119	16.772	18.427
	RE	1.000	1.339	0.995	1.198	1.000	16.403	12.554	11.999
	$\hat{\beta}_1$	0.427	0.204	0.198	0.201	0.649	0.179	0.174	0.142
	Bias	0.227	0.004	0.002	0.001	0.449	0.021	0.026	0.058
	MSE	0.242	0.135	0.181	0.15	0.706	0.232	0.403	0.369
	RE	1.000	1.408	1.050	1.267	1.000	2.177	1.253	1.378
100	$\hat{\gamma}$	0.401	0.49	0.496	0.496	0.505	0.491	0.499	0.504
	Bias	0.099	0.01	0.004	0.004	0.005	0.009	0.001	0.004
	MSE	0.067	0.071	0.096	0.088	0.178	0.119	0.177	0.173
	RE	1.000	0.804	0.594	0.648	1.000	1.497	1.006	1.029
	$\hat{\beta}_0$	1.329	1.03	1.01	1.01	3.173	2.945	2.332	2.349
	Bias	0.329	0.03	0.01	0.01	2.173	1.945	1.332	1.349
	MSE	0.69	0.723	0.978	0.902	15.813	11.574	9.623	10.235
	RE	1.000	0.806	0.595	0.645	1.000	1.424	1.413	1.318
	$\hat{\beta}_1$	0.049	0.198	0.204	0.201	0.13	0.204	0.202	0.204
	Bias	0.151	0.002	0.004	0.001	0.07	0.004	0.002	0.004
	MSE	0.102	0.055	0.075	0.064	0.157	0.109	0.129	0.141
	RE	1.000	1.436	1.054	1.234	1.000	1.395	1.178	1.078

Conclusions

For normal errors, the estimated values of the compared M- estimators for three parameters converge. As sample size increases the performance of 2SLS improves so, it is the most efficient estimator for large sample size. Also, 2SH(2) and S-M(3) are the most efficient M- estimators among other M- estimators but for Lognormal errors, the S-M(3) is more efficient than 2SH (2) and 2SLS has the least biased for the intercept term while for Student-t errors, 2SH (2) is the most efficient estimator, followed by S-M (3) and the bias of 2SLS decreases as sample size increases and for Chi-square errors distribution, 2SLS is the most biased estimators compared with different M-estimators. Also, MSE of compared M-estimators decreases as sample sizes increase.

Generally, there is no one of the compared methods can eliminate the others. However, the properties of each one differ according to sample size and the type of errors distribution.

References

- 1- Bhar, L. Robust regression, Advances in Data Analytical Techniques.(eds.) R. Parsad, VK Gupta, LM Bhar, VK Bhatia. Indian Agricultural Statistics Research Institute, 1 (2010), III69-III78.
- 2- Brikes, D. and Dodge, Y. Alternative Methods of Regression. Wiley Series in Probabilities and Statistics (1993).
- 3- Flavin, M., Robust estimation of the joint consumption/asset demand decision. NBER, Cambridge, Working Paper No. 7011, (1999).
- 4- Fox, J. Robust Regression. <http://cran.r-project.org> (2002).
- 5- Green, W.H. Econometric Analysis. Pearsons Education, Inc(2008)
- 6- Huber, P.J. Robust estimation of a location parameter, The Annals of Mathematical Statistics, 35 (1964), 73-101.
- 7- Huber, P.J. Robust regression: Asymptotics, conjectures and Monte Carlo, The Annals of Statistics, 1(5) (1973), 799-821.
- 8- Kim,TH and Muller ,C. Two-stage Huber estimation. Journal of Statistical Planning and Inference, 137 (2007), 405 – 418

Received: April 25, 2013