Journal of Information Sciences and Computing Technologies(JISCT)
ISSN: 2394-9066

SCITECH
RESEARCH ORGANISATION|

Volume 5, Issue 1
November 17, 2015

Journal of Information Sciences and Computing Technologies
www.scitecresearch.com/journals

A new approach called Weighted Least Squares Ratio (WLSR) Method to M-estimators

Murat Yazici

JForce Information Technologies Inc.
Göztepe mah. Göksuevleri Sit. Sardunya Sk.
B212B Istanbul 34815, Turkey

Abstract

Regression Analysis (RA) is an important statistical tool that is applied in most sciences. The Ordinary Least Squares (OLS) is a tradition method in RA and there are many regression techniques based on OLS. The Weighted Least Squares(WLS) method is iteratively used in M-estimators. The Least Squares Ratio (LSR) method in RA gives better results than OLS, especially in case of the presence of outliers. This paper includes a new approach to M-estimators, called Weighted Least Squares Ratio (WLSR), and comparison of WLS and WLSR according to mean absolute errors of estimation of the regression parameters (mae ß) and dependent value (mae y).

Keywords: Outliers– Least squares ratio (LSR) method – Weighted least squares ratio (WLSR) method – Robust statistics – M-estimators.

1. Introduction

The theory of robustness developed by Huber and Hampel (1960) laid the foundation for finding practical solutions too many problems, when statistical concepts were vague to serve the purpose. Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is and whether this should be thrown. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. Robust regression is an important tool for analyzing data that are contaminated with outliers. It can be used to detect outliers and to provide resistant results in the presence of outliers. Many methods have been developed for these problems. Many researchers have worked in this field and described the methods of robust estimators. The class of robust estimators includes M-, L- and R-estimators. The M-estimators are most flexible ones, and they generalize straightforwardly to multiparameter problems, even though they are not automatically scale invariant and have to be supplemented for practical applications by an auxiliary estimate of scale any estimate (Muthukrishnan and Radha, 2010).

The iteratively ordinary least squares approach is used in M-estimators during the calculation of the regression parameters. In this approach, the weighted errors are calculated by using a weighting function in each iterative step. Employing the weighting function in OLS, we get weighted least squares (WLS) method. This paper includes a new approach called Weighted Least Squares Ratio (WLSR) Method to M-estimators as an alternative to WLS Method. In this method, the errors are calculated by using the LSR method instead of the OLS method. Then, the weighted errors are calculated by using a weighting function in each iterative step.

In this study, it was shown which method (WLS and WLSR) gives better resultsin M-estimation including Huber, Tukey, Andrew and Ramsay's functionsaccording to the mean absolute errors (MAE) of the estimated regression parameters and dependent value via a simulation study using different sample sizes and error variances. Based on the simulation results, apart from Andrew's function, we can say that WLSR generally gives better estimates than WLSin case of increasingthe number of outliers and the error variance.

2. The LSR Method

The Least Squares Ratio (LSR) method is one of the forecasting techniques in regression analysis. LSR aims to estimate observed values with zero error ($Y = \hat{Y}$, or $Y - \hat{Y} = 0$). It starts with the same goal $Y = \hat{Y}$ as in Ordinary Least Squares. However, it proceeds by dividing through by Y and so $\hat{Y}/Y = 1$ is obtained under an assumption of $Y \neq 0$. Hence, it is obvious that, equations $1-(\hat{Y}/Y) = 0$ and $(Y-\hat{Y})/Y = 0$ are raised by basic mathematical operations. This final equation is taken into account as the origin of the LSR which minimizes the sum of $[(Y-\hat{Y})/Y]^2$. Consequently the aim of LSR can be written mathematically as follows (Akbilgic and Akinci, 2009):

$$\min_{\beta} \sum_{i=1}^{n} \left(\frac{Y - \hat{Y}}{Y} \right)^2 \tag{1}$$

The matrix representation of the regression model is as follows;

$$\mathrm{Y} = \beta \mathrm{X} + \mathrm{e} \tag{2}$$

where Y is an n × 1 vector of observed values; X is an n × p vector of the values of dependent variables; n is the number of observations; p is the number of unknown parameters, β is the p × 1 vector of regression coefficients; e is an n × 1 vector of error values.

Formula 1 can also be written as in formula 3, by using Eq. 2:

$$\min_{\hat{\beta}_{LSR}} \sum_{i=1}^{n} \left(\frac{Y - \hat{\beta}_{lsr} X}{Y} \right)^2 \tag{3}$$

If *rank(X)* is equal to p, the formula for estimating β appears as in Eq. 4 (Akbilgic and Akinci, 2009):

$$\hat{\beta}_{lsr} = \left[\left(\frac{X}{Y} \right)' \left(\frac{X}{Y} \right) \right]^{-1} \left(\frac{X}{Y^2} \right)' Y. \tag{4}$$

The matrix X/Y is obtained by dividing the values x_{ij} by y_i for $j = 1, 2, ..., p$, and X/Y^2 is computed by dividing the values x_{ij} by y_i^2 for $j = 1, 2, ..., p$.

3. M-Estimatorsandthe proposedWLSR Method

First proposed by Huber (1964, 1973, 2004), M-estimation for regression is a relatively straightforward extension of M-estimation for location. It represent one of the first attemps at a compromise between the efficiency of the least squares estimators and the resistance of the LAV estimators, both of which can be seen as special cases of M-estimation. In simplest terms, the M-estimator minimizes some function of the residuals. As in the case of M-estimation location, the robustness of the estimator is determined by the choice of weight function (Andersen, 2007).

The M-estimate $T_n(X_1,........,X_n)$ for the function ρ and the sample $x_1,........,x_n$ is the value of t that maximizes the objective function $\sum_{i=1}^{n}\rho(x_i;t)$. When ρ can be differentiated with respect to t, a function (which except for a multiplicative constant) we denote by ψ, $\left(i.e.\ \ \psi(x_i;t)=\frac{\partial\rho(x_i;t)}{\partial t}\right)$, we may find it more convenient to calculate T_n by finding the value of t that satisfies $\sum_{i=1}^{n}\psi(x_i;t)=0$. The corresponding w -function (weight function) for any ρ is then defined as follows (Ali and Qadir, 2005);

$$w(x_i;t)=\frac{\psi(x_i;t)}{t}. \tag{5}$$

Employing this w-function in OLS, we get weighted least squares (WLS) method and the resulting estimates are then called the weighted estimates. The weighted estimates are computed by solving the following equation (Hoaglin et al., 1983);

$$\hat{\beta}=\left(X'WX\right)^{-1}X'Wy \tag{6}$$

where W is a n x n diagonal square matrix having the diagonal elements as weights.

When we use the w-function in LSR, we get weighted least squares ratio.This method is named as weighted least squares ratio (WLSR) method. And, the weighted estimates are calculated by solving the following equation;

$$\hat{\beta}_{WLSR}=\left[\left(\frac{X}{Y}\right)'W\left(\frac{X}{Y}\right)\right]^{-1}\left(\frac{X}{Y^2}\right)'WY \tag{7}$$

M-estimators minimize objective function more general than the familiar sum of squared residuals associated with the sample mean. Instead of squaring the deviations of each observation x_i from the estimate t, we apply the function $\rho(x_i;t)$; and form the objective function by summing over the sample: $\sum_{i=1}^{n}\rho(x_i;t)$. The nature of $\rho(x_i;t)$; determines the properties of the M-estimator (Hoaglin et al., 1983).

Huber's M-estimator uses the following ψ -function;

$$\psi(t)=\begin{cases}-a, & t<-a\\ t, & -a\le t\le a\\ a, & t>a\end{cases} \tag{8}$$

Both the least squares and Huber objective functions increase without bound as the residual departs from 0, but the least-squares objective function increases more rapidly. Least squares assigns equal weight to each observation; the weights for the Huber estimator decline when $|t| > a$. The Huber's ψ-function takes into account the neighborhood of a normal model in a linear way. It has a *constant-linear-constant* behavior, i.e. it is constant beyond the specified bound (-a to a).Like the OLS it assigns equal weights to all observations within its bound, which surely will result in its high efficiency but distant outliers still have a maximum influence (in the form of constant a), which lead to the efficiency losses of about 10-20 percent in typical cases with outliers (Hampel et al 1986). To cope with this problem redescending M-estimators were introduced.

4. Redescending M-Estimators

Redescending M-estimators are very popular ψ-type M-Estimator which has ψ functions that are non-decreasing near the origin, but decreasing toward 0 far from the origin. Their ψ functions can be chosen to redescend smoothly to zero, so that they usually satisfy $\psi(x) = 0$ for all x with $|X| > k$,where k is referred to as the minimum rejected point. When choosing a redescending ψ functions we must take care that it does not descend too steeply, which may have a very bad influence on the denominator in the expression for the asymptotic variance

$$\frac{\int \psi^2 dF}{\left(\int \left(\psi' dF\right)\right)^2} \tag{9}$$

where F is the mixture model distribution. This effect is particularly harmful when a large negative values of $\psi'(x)$ combines with a large positive values $\psi^2(x)$, and there is a cluster of outliers near x (Muthukrishnan and Radha, 2010).

Huber's influence function is as follows;

$$\psi_H(x) = \begin{cases} k & x > k \\ x & x \le k \\ -k & x < -k \end{cases} \tag{10}$$

In Huber's function, the weighting of errors in M-estimation are calculated by the following function;

$$w_H(x) = \begin{cases} 1 & x \le k \\ k/|x| & x > k \end{cases} \tag{11}$$

Andrew's influenceand weight functions:

$$\psi_A(x) = \begin{cases} \sin(x/k) & |x| \le k \\ 0 & |x| > k \end{cases}. \tag{12}$$

$$w_A(x) = \begin{cases} \dfrac{\sin(x/k)}{(x/k)} & |x| \le k \\ 0 & |x| > k \end{cases} \tag{13}$$

Tukey's biweight M-estimator have ψ functions for any positive k, which defined by

$$\psi_T(x)=\begin{cases} x\left[1-\left(\frac{x}{k}\right)^2\right]^2 & |x|\le k \\ 0 & |x|>k \end{cases}. \quad (14)$$

$$w_T(x)=\begin{cases} \left[1-\left(\frac{x}{k}\right)^2\right]^2 & |x|\le k \\ 0 & |x|>k \end{cases} \quad (15)$$

Ramsay's influence and weight functionsare as follows;

$$\psi_R(x)=xe^{-k|x|}\left(\text{maximum at } k^{-1}\right) \quad (16)$$

$$w_R(x)=e^{-k|x|} \quad (17)$$

Figure 1 indicates a comparison M-estimator weight functions and the mean.

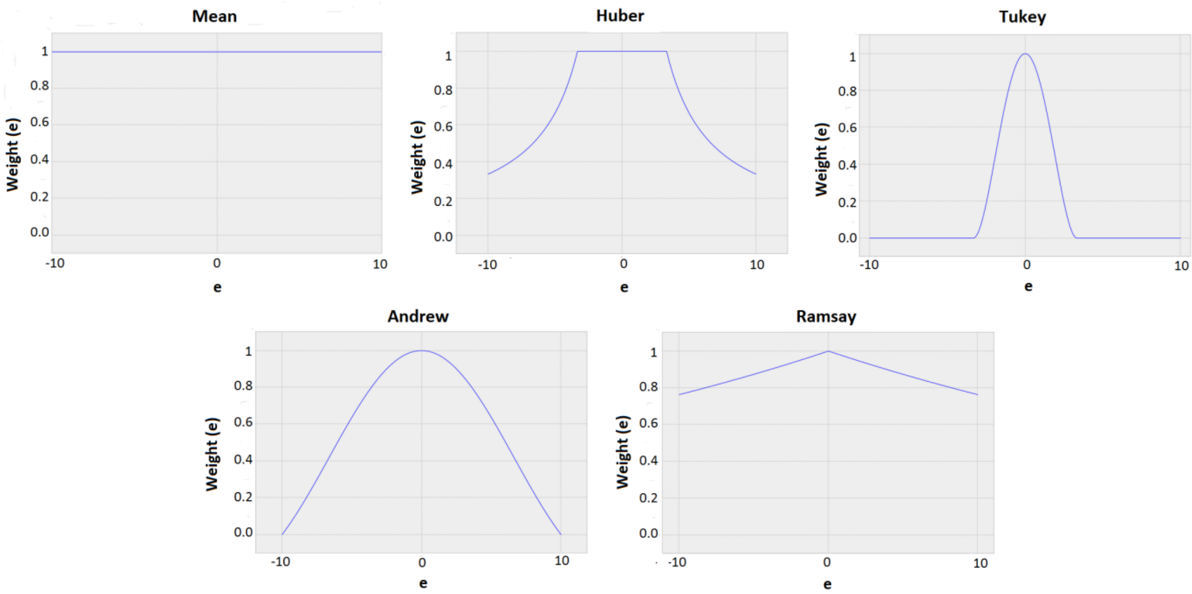

Figure 1. M-Estimator Weight Functions Compared to the Mean

For regression analysis, some of the redecending M-estimators can attain the maximum breakdown point. Moreover, some of them are the solutions of the problem of maximizing the efficiency under bounded influence function when the regression coefficient and the scale parameter are estimated simultaneously. Hence redecending M-estimators satisfy several outlier robustness properties (Muthukrishnan and Radha, 2010).

5. The Simulation Study

The simulation study evaluates linear multiple regression analysis with two independent variables as shown in (18). WLS and WLSR methods are compared according to the MAE of β and the MAE of y :

$$y=\beta_0+\beta_1x_1+\beta_2x_2+e, \quad (18)$$

where y is the dependent variable, x_1 and x_2 are independent variables, e is the error,and β_i are the true regression parameters.For OLS we have $\hat{\beta}_{ols} = \left[\hat{\beta}_{0,ols} \ \hat{\beta}_{1,ols} \ \hat{\beta}_{2,ols} \right]$; and also, LSR gives

$$\hat{\beta}_{lsr} = \left[\hat{\beta}_{0,lsr} \ \hat{\beta}_{1,lsr} \ \hat{\beta}_{2,lsr} \right].$$

In the simulation process, the independent variables x_1 and x_2 are randomly generated from a normal distribution with $\mu = 100$ and $\sigma^2 = 100$; β_0, β_1 and β_2 are equal to 1, so $\beta_i = [1\,1\,1]$. Thus, the regression model becomes as follows:

$$y = 1 + 1x_1 + 1x_2 + e\,. \tag{19}$$

Finally, errors are randomly generated as Gaussian white noise with variance δ_e^2. Therefore, the dependent variable has a normal distribution with mean 201 and variance 200 + δ_e^2.

The simulations were performed by R, using different sample sizes and error variances.During calculation of m estimators, OLS and LSR methods were used to fit initial regression model; initial residuals were found, and they were scaled by MAD; a chosen weight function was applied to obtain preliminary weights.The preliminary weights were used in iteratively reweighted least squares and iteratively reweighted least squares ratio methods to obtain regression parameters; secondary residuals were found during the first iteration. In the second and other iterations, the residuals were scaled by Huber proposal 2 untill the best model was found. The following criteria were used to obtain the final estimates;

$$\frac{\left| \hat{\beta}^{(q+1)} - \hat{\beta}^{(q)} \right|}{\left| \hat{\beta}^{(q+1)} \right|} < \varepsilon \tag{20}$$

where q refers to the number of iterations; ε indicates a very small positive number. In this study, ε took the value of 0.0001.

Table 1. Comparison of WLS and WLSR for sample size 30 with non-outlier

δ_e^2	n = 30		β_0	β_1	β_2	mae β	mae y
1	ols		0.98181	1.00024	0.99993	5339	5365
	lsr		0.96855	1.00023	0.99998	4661	4635
	huber	wls	0.94130	1.00023	0.99992	5403	5971
		wlsr	0.96148	1.00006	0.99981	4597	4029
	tukey	wls	0.97795	1.00022	0.99998	5415	5541
		wlsr	0.96257	1.00022	1.00004	4585	4459
	andrew	wls	0.98056	1.00008	1.00012	5308	5606
		wlsr	0.95407	1.00021	1.00015	4692	4394
	ramsay	wls	0.98030	1.00022	0.99996	5412	5569
		wlsr	0.96498	1.00021	1.00003	4588	4431
9	ols		1.02658	0.99970	1.00000	5374	5696
	lsr		0.90453	0.99993	1.00017	4626	4304
	huber	wls	0.88346	0.99985	1.00003	5471	6643
		wlsr	0.86118	0.99956	0.99972	4529	3357
	tukey	wls	1.01034	0.99978	1.00007	5363	5849
		wlsr	0.87606	1.00005	1.00031	4637	4151
	andrew	wls	1.05022	0.99938	1.00011	5244	5632
		wlsr	0.85929	0.99974	1.00075	4756	4368
	ramsay	wls	1.02464	0.99967	1.00004	5401	5867
		wlsr	0.89351	0.99993	1.00025	4599	4133
25	ols		0.95402	1.00000	1.00050	5347	5816
	lsr		0.55820	1.00089	1.00133	4653	4184
	huber	wls	0.73218	0.99997	1.00070	5507	6962
		wlsr	0.49073	1.00006	1.00073	4493	3038
	tukey	wls	0.99429	0.99981	1.00032	5367	6069
		wlsr	0.55972	1.00087	1.00126	4633	3931
	andrew	wls	1.14786	0.99898	0.99957	5281	5939
		wlsr	0.55788	1.00080	1.00120	4719	4061
	ramsay	wls	0.97549	0.99990	1.00041	5371	6073
		wlsr	0.54213	1.00095	1.00136	4629	3927
100	ols		0.83906	1.00111	1.00065	5481	6378
	lsr		-0.61739	1.00420	1.00308	4519	3622
	huber	wls	0.37311	1.00113	1.00116	5734	7666
		wlsr	-0.81007	1.00279	1.00211	4266	2334
	tukey	wls	0.76929	1.00137	1.00113	5472	6723
		wlsr	-0.84323	1.00511	1.00409	4528	3277
	andrew	wls	0.22183	1.00245	1.00548	5353	6591
		wlsr	-1.56419	1.00698	1.00877	4647	3409
	ramsay	wls	0.75614	1.00136	1.00125	5475	6765
		wlsr	-0.84465	1.00506	1.00418	4525	3235

Table 1 indicates the comparison of WLS and WLSR for sample size 30 with non-outlier and different error variances. According to mae β and mae y, we can say that WLS is a little more successful than WLSR in case of non-outlier and increasing error variance.

Table 2. Comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 1$

Outliers	n = 30		β_0	β_1	β_2	mae β	mae y
0%	ols		0.98181	1.00024	0.99993	5339	5365
	lsr		0.96855	1.00023	0.99998	4661	4635
	huber	wls	0.94130	1.00023	0.99992	5403	5971
		wlsr	0.96148	1.00006	0.99981	4597	4029
	tukey	wls	0.97795	1.00022	0.99998	5415	5541
		wlsr	0.96257	1.00022	1.00004	4585	4459
	andrew	wls	0.98056	1.00008	1.00012	5308	5606
		wlsr	0.95407	1.00021	1.00015	4692	4394
	ramsay	wls	0.98030	1.00022	0.99996	5412	5569
		wlsr	0.96498	1.00021	1.00003	4588	4431
10%	ols		56.23785	0.88867	0.85761	13	0
	lsr		2.03117	1.02660	1.01902	9987	10000
	huber	wls	11.35493	0.98558	0.97006	583	0
		wlsr	0.84122	1.00166	1.00153	9417	10000
	tukey	wls	1.41944	1.00081	0.99868	5141	6184
		wlsr	0.93510	1.00021	1.00033	4859	3816
	andrew	wls	0.94564	1.00020	1.00032	5322	6376
		wlsr	0.93510	1.00021	1.00033	4678	3624
	ramsay	wls	9.64222	0.98825	0.97242	3685	4445
		wlsr	0.93507	1.00022	1.00032	6315	5555
20%	ols		97.58009	0.83159	0.80078	33	0
	lsr		0.81191	1.06734	1.05876	9967	10000
	huber	wls	48.35211	0.99252	0.96596	11	0
		wlsr	-3.79216	1.06274	1.05737	9989	10000
	tukey	wls	51.18214	0.97983	0.95243	13	0
		wlsr	-5.41910	1.06980	1.06453	9987	10000
	andrew	wls	-0.82361	1.00948	1.00947	5260	6869
		wlsr	0.95270	1.00028	1.00011	4740	3131
	ramsay	wls	55.89915	0.94295	0.91542	10	0
		wlsr	-2.21081	1.05047	1.04559	9990	10000

Table 2 indicates the comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 1$. According to mae β and mae y, we can say that WLS is a little bit successful than WLSR in case of non-outlier. Except for Andrew's function, we can also say that WLSR gives better results than WLS in case of increasing the outlier.

Table 3. Comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 9$

Outliers	n = 30		β_0	β_1	β_2	mae β	mae y
0%	ols		1.02658	0.99970	1.00000	5374	5696
	lsr		0.90453	0.99993	1.00017	4626	4304
	huber	wls	0.88346	0.99985	1.00003	5471	6643
		wlsr	0.86118	0.99956	0.99972	4529	3357
	tukey	wls	1.01034	0.99978	1.00007	5363	5849
		wlsr	0.87606	1.00005	1.00031	4637	4151
	andrew	wls	1.05022	0.99938	1.00011	5244	5632
		wlsr	0.85929	0.99974	1.00075	4756	4368
	ramsay	wls	1.02464	0.99967	1.00004	5401	5867
		wlsr	0.89351	0.99993	1.00025	4599	4133
10%	ols		51.99734	0.872518	0.916838	23	0
	lsr		1.152778	1.021049	1.032742	9977	10000
	huber	wls	12.5887	0.968695	0.994425	593	0
		wlsr	0.549853	1.003117	1.005678	9407	10000
	tukey	wls	2.31577	0.99045	1.00062	5079	7220
		wlsr	0.90984	0.99922	1.00081	4921	2780
	andrew	wls	1.03646	0.99900	1.00056	5272	7491
		wlsr	0.90983	0.99922	1.00081	4728	2509
	ramsay	wls	8.45735	0.97592	1.00236	2907	4778
		wlsr	0.69591	1.00059	1.00229	7093	5222
20%	ols		94.19333	0.82663	0.83952	36	0
	lsr		-0.01883	1.06422	1.06963	9964	10000
	huber	wls	44.52814	0.98791	1.00927	22	0
		wlsr	-4.40365	1.06107	1.06461	9978	10000
	tukey	wls	47.03627	0.97637	0.99758	24	0
		wlsr	-6.05569	1.06824	1.07191	9976	10000
	andrew	wls	-0.17626	1.00617	1.00642	5216	8489
		wlsr	0.85186	1.00106	0.99969	4784	1511
	ramsay	wls	52.62088	0.93698	0.95456	25	0
		wlsr	-2.74006	1.04900	1.05195	9975	10000

Table 3 indicates the comparison of WLS and WLSR for sample size 30 with δ_e^2 =9. According to mae β and mae y, we can say again that WLS is a little bit successful than WLSR in case of non-outlier. Except for Andrew's function, we can also say that WLSR gives better results than WLS in case of increasing the outlier.

Table 4. Comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 25$

Outliers	n = 30		β_0	β_1	β_2	mae β	mae y
0%	ols		0.95402	1.00000	1.00050	5347	5816
	lsr		0.55820	1.00089	1.00133	4653	4184
	huber	wls	0.73218	0.99997	1.00070	5507	6962
		wlsr	0.49073	1.00006	1.00073	4493	3038
	tukey	wls	0.99429	0.99981	1.00032	5367	6069
		wlsr	0.55972	1.00087	1.00126	4633	3931
	andrew	wls	1.14786	0.99898	0.99957	5281	5939
		wlsr	0.55788	1.00080	1.00120	4719	4061
	ramsay	wls	0.97549	0.99990	1.00041	5371	6073
		wlsr	0.54213	1.00095	1.00136	4629	3927
10%	ols		51.38173	0.89167	0.90347	44	0
	lsr		0.85330	1.02545	1.03008	9956	10000
	huber	wls	14.95751	0.97630	0.98124	606	0
		wlsr	0.15691	1.00620	1.00878	9394	10000
	tukey	wls	1.83593	0.99739	0.99956	5148	7751
		wlsr	0.76898	0.99896	1.00131	4852	2249
	andrew	wls	1.09696	0.99857	1.00066	5380	8094
		wlsr	0.76898	0.99896	1.00131	4620	1906
	ramsay	wls	10.77944	0.98687	0.99202	1524	944
		wlsr	-0.00431	1.00485	1.00721	8476	9056
20%	ols		98.21293	0.83160	0.79301	57	0
	lsr		0.63358	1.06794	1.05741	9943	10000
	huber	wls	48.16742	1.00182	0.95824	39	0
		wlsr	-4.13248	1.06331	1.05780	9961	10000
	tukey	wls	50.86917	0.99064	0.94433	40	0
		wlsr	-5.80270	1.07056	1.06523	9960	10000
	andrew	wls	-0.02235	1.00412	1.00643	5316	9142
		wlsr	0.54983	1.00004	1.00207	4684	858
	ramsay	wls	56.29212	0.94763	0.90678	49	0
		wlsr	-2.54301	1.05114	1.04627	9951	10000

Table 4 indicates the comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 25$. According to mae β and mae y, we can say again that WLS is more successful than WLSR in case of non-outlier. Except for Andrew's function, we can also say that WLSR gives better results than WLS in case of increasing the outlier.

Table 5. Comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 100$

Outliers	n = 30		β_0	β_1	β_2	**mae** β	**mae** y
0%	ols		0.83906	1.00111	1.00065	5481	6378
	lsr		-0.61739	1.00420	1.00308	4519	3622
	huber	wls	0.37311	1.00113	1.00116	5734	7666
		wlsr	-0.81007	1.00279	1.00211	4266	2334
	tukey	wls	0.76929	1.00137	1.00113	5472	6723
		wlsr	-0.84323	1.00511	1.00409	4528	3277
	andrew	wls	0.22183	1.00245	1.00548	5353	6591
		wlsr	-1.56419	1.00698	1.00877	4647	3409
	ramsay	wls	0.75614	1.00136	1.00125	5475	6765
		wlsr	-0.84465	1.00506	1.00418	4525	3235
10%	ols		53.69994	0.88899	0.88326	115	0
	lsr		0.03277	1.02955	1.02729	9885	10000
	huber	wls	19.63308	0.97430	0.96986	645	0
		wlsr	-1.61438	1.01787	1.01650	9355	10000
	tukey	wls	-6.34791	1.05044	1.04984	3892	6370
		wlsr	-1.76901	1.01084	1.00974	6108	3630
	andrew	wls	0.92374	1.00075	1.00011	5410	8755
		wlsr	-0.53175	1.00369	1.00278	4590	1245
	ramsay	wls	19.24307	0.97584	0.97105	680	1
		wlsr	-1.97057	1.01887	1.01757	9320	9999
20%	ols		98.60902	0.80032	0.82071	104	0
	lsr		-0.90926	1.06384	1.07038	9896	10000
	huber	wls	48.83526	0.96748	0.98632	112	0
		wlsr	-5.65918	1.06288	1.06712	9888	10000
	tukey	wls	51.55209	0.95618	0.97289	110	0
		wlsr	-7.33481	1.07037	1.07443	9890	10000
	andrew	wls	-0.60457	1.00933	1.00781	5479	9623
		wlsr	-1.00785	1.00538	1.00628	4521	377
	ramsay	wls	56.49293	0.91747	0.93626	131	0
		wlsr	-4.16744	1.05211	1.05607	9869	10000

Table 5 indicates the comparison of WLS and WLSR for sample size 30 with $\delta_e^2 = 100$. According to mae β and mae y, we can say again that WLS is more successful than WLSR in case of non-outlier. Except for Andrew's function, we can also say that WLSR gives better results than WLS in case of increasing the outlier.According to the first five tables, WLS has better performance than WLSR incase of non-outlier.Also, WLSR gives better results than WLS in case of increasing the outlier and increasing the variance.

Table 6. Comparison of WLS and WLSR with non-outlier

n			30		50		100		500		1000	
δ_e^2			mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y
1	ols		5339	5365	5324	5470	5385	5595	5408	5840	5325	6021
	lsr		4661	4635	4676	4530	4615	4405	4592	4160	4675	3979
	huber	wls	5403	5971	5377	6283	5419	6837	5470	8961	5416	9708
		wlsr	4597	4029	4623	3717	4581	3163	4530	1039	4584	292
	tukey	wls	5415	5541	5354	5643	5377	5798	5407	5989	5300	6196
		wlsr	4585	4459	4646	4357	4623	4202	4593	4011	4700	3804
	andrew	wls	5308	5606	5301	5624	5212	5620	5319	5826	5339	5988
		wlsr	4692	4394	4699	4376	4788	4380	4681	4174	4661	4012
	ramsay	wls	5412	5569	5326	5662	5387	5790	5413	6041	5295	6206
		wlsr	4588	4431	4674	4338	4613	4210	4587	3959	4705	3794
9	ols		5374	5696	5316	5726	5322	5882	5401	6787	5520	7435
	lsr		4626	4304	4684	4274	4678	4118	4599	3213	4480	2565
	huber	wls	5471	6643	5457	7063	5499	7870	5472	9717	5641	9958
		wlsr	4529	3357	4543	2937	4501	2130	4528	283	4359	42
	tukey	wls	5363	5849	5358	5914	5376	6146	5381	7058	5468	7760
		wlsr	4637	4151	4642	4086	4624	3854	4619	2942	4532	2240
	andrew	wls	5244	5632	5309	5864	5374	5987	5273	6550	5235	6984
		wlsr	4756	4368	4691	4136	4626	4013	4727	3450	4765	3016
	ramsay	wls	5401	5867	5377	5939	5376	6168	5384	7117	5470	7845
		wlsr	4599	4133	4623	4061	4624	3832	4616	2883	4530	2155
25	ols		5347	5816	5405	6027	5491	6427	5464	7817	5532	8602
	lsr		4653	4184	4595	3973	4509	3573	4536	2183	4468	1398
	huber	wls	5507	6962	5525	7571	5671	8428	5592	9873	5699	9996
		wlsr	4493	3038	4475	2429	4329	1572	4408	127	4301	4
	tukey	wls	5367	6069	5426	6303	5490	6689	5398	8190	5462	8954
		wlsr	4633	3931	4574	3697	4510	3311	4602	1810	4538	1046
	andrew	wls	5281	5939	5272	6131	5298	6340	5317	7334	5257	8086
		wlsr	4719	4061	4728	3869	4702	3660	4683	2666	4743	1914
	ramsay	wls	5371	6073	5436	6328	5463	6736	5395	8283	5466	9026
		wlsr	4629	3927	4564	3672	4537	3264	4605	1717	4534	974
100	ols		5481	6378	5529	6864	5573	7570	5659	9385	5762	9866
	lsr		4519	3622	4471	3136	4427	2430	4341	615	4238	134
	huber	wls	5734	7666	5748	8380	5856	9183	5917	9991	5986	10000
		wlsr	4266	2334	4252	1620	4144	817	4083	9	4014	0
	tukey	wls	5472	6723	5539	7286	5517	7975	5583	9654	5704	9952
		wlsr	4528	3277	4461	2714	4483	2025	4417	346	4296	48
	andrew	wls	5353	6591	5378	6889	5303	7404	5359	8968	5360	9617
		wlsr	4647	3409	4622	3111	4697	2596	4641	1032	4640	383
	ramsay	wls	5475	6765	5513	7318	5524	8046	5569	9700	5703	9959
		wlsr	4525	3235	4487	2682	4476	1954	4431	300	4297	41

Table 6 indicates the comparison of WLS and WLSR with non-outlier for different sample sizes and error variances. According to mae β and mae y, we can say that WLS method outperforms than WLSR in case of increasing sample size and variance.

Table 7. Comparison of WLS and WLSR with 10%*n outliers, which are equal to 500

n			30		50		100		500		1000	
δ_e^2			mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y
1	ols		13	0	24	0	41	0	214	0	364	0
	lsr		9987	10000	9976	10000	9959	10000	9786	10000	9636	10000
	huber	wls	583	0	718	0	802	0	1022	0	1203	0
		wlsr	9417	10000	9282	10000	9198	10000	8978	10000	8797	10000
	tukey	wls	5141	6184	5299	6903	5385	7938	5387	9917	5366	9999
		wlsr	4859	3816	4701	3097	4615	2062	4613	83	4634	1
	andrew	wls	5322	6376	5304	6920	5385	7937	5385	9917	5365	9999
		wlsr	4678	3624	4696	3080	4615	2063	4615	83	4635	1
	ramsay	wls	3685	4445	4825	6392	5381	7960	5371	9919	5377	9999
		wlsr	6315	5555	5175	3608	4619	2040	4629	81	4623	1
9	ols		23	0	43	0	56	0	231	0	387	0
	lsr		9977	10000	9957	10000	9944	10000	9769	10000	9613	10000
	huber	wls	593	0	722	0	905	0	1075	0	1184	0
		wlsr	9407	10000	9278	10000	9095	10000	8925	10000	8816	10000
	tukey	wls	5079	7220	5387	8293	5411	9325	5521	10000	5451	10000
		wlsr	4921	2780	4613	1707	4589	675	4479	0	4549	0
	andrew	wls	5272	7491	5400	8317	5411	9325	5522	10000	5450	10000
		wlsr	4728	2509	4600	1683	4589	675	4478	0	4550	0
	ramsay	wls	2907	4778	4141	7631	4886	9470	4992	9999	4876	0
		wlsr	7093	5222	5859	2369	5114	530	5008	1	5124	10000
25	ols		44	0	45	0	78	0	251	0	427	0
	lsr		9956	10000	9955	10000	9922	10000	9749	10000	9573	10000
	huber	wls	606	0	736	0	931	0	1105	0	1242	0
		wlsr	9394	10000	9264	10000	9069	10000	8895	10000	8758	10000
	tukey	wls	5148	7751	5296	8815	5434	9705	5528	10000	5514	10000
		wlsr	4852	2249	4704	1185	4566	295	4472	0	4486	0
	andrew	wls	5380	8094	5316	8846	5433	9705	5528	10000	5514	10000
		wlsr	4620	1906	4684	1154	4567	295	4472	0	4486	0
	ramsay	wls	1524	944	2232	2325	2709	4065	2885	6916	2699	0
		wlsr	8476	9056	7768	7675	7291	5935	7115	3084	7301	10000
100	ols		115	0	138	0	150	0	356	0	569	0
	lsr		9885	10000	9862	10000	9850	10000	9644	10000	9431	10000
	huber	wls	645	0	738	0	910	0	1208	0	1435	0
		wlsr	9355	10000	9262	10000	9090	10000	8792	10000	8565	10000
	tukey	wls	3892	6370	5012	9042	5505	9876	5699	10000	5710	10000
		wlsr	6108	3630	4988	958	4495	124	4301	0	4290	0
	andrew	wls	5410	8755	5500	9440	5545	9883	5701	10000	5711	10000
		wlsr	4590	1245	4500	560	4455	117	4299	0	4289	0
	ramsay	wls	680	1	799	0	1034	0	1324	0	1497	0
		wlsr	9320	9999	9201	10000	8966	10000	8676	10000	8503	10000

Table 7 indicates the comparison of WLS and WLSR with 10%*n outliers for different sample sizes and error variances. According to mae β and mae y, except for Andrew's and Tukey's functions we can say that WLSR gives better results than WLS in case of increasing the sample size and the error variance.

Table 8. Comparison of WLS and WLSR with 20%*n outliers, which are equal to 500

n			30		50		100		500		1000	
δ_e^2			mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y
1	ols		33	0	68	0	142	0	517	0	873	0
	lsr		9967	10000	9932	10000	9858	10000	9483	10000	9127	10000
	huber	wls	11	0	25	0	58	0	282	0	522	0
		wlsr	9989	10000	9975	10000	9942	10000	9718	10000	9478	10000
	tukey	wls	13	0	32	0	60	0	311	0	604	0
		wlsr	9987	10000	9968	10000	9940	10000	9689	10000	9396	10000
	andrew	wls	5260	6869	5371	7793	5353	8915	5428	9996	5462	10000
		wlsr	4740	3131	4629	2207	4647	1085	4572	4	4538	0
	ramsay	wls	10	0	36	0	50	0	258	0	457	0
		wlsr	9990	10000	9964	10000	9950	10000	9742	10000	9543	10000
9	ols		36	0	74	0	135	0	514	0	812	0
	lsr		9964	10000	9926	10000	9865	10000	9486	10000	9188	10000
	huber	wls	22	0	33	0	69	0	244	0	564	0
		wlsr	9978	10000	9967	10000	9931	10000	9756	10000	9436	10000
	tukey	wls	24	0	38	0	80	0	290	0	623	0
		wlsr	9976	10000	9962	10000	9920	10000	9710	10000	9377	10000
	andrew	wls	5216	8489	5392	9334	5353	9894	5398	10000	5405	10000
		wlsr	4784	1511	4608	666	4647	106	4602	0	4595	0
	ramsay	wls	25	0	26	0	64	0	268	0	476	0
		wlsr	9975	10000	9974	10000	9936	10000	9732	10000	9524	10000
25	ols		57	0	84	0	141	0	584	0	868	0
	lsr		9943	10000	9916	10000	9859	10000	9416	10000	9132	10000
	huber	wls	39	0	51	0	72	0	284	0	531	0
		wlsr	9961	10000	9949	10000	9928	10000	9716	10000	9469	10000
	tukey	wls	40	0	51	0	82	0	329	0	584	0
		wlsr	9960	10000	9949	10000	9918	10000	9671	10000	9416	10000
	andrew	wls	5316	9142	5402	9725	5449	9981	5502	10000	5435	10000
		wlsr	4684	858	4598	275	4551	19	4498	0	4565	0
	ramsay	wls	49	0	50	0	79	0	247	0	490	0
		wlsr	9951	10000	9950	10000	9921	10000	9753	10000	9510	10000
100	ols		104	0	141	0	208	0	655	0	965	0
	lsr		9896	10000	9859	10000	9792	10000	9345	10000	9035	10000
	huber	wls	112	0	125	0	179	0	372	0	630	0
		wlsr	9888	10000	9875	10000	9821	10000	9628	10000	9370	10000
	tukey	wls	110	0	124	0	181	0	400	0	683	0
		wlsr	9890	10000	9876	10000	9819	10000	9600	10000	9317	10000
	andrew	wls	5479	9623	5507	9948	5511	10000	5684	10000	5622	10000
		wlsr	4521	377	4493	52	4489	0	4316	0	4378	0
	ramsay	wls	131	0	137	0	185	0	360	0	649	0
		wlsr	9869	10000	9863	10000	9815	10000	9640	10000	9351	10000

Table 8 indicates the comparison of WLS and WLSR with 20%*n outliers for different sample sizes and error variances. According to mae β and mae y, except for Andrew's function we can say that WLSR gives better results than WLS in case of increasing the sample size and the error variance.

Table 9. Comparison of the success of WLS and WLSR Methods, except for Andrew's function

	n	30		50		100		500		1000	
	δ_e^2	mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y	mae β	mae y
Non-Outlier	1	wls	wls	wls	wls	wls	wls	wls	wls	wls	wls
	9	wls	wls	wls	wls	wls	wls	wls	wls	wls	wls
	25	wls	wls	wls	wls	wls	wls	wls	wls	wls	wls
	100	wls	wls	wls	wls	wls	wls	wls	wls	wls	wls
10% Outliers	1	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
	9	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
	25	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
	100	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
20% Outliers	1	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
	9	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
	25	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr
	100	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr	wlsr

Table 9 indicates the comparison of the success of WLS and WLSR Methods with different error variances and outliers rates. We can say that WLS Method outperform than WLSR Method in case of non-outlier. We can also say that WLSR Method gives better results than WLS Method in case of increasing outlier ratios, and the error variance.

6. Conclusions and Future Work

In this study, it is shown which method (WLS and WLSR) gives better results in M-estimation according to the mean absolute errors (MAE) of the estimated regression parameters and dependent value via a simulation study using different sample sizes and error variances. It was studied on Huber, Tukey, Andrew and Ramsay's weighting functions in this paper. Based on the simulation results, we can say that WLSR gives better estimates than WLS in case of the presence of outliers and increased error variance apart from Andrew's weighting function. For future work, other weighting functions in the literature can be examined for which method gives better results.

Acknowledgements

I would like to thank JForce Information Technologies Inc. Istanbul, Turkey for supporting this study. I would also like to Dr. Oguz Akbilgic for sharing his knowledge with me.

References

[1] Akbilgic O., Akinci E. D., 2009. A novel regression approach: Least squares ratio. Communications in Statistics - Theory and Methods, 38:9, 1539-1545,

[2] Andersen R., 2007. Modern Methods for robust regression. Sage Publications.

[3] Andrews D. F., 1974. A Robust method for multiple linear regression. Technometrics, Vol.16, 523-531.

[4] Andrews D. F., Bickel P. J., Hampel F. R., Huber P. J., Rogers W. H., and Tukey J. W., 1972. Robust estimates of location. (Princeton Univ. Press, Princeton).

[5] Banas M., Ligas M., 2014. Empirical tests of performance of some M – estimators. Geodesy and Cartography. Vol. 63, No 2, 2014, pp. 127-146

[6] Chatterjee S., Machler M., 1997. Robust regression:a weighted least squares approach. Communications in Statistics - Theory and Methods, 26(6), 1381-1394.

[7] Chang X.-W., Guo Y., 2005. Huber's m-estimation in relative GPS positioning: computational aspects. Journal of Geodesy, Volume 79, Issue 6-7, pp 351-362.

[8] Drapper N. R., Smith H., 1998. Applied regression analysis. New York: Wiley.

[9] Duchnowski, R. 2011. Sensitivity of robust estimators applied in strategy for testing stability of reference points. EIF approach. *Geodesy and Cartography*, 60(2), 123-134

[10] Hoaglin D. C, Mosteller F., and Tukey J. W., 1983. Understanding robust and exploratory data analysis. John Wiley and Sons, New York.

[11] Huber P. J., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35(1), 73-101.

[12] Koller M., Stahel W. A., 2011. Sharpening wald-type inference in robust regression for small samples. Computational Statistics & Data Analysis 55(8), 2504–2515.

[13] Muthukrishnan R., Radha M., 2010. M-estimators in regression models. Journal of Mathematics Research. Vol. 2, No. 4, pp. 23-27.

[14] Rousseeuw P. J., Leroy A. M., 2003. Robust regression and outlier detection. Wiley-Interscience; first edition.