Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 23, Özel Sayı, 119-130, 2019

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 23, Special Issue, 119-130, 2019

DOI: 10.19113/sdufenbed.538869

The Comparison of the Estimators for the Parameters of the General Linear Regression Model via Simulation and Two Real Life Data Examples

Nalan MUTLU[1], Hakan Savaş SAZAK*[2]

[1,2]Ege University, Faculty of Science, Department of Statistics, 35100, İzmir
[1](ORCID: https://orcid.org/0000-0002-7520-8805)
[2](ORCID: https://orcid.org/0000-0001-6123-1214)

(Alınış / Received: 17.05.2018, Kabul / Accepted: 24.12.2018, Online Yayınlanma / Published Online: 12.03.2019)

Keywords
General linear regression model,
Least squares,
M-estimators,
MM-estimators,
S-estimators,
Robust

Abstract: In this study we compared the efficiency and robustness of several estimators, namely, the least squares (LS) estimators, the Huber and Tukey M-estimators, the S-estimators and the MM-estimators for the parameters of the general linear regression (GLR) model via simulation. First, the programs for each method were written by using Matlab. Then, an extensive simulation study was conducted under several models. The results are consistent with the literature but some important points were also found to be remarked. As the literature suggests, in general, the MM-estimators are the most efficient estimators, and among the robust estimators discussed here, the S-estimators are the least efficient ones. Naturally, the LS estimators are badly affected by the deviations from the assumed model because of their sensitive nature. Moreover, it was found that while the LS estimator of the variance of the error term is unbiased, the robust estimators discussed here are generally biased. Additionally, the MM-estimator of the variance of the error term is less biased than the other robust estimators and its bias gets smaller faster as the sample size increases compared to the others. At the end of the study, to be more illustrative, two real life data examples were given with the related comments.

Genel Doğrusal Regresyon Modelinin Parametrelerine Yönelik Tahmin Edicilerin Simülasyon Yoluyla Karşılaştırılması ve İki Gerçek Hayat Veri Örneği

Anahtar Kelimeler
Genel doğrusal regresyon modeli,
En küçük kareler,
M-tahmin edicileri,
MM-tahmin edicileri,
S-tahmin edicileri,
Dayanıklı

Özet: Bu çalışmada genel doğrusal regresyon modelinin parametrelerine yönelik bir çok tahmin edicinin ki bunlar en küçük kareler (EKK) tahmin edicileri, Huber ve Tukey M-tahmin edicileri, S-tahmin edicileri ve MM-tahmin edicileri olmak üzere etkinlik ve dayanıklılıklarını simülasyon yoluyla karşılaştırdık. Öncelikle her bir yöntem için Matlab kullanılarak program yazıldı. Daha sonra bir çok model altında kapsamlı bir simülasyon çalışması yürütüldü. Sonuçlar literatürle uyumlu olmakla beraber üstünde durulması gereken bazı önemli noktalar da bulunmuştur. Literatürde önerildiği şekilde genel olarak MM-tahmin edicileri en etkin tahmin edicilerdir ve burada ele alınan dayanıklı tahmin ediciler arasında S-tahmin edicileri en az etkinliğe sahiptirler. Doğal olarak EKK tahmin edicileri hassas yapıları sebebiyle varsayılan modelden sapmalardan kötü bir şekilde etkilenmektedirler. Ayrıca hata teriminin varyansının EKK tahmin edicisi yansızken burada ele alınan dayanıklı tahmin edicilerinin genelde yanlı olduğu bulunmuştur. Bunun yanında hata teriminin varyansının MM-tahmin edicisi diğer dayanıklı tahmin edicilere göre daha az yanlıyken örneklem hacmi arttıkça da yan miktarı diğerlerine göre daha hızlı bir şekilde azalmaktadır. Çalışmanın sonunda daha aydınlatıcı olması için ilgili yorumlarıyla beraber iki gerçek hayat verisi örneği verilmiştir.

1. Introduction

The general linear regression (GLR) model covers many particular cases and for this reason it can be used for general purposes [1]. The GLR model can be given as follows

$$\underset{n\times 1}{Y} = \underset{n\times(p+1)}{X}\ \underset{(p+1)\times 1}{\beta} + \underset{n\times 1}{\varepsilon} \tag{1}$$

Here, Y is the vector of the response variable, X is the matrix of the independent variables, β is the vector of the parameters, ε is the vector of the error term, n is

* Corresponding author: ssazak@yahoo.com

the sample size and p is the number of slope parameters. The assumptions related to Eq. (1) are

$$E(\varepsilon)=0$$

$$Var(\varepsilon)=\sigma^2 I$$

$$rank(X)=p+1$$

where I is the identity matrix. The second assumption implies constant variance and the independence of the error terms (under normality). In some situations, instead of the second assumption, the following assumption can be made for flexibility about the variance of the error terms and their independence but at an expense of much more complicated estimation process.

$$Var(\varepsilon)=\Sigma$$

In real life, there are very rare situations where the parameters are known. Thus, the sample should be used effectively to estimate the parameters of interest. The least squares (LS) method is generally used for the estimation of the parameters since it is very easy and straightforward but also known to be very sensitive against deviations from the assumed models and distributions. Quite many estimators which are called robust have been proposed so far to compensate for the sensitive nature of the LS estimators but none of them were fully efficient under normality although some of them possess satisfactory efficiencies. In this study we intended to compare the efficiency and robustness of the most popular estimators related to the GLR model via simulation. In the literature their efficiencies are already known but there are more points to be revealed about their properties. For this purpose, we first wrote programs for each method by using Matlab and conducted an extensive simulation study by using our own programs. Using our own programs in the simulations is an important contribution in the area. We also found some interesting features of the robust estimators of the variance of the error term. Conducting the regression analysis of the real life data sets using several graphical and numerical tools is another merit of the study. This paper is organized as follows. In Section 2 we give more detailed information about the literature and the methods used in this study. Section 3 presents the simulation results and the related comments. In Section 4 two real life data examples are given to illustrate the usage of the mentioned methods in the estimation of the parameters of interest with the related graphics and comments. The final section includes discussion and some concluding remarks.

2. The Literature Review and the Methods Included in the Study

In this section we will introduce the methods used in this study in detail.

2.1. The least squares method

The LS method is based on finding the estimators which minimizes the error sum of squares and used in many areas because of its easiness in the estimation process. Specifically, it is defined as follows

$$min \sum_{i=1}^{n} \varepsilon_i^2 \tag{2}$$

It can also be defined by using the matrix format as

$$min\ \varepsilon'\varepsilon \tag{3}$$

Depending on Eq. (1), taking the derivative w.r.t. β gives the following LS estimator of β

$$\hat{\beta}=(X'X)^{-1}X'Y \tag{4}$$

The LS estimator of σ^2 is the minimized errors sum of squares divided by the degrees of freedom of the residuals which is given below for our case as

$$\hat{\sigma}^2=\frac{(Y-X\hat{\beta})'(Y-X\hat{\beta})}{n-(p+1)}=\frac{e'e}{n-(p+1)} \tag{5}$$

2.2. The weighted least squares method

The weighted least squares (WLS) method is based on finding the estimators minimizing the weighted error sum of squares. In fact both the WLS and the LS are special cases of the generalized least squares (GLS) which is based on finding the estimators minimizing the Mahalanobis distance of the errors. The WLS method with the given weights w_i is defined as

$$min \sum_{i=1}^{n} w_i\varepsilon_i^2 \tag{6}$$

If we define $\underset{n\times n}{W}=diag(w_i; i=1,\dots,n)$, we can also give its definition by using the matrix format as follows

$$min\ \varepsilon'W\varepsilon \tag{7}$$

Taking the derivative w.r.t. β gives the following WLS estimator

$$\hat{\beta}=(X'WX)^{-1}X'WY \tag{8}$$

2.3. The iteratively reweighted least squares algorithm

Most of the robust estimation methods require iteration since they cannot be obtained explicitly. In order to solve them by iteration, generally, the iteratively reweighted least squares (IRWLS) algorithm is utilized. Andersen [2] gave the definition of its steps as follows

Step 1: In the beginning, let iteration number q be zero, q=0. $\hat{\beta}^{(q)}$ is obtained as an initial estimate by using the LS method.

Step 2: At iteration number q, the residuals $e_i^{(q)}$ are obtained by using the estimate $\hat{\beta}^{(q)}$. By using the residuals $e_i^{(q)}$, we calculate the estimate for the standard deviation of the error term $\hat{\sigma}^{(q)}$ with the median absolute deviation (MAD) from the formulas given below as proposed by Hampel et al. [3].

$$\hat{\sigma}^{(q)} = 1.4826\,MAD \tag{9}$$

$$MAD(e^{(q)}) = median\left| e_i^{(q)} - median(e_i^{(q)}) \right| \tag{10}$$

The coefficient 1.4826 in Eq. (9) is used to make the estimator of the standard deviation unbiased under normality (see Hampel et al. [3] for details).

Step 3: The residuals are standardized with $\hat{\sigma}^{(q)}$ as $u_i^{(q)} = \frac{e_i^{(q)}}{\hat{\sigma}^{(q)}}$ and the weights $w_i^{(q)}$ are obtained by using $u_i^{(q)}$.

Step 4: The WLS estimator $\hat{\beta}^{(q+1)}$ is obtained which minimizes the weighted error sum of squares $\sum_{i=1}^{n} w_i^{(q)} \varepsilon_i^2$. When expressed in the matrix format, it can be given as below

$$\hat{\beta}^{(q+1)} = \left(X'W^{(q)}X\right)^{-1} X'W^{(q)}Y \tag{11}$$

where $\underset{n\times n}{W^{(q)}} = diag\left(w_i^{(q)}; i = 1,\ldots,n\right)$.

Step 5: The condition given in Eq. (12) is checked for convergence at iteration q for a prespecified small value δ. The iteration stops at convergence. Otherwise, we continue to Step 6.

$$\frac{\left\| \hat{\beta}^{(q+1)} - \hat{\beta}^{(q)} \right\|}{\left\| \hat{\beta}^{(q+1)} \right\|} < \delta \tag{12}$$

Step 6: Iteration number q is increased by 1 unit, q=q+1. Then, we go to Step 2.

2.4. The least median of squares

The least median of squares (LMS) method was found by Rousseeuw [4]. It is based on minimizing the median of the squares of the errors which is given below

$$min\ \ median(\varepsilon_i^2) \tag{13}$$

The LMS method is known to have high break down point (BDP) but low efficiency under normality [5].

2.5. The M-estimators

The M-estimators were introduced by Huber [6] as a result of a search to find a robust alternative for the LS method which is known to be very sensitive to possible shifts from the assumed model. The principle of the M-estimation is minimizing the sum of a selected ρ function of the errors instead of the sum of squares of them. Thus, in this sense, the LS method is a special case of the M-estimation method. More specifically, the definition of the M-estimators can be given by the following expression

$$min\ \sum_{i=1}^{n} \rho(\varepsilon_i) \tag{14}$$

There are many alternatives for ρ function serving different purposes which can be found in Türkay [7]. Though depending on the selection of ρ function, in general, the M-estimators are robust with low BDP and high efficiency w.r.t. the LMS estimators. In general, ρ functions are not linear, and for this reason, the estimation process requires iteration. It is quite common to use the IRWLS algorithm to obtain the estimators. Susanti et al. [8] gave an algorithm which can be used to obtain the M-estimators. Its only difference from the IRWLS algorithm is that it includes some tests about the validity of the regression model, the existence of outliers and the significance of the independent variables in the model. For the M-estimators, we used two weight functions, namely, the Huber and Tukey bisquare weight functions. The weight function of Huber is [7]

$$w_i = \begin{cases} 1 & \text{if } |u_i| \le c \\ c/|u_i| & \text{if } |u_i| > c \end{cases} \tag{15}$$

The weight function of Tukey bisquare is [7]

$$w_i = \begin{cases} \left(1 - \left(\frac{u_i}{c}\right)^2\right)^2 & \text{if } |u_i| \le c \\ 0 & \text{if } |u_i| > c \end{cases} \tag{16}$$

We used c=1.345 and c=4.685 in our study, respectively, for the cases of the Huber and Tukey M-estimation as suggested by Holland and Welsch [9] to maintain 95% asymptotic efficiency w.r.t. the LS estimators under normality.

2.6. The S-estimators

The S-estimators which possess high BDP were found by Rousseeuw and Yohai [10]. They are called the S-estimators because they are based on the scale estimation of the errors. It is the generalized form of the LMS method [11]. The S-estimation method minimizes the sum of the function ρ of the scaled errors satisfying the conditions defined in Rousseeuw and Yohai [10]. It is aimed to increase the efficiency of the LMS method by using a robust but more efficient scale estimator than the median [8]. In this sense, it is defined by the following expression

$$min \sum_{i=1}^{n} \rho(\frac{\varepsilon_i}{\sigma_s}) \qquad (17)$$

For a specific sample, according to Rousseeuw and Yohai [10], the following equation is solved to obtain the S-estimators

$$\frac{1}{n}\sum_{i=1}^{n} \rho(\frac{e_i}{\hat{\sigma}_s}) = K \qquad (18)$$

Here, K is the expected value of ρ under the standard normal distribution, $\hat{\sigma}_s$ is the S-estimator of scale for the error term and e_i are the residuals calculated by using the S-estimator of the β vector [5]. Susanti et al. [8] gave an algorithm to obtain the S-estimators as follows

Step 1: $\hat{\beta}$ is obtained as an initial estimate by using the LS method.

Step 2: The residuals $e_i = Y_i - \hat{Y}_i$ are obtained by using the latest estimate $\hat{\beta}$.

Step 3: By using the residuals obtained in the previous step, we calculate the estimate for the standard deviation of the error term by using the following formula

$$\hat{\sigma}_s = \begin{cases} 1.4826\,MAD & \text{for iteration} = 1 \\ \sqrt{\dfrac{1}{nK}\displaystyle\sum_{i=1}^{n} w_i e_i^2} & \text{for iteration} > 1 \end{cases} \qquad (19)$$

Step 4: The residuals are standardized as $u_i = \dfrac{e_i}{\hat{\sigma}_s}$.

Step 5: The weights w_i are obtained by the following formula

$$w_i = \begin{cases} \begin{cases} \left(1-\left(\dfrac{u_i}{c}\right)^2\right)^2 & \text{if } |u_i| \le c \\ 0 & \text{if } |u_i| > c \end{cases} & \text{for iteration} = 1 \\ \dfrac{\rho(u_i)}{u_i^2} & \text{for iteration} > 1 \end{cases} \qquad (20)$$

Step 6: $\hat{\beta}$ is obtained from Eq. (8) by using the weights w_i by utilizing the WLS method.

Step 7: The steps 2-6 are repeated till convergence between the latter and former estimates is established.

Rousseeuw and Yohai [10] suggested using c=1.547 so that the BDP of the S-estimators is 0.5 (50%). Stuart [12] stated that the following objective function $\rho(u_i)$ which is associated with the Tukey bisquare weight function can be used in obtaining the S-estimators

$$\rho(u_i) = \begin{cases} \dfrac{u_i^2}{2} - \dfrac{u_i^4}{2c^2} + \dfrac{u_i^6}{6c^4} & \text{if } |u_i| \le c \\ \dfrac{c^2}{6} & \text{if } |u_i| > c \end{cases} \qquad (21)$$

By taking numerical integration in Matlab, we obtained the corresponding BDP values for some specific values of c and K where K is directly related to the value of c. Rousseeuw and Yohai [10] also provided the asymptotic relative efficiencies of the S-estimators for some selected values of c w.r.t. the LS estimators under normality. Table 1 given here is in exact conformity with the corresponding tables in Rousseeuw and Yohai [10] and Stuart [12].

It was noted by Rousseeuw and Yohai [10] that the S-estimators with the tuning constant c=1.547 can hardly be used as a final estimate because of a very low asymptotic efficiency of 28.7% w.r.t. the LS estimators under normality but they can be used as an initial estimate because of the high BDP of 50%.

Table 1. The asymptotic relative efficiency and BDP of the S-estimators corresponding to some selected values of c and K for the Tukey bisquare function

BDP	Efficiency	c	K
50%	28.7%	1.547	0.1995
45%	37.0%	1.756	0.2312
40%	46.2%	1.988	0.2634
35%	56.0%	2.251	0.2957
30%	66.1%	2.560	0.3278
25%	75.9%	2.973	0.3593
20%	84.7%	3.420	0.3899
15%	91.7%	4.096	0.4194
12%	95.0%	4.685	0.4368
10%	96.6%	5.182	0.4475

2.7. The MM-estimators

The MM-estimators were found by Yohai [13] to maintain both high efficiency under normality and robustness with high BDP at the same time. He suggested using the S-estimators with the tuning constant c=1.547 in the early stage to maintain a BDP of 50%, and using the S-estimators for the standard deviation and the M-estimators for β both with the tuning constant c=4.685 in the remaining stages to maintain an asymptotic relative efficiency of 95% w.r.t. the LS estimators under normality using the Tukey bisquare function in all stages. The steps of the algorithm to obtain the MM-estimators are [8]

Step 1: $\hat{\beta}$ is obtained as an initial estimate by using the S-estimation method with the tuning constant c=1.547.

Step 2: The residuals $e_i = Y_i - \hat{Y}_i$ are obtained by using the latest estimate $\hat{\beta}$.

Step 3: By using the residuals obtained in the previous step, we calculate the estimate for the standard deviation $\hat{\sigma}_s$ by using the S-estimation method but with the tuning constant c=4.685.

Step 4: The residuals are standardized as $u_i = \dfrac{e_i}{\hat{\sigma}_s}$.

Step 5: The weights w_i are obtained by using the Tukey bisquare function as in Eq. (16) with c=4.685.

Step 6: $\hat{\beta}$ is obtained from Eq. (8) by using the weights w_i by utilizing the WLS method.

Step 7: The steps 2-6 are repeated till convergence between the latter and former estimates is established.

3. The Simulation Results

In order to compare the efficiency and robustness of the estimators mentioned in this paper, namely, the LS estimators, the Huber and Tukey M-estimators, the S-estimators and the MM-estimators, an extensive simulation study was conducted including several models. Although all the programs in this study were written in Matlab for the GLR model given in Eq. (1), for easy interpretation and commentary, the simulations were conducted for the simple linear regression model which is a special case of the GLR model and also given below (see Mutlu [14] for details)

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (22)$$

In this model, β_0 is the intercept, β_1 is the slope parameter and ε is the error term. Without loss of generality, $\beta_0 = 0$, $\beta_1 = 1$ and $\sigma^2 = 1$. The sample sizes were taken as n=30, 50 and 100 for nn=[300000/n] Monte Carlo runs. The models (distributions) included in this study are

Model 1: Bivariate normal;

$$X \sim N(0,1),\ \varepsilon \sim N(0,1)$$

Model 2: Design variable X and normal error term;

$$X = 1,\ldots,n,\ \varepsilon \sim N(0,1)$$

Model 3: Mixture type outlier model;

$$X = 1,\ldots,n,\ \varepsilon \sim \pi N(0,k^2) + (1-\pi)N(0,1),\ \pi = 0.1$$

Model 4: Dixon's outlier model;

$$X = 1,\ldots,n,\ \varepsilon \sim r\ N(0,k^2) + (n-r)\ N(0,1), r = [0.5 + 0.1n]$$

Remark: For the models 2-4, X is a design variable instead of a normally distributed (or a stochastic) variate. Although this is not realistic in most of the situations in real life, X is generally assumed to be a design variable in regression analysis for easiness in theoretical inferences. See Sazak et al. [15] for a detailed discussion on this topic. Since this is quite common in the literature, we took X as a design variable for these models.

We produced $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ for all the estimators we mentioned before and obtained their simulated means, biases, variances and mean square errors (mse), and calculated the relative efficiency (REff) of the estimators w.r.t. the LS estimators. The REff of $\hat{\theta}_1$ w.r.t. $\hat{\theta}_2$, both being the estimators of θ, can be given by the following formula

$$\text{REff}\ (\hat{\theta}_1 \mid \hat{\theta}_2) = 100\frac{mse(\hat{\theta}_2)}{mse(\hat{\theta}_1)} \qquad (23)$$

For all the iterative algorithms, we used $\delta = 0.00001$ in Eq. (12) to check whether the convergence is established. For easy interpretation, we scaled the errors by dividing them by $\sqrt{1-\pi+\pi k^2}$ and $\sqrt{1-(r/n)+(r/n)k^2}$ for model 3 and 4, respectively, so that $\sigma^2 = 1$. For these models, we took k=3.

Table 2. The simulated values for the model 1 with n=30

		β_0	β_1	σ^2
mean	LS	0.001489	1.000279	0.999613
	Huber M	0.001497	0.999944	0.944790
	Tukey M	0.001333	0.999670	0.941852
	S	0.000706	1.000546	0.878201
	MM	0.001425	1.000059	0.937580
bias	LS	0.001489	0.000279	-0.000387
	Huber M	0.001497	-0.000056	-0.055210
	Tukey M	0.001333	-0.000330	-0.058148
	S	0.000706	0.000546	-0.121799
	MM	0.001425	0.000059	-0.062420
nxvar	LS	1.033659	1.121631	2.145740
	Huber M	1.094252	1.184038	5.102216
	Tukey M	1.123756	1.216181	5.087551
	S	1.436233	1.590554	3.311405
	MM	1.084026	1.173604	1.923670
nxmse	LS	1.033726	1.121633	2.145745
	Huber M	1.094319	1.184038	5.193660
	Tukey M	1.123810	1.216184	5.188986
	S	1.436248	1.590563	3.756454
	MM	1.084087	1.173604	2.040559
REff	LS	100.0000	100.0000	100.0000
	Huber M	94.4629	94.7295	41.3147
	Tukey M	91.9840	92.2256	41.3519
	S	71.9740	70.5180	57.1216
	MM	95.3545	95.5717	105.1547

Table 3. The simulated values for the model 1 with n=50

		β_0	β_1	σ^2
mean	LS	0.001380	1.000013	0.999432
	Huber M	0.001154	0.999671	0.968208
	Tukey M	0.000986	0.999679	0.965943
	S	0.000897	0.999442	0.926836
	MM	0.001103	0.999801	0.962087
bias	LS	0.001380	0.000013	-0.000568
	Huber M	0.001154	-0.000329	-0.031792
	Tukey M	0.000986	-0.000321	-0.034057
	S	0.000897	-0.000558	-0.073164
	MM	0.001103	-0.000199	-0.037913
nxvar	LS	1.026472	1.072909	2.099407
	Huber M	1.075403	1.122060	5.187122
	Tukey M	1.089619	1.137526	5.162613
	S	1.406942	1.467706	3.499878
	MM	1.071375	1.117582	1.986270
nxmse	LS	1.026567	1.072909	2.099423
	Huber M	1.075470	1.122066	5.237660
	Tukey M	1.089667	1.137531	5.220606
	S	1.406983	1.467722	3.767527
	MM	1.071436	1.117584	2.058141
REff	LS	100.0000	100.0000	100.0000
	Huber M	95.4529	95.6191	40.0832
	Tukey M	94.2092	94.3191	40.2142
	S	72.9623	73.1003	55.7242
	MM	95.8122	96.0025	102.0058

Tables 2-4 include the simulation results for the model 1 for n=30, 50 and 100, respectively. Depending on the simulation results, in general, we see a quite good efficiency of the MM-estimators even for the bivariate normal distribution. We see that all the estimators of β_0 and β_1 produce unbiased estimates. When we investigate the estimation of σ^2, all the estimators have downward bias other than the LS estimator. In general, the biases of the robust estimators tend to decrease as the sample size increases but the bias of the S-estimator of σ^2 is the largest of all. For this model, as expected, the best performance was shown by the LS estimators but the MM-estimators have quite high efficiencies compared to the other robust estimators. The MM-estimator of σ^2 has higher efficiency than the LS counterpart with the advantage of producing a biased estimator. Although this may sound weird, some methods deliberately produce bias to have smaller variance and mse such as ridge regression. As the sample size increases, the bias of the MM-estimator of σ^2 becomes almost zero which makes it slightly less efficient than the LS estimator but its efficiency is still quite impressive such as 99.71%. The worst performance was shown by the S-estimators for the model 1. This is not a surprising result since the S-estimators are generally used for the initial stages in iterations.

Table 4. The simulated values for the model 1 with n=100

		β_0	β_1	σ^2
mean	LS	0.001333	1.000241	0.999707
	Huber M	0.001097	0.999719	0.982784
	Tukey M	0.000892	0.999861	0.981396
	S	0.000767	0.999654	0.962741
	MM	0.000974	0.999879	0.980696
bias	LS	0.001333	0.000241	-0.000293
	Huber M	0.001097	-0.000281	-0.017216
	Tukey M	0.000892	-0.000139	-0.018604
	S	0.000767	-0.000346	-0.037259
	MM	0.000974	-0.000121	-0.019304
nxvar	LS	1.002436	1.043011	2.049762
	Huber M	1.041683	1.090182	5.204749
	Tukey M	1.045447	1.094382	5.199698
	S	1.356219	1.396886	3.516815
	MM	1.039586	1.086767	2.018372
nxmse	LS	1.002614	1.043017	2.049771
	Huber M	1.041804	1.090190	5.234387
	Tukey M	1.045527	1.094384	5.234308
	S	1.356278	1.396898	3.655640
	MM	1.039681	1.086769	2.055635
REff	LS	100.0000	100.0000	100.0000
	Huber M	96.2383	95.6729	39.1597
	Tukey M	95.8956	95.3063	39.1603
	S	73.9239	74.6667	56.0715
	MM	96.4348	95.9742	99.7147

We give the simulation results for the model 2 in Tables 5-7. The results are quite similar with those in the model 1.

Table 5. The simulated values for the model 2 with n=30

		β_0	β_1	σ^2
mean	LS	0.003337	0.999842	1.003854
	Huber M	0.003404	0.999847	0.949110
	Tukey M	0.003845	0.999846	0.945773
	S	0.005157	0.999749	0.882981
	MM	0.003903	0.999836	0.942574
bias	LS	0.003337	-0.000158	0.003854
	Huber M	0.003404	-0.000153	-0.050890
	Tukey M	0.003845	-0.000154	-0.054227
	S	0.005157	-0.000251	-0.117019
	MM	0.003903	-0.000164	-0.057426
nxvar	LS	4.115740	0.013261	2.158996
	Huber M	4.342526	0.013989	5.100173
	Tukey M	4.458823	0.014340	5.072576
	S	5.809854	0.018699	3.349994
	MM	4.307334	0.013878	1.948387
nxmse	LS	4.116074	0.013262	2.159442
	Huber M	4.342873	0.013990	5.177867
	Tukey M	4.459266	0.014341	5.160792
	S	5.810652	0.018701	3.760800
	MM	4.307791	0.013879	2.047320
REff	LS	100.0000	100.0000	100.0000
	Huber M	94.7777	94.7948	41.7052
	Tukey M	92.3038	92.4763	41.8432
	S	70.8367	70.9162	57.4198
	MM	95.5495	95.5550	105.4765

Table 6. The simulated values for the model 2 with n=50

		β_0	β_1	σ^2
mean	LS	-0.000948	1.000072	1.003932
	Huber M	-0.000097	1.000051	0.972852
	Tukey M	-0.000329	1.000068	0.970896
	S	0.001156	1.000007	0.931397
	MM	-0.000342	1.000068	0.967362
bias	LS	-0.000948	0.000072	0.003932
	Huber M	-0.000097	0.000051	-0.027148
	Tukey M	-0.000329	0.000068	-0.029104
	S	0.001156	0.000007	-0.068603
	MM	-0.000342	0.000068	-0.032638
nxvar	LS	3.975830	0.004751	2.089644
	Huber M	4.228725	0.005057	5.243984
	Tukey M	4.272772	0.005109	5.211853
	S	5.635368	0.006741	3.504558
	MM	4.194087	0.005018	1.987915
nxmse	LS	3.975875	0.004752	2.090417
	Huber M	4.228726	0.005058	5.280835
	Tukey M	4.272778	0.005109	5.254204
	S	5.635435	0.006741	3.739876
	MM	4.194093	0.005018	2.041177
REff	LS	100.0000	100.0000	100.0000
	Huber M	94.0206	93.9492	39.5850
	Tukey M	93.0513	93.0001	39.7856
	S	70.5513	70.4876	55.8954
	MM	94.7970	94.6913	102.4124

Table 7. The simulated values for the model 2 with n=100

		β_0	β_1	σ^2
mean	LS	-0.001105	1.000039	1.003695
	Huber M	-0.000274	1.000037	0.986725
	Tukey M	-0.000314	1.000040	0.986074
	S	-0.001225	1.000059	0.967540
	MM	-0.000654	1.000044	0.985599
bias	LS	-0.001105	0.000039	0.003695
	Huber M	-0.000274	0.000037	-0.013275
	Tukey M	-0.000314	0.000040	-0.013926
	S	-0.001225	0.000059	-0.032460
	MM	-0.000654	0.000044	-0.014401
nxvar	LS	3.913309	0.001211	2.023032
	Huber M	4.133156	0.001274	5.267381
	Tukey M	4.141214	0.001278	5.251880
	S	5.364764	0.001626	3.653573
	MM	4.098923	0.001267	2.019453
nxmse	LS	3.913431	0.001212	2.024397
	Huber M	4.133164	0.001274	5.285004
	Tukey M	4.141224	0.001278	5.271272
	S	5.364914	0.001626	3.758937
	MM	4.098965	0.001267	2.040192
REff	LS	100.0000	100.0000	100.0000
	Huber M	94.6837	95.0919	38.3046
	Tukey M	94.4994	94.7722	38.4043
	S	72.9449	74.5178	53.8556
	MM	95.4736	95.6138	99.2258

In this part of the study we investigate the robustness of the estimators we mentioned before by using the model 3 and 4. The simulation results belonging to the model 3 are given in Tables 8-10. This model represents the mixture type outlier model. For β_0 and β_1, although there are no big differences between the M-estimators and the MM-estimators, in general, we can say that the MM-estimators outperform the M-estimators. The LS estimators are the least efficient ones because of their sensitive nature. The S-estimators show the worst performance among the robust estimators mentioned here. For σ^2, we see very interesting results. The only unbiased estimates are produced by the LS estimator and all the robust estimators produce downward bias and this downward bias does not get smaller as the sample size increases. This fact makes the LS estimator of σ^2, asymptotically the most efficient one. Although the robust estimators are taking advantage of their bias compared to the LS estimator for small sample sizes (such as 30), as the sample size increases, their relative efficiencies drop steadily and dramatically. Most of the robust estimators of σ^2 get worse than the LS estimator after the sample size of 30. The only robust estimator of σ^2 which survives after 30 is the MM-estimator. Even at the sample size of 100, it is 113.61% efficient w.r.t. the LS estimator of σ^2 but surely, it will not last so long since the bias stays the same.

Table 8. The simulated values for the model 3 with *n*=30

		β_0	β_1	σ^2
mean	LS	0.001812	0.999928	1.004437
	Huber M	0.002853	0.999862	0.625516
	Tukey M	0.003889	0.999802	0.622030
	S	0.004803	0.999774	0.589114
	MM	0.002743	0.999852	0.795836
bias	LS	0.001812	-0.000072	0.004437
	Huber M	0.002853	-0.000138	-0.374484
	Tukey M	0.003889	-0.000198	-0.377970
	S	0.004803	-0.000226	-0.410886
	MM	0.002743	-0.000148	-0.204164
nxvar	LS	4.146446	0.013257	7.600526
	Huber M	3.042198	0.009834	2.342228
	Tukey M	3.049290	0.009853	2.331207
	S	3.612089	0.011630	1.622072
	MM	3.023759	0.009782	3.167211
nxmse	LS	4.146544	0.013257	7.601116
	Huber M	3.042442	0.009835	6.549373
	Tukey M	3.049744	0.009854	6.617053
	S	3.612781	0.011631	6.686889
	MM	3.023985	0.009782	4.417699
REff	LS	100.0000	100.0000	100.0000
	Huber M	136.2900	134.7958	116.0587
	Tukey M	135.9637	134.5325	114.8716
	S	114.7743	113.9761	113.6719
	MM	137.1219	135.5210	172.0605

Table 9. The simulated values for the model 3 with *n*=50

		β_0	β_1	σ^2
mean	LS	-0.001575	1.000089	1.004148
	Huber M	-0.000906	1.000069	0.637626
	Tukey M	-0.000779	1.000069	0.634845
	S	0.000942	1.000013	0.617546
	MM	-0.001254	1.000074	0.800916
bias	LS	-0.001575	0.000089	0.004148
	Huber M	-0.000906	0.000069	-0.362374
	Tukey M	-0.000779	0.000069	-0.365155
	S	0.000942	0.000013	-0.382454
	MM	-0.001254	0.000074	-0.199084
nxvar	LS	3.999825	0.004730	7.563032
	Huber M	2.917305	0.003494	2.292132
	Tukey M	2.893122	0.003476	2.300236
	S	3.498770	0.004157	1.660401
	MM	2.867689	0.003459	2.995086
nxmse	LS	3.999949	0.004730	7.563892
	Huber M	2.917346	0.003495	8.857890
	Tukey M	2.893153	0.003476	8.967129
	S	3.498814	0.004157	8.973945
	MM	2.867768	0.003459	4.976805
REff	LS	100.0000	100.0000	100.0000
	Huber M	137.1091	135.3626	85.3916
	Tukey M	138.2557	136.0714	84.3513
	S	114.3230	113.7934	84.2873
	MM	139.4795	136.7475	151.9829

Table 10. The simulated values for the model 3 with *n*=100

		β_0	β_1	σ^2
mean	LS	-0.000890	1.000031	1.003803
	Huber M	-0.000530	1.000035	0.644339
	Tukey M	-0.000420	1.000031	0.643750
	S	-0.000734	1.000051	0.638170
	MM	-0.001269	1.000040	0.804471
bias	LS	-0.000890	0.000031	0.003803
	Huber M	-0.000530	0.000035	-0.355661
	Tukey M	-0.000420	0.000031	-0.356250
	S	-0.000734	0.000051	-0.361830
	MM	-0.001269	0.000040	-0.195529
nxvar	LS	3.972222	0.001215	7.619661
	Huber M	2.820909	0.000866	2.335945
	Tukey M	2.764601	0.000854	2.355297
	S	3.289559	0.000987	1.696334
	MM	2.765930	0.000857	2.884842
nxmse	LS	3.972301	0.001215	7.621107
	Huber M	2.820937	0.000866	14.985453
	Tukey M	2.764619	0.000854	15.046700
	S	3.289613	0.000987	14.788463
	MM	2.766091	0.000858	6.707997
REff	LS	100.0000	100.0000	100.0000
	Huber M	140.8150	140.3023	50.8567
	Tukey M	143.6835	142.2524	50.6497
	S	120.7528	123.0198	51.5341
	MM	143.6071	141.6382	113.6123

Tables 11-13 include the simulation results for the model 4 which represents the Dixon's outlier model. For this model, the effect of the outliers on the LS estimators is more devastating than that in the model 3. Here, for β_0 and β_1, the M-estimators are performing better than the MM-estimators. Among the M-estimators, there are no big differences. Again for β_0 and β_1, compared to the model 3, the S-estimators are doing better, at least not as bad as they are doing in the model 3. When we investigate the situation in the estimation of σ^2, we see that, similar to the model 3, all the robust estimators have downward bias and it stays the same regardless of the sample size but the effect of this bias is much more devastating than that in the model 3. Even at the sample size of 30, only the MM-estimator could survive since only the MM-estimator is better than the LS estimator of σ^2. For the large sample size (*n*=100), even the efficiency of the MM-estimator drops below 100% (91.89%) w.r.t. the LS estimator. The M-estimators and the S-estimator of σ^2 are very badly affected from the bias similar to the model 3 but here the situation is much worse. This simulation result also shows the asymptotic superiority of the LS estimator of σ^2 because of its unbiasedness, surely for the distributions with finite mean and variance.

Table 11. The simulated values for the model 4 with n=30

		β_0	β_1	σ^2
mean	LS	0.006609	0.999687	0.975827
	Huber M	0.004079	0.999825	0.628920
	Tukey M	0.002677	0.999897	0.622735
	S	0.003628	0.999827	0.582426
	MM	0.004482	0.999802	0.775523
bias	LS	0.006609	-0.000313	-0.024173
	Huber M	0.004079	-0.000175	-0.371080
	Tukey M	0.002677	-0.000103	-0.377265
	S	0.003628	-0.000173	-0.417574
	MM	0.004482	-0.000198	-0.224477
nxvar	LS	8.527658	0.021458	5.368042
	Huber M	5.103122	0.013861	2.238542
	Tukey M	5.135022	0.014034	2.217537
	S	5.385578	0.015130	1.467573
	MM	5.208504	0.014078	1.958859
nxmse	LS	8.528968	0.021461	5.385571
	Huber M	5.103621	0.013862	6.369550
	Tukey M	5.135237	0.014035	6.487409
	S	5.385973	0.015131	6.698603
	MM	5.209106	0.014080	3.470560
REff	LS	100.0000	100.0000	100.0000
	Huber M	167.1160	154.8168	84.5518
	Tukey M	166.0871	152.9132	83.0158
	S	158.3552	141.8333	80.3984
	MM	163.7319	152.4235	155.1788

Table 12. The simulated values for the model 4 with n=50

		β_0	β_1	σ^2
mean	LS	-0.000190	1.000042	0.995350
	Huber M	-0.000022	1.000036	0.637060
	Tukey M	0.000774	1.000021	0.632556
	S	-0.001331	1.000069	0.612412
	MM	-0.000125	1.000043	0.792943
bias	LS	-0.000190	0.000042	-0.004650
	Huber M	-0.000022	0.000036	-0.362940
	Tukey M	0.000774	0.000021	-0.367444
	S	-0.001331	0.000069	-0.387588
	MM	-0.000125	0.000043	-0.207057
nxvar	LS	8.405250	0.007788	5.437603
	Huber M	4.909027	0.004933	2.244378
	Tukey M	4.906349	0.004928	2.233783
	S	5.144559	0.005396	1.533913
	MM	5.019659	0.004994	1.974759
nxmse	LS	8.405252	0.007788	5.438684
	Huber M	4.909027	0.004933	8.830636
	Tukey M	4.906379	0.004928	8.984543
	S	5.144647	0.005396	9.045138
	MM	5.019659	0.004994	4.118397
REff	LS	100.0000	100.0000	100.0000
	Huber M	171.2203	157.8654	61.5888
	Tukey M	171.3127	158.0146	60.5338
	S	163.3786	144.3154	60.1283
	MM	167.4467	155.9370	132.0583

Table 13. The simulated values for the model 4 with n=100

		β_0	β_1	σ^2
mean	LS	-0.002047	1.000047	0.997094
	Huber M	-0.000643	1.000034	0.644137
	Tukey M	-0.000302	1.000030	0.642874
	S	-0.002288	1.000065	0.636052
	MM	0.000028	1.000023	0.800574
bias	LS	-0.002047	0.000047	-0.002906
	Huber M	-0.000643	0.000034	-0.355863
	Tukey M	-0.000302	0.000030	-0.357126
	S	-0.002288	0.000065	-0.363948
	MM	0.000028	0.000023	-0.199426
nxvar	LS	8.095279	0.001940	5.380468
	Huber M	4.702883	0.001223	2.313005
	Tukey M	4.691248	0.001213	2.333000
	S	4.892428	0.001298	1.604173
	MM	4.816244	0.001240	1.878930
nxmse	LS	8.095698	0.001940	5.381312
	Huber M	4.702925	0.001223	14.976870
	Tukey M	4.691257	0.001214	15.086909
	S	4.892952	0.001299	14.849970
	MM	4.816244	0.001240	5.855995
REff	LS	100.0000	100.0000	100.0000
	Huber M	172.1418	158.6241	35.9308
	Tukey M	172.5699	159.8695	35.6688
	S	165.4563	149.3665	36.2379
	MM	168.0915	156.4915	91.8941

4. Real Life Data Examples

In this section we illustrate the methods explained in this study with two real life data examples. For both examples we will explore the data sets in detail using several graphical and numerical tools and give the results with the related comments.

4.1. The real life data example 1

For the first example we will work on the R air quality data set containing 153 daily readings of air quality values between May 1 and September 30 in 1973 from New York including 6 variables. Because of the missing values in the data set, only 111 observations will be used in the regression analysis [12]. The response variable Y is the mean ozone concentration (in parts per billion) from 13:00 to 15:00 hours at Roosevelt Island. There are three independent variables which are X_1; the solar radiation in Langleys in the frequency band 4000-7700 from 08:00 to 12:00 hours at Central Park, X_2; the average wind speed (in miles per hour) between 07:00 and 10:00 hours at LaGuardia Airport and X_3; the maximum daily temperature in degrees Fahrenheit again at LaGuardia Airport.

First we assume that the GLR model given in Eq. (1) and the very general conditions related to this model given just after it are right which we will check in the following stages. Then, we obtain and give the parameter estimates produced by the estimators mentioned in this study in Table 14. It is quite surprising to see that the S-estimators produced the smallest MSE and the largest R^2 values with great differences from the corresponding values produced by the other estimators despite the known low efficiency of the S-estimators in the literature. While the Huber and Tukey M-estimators and the MM-estimators produced close results, the Huber M-estimators are slightly worse. The LS estimators produced the poorest result regarding the MSE and R^2. Now we will investigate the validity of the model, the assumptions and possible outliers if any. Please note that the estimate of σ^2 and the MSE value for the LS method are the same while for the other methods they are different.

Table 14. The regression estimates for the example 1

	LS	Hub. M	Tuk. M	S	MM
β_0	-64.342	-78.862	-84.704	-83.902	-84.107
β_1	1.652	1.750	1.782	1.799	1.780
β_2	-3.334	-2.656	-2.298	-2.768	-2.344
β_3	0.060	0.049	0.045	0.049	0.046
σ^2	448.624	314.227	324.661	281.779	361.789
MSE	448.624	291.179	236.205	58.324	246.935
R^2	0.606	0.657	0.679	0.765	0.675

In order to verify the current model, we conduct residual analysis based on the LS estimates. The plot of $\hat{Y}$ vs. the studentized deleted residuals ($t_{i(i)}$) is given in Figure 1 (see Neter et al. [1] for the definitions).

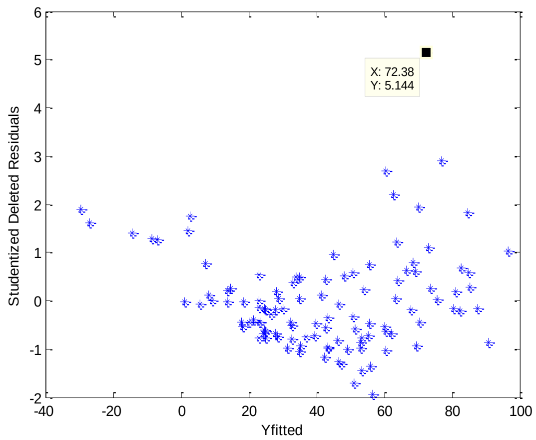

Figure 1. The plot of $\hat{Y}$ vs. $t_{i(i)}$ for the example 1

Figure 1 shows no sign of a mispecified model or invalidity of the assumptions since the residuals are randomly distributed without any systematic band while showing a couple of positive outliers one of which being very extreme and tagged on the plot. The boxplot of $t_{i(i)}$ given in Figure 2 also supports the existence of the outliers.

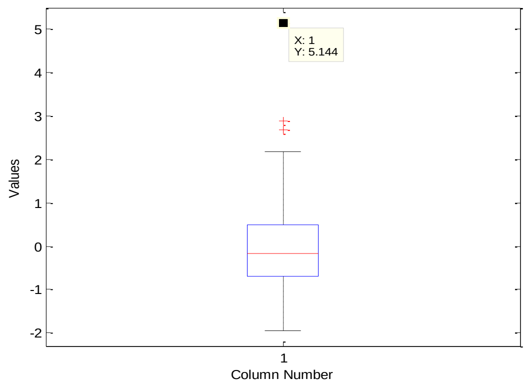

Figure 2. The boxplot of $t_{i(i)}$ for the example 1

The quantile-quantile (Q-Q) plot of $t_{i(i)}$ under normality is given in Figure 3 which is in fact needed just for the hypothesis testing process and the possible full efficiency of the LS estimators. It also shows the existence of the outliers one of which being very extreme while also showing the approximate normality of the remaining residuals.

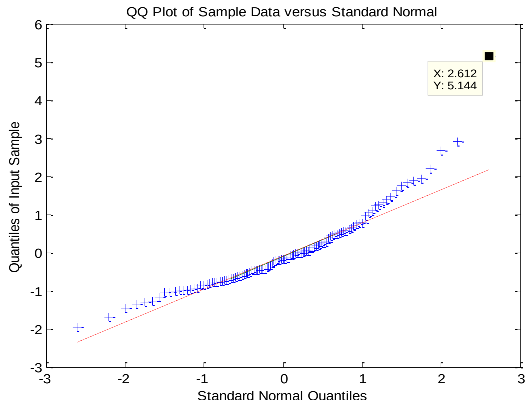

Figure 3. The Q-Q plot of $t_{i(i)}$ for the example 1

From all the plots obtained here, we can observe that there are totally 3 outliers one of which being very extreme. These are, from the largest magnitude to the smallest, the 77th ($t_{i(i)}$ =5.1440), the 34th ($t_{i(i)}$ =2.8963) and the 23rd ($t_{i(i)}$ =2.6813) observations. One should discard these observations unless a robust method is utilized. The existence of the outliers may be the reason of the high performance of the S-estimators for this real life data set.

4.2. The real life data example 2

For the second example we will work on the data set obtained from 22 patients who applied to Hacettepe University Hospital in Ankara, the capital city of Turkey [16]. The data set contains 3 independent variables, X_1 ; the amount of the osteocalcin hormone, X_2 ; the amount of the parathyroid hormone and X_3 ; the age. The response variable Y is the bone mineral density. Again, we assumed the GLR model and the related conditions in the first place and obtained the estimates depending on them. The estimates are given in Table 15. It is again surprising to see that the best result is obtained by using the S-estimators regarding the *MSE* and R^2 although the S-estimators are known to be inefficient compared to the other robust estimators. The *MSE* and R^2 values of the other robust estimators are very close to each other. The LS estimators show the poorest performance producing the largest *MSE* and the lowest R^2 values.

Table 15. The regression estimates for the example 2

	LS	Hub. M	Tuk. M	S	MM
β_0	1.0874	1.0772	1.0856	1.0446	1.0853
β_1	0.0288	0.0335	0.0340	0.0402	0.0336
β_2	-0.0022	-0.0024	-0.0024	-0.0025	-0.0024
β_3	-0.0060	-0.0061	-0.0062	-0.0059	-0.0062
σ^2	0.0039	0.0031	0.0030	0.0022	0.0032
MSE	0.0039	0.0031	0.0029	0.0005	0.0029
R^2	0.4029	0.4842	0.5047	0.7424	0.4970

Now we will investigate the validity of the model and the assumptions accepted in the beginning of the study. The plot of $\hat{Y}$ vs. $t_{i(i)}$ is given in Figure 4. This plot verifies that the model is right and the assumptions are valid since there is no systematic behavior and any change in the variability of the residuals. We just observe one outlier on the negative side of the residuals, possibly showing an observation which is much smaller than the others. This outlier is also detected by the boxplot given in Figure 5.

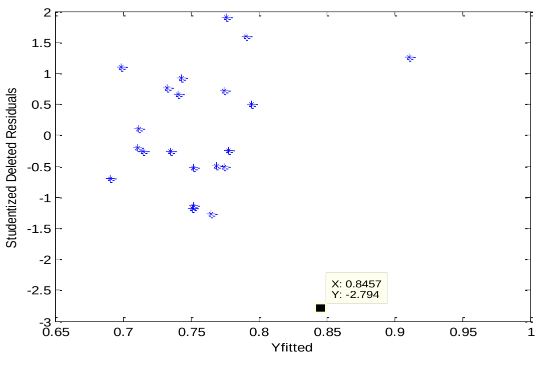

Figure 4. The plot of $\hat{Y}$ vs. $t_{i(i)}$ for the example 2

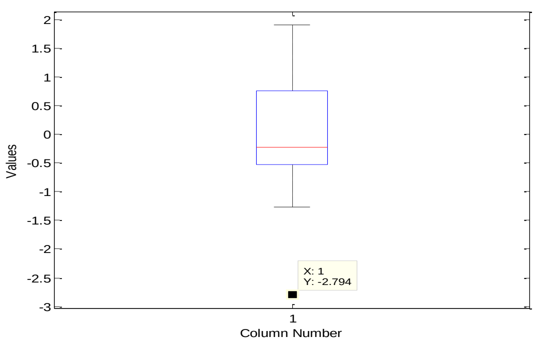

Figure 5. The boxplot of $t_{i(i)}$ for the example 2

The Q-Q plot of $t_{i(i)}$ under normality is given in Figure 6. The Q-Q plot shows that there is one outlier and the rest of the residuals have almost a perfect normal distribution. The outlier is the 7th observation with $t_{i(i)}$ =-2.7939.

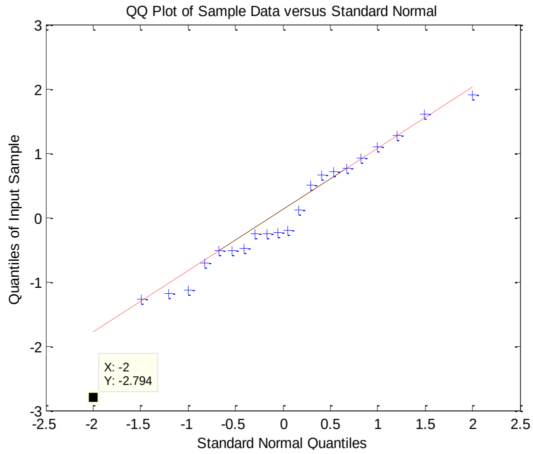

Figure 6. The Q-Q plot of $t_{i(i)}$ for the example 2

Again for this real life data set, we can say that the S-estimators may be better because of the existing outlier since they are known with their extreme robustness despite their inefficiency compared to the other robust estimators.

5. Discussion and Conclusion

In this study we compared several estimators for the parameters of the GLR model including the LS estimators, the Huber and Tukey M-estimators, the S-estimators and the MM-estimators via simulation by using our own programs written in Matlab. We obtained results consistent with the literature but also found some interesting results to be remarked. The MM-estimators are, in general, the most efficient ones as expected. The S-estimators are the least efficient ones among the robust estimators studied in this paper. The LS estimators are naturally the most efficient ones under normality but too sensitive to the deviations from the assumed models. We have also found that the robust estimators of the variance of the error term are generally biased and in some situations they stay biased despite the increase in the sample size whereas the LS estimator of the variance of the error term is always unbiased. Among the robust estimators of the variance of the error term, the MM-estimator is less biased than the others and its bias tends to get smaller faster compared to the others in most of the situations. In order to be more illustrative, we gave two real life data examples using some extra statistical measures and graphics to check the validity of the GLR model and the assumptions made. We also gave the values of the *MSE* and R^2 for the comparison of the estimators mentioned here. Surprisingly, the S-estimators showed brilliant performance in both examples despite their known low efficiency compared to the other robust estimators. The existence of the outliers may be the reason of this performance. This fact also made the LS estimators the worst of all. The case studies are beneficial for illustration but they are also useful to see that all the real life data sets are original and they have to be considered on their own.

References

[1] Neter, J., Kutner, M. H., Nachstheim, C .J., Wasserman, W. 1996. Applied Linear Statistical Models. McGraw-Hill, USA.

[2] Andersen, R. 2008. Modern Methods for Robust Regression. Thousand Oaks: SAGE Publications.

[3] Hampel, F. R, Ronchetti, E. M., Rousseeuw P. J. 1986. Robust Statistics. Wiley, New York.

[4] Rousseeuw, P. 1984. Least Median of Squares Regression. Journal of the American Statistical Association, 79, 871-880.

[5] Rousseeuw, P., Leroy, M. 1987. Robust Regression and Outlier Detection. Wiley, New York.

[6] Huber, P. J. 1964. Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics, 35, 73-101.

[7] Türkay, H. 2004. Doğrusal Regresyon Analizinde M Tahminciler ve Ekonometrik Bir Uygulama. Doğu Anadolu Bölgesi Araştırmaları, 106-115.

[8] Susanti, Y., Pratiwi, H., Sulistijowati, S., Liana, T. 2014. M estimation, S estimation, and MM estimation in robust regression. International Journal of Pure and Applied Mathematics, 91(3), 349-360.

[9] Holland, P. W., Welsch, R. E. 1977. Robust Regression Using Iteratively Reweighted Least-Squares. Communications in Statistics-Theory and Methods, 6(9), 813-827.

[10] Rousseeuw, P., Yohai, V. 1984. Robust Regression by Means of S-Estimators. Robust and Nonlinear Time Series Analysis, edited by J. Franke, W. Hardle, D. Martin, Lecture Notes in Statistics, 26, 256-272, Springer Verlag, Berlin/New York.

[11] Çetin, M., Toka, O. 2011. The Comparing of S-Estimator and M-Estimators in Linear Regression. Gazi University Journal of Science, 24, 747-752.

[12] Stuart, C. 2011. Robust Regression. Department of Mathematical Sciences, Durham University.

[13] Yohai, V. J. 1987. High Breakdown Point and High Efficiency Robust Estimates for Regression. The Annals of Statistics, 15(2), 642-656.

[14] Mutlu, N. 2018. The Comparison of the Iteration Methods Used in Parameter Estimation, M.Sc. Thesis, Ege University, İzmir.

[15] Sazak, H. S., Tiku, M. L., Islam, M. Q. 2006. Regression Analysis with a Stochastic Design Variable. International Statistical Review, 74(1), 77-88.

[16] Çetin, M. C., Orsoy, A. 2001. Doğrusal Regresyonda Sağlam Tahmin Ediciler ve Bir Uygulama. Anadolu University Journal of Science and Technology, 2(2), 265-270.