Open Journal of Statistics, 2022, 12, 261-276
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

A Comparison of the Estimators of the Scale Parameter of the Errors Distribution in the L_1 Regression

Carmen D. Saldiva de André, Silvia Nagib Elian*

Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil
Email: *selian@ime.usp.br

How to cite this paper: de André, C.D.S. and Elian, S.N. (2022) A Comparison of the Estimators of the Scale Parameter of the Errors Distribution in the L_1 Regression. *Open Journal of Statistics*, **12**, 261-276.
https://doi.org/10.4236/ojs.2022.122018

Received: March 18, 2022
Accepted: April 21, 2022
Published: April 24, 2022

Abstract

The L_1 regression is a robust alternative to the least squares regression whenever there are outliers in the values of the response variable, or the errors follow a long-tailed distribution. To calculate the standard errors of the L_1 estimators, construct confidence intervals and test hypotheses about the parameters of the model, or to calculate a robust coefficient of determination, it is necessary to have an estimate of a scale parameter τ. This parameter is such that τ^2/n is the variance of the median of a sample of size n from the errors distribution. [1] proposed the use of $\hat{\tau}$, a consistent, and so, an asymptotically unbiased estimator of τ. However, this estimator is not stable in small samples, in the sense that it can increase with the introduction of new independent variables in the model. When the errors follow the Laplace distribution, the maximum likelihood estimator of τ, say $\hat{\tau}^*$, is the mean absolute error, that is, the mean of the absolute residuals. This estimator always decreases when new independent variables are added to the model. Our objective is to develop asymptotic properties of $\hat{\tau}^*$ under several errors distributions analytically. We also performed a simulation study to compare the distributions of both estimators in small samples with the objective to establish conditions in which $\hat{\tau}^*$ is a good alternative to $\hat{\tau}$ for such situations.

Keywords

Minimum Sum of Absolute Errors Regression, Multiple Linear Regression, Variable Selection, Heavy Tail Distributions, Asymptotic Theory

1. Introduction

Consider the multiple linear regression model

$$y = X\beta + \varepsilon,$$

where

y is an $n \times 1$ vector of values of the response variable corresponding to X, an $n \times k$ matrix of predictor variables that may include a column of ones for the intercept term;

β is a $k \times 1$ vector of unknown parameters and;

ε is an $n \times 1$ vector of unobservable random errors.

The components of ε are independent and identically distributed random variables with cumulative distribution function F. Suppose that F has a unique median equal to zero and a continuous derivative f in the neighborhood of zero such that $f(0) > 0$. The scale parameter of f is defined as

$$\tau = \left(2f(0)\right)^{-1}. \quad (1.1)$$

So τ^2/n is the variance of the median in a sample of size n from the error distribution.

The L_1 estimator $\hat{\beta}$ of β, minimizes $\sum_{i=1}^{n}\left|y_i - x_i\beta\right|$ for all values of β, where y_i is the i-th element of the vector y and x_i is the i-th row of the matrix X.

The L_1 criterion is a robust alternative to the least squares regression whenever the data contains outliers or the errors follow a long tailed distribution such as Laplace or Cauchy.

It is well known that when the errors follow Laplace distribution, the L_1 estimators of β are maximum likelihood estimators and so, they are asymptotically unbiased and efficient. [2] proved that the L_1 estimator is asymptotically unbiased, consistent and follows a multinormal distribution with covariance matrix $\tau^2(X'X)^{-1}$. An important implication of this result is that the L_1 estimator of β has a smaller confidence ellipsoid than the least squares estimator for any error distribution for which the sample median is a more efficient estimator than the sample mean.

Based on the asymptotic distribution results, formulae for constructing confidence intervals and testing hypotheses on the parameters of the model have been developed [3] [4]. To apply these formulae and also to compute the standard errors of the estimators of β, it is necessary to have an estimate of the parameter τ. Several estimators of τ were proposed [1]. They recommend the consistent estimator

$$\hat{\tau} = \sqrt{n^*}\left(e_{\left(n^*-m+1\right)} - e_{(m)}\right)\Big/4,$$

where

$$m = \left(n^* + 1\right)\Big/2 - \sqrt{n^*},$$

n^* is the number of non-zero residuals and $e_{(1)}, \cdots, e_{(n^*)}$ are the non-zero residuals arranged in ascending order.

It is important to observe that $\hat{\tau}$ is a measure of the variability of the residuals and, although is influenced by all of them, it is determined by only two of them.

A consistent estimator of τ is also needed to calculate the robust coefficient of determination R proposed by [5]. This coefficient is an informal measure of goodness of fit of a model and it is given by

$$R_2 = RSEA/\left(RSEA + (n-p-1)(\hat{\tau}/2)\right),$$

where

$$RSEA = \sum_{i=1}^{n}\left|y_i - \hat{\beta}_0\right| - \sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|,$$

$$\hat{\beta}_0 = \text{median}\left(y_1, \cdots, y_n\right),$$

$\hat{y}_i$ is the predicted value of the response variable in the i-th observation, that is $\hat{y}_i = x_i\hat{\beta}$;

$\hat{\beta}$ is the L_1 estimator of the regression coefficients of the model.

A desirable property for a coefficient of determination is that it increases when passing from a reduced to a full model, that is, when new predictor variables are included in the model [6]. For R_2 this property is true only if $\hat{\tau}$ decreases as new variables are included in the model and this might not happen, as shown in Example 1.

Example 1—In this example, we use the real state data from [7]. The predictor variables are taxes, in hundred dollars (X_1), lot area, in thousand squares feet (X_2), living space, in thousand squares feet (X_3), age of the home, in years (X_4). The response variable (Y) is the selling price of the home, in thousands of dollars.

In **Table 1**, we present all possible linear models obtained with the four predicted variables, the number of parameters (k), the estimates $\hat{\tau}$ and the values of R_2 for each model. In this table, we see that the value of $\hat{\tau}$ for the model with variable X_1 only (4.0079) is smaller than the observed value of this statistic in the model with X_1 and X_3 (5.4301), and then the value of R_2 in the model containing only X_1 as predictor is larger than in the model with X_1 and X_3. However, the contribution of X_3 given that X_1 is already in the model is significant (p-value less than 0.01).

So, it may happen that the value of $\hat{\tau}$ increases even with the introduction of a variable with significant contribution in the model. This fact will decrease the value of R_2 and the new model might not be selected if the coefficient of determination is the criterion to select a model.

This instable behavior of $\hat{\tau}$ may be explained by the fact that it is determined by only two residuals and so we can expect that it happens more frequently in small samples.

When errors follow the Laplace distribution, the maximum likelihood estimator of τ is the mean absolute error [8], given by

$$\hat{\tau}^* = SEA/n$$

where

$$SEA = \sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|.$$

Table 1. Number of parameters (k), $\hat{\tau}$ and R_2 observed values for all possible regression models for the state data.

Variables in the model	k	$\hat{\tau}$	R_2
nothing	1	10.9629	0.0000
X_1	2	4.0079	0.7105
X_2	2	8.0285	0.3830
X_3	2	9.8973	0.4681
X_4	2	11.8750	0.1164
$X_1 \cdot X_2$	3	3.9590	0.7215
$X_1 \cdot X_3$	3	5.4301	0.6860
$X_1 \cdot X_4$	3	3.9636	0.7212
$X_2 \cdot X_3$	3	8.0913	0.5470
$X_2 \cdot X_4$	3	6.5918	0.4576
$X_3 \cdot X_4$	3	6.6577	0.6098
$X_1 \cdot X_2 \cdot X_3$	4	6.1331	0.6701
$X_1 \cdot X_2 \cdot X_4$	4	7.3275	0.5937
$X_1 \cdot X_3 \cdot X_4$	4	5.4875	0.7011
$X_2 \cdot X_3 \cdot X_4$	4	4.0022	0.7410
$X_1 \cdot X_2 \cdot X_3 \cdot X_4$	5	4.0246	0.7704

Although the usual regularity conditions do not hold for Laplace distribution, $\hat{\tau}^*$ is a consistent estimator of τ [9]. This estimator is a measure of variability of the residuals, and it has the property of decreasing when new predictor variables are included in the model. Using this estimator, it is possible to construct a robust coefficient of determination that satisfies the desirable conditions in [6]. It is possible also to calculate the coefficient of determination adjusted by the number of predictor variables proposed in [10].

Our objective is to study the possibility of using $\hat{\tau}^*$ as an alternative to $\hat{\tau}$ when the errors follow a distribution other than Laplace. We have special interest in small sample sizes because of the instable behavior of $\hat{\tau}$ in such cases.

[11] pointed out the importance of the L_1 method of estimation, presenting many practical situations in which its application is recommended. So, the search of procedures that make its use more efficient gives important contribution to the statistical theory.

The paper is organized as follows. Initially, the asymptotic distributions of $\hat{\tau}^*$ were derived analytically assuming errors with Normal, Mixture of Normals, Laplace and Logistic distributions. These results allowed to compute the asymptotic bias and mean squared error of this estimator. Then, we performed a simulation study and generated empirical distributions of $\hat{\tau}^*$ and $\hat{\tau}$ in small samples, after the fitting of models with one predictor variable and errors with the same distributions considered previously and Cauchy distribution also. The distributions considered in this study were characterized according to the weight of their tails [12]. The results obtained in this study allowed indicating situations in which $\hat{\tau}^*$ can be used as an alternative to estimate τ.

2. Asymptotic Distribution of $\hat{\tau}^*$

In this section, we derive analytically the asymptotic distribution of $\hat{\tau}^*$, considering four different distributions for the errors. We assume errors with normal (0, σ^2) distribution, mixture of Normals when random variables are selected from a normal (0, 1) with probability p and of a normal (0, σ^2) with probability 1 − p, Logistic distribution with mean zero and variance $\gamma^2\pi^2/3$ and Laplace distribution with mean zero and variance $2\sigma^2$.

First, we note that $\hat{\tau}^*$ may be written as

$$\hat{\tau}^* = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - x_i\hat{\beta}\right| = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - x_i\beta - \left(x_i\hat{\beta} - x_i\beta\right)\right|$$

where

$\hat{\beta}$ is the L_1 estimator of β.

Since $x_i\hat{\beta}$ is a consistent estimator of $x_i\beta$ [2], the asymptotic distribution of $\hat{\tau}^*$ is the same of $\frac{1}{n}\sum_{i=1}^{n}\left|y_i - x_i\beta\right|$, and this quantity is equal to $\frac{1}{n}\sum_{i=1}^{n}\left|\varepsilon_i\right|$, where ε_i is the i-th element of the vector of errors of the model. Next, we study the asymptotic distribution of this random variable for different errors distributions.

3. Errors with Normal (0, σ^2) Distribution

Using the fact that $\frac{\varepsilon_i}{\sigma} \sim N(0,1)$ we observe (see Appendix A) that

$$E\left(\frac{\left|\varepsilon_i\right|}{\sigma}\right) = \frac{\sqrt{2}}{\sqrt{\pi}} \text{ and } Var\left(\frac{\left|\varepsilon_i\right|}{\sigma}\right) = \frac{\pi - 2}{\pi}.$$

Further, the random variables $\frac{\left|\varepsilon_i\right|}{\sigma}$, $i = 1,\cdots,n$, are independent and identically distributed and, by the Central-limit theorem

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\varepsilon_i\right|}{\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}}}{\frac{\sqrt{\pi - 2}}{\sqrt{\pi}\sqrt{n}}} \xrightarrow{D} N(0,1).$$

So, it follows that

$$W_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - x_i\hat{\beta}\right|}{\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}}}{\frac{\sqrt{\pi - 2}}{\sqrt{\pi}\sqrt{n}}} = \frac{\frac{\hat{\tau}^*}{\sigma} - \frac{\sqrt{2}}{\sqrt{\pi}}}{\frac{\sqrt{\pi - 2}}{\sqrt{\pi}\sqrt{n}}} \xrightarrow{D} N(0,1).$$

Finally, since $\varepsilon_i \sim \text{Normal}\left(0,\sigma^2\right)$ then $\tau = \frac{\sqrt{2\pi}}{2}\sigma$, which implies that

$$W_n = \left(\frac{\pi}{\sqrt{2(\pi-2)}}\sqrt{n}\frac{\hat{\tau}^*}{\tau} - \frac{\sqrt{2n}}{\sqrt{\pi-2}}\right) \xrightarrow{D} N(0,1),$$

or that

$$W_n = \frac{\sqrt{n}\pi}{\sqrt{2(\pi-2)}}\left(\frac{\hat{\tau}^*}{\tau} - \frac{2}{\pi}\right) \xrightarrow{D} N(0,1).$$

4. Errors with Mixture of Normal Distributions

In this case, we assume that the errors distribution is a mixture of two normal distributions: a Normal (0, 1) selected with probability p and a Normal (0, σ^2) selected with probability ($1 - p$). Hence, the probability density function of ε_i is

$$f(\varepsilon_i) = \frac{p}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon_i^2}{2}\right) + \frac{1-p}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right), \quad -\infty < \varepsilon_i < \infty.$$

It is not very difficult to see that
$E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = p + (1-p)\sigma^2$ and that the parameter τ is

$$\tau = \frac{\sqrt{2\pi}\sigma}{2(p\sigma + 1 - p)}$$

Furthermore, $|\varepsilon_i| = |y_i - x_i\beta|$, $i = 1,\cdots,n$ are independent and identically distributed random variables with mean and variance (see Appendix A) given by

$$E(|\varepsilon_i|) = \frac{2p}{\sqrt{2\pi}} + \frac{2(1-p)\sigma}{\sqrt{2\pi}} \quad \text{and}$$

$$Var(|\varepsilon_i|) = \bar{\sigma}^2 = \frac{\left(\pi(1-p) - 2(1-p)^2\right)\sigma^2 - 4p(1-p)\sigma + p\pi - 2p^2}{\pi}.$$

So, as the same way that in the Normal errors distribution case

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\beta| - \frac{2p + 2(1-p)\sigma}{\sqrt{2\pi}}}{\frac{\bar{\sigma}}{\sqrt{n}}} \xrightarrow{D} N(0,1),$$

and

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\hat{\beta}| - \frac{2p + 2(1-p)\sigma}{\sqrt{2\pi}}}{\frac{\bar{\sigma}}{\sqrt{n}}} \xrightarrow{D} N(0,1),$$

that is,

$$U_n = \frac{\sqrt{n}\frac{\hat{\tau}^*}{\tau} - \frac{2(p + (1-p)\sigma)\sqrt{n}}{\sqrt{2\pi}\tau}}{\frac{\bar{\sigma}}{\tau}} \xrightarrow{D} N(0,1).$$

5. Errors with Logistic Distribution

Let us suppose that the errors follow a Logistic distribution with probability density function

$$f(\varepsilon_i) = \frac{\exp(-\varepsilon_i/\gamma)}{\gamma\left[1+\exp(-\varepsilon_i/\gamma)\right]^2}, \quad -\infty < \varepsilon_i < \infty,$$

such that $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \gamma^2\pi^2/3$, and, therefore, $\tau = 2\gamma$.

Also, because $|\varepsilon_i|$, $i = 1,\cdots,n$, are independent and identically distributed random variables, it is proved in the Appendix A that the mean and the variance of these variables are 1.386γ and $1.37\gamma^2$.

So, by the Central-limit theorem

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\beta| - 1.386\gamma}{\frac{\gamma\sqrt{1.37}}{\sqrt{n}}} \xrightarrow{D} N(0,1).$$

Using the same arguments of the previous demonstrations, it follows that

$$\frac{\hat{\tau}^* - 1.386\gamma}{\frac{\gamma\sqrt{1.37}}{\sqrt{n}}} = \frac{\frac{\hat{\tau}^*}{\tau} - \frac{1.386\gamma}{\tau}}{\frac{\gamma\sqrt{1.37}}{\tau\sqrt{n}}} \xrightarrow{D} N(0,1)$$

Since in this case $\tau = 2\gamma$, we have

$$\frac{2\sqrt{n}}{\sqrt{1.37}}\left(\frac{\hat{\tau}^*}{\tau} - 0.693\right) \xrightarrow{D} N(0,1).$$

6. Errors with Laplace Distribution

When the errors follow a Laplace distribution with mean zero and variance $2\sigma^2$ then

$$\tau = \sigma, \quad E(|\varepsilon_i|) = \sigma \quad \text{and} \quad Var(|\varepsilon_i|) = \sigma^2.$$

Therefore, because of the Central-limit theorem

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i\beta| - \sigma}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1),$$

and hence

$$\frac{\hat{\tau}^* - \tau}{\frac{\tau}{\sqrt{n}}} = \sqrt{n}\left(\frac{\hat{\tau}^*}{\tau} - 1\right) \xrightarrow{D} N(0,1).$$

Remark: Based on the asymptotic distribution of $\hat{\tau}^*$, confidence intervals for τ can be developed. In the Normal errors case, an asymptotic confidence interval for τ is

$$\left(k_1\hat{\tau}^*, k_2\hat{\tau}^*\right)$$

where, $k_1 = \dfrac{\sqrt{n}\pi}{2\sqrt{n} + z\sqrt{2(\pi-2)}}$, $k_2 = \dfrac{\sqrt{n}\pi}{2\sqrt{n} - z\sqrt{2(\pi-2)}}$ and z is the percentile of order $(1 + \gamma)/2$ of the standard Normal distribution and γ is the confidence coefficient of the interval.

This confidence interval enables us to test hypothesis like H: $\tau = \tau_0$ at a significance level $\alpha = (1 - \gamma)$.

7. Asymptotic Bias of $\hat{\tau}^*$

Because the sample mean of the values of the absolute residuals is a continuous and limited function, by the Helly-Bray Lemma [13], the expectation of $\hat{\tau}^*$ converges to the mean of its asymptotic distribution. Therefore, it is possible to calculate the asymptotic bias of this estimator.

The analysis of the results presented in the previous section shows that the bias of this estimator is different of zero for every errors distribution considered, except the Laplace distribution.

When the errors follow the Normal $(0, \sigma^2)$ distribution, the asymptotic bias is

$$\frac{\tau(2-\pi)}{\pi},$$

that is negative, and so, in average, $\hat{\tau}^*$ sub-estimates τ.

For the mixture of Normal distribution errors, the asymptotic bias is

$$\frac{p(1-p)\left(\sigma^2-1\right)^2 + (1-\pi)\sigma}{\sqrt{2\pi}\left(p\sigma + (1-p)\right)},$$

that is negative if $p(1-p)\left(\sigma^2-1\right)^2 < (\pi-1)\sigma$, and is positive otherwise.

In the Logistic distribution, the asymptotic bias is given by

$$-0.614\gamma = -0.307\tau,$$

and so, it is always negative.

8. Simulation Study

In this section, we perform a simulation study with the objective of generate empirical distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ considering small sample sizes, under the following distributions of the errors ε_i:

- Normal (0, 1) ($\tau = 1.253$);
- Logistic with mean zero and variance $\pi^2/3$ ($\tau = 2.00$);
- Laplace with mean zero and variance 2 ($\tau = 1.00$);
- Mixture of Normals (NM 85-15) when random variables are selected from a Normal (0, 1) with probability 0.85 and a N (0, 49) with probability 0.15 ($\tau =$ 1.439);
- Mixture of Normals (NM80-20) when random variables are selected from a Normal (0, 1) with probability 0.80 and a N (0, 49) with probability 0.20 ($\tau =$

1.513) and

- Cauchy with median zero and scale parameter 1 ($\tau = 1.571$).

Our objective is to find situations determined by errors distributions and sample sizes, under which $\hat{\tau}^*$ has empirically a better behavior than $\hat{\tau}$ in terms of bias and mean squared error.

The simulation study was designed as follows.

- We considered regression models with one independent variable generated from a Normal (0, 1) distribution, independently of the errors. Without loss of generality, the true parameters β_0 and β_1 were fixed equal to 1;
- The sample sizes (n) were set as 10, 20, 30, 50, 100 and 200;
- For each combination of sample size and errors distribution, 1000 sets of data were generated;
- Using the L_1 method, a regression model was fitted for each set of data and the values of $\hat{\tau}^*$ and $\hat{\tau}$ were calculated. So, this procedure generated 1000 values of $\hat{\tau}^*$ and $\hat{\tau}$.

The computations were performed using a special routine constructed in S-Plus 4.5.

The results obtained in this study are summarized in **Tables B1-B6** in Appendix B. They suggest that

- $\hat{\tau}^*$ is a good alternative to $\hat{\tau}$ when the errors follow Laplace, NM 85-15 or NM 80-20 distributions. In these cases, $\hat{\tau}^*$ has bias and mean squared error smaller or of the same order than $\hat{\tau}$;
- When the errors follow Normal or Logistic distribution, $\hat{\tau}^*$ tends to sub-estimate τ. The means of $\hat{\tau}$ distributions generated in the study are closer to the parameter value and its mean squared errors are in general uniformly smaller than that of $\hat{\tau}^*$, for all considered sample sizes.
- For the Cauchy distribution errors, $\hat{\tau}^*$ tends to super-estimate τ. This result may be a consequence of the fact that all the residuals are considered in the computation of this estimator. Although $\hat{\tau}$ has smaller bias and mean squared error than $\hat{\tau}^*$, $\hat{\tau}$ does not seem to be a good estimator of τ for sample sizes smaller or equal to 30.

9. Some Characteristics of the Distributions in the Study

The distributions considered in the previous sections are symmetrical about zero and can be ordered by the weight of their tails [12]. For $0 \le \alpha \le 1$, an appropriate coefficient that can be used with this objective is

$$b_2(\alpha) = \left[F^{*-1}(1-\alpha) + F^{*-1}(\alpha) - 2\nu^*\right] \Big/ \left[F^{*-1}(1-\alpha) + F^{*-1}(\alpha)\right],$$

where

$$F^*(x) = 2F(x) - 1,$$

$F(x)$ is the distribution function of the errors and

ν^* is the median of the density function associated to F^*.

This coefficient has the following properties

Table 2. Values of $b_2(\alpha)$ for the distributions in the study.

Distribution	$b_2(0.10)$	$b_2(0.05)$
Normal	0.2775	0.3551
Logistic	0.3455	0.4396
Laplace	0.4650	0.5641
NM 85-15	0.5618	0.7848
NM 80-20	0.6972	0.8076
Cauchy	0.7265	0.8541

- $-1 \le b_2(\alpha) \le 1$;
- Its computation does not require that the errors distribution have any finite moment;
- Its value is independent of the parameters of location and scale.

Large values of $b_2(\alpha)$ indicate that the distribution has heavy tails.

In Table 2 we present the values of $b_2(\alpha)$ for the distributions considered in this study, taking $\alpha = 0.10$ and $\alpha = 0.05$. For these values of α, it is clear that the ordering of the distribution according to its tails weights is: Normal, Logistic, Laplace, NM 85-15, NM 80-20 and Cauchy.

10. Concluding Remarks

In this paper, we studied the behavior of the estimator $\hat{\tau}^*$ with the objective of using it as an alternative to $\hat{\tau}$. We also determined analytically its asymptotic distribution under different distributions of the errors of the model. It was observed that, in general, $\hat{\tau}^*$ is asymptotically biased, with asymptotic bias equal to zero when the errors follow the Laplace distribution. In this case, the absence of asymptotic bias was already expected, since $\hat{\tau}^*$ is the maximum likelihood estimator of τ when the errors follow the Laplace distribution.

Performing a simulation study, the two estimators were compared empirically by their bias and mean squared error, under distributions with different tails weights and considering sample sizes varying from 10 to 200. The results suggest that $\hat{\tau}^*$ is a good alternative to $\hat{\tau}$ when the errors in the model follow Laplace or Mixture of Normal distributions with the values of the parameters fixed in the study; when the errors have Normal or Logistic distributions (lighter tails) or Cauchy distribution (heavy tails), $\hat{\tau}$ presented the best performance for every considered sample sizes. However, in the Cauchy distribution case, although $\hat{\tau}$ seemed to be better than $\hat{\tau}^*$, its use is not recommended in samples of size smaller or equal to 30 because of the bias of this estimator.

The results of the study indicate that $\hat{\tau}^*$ should be used when the distribution of the errors is close to the Laplace distribution, whatever the sample size. By the properties of this estimator mentioned in Section 1, we suggest that the fit of the data to the Laplace distribution be analyzed by the construction of a Q-Q plot of the residuals of the model. If there are not serious deviations, $\hat{\tau}^*$ should be used. Otherwise, Box-Cox transformations can be applied following [14]. After this, in the analysis of the transformed data, $\hat{\tau}^*$ can be used to construct confidence intervals and hypotheses tests about the parameters of the model and in the computation of robust coefficients of determination with and without a correction by the number of independent variables in the model.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] McKean, J.W. and Schrader, R.M. (1987) Least Absolute Errors Analysis of Variance. In: Dodge, Y., Eds., *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, Elsevier Science Publishers B. V., Amsterdam, 297-305.

[2] Basset, G. and Koenker, R. (1978) Asymptotic Theory of Least Absolute Error Regression. *Journal of the American Statistical Association*, **73**, 618-622. https://doi.org/10.1080/01621459.1978.10480065

[3] Dielman, T. and Pfaffenberger, R. (1982) LAV (Least Absolute Value) Estimation in the Regression Model: A Review. *TIMS Studies in the Management Sciences*, **19**, 31-52.

[4] Narula, S.C. (1987) The Minimum Sum of Absolute Errors Regression. *Journal of Quality Technology*, **19**, 37-45. https://doi.org/10.1080/00224065.1987.11979031

[5] McKean, J.W. and Sievers, G.L. (1987) Coefficient of Determination for Least Absolute Deviation Analysis. *Statistics and Probability Letters*, **5**, 49-54. https://doi.org/10.1016/0167-7152(87)90026-5

[6] Kvalseth, T.O. (1985) Cautionary Note about R^2. *American Statistician*, **39**, 279-285. https://doi.org/10.1080/00031305.1985.10479448

[7] Narula, S.C. and Wellington, J.F. (1977) Prediction, Linear Regression and Minimum Sum of Absolute Errors. *Technometrics*, **19**, 185-190. https://doi.org/10.1080/00401706.1977.10489526

[8] Elian, S.N., André, C.D.S. and Narula, S.C. (2000) Influence Measure for the L_1 Regression. *Communication in Statistics—Theory and Methods*, **29**, 837-849. https://doi.org/10.1080/03610920008832518

[9] Engelhardt, M. and Bain, L.J. (1973) Interval Estimation for the Two-Parameter Double Exponential Ditribution. *Technometrics*, **15**, 875-887. https://doi.org/10.1080/00401706.1973.10489120

[10] André, C.D.S., Elian, S.N., Narula, S.C. and Tavares, R.A. (2000) Coefficients of Determination for Variable Selection in the MSAE Regression. *Communication in Statistics—Theory and Methods*, **29**, 623-642. https://doi.org/10.1080/03610920008832506

[11] Portnoy, S. and Koenker, R. (1997) The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. *Statistical Science*, **12**, 279-300. https://doi.org/10.1214/ss/1030037960

[12] Groeneveld, R.A. and Meeden, G. (1984) Measuring Skewness and Kurtosis. *The Statistician*, **33**, 391-399. https://doi.org/10.2307/2987742

[13] Sen, P.K. and Singer, J.M. (1993) Large Sample Methods Is Statistics—An Introduction with Applications. Chapman and Hall, New York.

[14] Parker, I. (1988) Transformations and Influential Observations in Minimum Sum of Absolute Errors Regression. *Technometrics*, **30**, 215-220. https://doi.org/10.1080/00401706.1988.10488369

Appendix A. Results Used in Section 2

In Section 2, when we obtained the asymptotic distributions of $\hat{\tau}^*$, the distributions of the errors were symmetrical about zero. It is easy to see that if X is a random variable with values in the interval $]-\infty, \infty[$, symmetric about zero and with density $f(x)$, then the density of $Y = |X|$ is

$$g(y) = \begin{cases} 2f(y), & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Using this fact, we got $E(|\varepsilon_i|)$ for ε_i with Normal, Mixture of Normals, Logistic or Laplace distribution.

If the errors follow a **Normal** distribution, that is, $\varepsilon_i \sim N(0, \sigma^2)$, then $W_i = |\varepsilon_i/\sigma|$ has the density

$$g(w_i) = \begin{cases} \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} w_i^2\right), & w_i \geq 0 \\ 0, & w_i < 0 \end{cases}$$

and thus

$$E(W_i) = \int_0^\infty \frac{2w_i}{2\pi} \exp\left(-\frac{1}{2} w_i^2\right) \mathrm{d}w_i = \frac{2}{\sqrt{2\pi}} \left[-\exp\left(-\frac{1}{2} w_i^2\right)\right]_0^\infty = \frac{2}{\sqrt{2\pi}}.$$

Also, $E(W_i^2) = E\left(\frac{\varepsilon_i^2}{\sigma^2}\right) = \frac{1}{\sigma^2} E(\varepsilon_i^2) = Var(\varepsilon_i)/\sigma^2 = 1$, and so $Var(W_i) = (\pi - 2)/\pi$.

When the errors follow a **mixture of Normal** distribution, the probability density function of $U_i = |\varepsilon_i|$ is given by

$$f(u_i) = \begin{cases} \frac{2p}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u_i^2\right) + \frac{2(1-p)}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} u_i^2\right), & u_i \geq 0 \\ 0, & u_i < 0 \end{cases}$$

Therefore

$$\begin{aligned} E(U_i) &= \int_0^\infty \frac{2p}{\sqrt{2\pi}} u_i \exp\left(-\frac{1}{2} u_i^2\right) \mathrm{d}u_i + \int_0^\infty \frac{2p(1-p)}{\sqrt{2\pi}\sigma} u_i \exp\left(-\frac{1}{2\sigma^2} u_i^2\right) \mathrm{d}u_i \\ &= \frac{2p}{\sqrt{2\pi}} \left[-\exp\left(-\frac{1}{2} u_i^2\right)\right]_0^\infty + \frac{2(1-p)}{\sqrt{2\pi}\sigma} \left[-\sigma^2 \exp\left(-\frac{1}{2\sigma^2} u_i^2\right)\right]_0^\infty \\ &= \frac{2p}{\sqrt{2\pi}} + \frac{2(1-p)\sigma}{\sqrt{2\pi}} \end{aligned}$$

and

$$E(U_i^2) = E(\varepsilon_i^2) = Var(\varepsilon_i) = p + (1-p)\sigma^2$$

that is the variance of a random variable with mixture of Normal distributions with parameters p, $(1 - p)$, means equal to zero and variances 1 and σ^2 respectively.

Consequently,

$$Var(U_i) = \frac{\left[\pi(1-p) - 2(1-p)^2\right]\sigma^2 - 4p(1-p)\sigma + p\pi - 2p^2}{\pi}.$$

When ε_i has **Logistic** distribution with mean zero and variance $\gamma^2\pi^2/3$.

$$E(|\varepsilon_i|) = \int_0^\infty \frac{2x\exp\left(\frac{-x}{\gamma}\right)}{\lambda\left[1+\exp\left(\frac{-x}{\gamma}\right)\right]^2}dx = 2\gamma\int_0^\infty \frac{z\exp(-z)}{\left[1+\exp(-z)\right]^2}dz = 1.386\gamma,$$

because $\int_0^\infty \frac{z\exp(-z)}{\left[1+\exp(-z)\right]^2}dz = 0.693$.

Also $E(Z_i^2) = E(\varepsilon_i^2) = \gamma^2\pi^2/3$, and so

$Var(Z_i) = Var(|\varepsilon_i|) = 0.3289\gamma^2 - 1.386^2\gamma^2 = 1.37\gamma^2$.

For ε_i with **Laplace** distribution with zero mean and variance $2\sigma^2$, $|\varepsilon_i|$ has exponential distribution with mean equal to σ and variance equal to σ^2. Thus, $E(|\varepsilon_i|) = \sigma = \tau$ and $Var(|\varepsilon_i|) = \sigma^2 = \tau^2$.

Appendix B. Tables

Table B1. Values of descriptive statistics observed in the distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ generated in the simulation study for models with Normal (0, 1) errors ($\tau = 1.253$) and different sample sizes.

Sample size (n)	mean		median		mean squared error		minimum		maximum	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	1.233	0.675	1.193	0.661	0.198	0.366	0.276	0.001	2.893	1.073
20	1.316	0.735	1.289	0.730	0.149	0.286	0.175	0.351	2.681	1.210
30	1.252	0.758	1.216	0.753	0.109	0.257	0.368	0.477	2.404	1.117
50	1.138	0.776	1.123	0.773	0.097	0.051	0.447	0.501	2.250	1.030
100	1.164	0.785	1.158	0.783	0.069	0.223	0.415	0.611	2.014	0.978
150	1.244	0.790	1.229	0.790	0.058	0.217	0.626	0.646	2.071	0.951
200	1.246	0.788	1.221	0.786	0.052	0.218	0.680	0.635	2.425	0.933

Table B2. Values of descriptive statistics observed in the distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ generated in the simulation study for models with Logistic errors ($\tau = 2.00$) and different sample sizes.

Sample size (n)	mean		median		mean squared error		minimum		maximum	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	2.120	1.189	2.019	1.165	0.722	0.784	0.253	0.386	5.459	2.675
20	2.206	1.281	2.146	1.263	0.500	0.580	0.624	0.597	4.950	2.146
30	2.039	1.311	1.990	1.297	0.290	0.519	0.256	0.746	4.144	2.195
50	1.814	1.352	1.797	1.340	0.270	0.446	0.619	0.900	3.462	1.910
100	1.866	1.363	1.831	1.357	0.183	0.420	0.779	0.986	3.383	1.742
150	2.004	1.372	1.987	1.379	0.141	0.403	1.030	1.078	3.190	1.654
200	2.035	1.380	2.017	1.376	0.128	0.391	1.070	1.155	3.264	1.668

Table B3. Values of descriptive statistics observed in the distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ generated in the simulation study for models with Laplace errors ($\tau = 1.00$) and different sample sizes.

Sample size (n)	mean		median		mean squared error		minimum		Maximum	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	1.461	0.872	1.368	0.843	0.705	0.112	0.162	0.238	4.413	2.207
20	1.442	0.936	1.401	0.917	0.443	0.052	0.373	0.482	3.479	1.942
30	1.314	0.975	1.263	0.960	0.261	0.034	0.462	0.481	3.105	1.621
50	1.072	0.979	1.042	0.974	0.100	0.021	0.382	0.616	2.351	1.453
100	1.076	0.996	1.058	0.992	0.068	0.009	0.403	0.711	2.025	1.393
150	1.118	0.991	1.106	0.989	0.065	0.007	0.555	0.767	1.993	1.263
200	1.099	0.993	1.093	0.992	0.053	0.005	0.557	0.788	1.942	1.223

Table B4. Values of descriptive statistics observed in the distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ generated in the simulation study for models with 0.85 Normal (0, 1) + 0.15 Normal (0, 49) errors ($\tau = 1.439$) and different sample sizes.

Sample size (n)	mean		median		mean squared error		minimum		Maximum	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	1.871	1.340	1.550	1.152	1.884	0.568	0.232	0.293	9.958	6.001
20	1.640	1.468	1.560	1.387	0.393	0.293	0.425	0.351	8.533	3.838
30	1.518	1.498	1.501	1.443	0.192	0.206	0.530	0.475	3.465	3.394
50	1.325	1.484	1.282	1.455	0.144	0.121	0.314	0.636	2.696	2.935
100	1.338	1.490	1.313	1.479	0.104	0.058	0.561	0.883	2.501	2.334
150	1.450	1.500	1.435	1.493	0.072	0.043	0.721	0.942	2.558	2.120
200	1.451	1.502	1.441	1.500	0.067	0.031	0.706	0.975	2.432	2.053

Table B5. Values of descriptive statistics observed in the distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ generated in the simulation study for models with 0.80 Normal (0, 1) + 0.20 Normal (0, 49) errors ($\tau = 1.513$) and different sample sizes.

Sample size (n)	mean		median		mean squared error		minimum		Maximum	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	2.182	1.624	1.716	1.414	2.933	0.638	0.259	0.249	12.500	5.204
20	1.808	1.690	1.677	1.626	0.684	0.427	0.393	0.541	9.292	4.419
30	1.610	1.498	1.558	1.498	0.221	0.270	0.606	0.475	4.200	3.394
50	1.410	1.727	1.399	1.697	0.153	0.193	0.480	0.803	2.883	3.244
100	1.405	1.739	1.367	1.719	0.111	0.126	0.671	1.069	2.738	2.830
150	1.519	1.737	1.499	1.729	0.083	0.098	0.759	1.146	2.488	2.617
200	1.523	1.745	1.511	1.740	0.075	0.092	0.784	1.212	2.383	2.480

Table B6. Values of descriptive statistics observed in the distributions of the estimators $\hat{\tau}^*$ and $\hat{\tau}$ generated in the simulation study for models with Cauchy errors ($\tau = 1.571$) and different sample sizes.

Sample size (n)	mean		median		mean squared error		minimum		maximum	
	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$	$\hat{\tau}$	$\hat{\tau}^*$
10	3.480	7.390	2.458	2.070	20.150	2594.0	0.294	0.340	48.505	1203.7
20	2.375	22.800	2.147	2.800	1.794	182089.0	0.483	0.400	8.343	13019.1
30	1.931	14.740	1.863	3.010	0.587	51611.0	0.533	0.740	5.191	7064.1
50	1.578	6.902	1.543	3.608	0.168	348.0	0.533	0.941	3.920	285.1
100	1.530	41.700	1.509	3.700	0.117	1281096.0	0.684	1.500	2.790	35792.4
150	1.629	7.866	1.620	3.955	0.097	581.0	0.816	1.689	2.782	488.8
200	1.615	7.747	1.603	4.205	0.086	375.0	0.819	1.819	2.787	335.2