Pertanika 12(1), 89-98 (1989)

On Robust Alternatives to the Maximum Likelihood Estimators of a Linear Functional Relationship

MOKHTAR B. ABDULLAH
Department of Statistics
Centre for Quantitative Studies,
Universiti Kebangsaan Malaysia.
43600 UKM Bangi, Malaysia

Key words: Outlier; linear functional relationship; Theil's estimators; L1-norm estimator.

ABSTRAK

Kertas ini membincangkan beberapa calon penganggar kukuh bagi menganggar hubungan fungsian linear ringkas (SLFR). Penganggar kebolehjadian maksimum klasik bagi SLFR boleh terjejas dengan kehadiran data terpencil. Ini ialah kerana ianya merupakan penganggar yang berasaskan min. Beberapa penganggar yang berasaskan median bagi SLFR diselidiki. Dari calon-calon yang dipertimbangkan itu penganggar jenis Theil dan penganggar norma-L1 terubahsuai didapati paling kukuh (iaitu, tak peka kepada data terpencil).

ABSTRACT

This paper discusses a number of candidates for robust estimators of a simple linear functional relationship (SLFR). The classical maximum likelihood estimators of the SLFR can be affected by the presence of possible outliers. This is due to the fact that they are mean-based estimators. Some median-based estimators of the SLFR are examined. Among those considered Theil-type estimators and the modified L1-norm estimator are found to be most robust (i.e insensitive to the outliers).

1. INTRODUCTION: MODEL, ASSUMPTIONS AND MAXIMUM LIKELIHOOD (ML) ESTIMATORS

The model considered in this paper is a simple linear functional relationship (SLFR). It specifies that two mathematical variable ξ and η are linearly related but observed with mutually independent, normally distributed errors δ and ε, respectively. That is we observe

$$\begin{aligned} x_i &= \xi_i + \delta_i, \quad \delta_i \sim N(0,\sigma^2) \\ y_i &= \eta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \tau^2) \end{aligned} \tag{1.1}$$

where

$$\eta_i = \alpha + \beta\xi_i, i = 1, \ldots, n$$

There are $(n + 4)$ parameters to be estimated, i.e, α, β, σ^2, τ^2, and $\xi_1, \ldots, \xi_n$ (or equivalently $\eta_1, \ldots, \eta_n$). The presence of the *incidental parameters* ξ_i, i = 1, ..., n leads to inconsistencies of the classical ML estimates unless an additional assumption is made about the variance ratio $\lambda = \tau^2/\sigma^2$. That is, if λ is known there exists a consistent ML estimate of the SLFR parameters. The main results concerning functional relationship were discussed by Lindley (1947), Madansky (1959), Sprent (1969), Solari (1969), Moran (1971), Kendall and Stuart (1973), and Anderson (1980).

With the assumptions on the error terms in (1.1), the log-likelihood function may be written as

$$L = \text{constant } n \log\sigma^2 - 1/2\sigma^2$$

$$\left\{\sum_{i=1}^{n}(X_i - \xi_i)^2 + \lambda^{-1}\sum_{i=1}^{n}(y_i - \alpha - \beta\xi_i)^2\right\}. \tag{1.2}$$

Let

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \qquad (1.2)$$

where

$$\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$$

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$$

Then the maximum likelihood estimates of β, α, and ξ_i are

$$\hat{\beta} = \frac{(S_{yy} - \lambda S_{xx}) + \left((S_{yy} - \lambda S_{xx})^2 + 4\lambda S_{xy}^2\right)^{\frac{1}{2}}}{2S_{xy}}$$

(1.4) (i)

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \qquad (1.4)\ (ii)$$

and $$\hat{\xi}_i = \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{(\lambda + \hat{\beta}^2)}$$

and the maximized log – likelihood function is

$$L_{\max} = \text{constant} - n \log \frac{\sum_i r_i^2}{2\,nh}$$

where $r_i = y_i - \hat{\alpha} - \hat{\beta}x_i$ (1.4) (iii)

$$h = \lambda + \hat{\beta}^2$$

It should be noted from (1.3) that for $\lambda = +\infty$ or equivalently $\sigma^2 = 0$ or $\xi_i = x_i$ model (1.1) reduces to a regression of y on x while for $\lambda = 0$ or $\tau^2 = 0$ model (1.1) becomes a regression of x on y.

2. OUTLIERS IN THE SLFR

In the regression (of y on x) model, the presence of a possible outlier is always associated with a contamination in the dependent variable y. A contamination that occurs in y_r may be due to the error term y_r being drawn from some heavy-tailed distribution, e.g, $\varepsilon_r \sim N(0, h_\varepsilon^2)$, $h_\varepsilon^2 > \tau^2$. Since no distributional assumption is made for the regressor x (i.e, x is fixed) then the occurence of a possible outlier in the variable may only be caused by a recording or typing error.

In the SLFR an outlier may be generated from one of three types of contaminations, i.e,

(a) a contamination in the x's where the contaminated error term ξ_r is drawn from $N(0, h_\delta^2)$, $h_\delta^2 > \sigma^2$ and
$\delta_{i \neq r}$, $i = 1, \ldots, n \sim N(0, \sigma^2)$

or

(b) a contamination in the y's where
$\varepsilon_r \sim N(0, h_\varepsilon^2)$, $h_\varepsilon^2 > \tau^2$ and
$\varepsilon_i \neq i = 1, \ldots, n \sim N(0, \tau^2)$

or

(c) simultaneous contaminations in both the x's and y's.

The presence of the outlying observation can cause the fitted line to be dictated by the wild observation resulting in failure to pass through the bulk of the data (x_i, y_i) $i = 1, \ldots, n$. The nonrobustness (i.e, sensitiveness to the bad observation) is due to the fact that $\hat{\alpha}$ and $\hat{\beta}$ given in (1.4) (i)-(ii) are based on the means of the observed values. Hampel (1974) pointed out that the mean (as an estimate of a location parameter) is nonrobust because its influence function is unbounded. Using a similar (i.e influence function) approach Kelly (1984) demonstrates that the influence functions of $\hat{\alpha}$ and $\hat{\beta}$ are unbounded and, therefore, these ML estimates are nonrobust. In contrast, it is well known that in the location problem the median is more robust than the mean because its influence function is bounded. Therefore, median-based estimators of the SLFR may provide good alternatives to the ML estimators.

In the next sections, we present several candidates for the estimators of the SLFR which are based on the median of the observations. A simulation study is performed to examine the performance of these candidates and their improvement over the mean-based estimators.

3. ESTIMATORS BY GROUPING METHODS

In the history of SLFR, grouping methods have been developed as an alternative to the classical maximum likelihood estimation procedure. The main advantage of the grouping methods over the classical one is the simplicity of the calculations and the dropping of the assumption of normality.

One of the well known grouping methods suggested in the literature was that of Bartlett (1949). The method consists of arranging the y-values according to the ordered x-values, forming three groups, omitting observations in the middle group, and joining the two centre of gravity (i.e, the means) of the remaining groups by a straight line to get the desired slope estimator. In an obvious notation, the 3-grouped-mean estimators of α and β proposed by Bartlett (1949) are defined as

$$\hat{\beta}_B = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1} \quad (3.1)\,(i)$$

and

$$\hat{\alpha}_B = \bar{y} - \hat{\beta}_B \bar{x} \quad (3.2)\,(ii)$$

$$\bar{y}_1 = n_1^{*-1} \sum_{i=1}^{n_1} y_i$$

$$\bar{x}_1 = n_1^{*-1} \sum_{i=1}^{n_1} x_i$$

$$\bar{y}_3 = n_3^{*-1} \sum_{i=1}^{n_2} y_i$$

$$\bar{x}_3 = n_3^{*-1} \sum_{i=1}^{n_2} x_i$$

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$$

$$\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$$

and the y-values are first ordered according to the magnitude of the x's, and

$$n_1^* = k \quad (3.2)\,(iii)$$

$$n_3^* = n - k + 1$$

where k is the number of observations in the first and third groups.

It is not necessary to take an equal number of observations in each group especially when n is indivisible by three. However, for the special case of equally spaced ξ_i-values, Bartlett shows that the optimal choice is $k = n/3$ as this will minimize the sampling variance of $\hat{\beta}_B$.

Bartlett also shows that the estimators are consistent if the grouping is unaffected by the errors δ_i, i.e, the ordering according to the x_i is identical to the ordering according to the x_i. In other words, for

$$c = \min\{|\xi_{i+1} - \xi_i|\}, = i = 1, \ldots, n$$

it is assumed that

$$\text{Prob}\{|\delta_i| \geq c/2\} = 0 \quad (3.3)$$

so that the estimators in (3.1) (i)-(ii) are consistent.

In the formulation of the Bartlett 3-group-mean estimators in (3.1) it was assumed that there is no outlier in the x- or in the y- values.

However, some difficulties may arise in using the grouping method if there is a possible outlier in the data set. The possible difficulties are

(a) the presence of the outlying observation may cause the ordering of the x-values no longer identical to the ordering of the ξ_i-values, i.e, the assumption in (3.3) may no longer hold.

(b) the fact that the mean was used as the centre of gravity of each group may result in the nonrobustness of the Bartlett estimators.

In practice, however, it is difficult to determine whether problem (a) does or does not arise since the true values ξ_i are unlikely to be known. Even if ξ_i are known, it is difficult to guarantee that this problem does not occur when an outlier is present. For problem (b), the most appropriate choice is to adopt the median as the centre of gravity of each subgroup since it is more robust than the mean. However, the robustness of the grouped-median estimators may still not be guaranteed if the presence of the outlier changes the ordering of the x-values. In this situation, only a proper allocation of the observations in the first and third groups can avoid the estimates from being influenced by the outlier. Unfortunately, the most appropriate choice of the allocation, k, in this situation is not known. The choice of $k = n/3$ was suggested for the no-outlier situation and may not always be appropriate for the case where an extreme outlier may be present.

The median-versions of (3.1)(i) and (3.1(ii) may now be written as

$$\tilde{\beta}_B = \frac{\tilde{y}_3 - \tilde{y}}{\tilde{x}_3 - \tilde{x}_1} \quad (3.4)(i)$$

and

$$\tilde{\alpha}_B = \text{med}\ \{y_i - \tilde{\beta}x_i\}, i = 1, \ldots, n \quad (3.4)(ii)$$

where

$$\tilde{y}_3 = \text{med}\ \{y_i\}, \quad n_3^* \le i \le n$$
$$\tilde{y}_1 = \text{med}\ \{y_i\}, \quad 1 \le i \le n_1^*$$
$$\tilde{x}_3 = \text{med}\ \{x_i\}, \quad n_3^* \le i \le n$$
$$\tilde{x}_1 = \text{med}\ \{x_i\}, \quad 1 \le i \le n_1^*$$

n_1^* and n_3^* are given by (3.2)(iii) and again the y-values are arranged in accordance with the ordered values of the x's.

4. THEIL'S AND REPEATED-MEDIAN ESTIMATORS

Perhaps one of the simplest methods and yet may produce estimates which are more robust than the classical estimates is the one proposed by Theil (1950).

This method begins with all possible pairs of data points, calculates the slope obtained from each point, and finally determines the median of these slopes. Assuming that the $x_i^1 s$ are all distinct, define

$$\tilde{\beta}_{ij} = (y_j - y_i)/(x_j - x_i), 1 \le i \le n \quad (4.1)$$

This yields $\binom{n}{2}$ slope values. The estimator of the slope is then

$$\tilde{\beta}_T = \text{median}\ \{\tilde{\beta}_{ij}\} \quad (4.2)(i)$$

The estimator $\hat{\alpha}$ is then given by

$$\tilde{\alpha} = \text{median}\ \{y_i - \tilde{\beta}_T x_i\} \quad (4.2)(ii)$$

Alternatively, in $\tilde{\alpha}_T$ (4.2)(ii) can be obtained from a procedure proposed by Maritz (1979) which is similar to the one that leads to (4.1). Maritz derives the estimate of the intercept term by calculating the intercepts for $\binom{n}{2}$ pairs, i.e,

$$a_{ij} = (x_i y_j - x_j y_i)/(x_i - x_j), 1 \le i \le j \le n \quad (4.3)$$

Then the intercept estimator is given by

$$\tilde{\alpha}_T = \text{median}\ \{a_{ij}\} \quad (4.4)$$

Another Theil-type estimator was devised by Siegel (1982) which is based on a repeated-median method. This approach starts with the pairwise slopes as in Theil's method, takes the medians in two stages; first, at each point and then across points. That is we find

(i) $\tilde{\beta}_i = \text{median}\ \{\tilde{\beta}_{ij}\}$

$\tilde{\beta}_{ij}$ is given by (4.1) yielding (n–1) slopes.

(ii) $\tilde{\beta}_S = \text{median}\ \{\tilde{\beta}_i\}$ (4.5)(i)

which is the median of a set consisting n slopes.

Finally,

$$\tilde{\alpha}_S = \text{median}\ \{y_i - \tilde{\beta}_S x_i\} \quad (4.5)(ii)$$

or

$$\tilde{\alpha}_S = \text{median}\ \{a_{ij}\} \quad (4.5)(iii)$$

where a_{ij} is computed from (4.3).

5. L1-NORM ESTIMATOR

Harvey (1977) pointed out that for a simple linear regression model, the L1-norm regression (of y on x) which consists of minimizing

$$\sum_{i=1}^{n} |y_i - \alpha - \beta x_i| \quad (5.1)$$

is perhaps the most natural generalisation of the sample median in the location problem.

We now extend this well known method to the problem of estimating the SLFR parameters.

For the SLFR model in (1.1) the L1-norm estimators of α and β are given by $\hat{\alpha}$ and $\hat{\beta}$ which minimize

$$R = \sum_{i=1}^{n} \{|x_i - \xi_i| + |y_i - \alpha - \beta\xi_i|\} \quad (5.2)$$

Brown (1982) suggests that the solution to (5.2) i.e the L1-estimates of a and b can be obtained as follows: perform L1-norm regressions of y on x and x on y, from which their respective estimates of a and b are obtained. By computing the residuals from each fitted line, the required estimates $\hat{\alpha}$ and $\hat{\beta}$ are determined, depending on which line i.e y on x or x on y, has the smaller sum of absolute deviation. That is, by performing the L1-norm regression of y on x we obtain $\hat{\alpha}_y$ and $\hat{\beta}_y$ which minimize R_y where

$$R_y = \sum_{i=1}^{n} \left| y_i - \hat{\alpha}_y - \hat{\beta}_y x_i \right| \qquad (5.3)\,(i)$$

and similarly, performing the L1-norm regression of x on y yields $\hat{\alpha}_x$ and $\hat{\beta}_x$ which minimize R_x, where

$$R_x = \sum_{i=1}^{n} \left| x_i - (y_i - \hat{\alpha}_x)/\hat{\beta}_x \right| \qquad (5.3)\,(ii)$$

The required estimates of $[\alpha, \beta]$ are then given by either $[\hat{\alpha}_y, \hat{\beta}_y]$ or $[\hat{\alpha}_x, \hat{\beta}_x]$ according to the smaller of

$$[R_x, R_y] \qquad (5.4)$$

However, Brown (1982) rejects the estimates obtained from the criterion (5.4) for the starting values because they may over- or under-estimate α and β, respectively. Over- or under-estimation of the parameters occurs when the estimates chosen by the criterion in (5.4) are the ones that are influenced by the outlier.

5.1. Modified L1-norm Estimator

In this section we propose a modification of the criterion in (5.4). The rationale behind this modification is to establish a new L1-norm criterion that will avoid the possibility of under- or over-estimation of the parameters.

Under- or over-estimation of the parameters was due to the inclusion of the outlier's residual in R_x and R_y, respectively. This has resulted in choosing incorrect estimates since the outlier's residual associated with the poor fitted line will always be smaller than the one associated with the best (or robust) fitted line.

To avoid such an incorrect choice for the estimates, we devise a procedure in which the largest residual is to be excluded from the overall sum of absolute residuals in each direction. By omitting the largest residuals which are usually associated with the possible outlier the comparison between R_x and R_y is now based on the corrected sum of absolute residuals associated with uncontaminated data points. Let the corrected sums of absolute residuals corresponding to R_y and R_x be R^*_y and R^*_x, respectively, where

$$R^*_y = \sum_{i=1}^{n} \left| y_i - \hat{\alpha}_y - \hat{\beta}_y x_i \right| - \max\left\{ \left| y_i - \hat{\alpha}_x - \hat{\beta}_x x_i \right| \right\} \qquad (5.5)\,(i)$$

$$R^*_x = \sum_{i=1}^{n} \left| x_i - (y_i - \hat{\alpha}_y)/\hat{\beta}_y \right| - \max\left\{ \left| x_i - (y_i - \hat{\alpha}_x)/\hat{\beta}_x \right| \right\} \qquad (5.5)\,(ii)$$

Then the required estimates of $[\alpha, \beta]$ are then given by either $[\hat{\alpha}_y, \hat{\beta}_y]$ or $[\hat{\alpha}_x, \hat{\beta}_x]$ according to the smaller of

$$[R^*_x, R^*_y] \qquad (5.6)$$

Note that the uncorrected sums of absolute residuals associated with a poor fitted line will always have a larger value than the sum of absolute residuals corresponding to the best fitted line. This is because the poor fitted line does not pass through the bulk of the data while the best one should pass through most of the data points. Therefore, the correct choice for the estimate comes from the fact that the smaller value of the sum of absolute residuals should always be associated with the best fitted line.

However, a proper choice for the estimates based on the two sums of absolute residuals may not matter very much if no contaminated data point or only a mild outlying observation is present in the data set. This is because choosing either one of the two sums of absolute residuals will still lead to reasonably good estimates.

In situations where several outliers are likely to be present (5.5) (i)-(ii) can be modified so that more than one largest absolute residual can be omitted. That is, if there are r possible outliers in the data set then (5.5) (i) can be written as

$$R^*_y = \sum_{i=1}^{n} \left| U_y \right|_i - \sum_{m=n-r+1}^{n} U_y^{(m)} \qquad (5.5)\,(iii)$$

where

$$U_y = y_i - \hat{\alpha}_y - \hat{\beta}_y x, \quad i = 1, \ldots, n$$

and

$U_y^{(m)}$ = the m-th largest absolute y-residuals, $\left| U_y \right|_i$

and r is the number of the largest absolute residuals to be eliminated.

Similarly, (5.5) (ii) can now be expressed in the form

$$R^*_x = \left| U_x \right|_i - \sum_{m=n-r+1}^{n} U_x^{(m)} \qquad (5.5)\,(iv)$$

where

$$U_x = x_i - (y_i - \hat{\alpha}_x)/\hat{\beta}_x$$

and

$U_x^{(m)}$ = the m-th largest absolute xi-residuals, $|U_{x_i}|$

We shall now compare the performance of the modified L1-norm estimator in (3.6) with the original L1-norm estimators in (3.4).

Example: (Contamination in the y's)

This example presents the analysis of a data set of 11 observations generated from the SLFR model in (1.1). The true parameter values are $\alpha = \beta = 1$ and $\delta_i, \varepsilon_i \sim N(0, 0.5^2)$, $i = 1, \ldots, 10$ and $\varepsilon_{11} \sim N(0, 10^2)$.

The data set are given in Table 1(a).

To compute the L1-norm estimates we use NAG-routine E02GAF (based on the algorithm of Barrodale and Roberts (1973)).

We now summarize the results for the estimates from the various methods in the following Table 1(b).

TABLE 1(b)

	Estimates	
Estimator	$\hat{\alpha}$	$\hat{\beta}$
SLFR	2.530	2.641
BMN	2.479	1.615
BDYX	0.909	1.114
OL1	3.050	2.586
ML1	0.819	1.013
THEIL	1.059	1.073
SIEGEL	0.935	1.107

The codes for the various estimators are as follows:

SLFR : The ML of SLFR given in (1.4)(i)-(ii)

BDYX : The Bartlett's group median estimators given in (3.4)(ii)

BMN : The Bartlett's 3-group mean estimators given in (3.1)(i)-(ii)

OL1 : The original L1-norm estimators defined by criterion (5.4)

ML1 : The modified L1-norm estimators defined by criterion (5.6)

THEIL : Theil's estimators given in (4.2)(i)-(ii)

SIEGEL: Siegel's estimators given in (4.5)(i)-(ii)

From the results in Table 1(b), the mean-based estimators, i.e, SLFR and BMN were badly affected by the extreme outlier in the data set. The median-based estimators such as ML1, BDYX, THEIL and SIEGEL performed reasonably well for both types of contamination.

For OL1, its performance against the extreme outlier suffers from the incorrect choices of the estimates $\hat{\alpha}$ and $\hat{\beta}$ which overestimate α and β, respectively.

In order to make better comparisons among the various starting estimators, let $\hat{\alpha}_l$ and $\hat{\beta}_l$, 1 = 1, ..., 100 represent the hundred values of the estimates of the SLFR parameters, and let the 'efficiency' of the various estimators to be measured by the 'empirical mean square error' criterion which is defined as

$$\text{MSE}\left(\hat{\theta}\right) = L^{-1}\sum_{l=1}^{L}\left(\hat{\theta}_l - \theta\right)^2$$

TABLE 1(a)

i	1	2	3	4	5	6	7
x_i	– 5.605	– 3.966	– 2.956	– 1.772	–1.215	– 0.442	1.199
y_i	-3.857	-3.197	-3.112	-1.378	-0.112	1.906	1.833
i	8	9	10	11			
x_i	2.984	3.200	3.675	4.288			
y_i	2.868	4.060	5.003	22.366			

where L=100, $\hat{\theta}$ is the estimate of the parameter θ.

The hundred samples were generated under the following sampling situations;

$\alpha = \beta = 1$, $n = 11, -5 \leq \xi_i \leq 5$, equally spaced and with increment 1.

The uncontaminated errors $\delta_i \sim N(0, \sigma^2)$ where $\sigma = 0.5$, and $\varepsilon_i \sim N(0, \tau^2)$, $\tau = 0.5$. For each contaminated sample it is assumed that the contamination occurs either in the x's or in the y's.

For a contamination in x_r, the values of the x_r outlying observation and the corresponding uncontaminated y_r observation are generated by

$$x_r = \xi_r + \delta_r \quad \text{where} \quad \delta_r \sim N(0, h^2)$$

and

$$y_r = 1 + \xi_r + \xi_r, \varepsilon_r \sim N(0, \tau^2), \quad \text{and}$$

$$h^2 > \tau^2.$$

Similarly, for a contamination in y_r, the values of the y_r outlying observation and the corresponding uncontaminated x_r are obtained from

$$x_r = \xi_r + \delta_r \quad \text{where} \quad \delta_r \sim N(0, \sigma^2)$$

$$y_r = 1 + \xi_r + \varepsilon_r, \varepsilon_r \sim N(0, h^2), \quad \text{and } h^2 > \sigma^2.$$

In each contaminated sample, it is assumed that only a single outlier is present and this outlier is located at an r-th position and r is either fixed or random. The selected value of a fixed r is $r = 11$, i.e, the outlier is to be located at the last data point. For the case where the location of the outlier is selected at random r can take any value between 1 and n. The selected values of h are

h = 2.0, 6.0 and 10.0

In this simulation study, the NAG-library subroutine G05DDF was used to generate the normal variates δ and ε, respectively. All computer programs were written in FORTRAN and executed on the DEC-10 computer system at the University of Dundee.

6. SIMULATION RESULTS AND DISCUSSION

Tables 2(a)-(b) demonstrate the performances of the various estimators for the case of contamination in x_r, while Tables 3(a)-(b) for the contamination in y_r.

The performance of an estimator is judged from its MSE value with respect to the true values of α and β, where $\alpha = \beta = 1$. Estimator A is said to perform better (or be more efficient or more robust) than estimator B if the MSE value associated with A is smaller than that of B.

The percentage efficiency between two estimators is defined as the reciprocal of the ratio $\dfrac{\text{(the lowest MSE)}}{\text{(the largest MSE)}} \times 100\%$ is also computed. For each estimator these efficiencies as well as their ranks (where the smallest rank corresponds to the smallest value of the efficiency, and (so forth) for $\hat{\alpha}$ and $\hat{\beta}'$ are displayed in the second and the third rows, respectively.

Inspection of the results in Tables 2-3 illustrates that in a situation where there is no possible outlier (i.e, $h = 0.5$), the classical estimators such as SLFR and BMN perform well.

In the presence of an extreme outlier (for $h \geq 6$) Theil-type estimators (i.e, THEIL and SIEGEL) and ML1 seems to be most robust among those considered. Its MSE values are relatively smaller (and the percentage efficiencies are higher) than those of the other estimators in all situations.

The rank averages of the percentage efficiencies for $\hat{\alpha}$ and $\hat{\beta}$ also agrees with the fact that the most robust estimators are SIEGEL and THEIL and they are closely followed by ML1.

In situations where less severe outliers are present the performance of OL1 seems to be comparable to that of ML1. However, OL1 tends to perform badly when more extreme outliers are likely to be present. This was due to the incorrect choices of estimates made by the OL1-criterion in (5.4) which either over- or under-estimates the true parameters. This can be seen from the poor performance of OL1 in Tables 3(a)-(b) compared to those of ML1 and Theil-type estimators.

The high values of MSE (correspondingly the small values of percentage efficiencies) for SLFR and BMN simply confirm the fact that these mean-based estimators are very sensitive to the presence of an extreme outlier.

TABLE 2(a): $\alpha = \beta = 1$, n - 11
x_r – contamination; r = 11, fixed.

h	No Outlier 0.5		2.0		Outlier - x_{11} 6.0		10.0		RANK AVE.		RANK AVE
Est.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\beta}$
SLFR MSE	.054	.004	.082	.013	.251	.060	.273	.133			
% eff	100	100	95	69	34	20	32	9			
Rank	6.5	6.5	4	1	1	1	2	1	3.4	2.4	3.9
BMN MSE	.054	.004	.084	.010	.230	.040	.332	.071			
% eff	100	100	93	90	37	30	26	17			
Rank	6.5	6.5	3	5	2	2	1	2	3.1	3.9	3.5
BDYX MSE	065	.008	.078	.012	.105	.041	.099	.041			
% eff	83	50	100	75	81	29	88	29			
Rank	5	1.5	7	3	4	3	6	4	5.5	2.9	4.2
OL1 MSE	.074	.008	.087	.013	.109	.023	.161	.067			
% eff	73	50	89	69	78	52	54	18			
Rank	2.5	1.5	1.5	2	3	4	3	3	2.5	2.6	2.6
ML1 MSE	.083	.007	.080	.011	.097	.015	.100	.012			
% eff	73	57	97	82	87	80	87	100			
Rank	2.5	3	6	4	5	5	5	6.5	4.6	4.6	4.6
THEIL MSE	.074	.005	.087	.009	.085	.013	.101	.015			
% eff	73	80	89	100	100	92	86	80			
Rank	2.5	5	1.5	7	7	6	4	5	4.0	5.8	4.9
SIEGEL MSE	.076	.006	.081	.010	.092	.012	.087	.012			
% eff	71	67	96	90	92	100	100	100			
Rank	1	4	5	6	6	7	7	6.5	4.8	5.9	5.3

TABLE 2(b): $\alpha = \beta = 1$, n = 11
x_r – contamination; r random.

h	No Outlier 0.5		2.0		Outlier - x_r 6.0		10.0		RANK AVE.		RANK AVE.
Est.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\beta}$
SLFR MSE	.054	.004	.103	.008	.239	.046	.264	.123			
% eff	100	100	81	100	35	24	32	8			
Rank	7	6.5	1	5	1.5	1	2	1	2.9	3.4	3.1
BMN MSE	.054	.004	.097	.008	.239	.038	.363	.068			
% eff	100	100	86	100	35	29	23	15			
Rank	6	6.5	2	5	1.5	2	1	2	3.1	3.9	3.5
BDYX MSE	.074	.008	.087	.012	.101	.037	.084	.023			
% eff	83	50	96	67	82	30	100	43			
Rank	5	1	5.5	1.5	5	3	7	4	5.6	2.4	4.0
OL1 MSE	.077	.007	.089	.012	.108	.017	.135	.051			
% eff	70	57	94	67	77	65	62	19			
Rank	2.5	2.5	3.5	1.5	3	4	3	3	3.0	2.8	2.9
ML1 MSE	.080	.007	.084	.011	.083	.011	.086	.012			
% eff	67	57	100	73	100	100	97	83			
Rank	1	2.5	7	3	7	6.5	6	6	5.3	4.5	4.9
THEIL MSE	.075	.005	.089	.008	.104	.015	.094	.014			
% eff	72	80	94	100	81	73	89	71			
Rank	4	5	3.5	5	4	5	4	5	3.9	5.0	4.5
SIEGEL MSE	.077	.006	.087	.009	.097	.011	.089	.010			
% eff	70	67	96	89	85	100	94	100			
Rank	2.5	4	5.5	4	6	6.5	5	7	4.8	5.4	5.1

As far as computational aspect is concerned, the Theil type estimators (i.e, THEIL and SIEGEL) and the Bartlett-median estimator, BDYX, are much easier to compute than the other estimators. In the case of L1-norm estimator (either OL1 or ML1) its formulation was originally based on a linear programming scheme (see Wagner 1959) which is computationally cumbersome. However, this difficulty has been reduced considerably with the presence of an efficient algorithm proposed by Barrodale and Roberts (1973).

7. CONCLUSION

The numerical evidence shows that the Theil-type estimators (i.e, THEIL and SIEGEL) and ML1 are most robust among those candidates presented in the study. The practical advantage of the Theil-type estimators over ML1 is their simple computations. On the other hand, ML1 estimates can be obtained efficiently when such an algorithm set out by Barrodale and Roberts (1973) is available.

These robust estimates can be used as starting values in other robust estimination procedures of the SLFR such as the M-estimation of Huber (1964).

ACKNOWLEDGEMENT

This investigation is a part of Mokhtar b. Abdullah's Ph.D thesis at the University of Dundee, Scotland. The research was supported by the Government of Malaysia.

REFERENCES

ABDULLAH M.B. 1987. Robust Estimation of a Linear Functional Relationship. Unpublished Ph.D thesis, Dept. of Mathematical Science, University of Dundee.

ANDERSON, T.M. 1984. Estimating Linear Statistical Relationships. *Ann. Statist.* **12:** 1-45.

BARRODALE, I. and F.D.K. ROBERTS. 1973. An Improved Algorithm for Discrete L1 Approximation. *SIAM J. Numer. Anal.* **10:** 839-848.

TABLE 3(a): $\alpha = \beta = 1$, n = 11
y_r - contamination; r = 11, fixed

	No Outlier				Outlier - y_r				RANK		RANK
h	0.5		2.0		6.0		10.0		AVE.		AVE.
Est.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha} + \hat{\beta}$
SLFR MSE	.048	.004	.101	.017	.368	2.058	2.538	108.5			
% eff	100	100	82	47	24	.5	3	.01			
Rank	6.5	7	3	1	2	1	1	1	3.1	2.5	2.8
BMN MSE	.048	.005	.100	.011	.384	.059	.889	.125			
% eff	100	80	83	73	23	18	10	7			
Rank	6.5	6.5	4	4	1	2	3	3	3.6	3.6	3.6
BDYX MSE	.061	.007	.083	.104	0.22	.092	0.26				
% eff	78	57	100	100	86	50	97	34			
Rank	5	3	7	6.5	5	4	5	4	5.5	4.4	4.9
OL1 MSE	.069	.008	.105	.016	.151	.051	1.403	1.279			
% eff	71	50	79	50	59	21	6	.7			
Rank	3.5	1.5	2	2	3	3	2	2	2.6	2.1	2.4
ML1 MSE	.069	.008	.108	.013	.092	.016	.089	.015			
% eff	69	50	77	61	98	69	100	60			
Rank	1.5	1.5	1	3	6	5	7	5	3.9	3.6	3.8
THEIL MSE	.069	.005	.088	.009	.107	.012	0.98	.011			
% eff	69	80	94	89	84	91	91	82			
Rank	1.5	5.5	5	5	4	6	4	6	3.6	5.6	4.6
SIEGEL MSE	.067	.006	.085	.008	.090	.011	.090	.009			
% eff	71	67	97	100	100	100	99	100			
Rank	3.5	4	6	6.5	7	7	6	7	5.6	6.1	5.9

TABLE 3(b): $\alpha = \beta = 1$, n = 11
y_r - contamination; r random

h	No Outlier 0.5		Outlier - y_r 2.0		6.0		10.0		RANK AVE.		RANK AVE.
Est.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
SLFR MSE	.055	.004	.087	.010	.351	.134	1.065	3.520			
% eff	100	100	84	70	23	5	8	.2			
Rank	6.5	7	1.5	4.5	1.5	1	2	1	2.9	3.8	3.4
BMN MSE	.055	.005	.087	.010	.343	.040	.936	.113			
% eff	100	80	84	70	23	17	10	7			
Rank	6.5	6	1.5	4.5	1.5	3	3	4	3.1	4.4	3.8
BDYX MSE	.067	.009	.079	.019	.159	.061	.438	.223			
% eff	82	44	92	37	50	11	21	3			
Rank	5	1	3.5	1	3	2	4	3	3.9	1.8	2.9
OL1 MSE	.072	.008	.074	.012	.092	.019	1.595	.810			
% eff	76	50	98	58	87	37	6	1			
Rank	4	2.5	6	2	6	4	1	2	4.3	2.6	3.5
ML1 MSE	.078	.008	.075	.011	.080	.012	.092	.011			
% eff	70	50	97	63	100	58	100	73			
Rank	1	2.5	5	3	7	5	7	5	5.0	3.9	4.5
THEIL MSE	.075	.006	.079	.007	.100	.007	.107	.008			
% eff	73	67	92	100	80	100	86	100			
Rank	2	5	3.5	7	4	7	5	7	3.6	6.5	5.1
SIEGEL MSE	.074	.007	.073	.008	.097	.009	.100	.009			
% eff	74	57	100	87	82	78	92	89			
Rank	3	4	7	6	5	6	6	6	5.3	5.5	5.4

BROWN, M.L. 1982. Robust Line Estimation with Errors in Both Variables. *J. Amer. Statist. Assoc.* **77:** 71-79.

HAMPEL, F.R. 1968. The Influence Curve and Its Role in Robust Estimation. *J. Amer. Statist. Assoc.* **69:** 383-393.

HARVEY, A.C. 1977. A Comparison of Preliminary Estimators for Robust Regression. *J. Amer Statist. Assoc.* **72:** 910-913.

HUBER, P.J. 1964. Robust Estimation of a Location Parameter. *Ann. Math. Statist.* **35:** 73-101.

KELLY, G. 1984. The Influence Function in the Errors in Variables Problem. *Ann. Statist.* **12:** 87-100.

KENDALL, M.G. and A. STUART. 1973. *The Advanced Theory of Statistics.* London: Griffin.

LINDLEY, D.V. 1947. Regression Lines and the Linear Functional Relationship. *J. Roy. Statist. Soc. Suppl.* **9:** 218-244.

MARITZ, J.S. 1979. On Theil's Method in Distribution-Free Regression. *Austral. J. Statist* **21:** 30-35.

MADANSKY, A. 1959. The Fitting of Straight Lines when Both Variables are Subject to Error. *J. Amer. Statist. Assoc.* **54:** 173-205.

MORAN, P.A. 1971. Estimating Structural and Functional Relationships. *J. Multivar. Anal.* :232-255.

SIEGEL, A.F. 1982. Robust Regression Using Repeated Medians. *Biometrika* **69:** 242-244.

SOLARI, M.E. 1969. The 'Maximum Likelihood Solution' of the Problem of Estimating a Linear Functional Relationship. *J.R. Statist. Soc. B* **31:** 372-375.

SPRENT, P. 1969. *Models in Regression and Related Topics.* London: Methuen.

THEIL, H. 1950. A Rank Method of Linear and Polynomial Regression Analysis. I. *Proc, Kon. Ned. Akad. Wetensh. A* **53:** 386-392.

WAGNER, H.M. 1959. Linear Programming Techniques for Regression Analysis, *J. Amer. Statist. Assoc.* **54:** 206-212.

(Received 10 June, 1988)