Listy Biometryczne - Biometrical Letters
Vol.XXV (1988), No.1,2, 3-24

L-ESTIMATORS IN LINEAR MODELS- FOUNDATIONS, COMPUTATIONAL ASPECTS, AN APPLICATION

Jaromir Antoch, Anna Bartkowiak

Department of Statistics, Charles University, Prague, Czechoslovakia
Institute of Computer Science, University of Wroclaw, Poland

Summary

This paper describes one class of robust estimators, the so called L-estimators. The main representants of this class, i.e., the L_1-norm estimator, the regression quantiles, the α-trimmed and α-Winsorized least squares estimators are described and some properties of them summarized. Computational aspects of the introduced estimators are considered and a short description of the implementation of the algorithms on the ODRA 1305 computer is given. Also some applications of the introduced estimators are presented. They are concerned with establishing a dependency between arterial blood pressure readings recorded automatically and in a traditional way.

1. INTRODUCTION

Let us consider the linear model of the form

$$y = XQ + e , \qquad (1.1)$$

where $y = (y_1,...,y_n)'$ is the vector of independent observations, $X = (x_{ij})$, $i = 1,...,n$, $j = 1,...,p$, is the nxp design matrix, $Q = (Q_1,...,Q_p)'$ is the vector of unknown parameters and $e = (e_1,...,e_n)'$ is the vector of independent, identically distributed (iid) errors. The distribution function *F* describing the probability distribution of e is generally unspecified, we assume only that it belongs to an appropriate family F of distribution functions. In the following we shall assume that $x_{i1}=1$, $i=1,...,n$, i.e., that we have a model with an intercept. Our main interest is in robust estimation of the vector of parameters Q with

---

Key words: linear model, L_1-estimator, regression quantiles, α-trimmed and α-Winsorized least squares estimator

emphasis on the so called distributional robustness. By this we mean estimators which are independent of the form of F, i.e. especially which are not so sensitive to longtailed distributions comprising possibly outlying observations ("outliers").

For the location submodel

$$y_i = Q + e_i \quad , \qquad i = 1,\ldots,n \quad , \tag{1.2}$$

three broad classes of robust estimators, i.e., M- , L- and R-estimators were introduced and intensively studied. For more informations see, e.g., the monographs of Huber (1981) or Hampel et al. (1986), where the finite sample as well as asymptotic properties of these estimators are described.

The M- and R-estimators for the location submodel were extended in a straightforward way to the linear model - what was not the case for the L-estimators though they are computationally so appealing.

Despite several attempts it seems to be only Koenker and Bassett's (1978) concept of regression quantiles as an extension of the sample quantile to the linear model which provides a reasonable solution and yields a reasonable robust estimator of the vector of parameters Q . The concept of regression quantiles offers a basis not only for the construction of robust L-estimators in a general linear model but also for the construction of robust tests of linear hypotheses and for a robust analysis of variance. The same authors suggested also the α-trimmed least squares estimator $L(\alpha)$ as an extension of the α-trimmed mean to the linear model. Later on Jurečková (1983 b) proposed and studied the α-Winsorized least squares estimator $W(\alpha)$ as an extension of the α-Winsorized mean.

The theoretical properties of regression quantiles, α-trimmed and α-Winsorized least squares estimators were intensively studied, among others, by Koenker and Bassett (1978), Ruppert and Carroll (1978, 1980), Koenker (1983), Jurečkowá (1983a, 1983b, 1984) and Antoch and Jurečkowá (1985). Numerical behavior based on broad Monte Carlo study is described in Antoch (1985) and Antoch et al.(1984). The construction of robust tests of hypotheses in linear model based on L-estimators was considered e.g, in the papers of Koenker and Bassett (1982), Jurečkowá (1983) and Ruppert and Carroll (1980). Fast algorithm for model choice in general linear model was described by Antoch (1986).

In the following we describe in section 2 the L_1-norm approach in regression. In section 3, we introduce the concept of regression quantiles and show the relation of this concept with the basic concept of M-estimators. In section 4 we introduce the concept of α-trimmed estimators and show some their asymptotic properties. In section 5 the α-Winsorized estimators are introduced. In section 6 a general discussion of the introduced estimators is given. In section 7 we show computational algorithms for evaluation of the estimators introduced in sections 2-4. In section 8 we show a medical example (blood pressure data) in which we evaluate and compare the considered estimators. Section 9 contains some final remarks relating to the use of the considered estimators when applied to the medical data.

2. L_1- NORM APPROACH IN REGRESSION

Let us consider the model (1.1). A L_1-norm estimator of Q is nothing else than a vector Q^o which minimizes the sum of absolute deviations from the expected value, i.e.,

$$Q^o = \arg\min_Q \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} x_{ij} Q_j \right| . \tag{2.1}$$

Due to historical and mnemotechnical reasons we shall denote this estimator in subsequent text also by L1.

It has been known for some time, see e.g. Armstrong (1979), that (2.1) is equivalent to the following linear programming problem: Minimize

$$\sum_{i=1}^{n} (P_i + N_i) \tag{2.2}$$

subject to

$$y_i - \sum_{j=1}^{p} x_{ij} Q_j + N_i - P_i = 0, \quad P_i \geq 0, \quad N_i \geq 0, \tag{2.3}$$

$$i = 1,\ldots,n \quad ,$$

where P_i and N_i are given by

$$P_i = \max\,(0, y_i - \sum_{j=1}^{p} x_{ij} Q_j), \tag{2.4}$$

$$N_i = \max\,(0, -(y_i - \sum_{j=1}^{p} x_{ij} Q_j)) \quad , \quad i = 1,\ldots,n \ .$$

A L_1-norm regression passes always through at least p points belonging to the considered cluster of points-individuals (p=rank of the matrix X in 1.1).

An elementary introduction to this problem may be found in Sposito, Smith and McCormick (1978). An efficient algorithm for solving (2.2) using a modified simplex method was given by Barrodale and Roberts (1973, 1974). Lately Armstrong, From and Kung (1979) gave another version of this algorithm. A method for computing linear regression in L_1-norm stepwise with downdating and updating the rows and columns was proposed by Peters and Willms (1983). Their program uses partially the procedure of Barrodale and Roberts (1974).

3. REGRESSION QUANTILES

Let us consider the model (1.1) and let α, $0<\alpha<1$, be a fixed constant. Koenker and Bassett (1978) introduced α-regression quantiles $Q(\alpha)$ as any solution of the minimization problem

$$\min_{Q} \sum_{i=1}^{n} [\alpha P_i + (1-\alpha) N_i] \quad , \tag{3.1}$$

where P_i and N_i, $i = 1,\ldots,n$, are given by (2.4). This is nothing else than to minimize the sum of very simply weighted residuals, i.e.,

$$\min_{Q} \sum_{i=1}^{n} w_i r_i \quad , \tag{3.2}$$

where

$$r_i = y_i - \sum_{j=1}^{p} x_{ij} Q_j \; , \quad w_i = \begin{cases} \alpha & \text{iff} \quad r_i > 0 \\ \alpha-1 & \text{iff} \quad r_i < 0 \end{cases} \; . \tag{3.3}$$

It is interesting to notice that an α-regression quantile $Q(\alpha)$ is in fact an M-estimator, i.e., an estimator which yields the solution of the minimization problem

$$\min_{Q} \sum_{i=1}^{n} \rho \, (y_i - \sum_{j=1}^{p} x_{ij} Q_j) \quad , \tag{3.4}$$

where

$$\rho(x) = \begin{cases} \alpha x & x \geq 0 \\ (\alpha-1)\, x & x < 0 \end{cases} \; . \tag{3.5}$$

This fact is very important because it enables to use a lot of general results derived for M-estimators in linear model when describing the behavior of regression quantiles and other estimators based on them.

For the case $\alpha = 0.5$, which gives the so called regression median, one can easily see that the solution of (3.1) coincides with the L_1-norm estimator.

Koenker and Bassett (1978) studied the asymptotic behavior of the regression quantiles and showed that they appear to have analogous properties to the ordinary sample quantiles of the location model. They established asymptotic normality of the regression quantiles and pointed out that these can be easily computed by a modification of the linear programming algorithm (in fact it is sufficient to modify slightly the procedure for the L_1-norm estimator). Later on Ruppert and Carroll (1980) and Jurečkowá (1984) derived similar asymptotic results under less restrictive conditions.

It is important to notice that all asymptotic results for regression quantiles have sense only in the case when we consider the model (1.1) with intercept. Otherwise it is necessary to supplement the model by an additional dummy intercept and to estimate it simultaneously with other components of the vector of parameters.

4. α-TRIMMED LEAST SQUARES ESTIMATOR

Koenker and Bassett (1978) suggested the α-trimmed least squares estimator $L(\alpha)$ in the following way. Let $0<\alpha<1/2$ be a fixed constant and let $Q(\alpha)$, resp. $Q(1-\alpha)$, be the corresponding regression quantiles. Remove every observation y_i such that

$$y_i - \sum_{j=1}^{p} x_{ij} Q_j(\alpha) \leq 0 \quad \text{or} \quad y_i - \sum_{j=1}^{p} x_{ij} Q_j(1-\alpha) \geq 0 , \tag{4.1}$$

where $Q(\alpha) = (Q_1(\alpha),\ldots,Q_p(\alpha))'$,

$Q(1-\alpha) = (Q_1(1-\alpha),\ldots,Q_p(1-\alpha))'$.

Then the α-trimmed least squares estimator $L(\alpha)$ is defined as the ordinary least squares estimator computed from remaining observations. In other words, if we construct the diagonal matrix $D=\text{diag}\,(d_{11},\ldots,d_{nn})$ such that $d_{ii}=0$ if (4.1) holds, resp. $d_{ii}=1$ otherwise, then the α-trimmed least squares estimator $L(\alpha)$ is the classical least squares estimator computed from the model

$$Dy = DXQ + De \tag{4.2}$$

instead from the model (1.1). It means that

$$L(\alpha) = (X'DX)^{-}X'DY \tag{4.3}$$

The idea of this estimator is graphically illustrated for the case of simple regression line in figures 1a and 2b.

The choice of a symmetrical trimming is not crucial. On the contrary, in the case of nonsymmetrical distribution of errors a nonsymmetrical trimming may be preferable. Simple modification of the procedure just described consists in fixing of two constants α_1,α_2, $0<\alpha_1<1/2<\alpha_2<1$ and computing the corresponding regression quantiles $Q(\alpha_1)$, $Q(\alpha_2)$. Then one will follow the scheme described just before with $Q(\alpha_1)$ and $Q(\alpha_2)$ instead with $Q(\alpha)$ and $Q(1-\alpha)$.

According to common experience based both on theoretical results and on simulations and practical applications the α-trimmed least squares estimator seems to be the most important representant of the class of robust estimators. Therefore, we shall present here in more details the conditions under which the distribution of this estimator as well as some its asymptotic properties, can be derived. These quite general

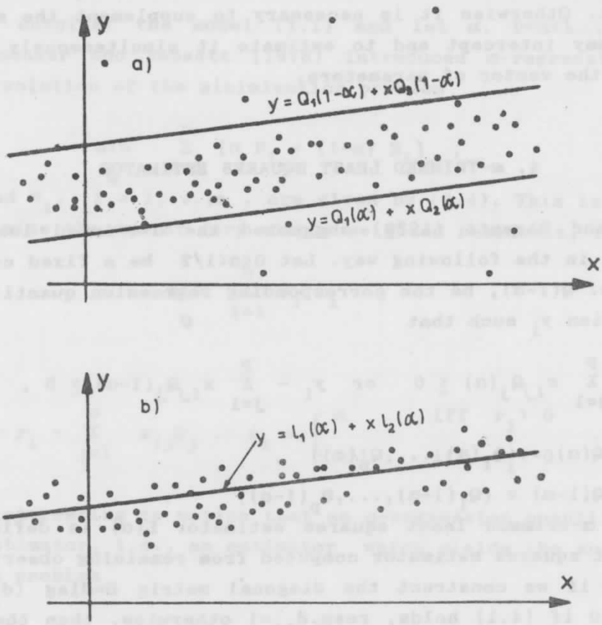

Fig. 1. a) Original data with corresponding regression lines evaluated from regression quantiles $Q(\alpha)=(Q_1(\alpha),Q_2(\alpha))$ and $Q(\alpha)=(Q_1(1-\alpha),Q_2(1-\alpha))$,
b) Data after trimming with the resulting α-trimmed least squares estimator $L(\alpha)=(l_1(\alpha),l_2(\alpha))$.

conditions are due to Jurečková (1984); for some more restrictive ones see, e.g., the paper of Ruppert and Carroll (1980).

Condition (A): Suppose that $F(x)$ is a distribution function of the error term e such that:

- $F(x)$ is absolutely continuous with density $f(x)$ such that $f(x)>0$ for $x \in [F^{-1}(\alpha)-\varepsilon,\ F^{-1}(1-\alpha)+\varepsilon]$, $\varepsilon > 0$, $F^{-1}(t) = \sup_{x}\{ x : F(x) \le t\}$, $0 < \alpha < 1/2$ fixed;
- the derivative $f'(x)$ of $f(x)$ exists and is bounded and positive in the neighbourhood of $F^{-1}(\alpha)$, $F^{-1}(1-\alpha)$.

Condition (B):

- suppose that there exists a positive definite matrix C such that $n^{-1} X'X \longrightarrow C$, as $n \longrightarrow \infty$;
- $\max_{1\le j\le p} n^{-1} \sum_{i=1}^{n} x_{ij}^4 = O_p(1)$ as $n \longrightarrow \infty$;

$$\max_{1\le i\le n, 1\le j\le p} n^{-1/2} \left| x_{ij} \right| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty \quad ;$$

$$x_{i1} = 1 \quad , \quad i=1,\dots,n \quad .$$

It was shown by Jurečkowá (1984) that under fulfilling of the conditions (A) and (B) the distribution of the α-trimmed least squares estimator is asymptotically normal, i.e.,

$$\mathcal{L}(n^{1/2}(L(\alpha) - Q)) \sim N_p(0, \sigma^2(\alpha,F) \; C^{-1}) \quad \text{as} \quad n \longrightarrow \infty \; , \qquad (4.4)$$

where $N_p(a, \Sigma)$ stands for the p-dimensional normal distribution with expectation a and covariance matrix Σ . More precisely, in our case

$$\sigma^2(\alpha,F) = (1-2\alpha)^{-2} \left[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x^2 dF(x) + 2\alpha \, (F^{-1}(\alpha))^2 \right] \qquad (4.5)$$

is the asymptotic variance of the α-trimmed mean in the location case and C^{-1} is the asymptotic covariance matrix derived from the cross-products of the columns of the design matrix.

Relations between L- and M-estimators in linear model were studied by Jurečkowá (1983a, 1983b). Under the fulfilling of the conditions (A) and (B) she has shown that

$$n^{1/2} \; \| \, L(\alpha) - M(\alpha) \, \| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty \; , \qquad (4.6)$$

where $M(\alpha)$ is Huber's M-estimator with ρ function of the form

$$\rho(x) = \begin{cases} |x| \; F^{-1}(\alpha) & |x| \ge F^{-1}(\alpha) \\ x^2 & |x| < F^{-1}(\alpha) \end{cases} \quad . \qquad (4.7)$$

It means, that $M(\alpha)$ minimizes (3.4) when we use $\rho(x)$ of the form (4.7).

A modification of this estimator which is moreover resistant to the existence of leverage points was described by Antoch and Jurečkowa (1985).

Some principal results describing the behavior of $L(\alpha)$ when applied to real data can be found in Ruppert and Carroll (1980) and Antoch and Jurečkowá (1985). More profound numerical Monte Carlo studies were performed by Antoch (1985) and Antoch et al. (1984).

5. α-WINSORIZED LEAST SQUARES ESTIMATOR

The concept of an α-Winsorized least squares estimator $W(\alpha)$ was introduced by Jurečkowá (1983 b) in the following way. Suppose that α, $0 < \alpha < 1/2$, is a fixed constant, $Q(\alpha)$ and $Q(1-\alpha)$ are the corresponding regression quantiles and $L(\alpha)$ is an α-trimmed least squares estimator.

Than the α-Winsorized least squares estimator $W(\alpha)$ is defined by

$$W(\alpha) = n^{-1} \{[n\alpha] (Q(\alpha)+Q(1-\alpha))+(n-2[n\alpha]) L(\alpha)\} . \qquad (5.1)$$

In other words, (5.1) is a direct generalization of the α-Winsorized mean in the location case (1.2), because we replace the trimmed observations by the corresponding regression quantiles with weights equal to the trimming proportions.

6. GENERAL DISCUSSION OF THE INTRODUCED ESTIMATORS

It is generally known, that the least squares (LSE) estimators of parameters in linear model are very sensitive to single outliers or some amount of points-individuals coming from other distribution than that which is assumed for the main part of data. Therefore, there is a need to search for estimators obtained by other methods than LSE, with the hope, that the employed methods will not be much influenced by the outliers.

The L_1-norm estimator was introduced historically first - as compared to other estimators considered in our paper. However, especially when the sample size is small, the estimate of the parameters Q , can be quite different from estimates obtained by other methods. This fact contributes to the opinion confirmed by Monte Carlo studies, that this estimator can be not reliable and satisfactory enough. When using this estimator for real data one should keep these facts in mind. For more details see, e.g., Antoch (1985).

The concept of α-trimmed least squares estimator is intuitively very appealing: We find from each side of the regression line or regression plane "remote" points which might give large residuals (see Fig.1a). These points are likely to distort the regression equation fitted to the main (central) part of the points contained in the considered cloud of data points. We remove these "remote" points from our data set and compute from the remainder the ordinary least squares estimator $L(\alpha)$ (see Fig 1b) which now is not influenced by the "remote" points removed from the data.

The α-Winsorized estimators $W(\alpha)$ can throw additional light on the "remote" points, if any. The computation of these estimators is trivial in the case when we have computed already the α-trimmed least squares estimators (the computational algorithms are given in the next sectin of this paper). It is highly recommendable to compute always $W(\alpha)$ together with $L(\alpha)$ for the additional control. A remarquable difference between them ought to be a signal for us to increase our attention and to reanalyse the data with the highest possible care.

The behavior of the estimators considered by us was studied in simulation experiments described by Antoch (1985) and Antoch et al. (1984). Various measures of errors like integrated mean square error etc.

between the true known model and estimated regression curves were calculated for the estimators described in previous sections. Generally, the results were the best for the α-trimmed least squares estimator L(α). Namely, the measures of errors for L(α) were in the same situations less or equal to those for the W(α) or the L_1-norm estimator. Nevertheless, it makes sense to compute W(α) always together with L(α) as an indicator of very remote outliers. The variant (iii) in section 8 provides a very nice illustration of the case. Considering the α-trimmed estimators let us remind, that the trimming proportions must be large enough, at least as large as the expected percentage of distorted data, but preferably slightly larger. It is always necessary to bear in mind that for fixed α just approximately [nα] observations will be trimmed off from "both sides" of the bulk of the data - also together 2[nα] observations. Therefore, it is preferable to overestimate the percentage of bad data than to underestimate it. This means that we should use preferably a larger α .

7. COMPUTATIONAL ASPECTS OF DESCRIBED METHODS

L_1 - NORM ESTIMATOR

Hitherto now the algorithm proposed by Barrodale and Roberts (1973) (and some of its generalizations) seems to be the most convenient one. It is in fact modification of the simplex method. The acceleration against the ordinary simplex method (algorithm) is caused by a trick enabling to pass through several vertices in one step. Some CPU times recorded on the ODRA 1305 computer are shown in Table 1. The solution is not generally unique, but this is a very rare case in practice. Nevertheless, in the case of multiple solutions some rule must be given which selects a unique one. Such a rule may be, e.g., that which choses the shortest in the L_2-norm among all avaiable solutions.

REGRESSION QUANTILES

Regression quantiles can be evaluated by an easy modification of the algorithm for the L_1-norm estimator, which consists in simple adaptation of the object function according to (3.1). It is sufficient to add to the formal parameters of the subroutine for calculation of the L_1-norm estimator an additional parameter ALPHA (corresponding to α, the desired portion of trimming) and to change adequately the form of the object function. CPU times are practically the same as in the case of a L_1-norm estimator.

Table 1. Times of run of the programs for problems of various sizes - CPU times of ODRA 1305 computer in minutes.
N - number of rows , P - number of columns in the design matrix X appearing in (1.1),
L1 - L_1-norm estimator,
L(.25) - α-trimmed LSE estimator with α =0.25
LSE1 - classical LSE estimator calculated by QR decomposition (Lawson-Hanson subroutine)
LSE2 - classical LSE estimator calculated by modified Gauss-Jordan pivots (SABA package)

N	P	L1	L(.25)	LSE1	LSE2
300	6	5.03	-	0.34	0.09
200	6	2.45	5.02	0.23	0.06
100	6	0.48	1.53	0.12	0.03
50	6	0.23	0.43	0.06	0.02
300	11	9.38	-	-	0.16
200	11	6.54	10.12	0.54	0.11
100	11	2.01	3.37	0.27	0.06
50	11	0.44	1.19	0.14	0.03
200	15	9.13	-	1.26	0.16
100	15	3.12	5.09	0.43	0.09
50	15	1.14	1.59	0.22	0.05

These times are to be read in the following manner: CPU = 1.59 means 1 minute and 59 seconds.

α-TRIMMED LEAST SQUARES ESTIMATOR

The algorithm for computation of an α-trimmed least squares estimator L(α) is the following:

(a) Fix α, $0 < \alpha < 1/2$;

(b) Compute α and (1-α) regression quantiles Q(α) and Q(-α) ;

(c) Remove "outliers" according to (4.1) ;

(d) Compute the classical least squares estimator from remaining observations.

α-WINSORIZED LEAST SQUARES ESTIMATOR

The algorithm for computation of an α-WINSORIZED least squares estimator W(α) is the following:

(e) Proceed (a) - (d) of the preceding algorithm for computation of the α-trimmed least squares estimator L(α) ;

(f) Use the formula (5.1) .

One can see that the computation of W(α) is trivial in the case when we have already computed L(α). CPU times for both these two estimators are practically the same.

THE CLASSICAL LEAST SQUARES ESTIMATOR

For the computation of the least squares estimator we have used an algorithm based on the QR decomposition of the design matrix X by means of Householder transformations. The program uses subroutines which were originally elaborated by Lawson and Hanson (1974) and slightly modified by Marazzi (1980) and Antoch (1986).

All these algorithms were implemented in FORTRAN IV on the ODRA 1305 computer in the Institute of Computer Science of the Wroclaw University in the form of three separate programs. The general pattern of the programs is similar to that of programs from the package SABA elaborated in the same place. Each program can run under several options establishing mainly the way of reading in the data and printing partial or full results. The options are chosen by declaring logical variables at the beginning of the program. The meaning of various variables and parameters is described in the formal description of each program. The calculations may be repeated for several sets of the data or their various subsets. For more details see the Report N-159 of Antoch, Bartkowiak and Pękalska (1986). Antoch prepared also a version of computing programs for the IBM PC compatible microcomputers.

In Table 1 we show CPU times of run of the programs on an ODRA 1035 computer. We consider computation times of the L_1-norm estimator L1, the α-trimmed least squares estimator L(.25) and the classical least squares estimator LSE1 computed using the QR decomposition.The programs have run for randomly generated data sets with p=6, 11, 15 variables and n = 50, 100, 200, 300 individuals.

For comparison we present also CPU times of calculations of the classical least squares estimator LSE2 computed using the program PD-ABA12 from the SABA package (this program uses modified Gauss-Jordan transformations).

Another presentation of CPU times of run of the program for evaluating a L_1-norm estimator was published by Armstrong et al. (1979), pp.178. They used a CDC 6600 computer. Their conclusion was that the number of parameters may not exceed 25 and the number of observations may not exceed 1000. Let us recall that they applied a different algorithm for solving the regression problem in L_1-norm then that we have used for our computations.

8. MEDICAL EXAMPLE

We shall consider data collected in Lower Silesia Medical Diagnostic Center in Wrocław. The data come from an experiment described in more details by Bartkowiak, Ruta and Włodarczyk (1985). Here we shall take into account only two variables, i.e.:

y - systolic (resp. diastolic) blood pressure recorded traditionally -

in the position "sitting";

x - systolic (resp. diastolic) blood pressure recorded automatically using the Avionics 1905 pressurometer - also in the position "sitting".

We want to establish a linear relationship $y = a + bx$ for prediction of y for given value of x. In this paper we do it for three groups of data (each group of data comprises the appropriate values of the variable x and y):

(a) group I comprises systolic blood pressure (BP) measurements for 117 women;

(b) group II comprises systolic BP measurements for 119 men;

(c) group III comprises diastolic BP measurements for 117 women (these women are the same as in group I).

The scatterdiagrams of points-individuals characterized by the observed values (x,y) revealed no big outliers. In this situation all the considered estimators should give similar results not too different from those yielded by the classical least squares method. In such circumstances we could not demonstrate the advantages of the α-trimmed and α-Winsorized least squares estimators. To do this, we introduced artifically into each group of data one outlier. In particular, we made the following distortions:

(i) in group I we changed the value $(x,y) = (105,95)$ which has been observed for the individual no. 1, to the value (105,195);

(ii) in group II we changed the value $(x,y) = (140,120)$, which has been observed for the individual no. 1, to the value (140,320);

(iii) in group III we changed the value $(x,y) = (70,70)$, which has been observed for the individual no. 1, to the value (70,700).

In this way we produced some errors which are likely to occur when punching the data or missinterpreting one's notices.

The scatterdiagrams of the distorted data are given in figures 2, 3 and 4.

For each group of the data we calculated the ordinary least squares estimator, the L_1-norm estimator, the α-trimmed and α-Winsorized least squares estimators with $\alpha = 0.01$ and $\alpha = 0.1$. This was done for the orginal (i.e. undistorted) and the distorted (according to the changes described above in points (i)-(iii)) data sets as well. The results are summarized in Tables 2, 3, 4. In the following part of this section we shall provide detailed discussion of the results.

Ad (i). The estimates of the parameters a and b obtained both for the original and the distorted data are summarized in Table 2.

Table 2. Set I . Estimates of the parameters a and b of the regression $y = a + bx$. Distortion: (x,y) = (105,95) was changed to (x,y') = (105,195).

Method of estimation	Original data		Distorted data	
	a	b	a	b
LSE	10.81	0.96	15.69	0.93
L_1-norm	5.00	1.00	5.00	1.00
0.1 - trimmed	8.57	0.97	8.57	0.97
0.1 - Winsorized	9.17	0.97	9.31	0.97
0.01 - trimmed	9.70	0.97	19.06	0.94
0.01 - Winsorized	9.91	0.97	14.15	0.94

Looking at the results for original data we see that they are similar with except of those corresponding to the L_1-norm estimator. Quite other situation is met in the case of distorted data. The least squares estimates of a and b changed in a natural way corresponding to the shift in the y value for the individual no. 1 . On the other hand, the L_1-norm estimate was not influenced by the outlier at all. To obtain the α-trimmed and the α-Winsorized least squares estimates, for $\alpha = 0.01$ two regression lines, i.e. $y = 0.83x$ and $y = 154.16 + 0.38x$, were constructed. These lines trimmed off four individuals (observations) which are marked in Fig.2 . Looking at this figure one can see that each of the regression lines has trimmed off only those two points, through which it passed through, i.e., two points from the top and two points from the bottom of the corresponding scatterdiagram. The α-trimmed and α-Winsorized least squares estimates for the original and distorted data differ, however this difference is much smaller than those for the least squares estimates.

For $\alpha = 0.1$ two regression lines, i.e. $y = 5.0+0.8x$ and $y = 20.0+x$, were constructed. They trimmed off, apart from the four individuals mentioned above another 23 individuals (observations). After removing all these 27 points we had still 90 observations for further calculations.

Looking at Table 2 and comparing the 0.1- trimmed and 0.1-Winsorized least squares estimates both for the original and the distorted data we state that they are practically the same in both situations; moreover, they are near to the classical least squares estimates for the original data.

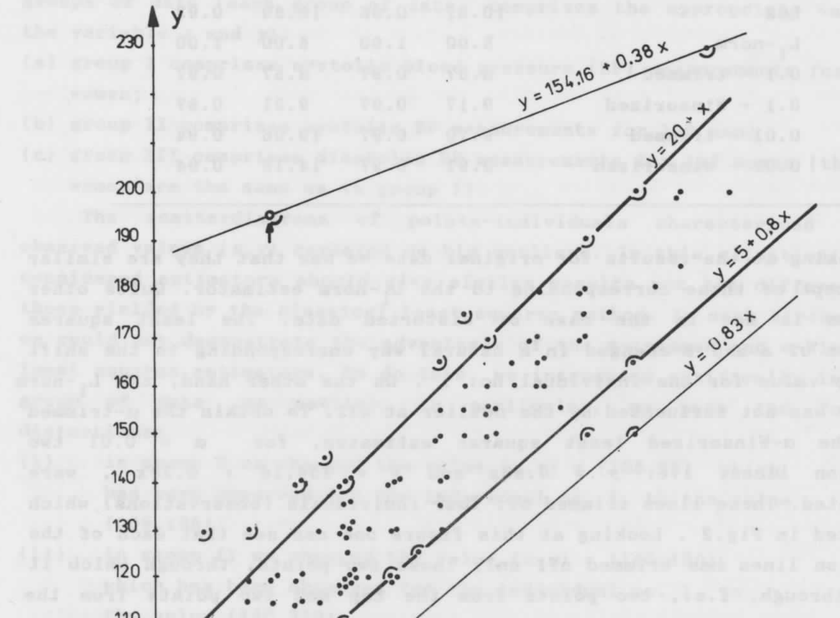

Fig. 2. Trimming off observations in group I. The distorted point (105,195) is indicated by an arrow.

Ad (ii). The estimates of the parameters *a* and *b* obtained both for the original and the distorted data are summarized in Table 3.

Table 3. Set II. Estimates of the parameters *a* and *b* of the regression $y = a + bx$. Distortion: $(x,y) = (140,120)$ was changed to $(x,y') = (140,320)$.

Method of estimation	Original data		Distorted data	
	a	*b*	*a*	*b*
LSE	21.16	0.87	22.03	0.87
L_1-norm	17.78	0.89	17.77	0.89
0.1 - trimmed	23.67	0.85	23.67	0.85
0.1 - Winsorized	24.14	0.85	24.14	0.85
0.01 - trimmed	19.69	0.88	19.78	0.88
0.01 - Winsorized	19.23	0.88	19.33	0.88

From the results for the original data one can see again that they are similar with except of those corresponding to the L_1-norm estimator. The situation for the distorted data is different than in the previous case. Despite the fact that the distortion was more serious (we have changed the value $(x,y) = (140,120)$ to the value $(140,320)$), the least square estimate of the parameter *b* remained the same and the estimate of the parameter *a* changed only a little. This seems to be caused by the fact that the *x* value of the distorted observation lies approximately in the middle of the range of values for *x* , so that it is solely the intercept which was slightly influenced, not the slope.

To obtain the α-trimmed and the α-Winsorized estimates, for $\alpha = 0.01$ two regression lines, i.e., $y = -3.0+x$, and $y = 15.38+1.15x$ were constructed. These lines trimmed off five individuals (observations) , which are marked in Fig.3. Looking at this figure one can see that the first regression line trimmed off two individuals from the bottom of the scatterdiagram, while the second one trimmed off three points from the top, including the outlier. In the last case the regression line does not pass through the outlier. The α-trimmed and α-Winsorized least squares estimates of *a* and *b* for the original and distorted data are practically the same.

For $\alpha = 0.1$ two other regression lines, i.e., $y = 16.0+0.8x$ and $y = 36.36+0.86x$, were constructed. They trimmed off, apart from the five individuals mentioned above, another 25 individuals. After removal all these 30 points we had still 89 individuals for further calculations. Again, comparing in Table 3 the 0.1-trimmed and the 0.1-Winsorized least

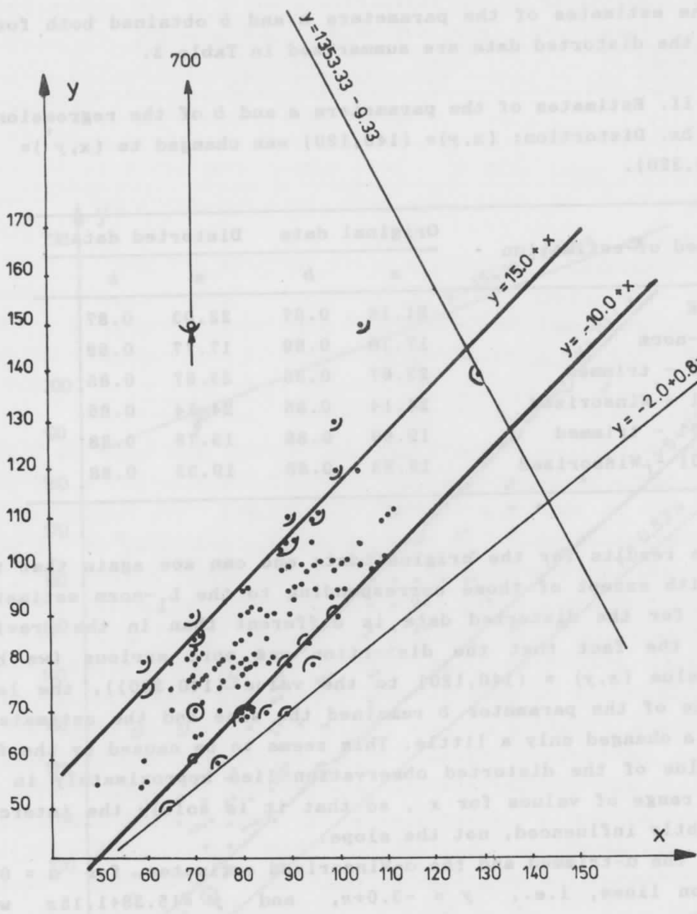

Fig. 4. Trimming off observations in group III. The distorted point (70,700) is indicated by an arrow.

remote observation) can influence the resulting estimate despite small weights put on the appropriate regrssion quantiles. A considerable difference between two 0.01-trimmed and the 0.01-Winsorized least squares estimates of the parameters a and b is thus really a very good indicator of peculiarities in the data.

For α = 0.1 two other regression lines, i.e., $y = -10+x$ and $y = 15+x$, were constructed. They trimmed off, apart of the four individuals mentioned above, 27 another individuals (observations). After removal all these 31 individuals we had still 86 of them for further calculations. Comparing in Table 4 the 0.1-trimmed and the 0.1-Winsorized least squares estimates both for the original and the distorted data, one can see that squares estimates both for the original and the distorted data, one can see that they are practically the same in both situations.

Ad (iii). In this group of data the distortion was really big. The estimates of the parameters *a* and *b* for the original and the distorted data are summarized in Table 4 .

Table 4. Set III. Estimates of the parameters *a* and *b* of the regression $y = a + bx$. Distortion: $(x,y) = (70,70)$ was changed to $(x,y') = (70,700)$.

Method of estimation	Original data		Distorted data	
	a	*b*	*a*	*b*
LSE	4.99	0.98	41.00	0.61
L_1-norm	12.22	0.89	12.22	0.89
0.1 - trimmed	10.38	0.91	10.67	0.91
0.1 - Winsorized	8.90	0.93	9.13	0.92
0.01 - trimmed	7.62	0.94	7.81	0.94
0.01 - Winsorized	7.36	0.95	19.23	0.86

We can see from the results obtained for the original data, that they differ very much in the estimate of the intercept and very little in the estimate of the slope. The data have generally a larger spread in the *y*-values (there seem to be two or three larger outliers which could influence the estimate of the intercept even for the original, undistorted data set) . It is again the L_1-norm estimate which differs substantially from all the other ones.

Quite other situation is met in the case of distorted data. The least squares estimates of *a* and *b* changed substantially in a natural way corresponding to the big shift in the *y* value for the individual no. 1 . On the other hand, the L_1-norm estimate was not influenced by the outlier at all. To obtain the α-trimmed least squares estimates for $\alpha = 0.01$ two regression lines, i.e., $y = 1353.3-9.3x$ and $y = -2.0+0.8x$, were constructed. They trimmed off four individuals.

Looking at Fig. 4 one can see that the first regression line has trimmed off two individuals, including the outliers from the top of the scatterdiagram, while the second one trimmed off two points from the bottom. Therefore the α-trimmed estimator was not influenced by the outlier and the values of the estimates *a* and *b* for the original and distorted data are practically the same.

On the contrary, the α-Winsorized least squares estimator constructed in this case can serve as a very good example of the fact mentioned in sections 5 and 6, namely, how the existence of only one big outlier (a

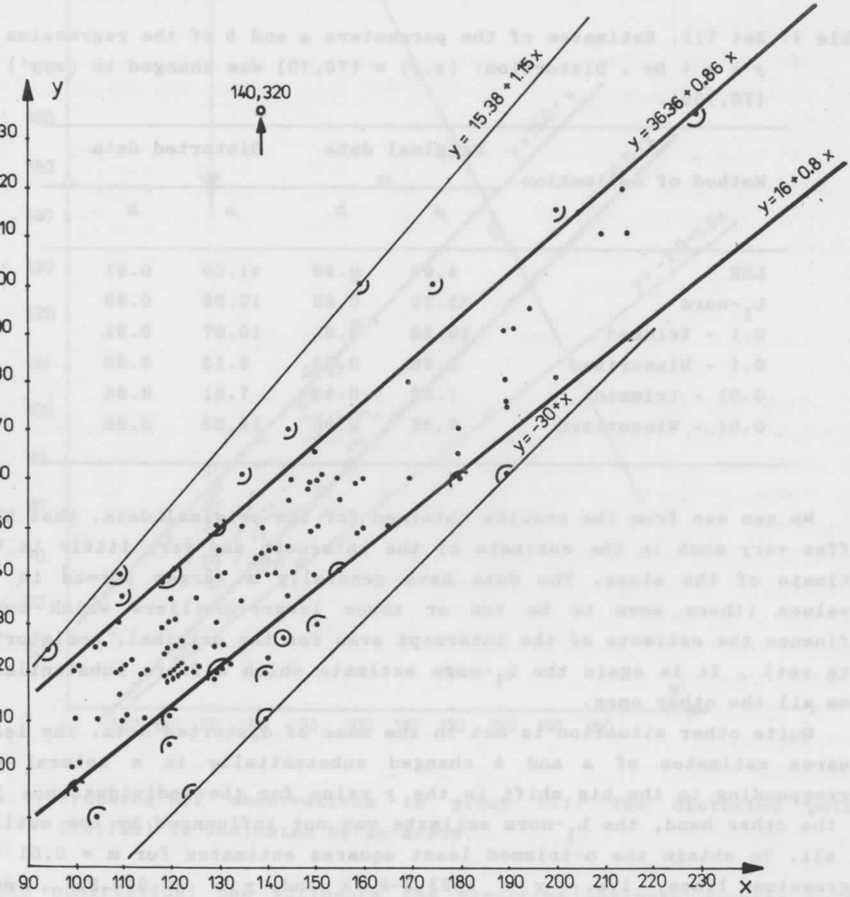

Fig.3. Trimming off observations in groupII. The distorted point (140,320) is indicated by an arrow.

they are practically the same in both situations.

Let us summarize several general conclusions following from our analysis.

The classical least squares estimator was again shown to be very sensitive to the existence of outliers - in our case to the existence of only one outlying observation. This is a known fact but one should keep it always in mind.

The L_1-norm estimator was not influenced by the outliers at all, but the estimates of *a* and *b* were quite different from all the other ones and were not enough reliable and satisfactory.

The α-trimmed least squares estimator yielded the best results. The α-Winsorized least squares estimator was shown to be a good complement to the α-trimmed least squares estimator serving as as indicator of big outliers in example (iii).

9. GENERAL SUMMARY AND CONCLUSIONS

In sections 2-6 of this paper several representants of the class of robust L-estimators, i.e., the L_1-norm estimator (the regression median), the α-trimmed and the α-Winsorized least squares estimators were introduced and some properties of them summarized. Their behavior was illustrated by an analysis of true medical data for which we considered distortions corresponding to different practical situations. All results obtained from the considered data fully coincide with those of Antoch (1985) and Antoch et al.(1986).

The estimators considered in this paper are generally not resistant against leverage points. In our data we did not have any leverage points - and therefore the obtained results are reasonable and show decidely the superiority of the introduced estimators in comparison to the classical least squares estimator when the data comprise big errors. In the case when there are some leverage points in the elaborated data, we advise that one should use modified estimators proposed by Antoch and Jurečková (1985).

We did not consider in this paper the problem of confidence intervals (confidence ellipsoid) for the introduced estimators nor the problem of variable selection. These problems were elaborated, a.o., by Antoch (1986).

REFERENCES

Abdelmalek N.N. (1974). On the discrete L_1 approximation and L_1 solution of overdetermined linear equations. *Journal of Approx. Theory* 11, 38-53.

Andrews D.F. (1974). A robust method for multiple linear regression. *Technometrics* 16, 523-531.

Antoch J. (1985). Numerical behavior of L-estimators in general linear model. *Comput. Stat. Quarterly* 2, 363-381.

Antoch J. (1986). Algorithmic development in variable selection procedures. In: *Proceedings of COMPSTAT 86*, Rizzi a. et all. eds, Physica Verlag Viena , 83-90.

Antoch J., Bartkowiak A., Pękalska J. (1986). Computing L_1-norm, α-trimmed and α-Winsorized regression on the ODRA 1305 computer. *Report N-159*, Institute of Computer Science of Wrocław University, March 1986.

Antoch J., Collomb G., Hassani S. (1984). Robustness in parametric and nonparametric regression estimation. An investigation by computer simulations. In: *Proceedings of COMPSTAT 84*, Havranek et al. eds., Physica Verlag Viena, 49-54 .

Antoch J., Jurečkowá J. (1985). Trimmed least squares estimator resistant to leverage points. *Comput. Stat. Quarterly* 2, 329-339.

Armstrong R.D., Frome E.L., Kung D.S.E. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem. *Commun. Statist.-Simul. Comput.* B 8, 175-190.

Barrodale I., Roberts F.D.K. (1973) . An improved algorithm for discrete L_1-norm approximation. *SIAM J. of Num. Analysis* 10 , 839-848.

Barrodale I., Roberts F.D.K. (1974) . Solution of an overdetermined system of equations in L_1-norm, *CACM* 19 , 319-320.

Bartkowiak A., Ruta R., Włodarczyk W. (1985) . Statistical relationships among arterial blood pressure measurements recorded automatically and in traditional way (in Polish) . In: *Człowiek, Populacja, Środowisko*, Prace Dolnośląskiego Centrum Diagnostyki Medycznej DOLMED, Wrocław, 12 , 49-92.

Bassett G., Koenker R. (1982). An empirical quantile function for linear models with iid errors. *J. Amer. Statist. Ass.* 77 , 407-415.

Bassett G., Koenker R. (1981). *Convergence of an empirical distribution function for the linear model.* Preprint.

Bickel P.J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Applied Statistics* 1 , 597-616.

Bickel P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Ass.* 70, 428-434.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986). *Robust statistics - The appr ach based on influence function.* J.Wiley, New York.

Huber P.J. (1981). *Robust statistics.* J.Wiley, New York

Jurečkowá J. (1983). Trimmed polynomial regression. *Comment. Math. Univ. Carolinae* 24 , 597-607.

Jurečkowá J. (1983 a). Robust estimators of location and regression parameters and their second order asymptotic relations. In: *Trans. of 9th Prague Conf. on Inform. Theory etc.*, Visek J.A. ed., Academia

Praha with Reidl Dordrecht, 19-32.

Jurečková J. (1983 b). Winsorized least squares estimator and its M-estimator counterpart. In: *Contributions to Statistics - Essays in Honour of N.L.Johnson,* Sen P.K. ed., North Holland Amsterdam, 237-245.

Jurečková J. (1984). Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika* 20 , 345-357.

Koenker R., Bassett G. (1978). Regression quantiles. *Econometrica* 46 , 33-50.

Koenker R., Bassett G. (1982). Tests of linear hypotheses and L_1 estimation. *Econometrica* 50 , 1577-1583.

Koenker R. (1983). *A note on L-estimates for linear model* . Bell Laboratories, Murray Hill , Preprint.

Koenker R., D'Orey V. (1985). *Computing regression quantiles.* Univ. of Illinois, Dept. of Economics, Preprint.

Lawson Ch.L., Hanson R.J. (1974). *Solving least squares problems.* Prentice Hall Inc., Englewood Cliffs , New Jersey.

Marazzi A. (1980). ROBETH - A subroutine library for robust statistical procedures . In: *Proceedings of COMPSTAT 80,* Barritt M.M. ed., Physica Verlag Viena , 577-583.

Narula S.C. (1982). Optimization techniques in linear regression : A review. *TIMS Studies in Man. Sciences* 19 , 11-29.

Narula S.C., Wellington J.F. (1982). The minimum sum of absolute errors regression - A state of the art survey . *Int. Stat. Rev.* 50 , 317-326.

Peters U., Willms C. (1983). Program for computing linear regression in L_1-norm with downdating and updating of rows and columns. *OR Spectrum* 5, 229-239.

Ruppert D., Carroll R.J. (1978). *Robust regression by trimmed least squares estimation.* Univ. of North Carolina at Chapel Hill, Dept. of Statistics , Preprint.

Ruppert D., Carroll R.J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Ass.* 75 , 828-838.

Sposito V., Smith W., McCormick G. (1978). *Minimizing the sum of absolute deviations.* Vanderhoeck & Ruprecht , Göttingen.

Received 1 September 1987: revised 20 January 1989

L - ESTYMATORY W MODELACH LINIOWYCH. PODSTAWY. ASPEKTY OBLICZENIOWE. PRZYKŁAD ZASTOSOWAŃ.

Streszczenie

W pracy przedstawiono kilka reprezentantów odpornych estymatorów parametrów modeli liniowych, a mianowicie estymatory w normie L_1 oraz α-obcięte i α-winsoryzowane. Omówiono algorytmy służące do obliczeń wymienionych estymatorów. Pokazano przykład medyczny w którym zastosowano omawiane metody estymacji parametrów dla równania regresyjnego wyznaczającego ciśnienie tętnicze krwi mierzone metodą tradycyjną w zależności od analogicznego ciśnienia wyznaczonego automatycznie za pomocą presurometru Avionics 1905.