QÜESTIIÓ, Vol 19, 1, 2, 3 pp. 107-121, 1995

FITTING A LINEAR REGRESSION MODEL BY COMBINING LEAST SQUARES AND LEAST ABSOLUTE VALUE ESTIMATION

SIRA ALLENDE*, CARLOS BOUZA* and ISIDRO ROMERO†

Robust estimation of the multiple regression is modeled by using a convex combination of Least Squares and Least Absolute Value criterions. A Bicriterion Parametric algorithm is developed for computing the corresponding estimates. The proposed procedure should be specially useful when outliers are expected. Its behavior is analyzed using some examples.

Key words: Outliers in regression, L_1 regression, bicriterion parametric algorithm.

1. INTRODUCTION

The main objective of many applications is to obtain a function which should describe the relationship between a vector of known variables $\underline{x}^T = (X_1, \ldots, X_m)$ with $X_1 = 1$ and a response variable Y. That is the case when the abundance of phytoplankton (Y) is studied. The salinity, temperature and other variables can be measured. The biologist's aims is to establish a functional and to use the adjusted model for studying the effect of the explanatory variables in the abundance of phytoplankton. A classic approach is to suppose that

$$Y_i = \sum_{j=1}^{m} X_{ij} B_j + \varepsilon_1, \qquad i = 1, \ldots, n \tag{1.1}$$

* Sira Allende and Carlos Bouza. Universidad de La Habana.

† Isidro Romero. Universidad Autónoma de Puebla

– Article rebut el desembre de 1992.

– Acceptat el desembre de 1994.

is an adequate model. The parametric vector $\underline{B}^T = (B_1, \ldots, B_m)$ should be estimated. We usually want to compute an estimate $\hat{\underline{B}}$ of $\underline{B}$ which minimizes a certain function of the vector $(\varepsilon_1, \ldots, \varepsilon_n)'$, where $\varepsilon_i = Y_i - \underline{x}\underline{B}$ for $i = 1, \ldots, n$ and $\underline{x}_i^T = (X_{i1}, \ldots, X_{im})$. The classic approach is to minimize the squared deviations which leads to Least Squares (LS) estimation. The optimization problem to be solved here is:

$$\text{P1}: \quad \min_{\underline{B}} \sum_{i=1}^{n} \varepsilon_i^2$$

$$(1.2) \qquad \text{Subject to}: \quad \sum_{j=1}^{m} X_{ij}\hat{B}_j + \varepsilon_i = Y_i; \quad i = 1, \ldots, n$$

However LS estimates are severly affected by outliers because of the weight given to each point in the solution of P1 is the same. Then, if a heavy tailed distribution generates some residuals, the influence of them may be very important. Therefore P1 does not provide adequate estimations of $\underline{B}$ in the study of phytoplankton because it is frequent to obtain some extreme data points. Several studies, see Dielman (1984), Dielman and Pfaffenberger (1988) and Efron (1988), have shown that Least Absolute Value (LAV) regression criterion provides improved estimates. The corresponding optimization problem is:

$$\text{P2}: \quad \min_{\underline{B}} \sum_{i=1}^{n} |\varepsilon_i|$$

Subject to (1.2)

From a numerical point of view P2 is solved by computing the solution of the Linear Programming Problem

$$\text{P3}: \quad \min_{\underline{B}} \sum_{i=1}^{n} (d_i^+ + d_i^-)$$

$$(1.3) \qquad \text{Subject to}: \quad \sum_{j=1}^{m} B_j X_{ij} + d_i^+ - d_i^- = Y_i; \quad i = 1, \ldots, n$$

$$(1.4) \qquad d_i^+ \geq 0, \quad d_i^- \geq 0; \quad i = 1, \ldots, n$$

We will denote $d_i^+ + d_i^- = e_i$.

The equivalence between LAV problems and Linear Programming was pointed out by Charnes *et al.* (1955). Special purpose algorithms, based on modifications of the Simplex Method (SM), increased the possibilities of using LAV. The consistency and asymptotic efficiency of this method with respect to LS was established, see Basset and Koenker (1978) and Dielman and Pfaffenberg (1988). The analysis of the behavior of LAV estimation plays a key role in the evaluation of regression equation fitting. LAV estimates are recommended as a good starting point in the search for a robust estimate of $\underline{B}$. As an example we can mention the results of Antonch and Bartkoviak (1988) in the study of the robustness of $\alpha-$trimmed and $\alpha-$winsorised estimators of $\underline{B}$. LAV and LS estimates may coincide but it follows from Jenssen's inequality that $||\underline{e}||_2 \leq ||\underline{e}||_1$, where $\underline{e}^T = (e_1, \dots, e_n)$. The idea of combining LAV and LS in a convex function is present in different previous publications. Arthanari and Dodge (1981) quoted the possibility of combining them for deriving a compromise estimator of $\underline{B}$. Dodge (1984) analysed the robustness of this procedure establishing that it provides the solution for a certain $M-$type estimator. Using the structure of the problem posed in that paper we have the optimization problem $Q(u)$ which is denoted as follows:

$$Q(u): \qquad \min_{\underline{B}} u \sum_{i=1}^{n} e_i^2 + (1-u) \sum_{i=1}^{n} e_i$$

Subject to (1.3)−(1.4)

The present paper follows a research of Allende and Bouza (1991). They studied the related optimization problem for $u \in [0,1]$ as appearing in $Q(u)$. Dodge (1984) solved a similar problem when u was known. The parameter u characterizes the contamination between an assumed normal distribution of the residuals and a heavy tailed one. We solve this problem formulating a Bicriterion Optimization Problem (BOP). A set of linear regression models is generated and one of them is selected on the basis of the analysis of values of a proposed coefficient. It has the property to increase with the goodness of the approximation of the equation to the observed data.

The Mathematical Model is discussed in the second section. The structure of the optimization problem is characterized by fixing the Karush-Kuhn-Tucker Conditions (KKTC). In the third section a model selection criterium is proposed. An algorithm for solving the problem, for a fixed value of u, is proposed. In Section 4 the results of a classic regression problems are used for evaluating the proposed procedure. A Monte Carlo experiment is performed and the behavior of the algorithm is discussed. The method is also used for fitting the regression equation of a set of data of the abundance of phytoplankton in the bay of Monterrey.

2. A CONVEX COMBINATION OF LS AND LAV

We consider the classic linear regression model $\underline{Y} = X\underline{B} + \varepsilon$. $\underline{Y} \in \mathbb{R}^n$ is an observable random vector and $\varepsilon \in \mathbb{R}^n$ a random unobservable vector with $E(\varepsilon) = \underline{0}$ and $\underline{B} \in \mathbb{R}^m$ is unknown.

A compromise between LS and LAV for estimating $\underline{B}$ was expressed by Allende and Bouza (1991) in terms of a Bicriterion Optimization Problem (BOP) given by:

$$\min_{\underline{\hat{B}}}\left\{\sum_{i=1}^{n} e_i^2, \qquad \sum_{i=1}^{n} e_i\right\}$$
Subject to: (1.3)–(1.4)

The definition of e_i fixes that $\underline{e}^T = (e_1, \ldots, e_n) \geq \underline{0}$.

An efficient point of BOP represents a regression model such that there does not exist another offering a smaller LAV and LS errors simultaneously. The set of efficient points of BOP coincides with the set of optimal solutions of $Q(m)$. See Lommatsch (1979) for a detailed discussion of this result. If $m \neq 1$ then $Q(m)$ can be reformulated as

$$Q(\theta): \qquad \min_{\underline{\hat{B}},\theta}\left\{\sum_{i=1}^{n} e_i^2 + e\sum_{i=1}^{n} e_i : \quad \theta = m/1-m\right\}$$
Subject to: (1.3)–(1.4)

θ will be called weight coefficient. $QP(\theta)$ is an one parameter Parametric Programming Problem and its special structure will be analyzed for obtaining a more efficient solution. The Karush-Kuhn-Tucker Conditions (KKTC) lead to a set of matrix equations which will be refered as the LCP–model or simply LCP. Taking $\underline{1}_m$ as a vector of m components equal to one, $\underline{g}(j)$, $j = 1, \ldots, 5$, and $\underline{B}(h)$, $h = 1,2$, as nonnegative vectors, $\underline{L}(h)$, $h = 1,2$, $\underline{g}(j)$, $j = 3,4,5$, and $\underline{e}$ belong to $\mathbb{R}^n$, $\underline{B}(h)$ and $\underline{g}(h)$, $h = 1,2$, belong to $\mathbb{R}^m$ and

(2.1) $$\underline{L}^T(1)\underline{g}(4) = \underline{L}^T(2)\underline{g}(5) = \underline{e}^T\underline{g}(3) = \underline{B}^T(1)\underline{g}(1) = \underline{B}^T(2)\underline{g}(2) = 0$$

The LC model is described as follows:

$$\begin{aligned}
\underline{X}^T\underline{L}(1) - \underline{X}^T\underline{L}(2) &= \underline{g}(1)\\
-\underline{X}^T\underline{L}(1) + \underline{X}^T\underline{L}(2) &= \underline{g}(2)\\
\theta\underline{1}_m + 2\underline{e} - \underline{L}(1) - \underline{L}(2) &= \underline{g}(3)\\
\underline{Y} - \underline{XB}(1) + \underline{XB}(2) + \underline{e} &= \underline{g}(4)\\
-\underline{Y} + \underline{XB}(1) - \underline{XB}(2) + \underline{e} &= \underline{g}(5)
\end{aligned}$$

Allende and Bouza (1991) proved that to solve $QP(\theta)$ is equivalent to obtain the solution of the LCP. The convexity of the objective function plays a key role in the proof.

An algorithm for estimating $\underline{B}$ should be able to determine, for each $\theta \geq 0$, an element of the set $G(\theta) = \left\{(\underline{x}, \underline{g}) \in \mathbb{N}(\theta) : \underline{x}^T\underline{g} = 0\right\}$, where $\underline{x} = (\underline{L}^T(1), \underline{L}^T(2), \underline{e}^T, \underline{b})^T$,

$\underline{b} = (\underline{B}^T(1), \underline{B}^T(2))$, $\mathbb{N}(\theta) = \{(\underline{x}, \underline{g}) : \underline{g} - \underline{W}\underline{x} = \underline{r}_0 + \theta \underline{r}_1\}$ and $\underline{g} = (\underline{g}^T(4), \underline{g}^T(5), \underline{g}^T(3), \underline{g}^T(1), \underline{g}^T(2))^T$. The vectors $\underline{r}_0$ and $\underline{r}_1$ are such that

$$r_{0i} = \begin{cases} 0 & \text{if} \quad i = 1, \ldots, 3n \\ Y_i & \text{if} \quad i = 3n+1, \ldots, 3n+m \\ -Y_i & \text{if} \quad i = 3n+m+1, \ldots, 3n+2m \end{cases} \tag{2.2}$$

$$r_{1i} = \begin{cases} 0 & \text{if} \quad i = 1, \ldots, 2n \\ 1 & \text{if} \quad i = 2n+1, \ldots, 3n \\ 0 & \text{if} \quad i = 3n+1, \ldots, 3n+m \end{cases} \tag{2.3}$$

Take $\underline{0}_{KH}$ as a matrix of zeros and $\underline{I}$ as the $m \times m$ identity matrix. (2.1) is a complementarity condition because, when the product of two variables is zero, each of them is called complement of the other one. The vector of the coefficient of the regression can be expressed by $\underline{B} = \underline{B}(1) - \underline{B}(2)$. The matrix $\underline{W}$ is given by:

$$\underline{W} = \begin{bmatrix} \underline{0}_{nn} & \underline{0}_{nn} & \underline{0}_{nm} & \underline{X}^T & -\underline{X}^T \\ \underline{0}_{nn} & \underline{0}_{nn} & \underline{0}_{nm} & -\underline{X}^T & \underline{X}^T \\ \underline{0}_{mn} & \underline{0}_{mn} & 2\underline{I} & -\underline{I} & -\underline{I} \\ -\underline{X} & \underline{X} & \underline{I} & \underline{0}_{mm} & \underline{0}_{mm} \\ \underline{X} & -\underline{X} & \underline{I} & \underline{0}_{mm} & \underline{0}_{mm} \end{bmatrix} \tag{2.4}$$

Using $\underline{W}$ the solution algorithm determines an element $(\underline{x}^*, \underline{g}^*)$ of the polyhedron $\mathbb{N}(\theta)$ with the property $\underline{x}^{*T}\underline{g}^* = 0$. Note that the points of $\mathbb{N}(\theta)$ that satisfy the constraint $-\underline{X}\underline{B} - \underline{e} \leq -\underline{Y}$ conforms $G(\theta)$.

Now we can formulate the following theorem:

Theorem 2.1

i) $G(\theta) \neq \Phi$ for all $\theta \geq 0$

ii) $G(\theta)$ is a closed face of the convex polyhedron $\mathbb{N}(\theta)$ or is equal to $\mathbb{N}(\theta)$.

Proof:

Take an arbitrary $\theta \geq 0$ and a particular vector $(\underline{x}_0, \underline{g})$ such that

$$\begin{aligned} \underline{e}_0 &= \underline{1}_n \theta / 2 \\ \underline{L}(1)_0 &= \underline{L}(2)_0 = \underline{g}(3)_0 = \underline{g}(4)_0 = \underline{g}(5)_0 = \underline{0} \end{aligned}$$

fixing

$$B(1)_{0i} = \left\{ \begin{array}{ll} Y_i - \theta/2 & \text{if } \theta \leq 2Y_i \\ 0 & \text{otherwise} \end{array} \right\}$$

$$B(2)_{0i} = \left\{ \begin{array}{ll} -Y_i - \theta/2 & \text{if } \theta > 2Y_i \\ 0 & \text{otherwise} \end{array} \right\}$$

As $\underline{g}(1)_0 = \underline{g}(2)_0 = \underline{0}$ the constraints of $G(\theta)$ are satisfied. For obtaining ii) consider the partition of $\underline{W}$ given by:

$$\underline{W} = \begin{bmatrix} \underline{0}_{(2n)2n} & \underline{0}_{(2n)m} & \underline{A}_{(2n)2m} \\ \underline{0}_{(m)2n} & \underline{0}_{(m)m} & \underline{I}^*_{(m)2m} \\ -\underline{A}^T_{(2m)2n} & -\underline{I}^{*T}_{(2m)m} & \underline{0}_{(2m)2m} \end{bmatrix}$$

where $\underline{I}^* = [-\underline{I}_{mm} \quad -\underline{I}_{mm}]$ and

$$\underline{A} = \begin{bmatrix} \underline{X}^T & -\underline{X}^T \\ -\underline{X}^T & \underline{X}^T \end{bmatrix}$$

Taking $\underline{v}^T = [\underline{v}^T(1), \underline{e}^T, \underline{v}^T(3)] \in \mathbb{R}^{3n+2m}$ the relation $\underline{v}^T \underline{W} \underline{v} = 2\underline{e}^T \underline{I} \underline{e} \geq 0$ permits to establish that $\underline{W}$ is positive semidefinite. From Theorem A.3.3 of Bank *et. al.* (1982) follows ii).

3. ALGORITHM FOR SELECTING THE REGRESSION EQUATION

We denote the prediction of Y_i, for a fixed weight coefficient θ, by $Y_{i\theta}$. An estimate of $\underline{B}$ is computed for each θ. Denote by $\underline{B}_\theta$ the corresponding estimator. The statistician needs to evaluate the behavior of the estimator with respect to various values of θ. The goodness of the regression equation obtained by the formula $B = \underline{XB}_\theta + \underline{e}$ is measured by

$$\frac{\sum_{i=1}^{n} (Y_{i\theta} - \overline{Y})^2 + \theta \sum_{i=1}^{n} |Y_{i\theta} - \overline{Y}|}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2 + \theta \sum_{i=1}^{n} |Y_i - \overline{Y}|} = R(\theta)$$

When $Y_{i\theta} \approx Y_i$, for any $i = 1, \ldots, n$, the coefficient $R(\theta)$ will be near to one and it will decrease with the $Y_{i\theta}$'s when they are very different from Y_i. Note that $R(\theta)$ plays a role similar to the determination coefficient in common LS regression problems.

The decision maker of the statistician can fix a threshold value of R_0 for judging whether a fitted regression equation is admissible or not. This criterion permits to obtain a set with the regression equations which can be used for the decision making. Take $\underline{B}_\theta$ as the estimate of $\underline{B}$ obtained for a θ fixed. Then $M = \underline{X}_\theta \underline{B}_\theta R(\theta)$ is such set.

Consider the sets

$$\mathbb{Z} = \left\{(\underline{x}, \underline{g}, \theta) \in \mathbb{R}^{3m+2n} \times \mathbb{R}^{3m+2n} \times \mathbb{R} : (\underline{x}, \underline{g}) \in \mathbb{N}(\theta)\right\}$$

and

$$\mathbb{C} = \left\{(\underline{x}, \underline{g}, \theta) \in \mathbb{Z} : \underline{x}' \underline{g} \geq 0\right\}$$

We will obtain a description of the vertices in $\mathbb{Z}$ contained in $\mathbb{C}$ using the results of Bank *et al.* (1982). Take $V = \left\{\underline{V}_n = {}^h\underline{x}, {}^h\underline{g}, {}^h\theta) \in \mathbb{Z} \cap \mathbb{C} \quad h = 0, 1, \ldots, p\right\}$.

The parametric problem $QP(\theta)$ is solved by taking into account those vertices. The statistician has a set $\Theta = \{\theta_1, \ldots, \theta_s\}$, $s > 0$ is an arbitrary integer, of weight coefficients and the solution of (3.2) for each $\theta_j \in \Theta$ permits to determine $\underline{B}_{\theta j}$. The 'best' equation is selected analyzing the corresponding $R(\theta_j)$'s. If $M = \Phi$ we consider that (1.1) is an inadequate model. The least permissible value R_0 of the $R(\theta_j)$, $j = 1, \ldots, s$, is fixed by the statistician. The appropiate value of θ is determined by the following procedure:

Procedure for selecting the best equation

START: Give $\Theta = \{\theta_1, \ldots, \theta_s\}, R_0$ and the observations $(y_i, x_{i1}, \ldots, x_{im})$, $i = 1, \ldots, n$

BEGIN {main}

(1) Construct a vertex $\underline{v}_0$ of the polyhedron $\mathbb{Z}$ such that $\underline{v}_0 \in \mathbb{C}$
Repeat:=true
i:=0

(2) while repeat do
begin
If there is $\underline{v}_{i+1} \in \mathbb{Z}$ adjacent to $\underline{v}_i$ with $\theta_{i+1} > \theta_i$ and $\underline{v}_{i+1} \in \mathbb{C}$
then
i:=i+1
else
begin
Repeat:=false
{describe the unbounded edge of $\mathbb{Z}$ contained in $\mathbb{C}$:
$(\underline{d}_1, \underline{d}_2, \underline{d}_3) \in \mathbb{R}^{3m+2n} \times \mathbb{R}^{3m+2n} \times \mathbb{R}$}
end
end; while (2)

$$\underline{v}_i = ({}^i\underline{x}, {}^i\underline{g}, {}^i\theta), \quad i = 1, \ldots, p \text{ was calculated}$$
$$\underline{B}(i) = {}^i\underline{B}(1) - {}^i\underline{B}(2)$$

Repeat:=true
$j = 1;\quad \mathbb{R}^* = -\infty$

(3) While repeat do
begin
Calculation of $\mathbb{R}(\theta_j)$ for each $j = 1, \ldots, s$

(4) While $j \leq s$ do
Determine:
$k = \max_{0 \leq i \leq p} \{i \mid {}^i\theta \leq \theta_j\}$

If $k < p$ then
begin

$$\begin{aligned} a &= 1 - \frac{\theta_j - k_\theta}{k + 1_\theta - k_\theta} \\ \underline{B}(j) &= a\underline{B}(k) + (1-a)\underline{B}(k+1) \\ \underline{e}(j) &= a^k\underline{e} + (1-a)^{k+1}\underline{e} \end{aligned}$$

end;
else
begin

$$\begin{aligned} a &= \theta_j / p_\theta \\ \underline{B}(j) &= a\underline{B}(p) \\ \underline{e}(j) &= a^p\underline{e} \end{aligned}$$

end;
Calculate $\mathbb{R}(\theta_j)$
If $\mathbb{R}(\theta_j) > \mathbb{R}^*$ then $\mathbb{R}^* = \mathbb{R}(\theta_j)$
$j := j + 1$
end: {while (1)}
If $\mathbb{R}^* \geq \underline{R}_0$ then $\underline{B}_0$ describes the linear regression model
else
begin
write "do you want to give another value of θ?"
if (answer="not") then repeat:=false
else
$s = s + 1$
end: {while (3)}
END {main}

This procedure permits to determine a set of adequate regression equations.

Since $\underline{r}_1$, defined in (2.3), is non null: $\mathbb{Z}$ is not empty.

Considering $\theta = 0$ the initial point $\underline{v}_0$ is obtained by applying the algorithm of Lemke (1970). It is an algorithm based in a Simplex tableau and a complementarity pivot technique. An artificial variable $\mathbb{Z}_0$ is introduced and a starting tableau is constructed by eliminating linear dependent rows and columns from $\underline{W}$. Then $\underline{W}'$ is obtained and an initial basic solution of the system $\underline{y}^* - \underline{W}'\underline{x} - \underline{x}_0 = \underline{r}_0$ is constructed with $\underline{y}$ as a basic value. By entering $\mathbb{Z}_0$ into the basis an index k is fixed by $r_{0k} = \min_i r_{0i}$ and y_k leaves the basis. The basic variables $y_j \quad (j \neq k)$ and $\mathbb{Z}_0$ have positive values. At each iteration the complement of the variable which lefts the basis in the previous iteration is entered. For the regression problem we can specify this algorithm in the following way:

T_0 : Starting Tableau

Basic variable	Non Basic Variable	$\mathbb{Z}_0$	
		-1	
		•	
$\underline{g}'$	$\underline{W}'$	•	$\underline{r}'_0$
		•	
		-1	

The solution of the initial vertex is obtained by using the following procedure:

Procedure for obtaining the initial solution (PIS)

Begin

{T_0 is not primal feasible because $r'_0 \ngeq 0$}

EV:=$\mathbb{Z}_0$ (variable $\mathbb{Z}_0$ enters into the basis)

Determine an index such that

$|y_{i0}| = \max_{i=1,\dots,m} |y_i|$

If $y_{i0} < 0$ then RV = $\underline{g}_{i_0}(4)$; {$\underline{g}_{i_0}(4)$ leaves the basis}

else

RV = $\underline{g}_{i_0}(5)$; {$\underline{g}_{i_0}$ leaves the basis }

T':= Pivoting T: {T' tableau is obtained after pivoting T}

Repeat:=true

While Repeat do

begin

EV:=comp (RV): { The complement variable of RV is introduced into the basis}

$q :=$ index of EV

(4) Determine an index i_0 such that

$$t'_{i_0} = \min_{t_{iq}>0} \frac{t'_{i_0}}{t_{i_q}}$$

RV:= Basic Variable (i_0) { The basic variable (BV) corresponding to the index i_0 leaves the basis }

If RV = $\mathbb{Z}_0$ then

Repeat:=false

T' := Pivoting T'

end; {while}

(5) Calculate

$t_\theta = f^{-1}\underline{r}_1$

{f denotes the basis corresponding to the tableau T' and f^{-1} its inverse}

Add the column t_θ to T'

{t_θ is the column corresponding to θ}

(6) Determine a Basic Index i such that $(i \in \text{IB}(T'))$

$$t_{q0}/t_{q\theta} = \min_{\substack{t'_{i\theta}>0 \\ i\in \text{IB}(T')}} t'_{i0}/t_{i\theta}$$

T'' = Pivoting $T'(t_{q\theta}$ = pivot)

Exchange rows of T' in such a way that the last row corresponds to θ

END

This procedure computes an initial solution of LCP. The following theorem establishes the importance of PIS.

Theorem 3.1

A solution of LCP(0) is obtained by using PIS.

Proof:

PIS is a specification of Lamke's algorithm. It solves the linear complementary problem, for a finite number of iterations, for any right hand side vector if the matrix is copositive plus. Then we need to prove that $\underline{W}'$ is copositive plus. That is to derive if the following results hold:

a) $\underline{v}^T\underline{W}'\underline{v} \geq \underline{0}, \quad \forall \quad \underline{v} \in \mathbb{R}_+^{n+3m}$

b) $\underline{v}^T\underline{W}'\underline{v} = \underline{0} \quad \Rightarrow \quad (\underline{W}'^T + \underline{W}')\underline{v} = \underline{0}$

a) is obtained by using a procedure similar to that applied in Theorem 2.1 for deriving that $\underline{W}$ is a semidefinite positive matrix.

Taking

$$\underline{W}'^T + \underline{W}' = \begin{bmatrix} \underline{0}_{nm} & \underline{0}_{nm} & \underline{0}_{n(2m)} \\ \underline{0}_{mn} & 4\underline{I}_{mm} & \underline{0}_{m(2n)} \\ \underline{0}_{(2m)n} & \underline{0}_{(2m)n} & \underline{0}_{(2m)2n} \end{bmatrix}$$

$$\underline{v}^T\underline{W}'\underline{v} = 2\underline{Ie} = \underline{0} \quad \Rightarrow \quad (\underline{W}'^T + \underline{W}')\underline{v} = \underline{0}.$$

Therefore b) is satisfied and $\underline{W}'$ is copositive plus.

In order to obtain a vertex of $\mathbb{Z}$ belonging to $\mathbb{C}$, θ is introduced in the tableau as a variable in Step 5. The coefficients is the column entering into the basis are not smaller or equal to zero, since LCP(0) has a feasible solution. Hence θ is exchanged with one of the variables of the basis. According to the rules of the SM exactly one index k^*, with $(\underline{x}_{k^*}, \underline{g}_{k^*})$, fixes the non basic variables in the obtained solution. The solution obtained by using PIS corresponds to a vertex of $\mathbb{Z} \cap \mathbb{C}$, see Bank *et al.* (1982).

For solving the parametric dependent LCP it is necessary to modify the parametric principal pivoting algorithm refered by Pang and Lee (1981), because θ and $\underline{r}_1$ has null components.

The procedure for solving parameter dependent LCP needs, as initial information, the tableau T'', the index sets IB and IS of the basic and non basic variables respectively. The following procedure solves parameter dependent LCP:

Procedure for solving Parameter Dependent LCP

Begin
    $T := T''$
    cont:=1
    EV:=$\underline{x}_{k^*}$
    $h := K^*$
    While cent ≤ 2 do
        begin
        $Q_0 = +\infty$
        Determine an index $d \in \{1, \ldots, 3m+n\}$ such that
        $Q_0 = t_{Q_0} = \min\limits_{t_{i_h}>0} \; t_{i_0}/t_{i_h}$
  If $Q_0 < \infty$ and BV$(Q) \neq \theta$ then
        begin
        T'' = Pivoting (T)
        $({}^{k+1}\underline{x}, {}^{k+1}\underline{g}, {}^{k+1}\theta)$: solution corresponding to T''

        If BV$(Q) = x_j$ then
            EV = g_j
            $h = j + n$
        If BV$(Q) = g_j$ then
            EV = x_j
            $h = j$

$T := T''$
$k = k + 1$
end

If $Q_0 = +\infty$ then
begin

$$z_h = \frac{\theta - k_\theta}{t_{3m+n,0}}$$

$BV(i) = BV(i) - t_{ih}, \quad i \in IB$
$NBV(j) = d_1, \quad j \neq h$

END

4. BEHAVIOR OF THE PROPOSED ALGORITHM

The proposed procedures were programmed in Fortran 77. The computer program uses the facilities of a standard Numerical Analysis Package (NAP) as MatLab for some calculations. Then, the accuracy of the approximations and the size of the solvable problem, depend of the characteristics of the MAP used. The code and more technical information is available from Dr. Romero. It can be implemented in an IBM 486 PC.

Dodge (1984) studied the robustness of the estimates derived by using $Q(u)$ for a fixed u. This procedure permits to compute M−type estimates of $\underline{B}$. The method proposed in this paper computes the best regression equation for a given set of values of $\theta = u/1 - u$. The values of θ reflect the ideas on the possible degree of contamination of the distribution of the residuals. For example, if it is close to a normal distribution function, with zero mean and variance σ^2, the selected model will have $\theta \approx 0$.

The proposed procedure is a generalization of the model of Dodge (1984). Its construction grants that the behavior of its solution is not worse than the use of LS or LAV as optimization criteria. Say that the estimates will not have a larger Residual Sum of Squares (RSS) and a Sum of Absolute Deviations (SAD) than any other derived by LS or LAV.

The data from the experiment of Hald (1952) used by Draper and Smith (1981) are reanalyzed. This is a classic example where LS is a good method. Table 4.1 gives the corresponding results. The optimal convex combination was obtained for $u = 0.05$. Note that observation number 6 is considered as an extreme data point by the three methods. For LAV observation number 8 is another outlier. The regression equation fitted by the algorithm has a smaller RSS than the equation derived by the use of LAV and its SAD is smaller than the corresponding to LS method. This is expected because of the theoretical properties of the methods. As the degree of contamination is small the results of the recomended model are similar to the obtained using LS. The computing time for this example was 18.24 seconds.

Table 4.1

Analysis of Hald's data

Estimated Coefficients	LS	LAV	Convex Combination
B_1	62.206	62.405	62.300
B_2	1.551	1.551	1.551
B_3	0.510	0.499	0.511
B_4	0.102	0.102	0.101
B_5	−0.144	−0.130	−0.140
Item		Residuals	
1	0.005	0.548	0.149
2	1.511	1.103	1.395
3	−1.671	−1.326	−1.685
4	−1.727	−2.040	−1.830
5	0.251	0.368	0.185
6	3.925*	4.557*	3.890*
7	−1.449	−0.741	−1.411
8	−3.175	−3.408	−3.250
9	1.375	1.671	1.366
10	0.281	0.443	0.248
11	1.991	1.990	1.946
12	0.973	1.542	0.984
13	−2.294	−1.703	−2.286
RSS	45.037	50.919	47.784
SAD	20.628	18.321	20.625

A study of the abundance of phytoplankton (Y) produced 120 measurements in the analysis of the effect of pollution in the bay of Monterrey, Mexico. The independent variables were salinity (X_1), temperature of the water (X_2), Ph (X_3), intensity of the light (X_4), $X_5 = X_1X_2$, $X_6 = X_1X_3$, $X_7 = X_1X_4$, $X_8 = X_2X_3$, $X_9 = X_2X_4$ and $X_{10} = X_3X_4$. The behavior of the three approaches, in the presence of outliers was analyzed using Monte Carlo experiments. Twelfe points of the collected data, were considered as outliers by the experts. A percent of the generated sample was selected from the set of outliers. In each Monte Carlo experiment 0, 6, 12 or 24 experiment points were generated by means of a random sampling with replacement mechanism from the outliers. The rest of the units were selected from the not outliers. One hundred samples of size fifty were generated for each percent of outliers. Table 4.2 contains the means of the calculated RSS and SAD.

Table 4.2

Average of RSS and SAD in 100 Monte Carlo experiments with the abundance of phytoplankton data for $\theta \in \{0, 0.01, \ldots, 0.99\}$

	% of outliers	Methods LS	LAV	Convex Combination
R	0	105.8	124.4	110.3
S	6	135.9	188.7	177.6
S	12	158.0	188.6	185.3
	24	274.6	198.5	179.5
S	0	75.4	51.3	80.5
A	6	93.5	53.2	66.7
D	12	137.5	53.9	66.5
	24	245.6	85.8	100.5

Note that the existence of outliers does not affect seriously the accuracy of the regressions obtained by the proposed Convex Combination of LS and LAV, in the Montecarlo experiments. The mean computing time for deriving the solutions was 46.33 seconds.

ACKNOWLEDGMENT

Contracts with CONACYT supported this research. The computation was developed at Colegio de Computación of Benmérita Universidad Autónoma de Puebla, México. The authors would like to thank helpful comments of the editor and referees which yield improvements in the quality of the original draft.

REFERENCES

[1] **Allende, S.M.** and **Bouza, C.N.** (1991). *A parametric programming approach to the estimation of the coefficients of the linear regression model.* Presented at "Parametric Optimization and Related Topics III" and "23 Jahrestagung Mathematische Optimisrung". Güstrow, Germany.

[2] **Antoch, J.** and **Barthowiak, A.** (1988). "$L-$estimation in linear models: Foundations and computational aspects. An application". *Biometrical Letters*, **25**, 3–24.

[3] **Arthanari, T.S.** and **Dodge, Y.** (1981). *Mathematical Programming in Statistis.* J. Wiley. N. York.

[4] **Bank, B., Guddat, J., Klatte, B.** and **Tammer, K.** (1982): *Nonlinear Parametric Optimization.* Akademie Verlag. Berlin.

[5] **Charnes, A., Cooper, W.W.** and **Ferguson, R.D.** (1955). "Optimal estimation of executive compensation by linear programming". *Management Sc.*, **1**, 138–151.

[6] **Dielman, T.E.** (1984). "Computational algorithms for calculating the Least Absolute Value and Chebyshov estimates for multiple regression". *American J. of Mathematical and Management Sc.*, **4**, 169–197.

[7] **Dielman, T.E.** and **Pfaffenberger, R.** (1988). "Least Absolute Value regression necessary sample size to use normal theory inference procedures". *Decision Sc.*, **19**, 734–743.

[8] **Dodge, Y.** (1984). "Robust estimation of regression coefficients by minimizing the convex combination of Least Squares and Least Absolute Deviation". *Computational Statistics Quart.*, **1**, 139–153.

[9] **Draper, N.R.** and **Smith, R.** (1981). *Applied Regression Analysis.* J. Wiley. N. York.

[10] **Efron, B.** (1978). "Computer intensive methods in statistical regression". *SIAM*, **30**, 421–449.

[11] **Hald, A.** (1952). *Statistical Theory Engineering Applications.* J. Wiley. N. York.

[12] **Lemke, C.E.** (1970). "Recent results on complementarity problems". In Rosen, J.B. *et al.* (eds.): *Non Linear programming Proceedings Symposium.* University of Wincosin, Madison. Academic Press. N. York, 349–384.

[13] **Lommatsch, K.** (1979). *Anwendung del Linearen Parametrichen Optimierung.* Akademie Verlag. Berlin.

[14] **Pang, J.S.** and **Lee, P.S.C.** (1981). "A parametric complementary technique for the computation of equilibrium prices in a single commodity spatial model". *Math. Programming*, **20**, 81–102.