Hacettepe Journal of Mathematics and Statistics
Volume 34 (2005), 95 – 102

A COMPARISON OF THREE LINEAR PROGRAMMING MODELS FOR COMPUTING LEAST ABSOLUTE VALUES ESTIMATES

E. Tuğçe Çerezci* and Fikri Gökpınar*

Received 16 : 08 : 2005 : Accepted 07 : 12 : 2005

Abstract

Several methods have been developed to estimate the regression model parameters using the Least Absolute Values method. In this study, three methods for finding Least Absolute Values Estimates developed by Charnes et al. (1995), Gonin and Money (1989) and Li (1998), are compared with respect to Central Process Unit (CPU) time and the size of the determination coefficients.

Keywords: Least Squares, Least Absolute Values, Linear Programming, Determination Coefficients, CPU Time.

1. Introduction

The Least Absolute Value (LAV) is used as an alternative to the Least Squares (LS) method for estimating regression model coefficients. This widely used method depends basically on estimating the coefficients by minimizing the absolute difference between observations and estimation values. In this context, the LAV estimation method could be modelled as a constrained optimization problem. The model will take a very long time to be solved if the number of independent variables and observations are too large. In this respect, several models have came into existence in order to reduce the CPU time. These models were developed and compared with respect to the CPU time, but they were not compared with respect to the determination coefficients of these models. To this end, the most popular three methods using the LAV method will be compared considering both the CPU time and the determination coefficients. Additionally, the estimators of these LAV methods will be compared with those of the LS method with respect to their determination coefficients.

*Gazi University, Faculty of Science and Art, Department of Statistics, Beşevler, Ankara, Turkey.

Let the linear regression model with n observations be identified as shown below;

$$y_i = \beta_0 + \sum_{i=1}^{m} x_{ij}\beta_j + \varepsilon_i,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m. \tag{1.1}$$

Let $(x_{i1}, x_{i2}, \ldots, x_{im}, y_i)$ be the ith observations and let $b_0, b_1, \ldots, b_m$, estimated by minimizing the overall absolute values of the differences between the values of $\widehat{y}$ and y, be the estimators of $\beta_0, \beta_1, \ldots, \beta_m$. So the LAV method is given as follows:

$$\min\left(\sum_{i=1}^{n} |y_i - \widehat{y}_i|\right). \tag{1.2}$$

Charnes *et. al.* [1] pointed out that this equation could be solved using a simplex method. Considering LAV, several algorithms were developed prior to 1990. Gentle *et. al.* [6] developed a LAV algorithm to be solved using the multi-regression method. The most popular three methods using the LAV method were developed by Charnes *et. al.* [2], Gonin and Money [4] and Li [5].

In this paper, the three models developed by Charnes *et. al.* [2], Gonin and Money [4] and Li [5] will be compared with respect to the CPU time, the regression parameters and alternative determination coefficients. Additionally, the estimators of these LAV methods will be compared with those of the LS method with respect to their determination and regression coefficients. In this context, in section 2, the three models will be defined and the determination coefficient to be used will be explained. In section 3, the three models will be compared under different situations using a MATLAB simulation programme.

2. Modelling Coefficient Estimation under the LAV Method

In this section, the most popular three models used to estimate the regression model parameters under the LAV method will be defined and a determination coefficient will be given.

The first model under consideration has been developed by Charnes *et. al.* [2] and Ignizo [3]. In this model, the aim is to minimize the overall absolute difference between observations and estimation values, in other words, minimizing the overall error terms. In this respect, goal programming is used to develop the linear programming model which will be used to minimize the overall total positive and negative deviations. That is, expression (2.1) is fulfilled. Most LAV regression algorithms are defined by the linear programming model developed by the goal programming techniques (1.2) as follows.

Model 1

$$\min\left(\sum_{i=1}^{n}(d_i^+ + d_i^-)\right)$$

(2.1) Restrict to;

$$y_i - (b_0 + \sum_{i=1}^{m} x_{ij}b_j + d_i^+ - d_i^-) = 0,$$
$$i = 1, 2, \ldots, n,$$
$$j = 1, 2, \ldots, m.$$

Here, d_i^+, $d_i^- \geq 0$ and b_j $(j = 1, 2, \ldots, m) \in F$, where F is the feasible set, and d_i^+, d_i^- are respectively the ith related positive and negative deviation variables.

This primary linear regression model is known to be the best model of the LAV regression algorithms, but two calculation difficulties should be mentioned.

i) Model 1 with the equality constraint can only be solved by Big-M or 2-phase methods. Considering linear programming, both methods require more time than the simplex method.
ii) Model 1 requires many deviation variables (d_i^+ and d_i^-). For n observations, $2n$ deviation variables must be used. These large data sets lead to high CPU times (Li [5]).

Gonin and Money [4] suggested an alternative LAV model where the equality of (2.1) is reformulated. Here, the number of deviation variables used is almost half of that used by model 1. The model can be identified as shown below.

Model 2

$$\min \sum_{i=1}^{n} d_i.$$

(2.2) Restrict to;

$$y_i - (b_0 + \sum_{i=1}^{m} x_{ij} b_j + d_i^+ - d_i^-) \geq 0,$$
$$i = 1, 2, \ldots, n,$$
$$j = 1, 2, \ldots, m,$$
$$y_i - (b_0 + \sum_{i=1}^{m} x_{ij} b_j + d_i^+ - d_i^-) \leq 0,$$
$$i = 1, 2, \ldots, n,$$
$$j = 1, 2, \ldots, m,$$
$$b_j \in F.$$

It is clearly seen that the number of deviation variables used in Model 1 is twice that used in Model 2.

Another model to be mentioned in this paper is the one developed by Li [5], which depends on the modified goal programming technique. Here, when developing this model, Li considered the fact that some of the observations are greater than zero and the others are less than zero.

Model 3

$$\min \sum_{i=1}^{k} \left(y_i - b_0 - \sum_{j=1}^{m} x_{ij} b_j + 2d_i \right) + \sum_{i=k+1}^{n} \left(-y_i + b_0 + \sum_{j=1}^{m} x_{ij} b_j + 2d_i \right),$$
$$i = 1, 2, \ldots, n,$$
$$j = 1, 2, \ldots, m.$$

(2.3) Restrict to;

$$b_0 + \sum_{i=1}^{m} x_{ij} b_j - d_i + s_i) = y_i,$$
$$i = 1, 2, \ldots, k,$$
$$- b_0 - \sum_{i=1}^{m} x_{ij} b_j - d_i + s_i) = -y_i,$$
$$i = k+1, k+2, \ldots, n,$$
$$d_i \geq 0, s_i \geq 0, y_i \geq 0,\ (i = 1, 2, .., k),\ y_i < 0,\ (k + i = k + 1, 2, .., n).$$

Here s_i denotes the dummy variables used to solve the $\leq$ inequality by using the simplex method to transform it into a standard form.

When Model 2 is compared with Model 3, the following results arise.

i) While Model 2 consists of $2n$ deviation variables, (d_i^+, and d_i^-), Model 3 has only n deviation variables (d_i).

ii) Model 2, which is solved by the Big-M method, takes a longer time than Model 3, which is solved by the simplex algorithm (Li, [5]).

While the advantage of Model 1 is its number of constraints and the advantage of Model 2 is its deviation variables, the advantages of Model 3 developed by Li are both the number of constraints and the deviation variables. Thus, less CPU time is used. But an evaluation of the models according to their CPU times is not enough to determine which of them is the best. It is important to measure how the dependent variables explain the determination variables. This is measured using the statistical determination coefficient. The Determination coefficient which is used in the LS method is

$$(2.4) \qquad R^2 = \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \widehat{y})^2}, \ i = 1, 2, \ldots, n.$$

However, due to the assumptions of the LS method, R^2 cannot be used in the LAV method. That is, R^2 cannot be used for the above mentioned models. Basically, the determination coefficient is the square of the correlation between the variables $\widehat{y}$ and y, and thus, instead of the equality (2.4),

$$(2.5) \qquad R^2 = r(y, \widehat{y})^2$$

could be used to compare the models in question, where $r(y, \widehat{y})$ is the correlation between $\widehat{y}$ and y (Weisberg, [7]). Regarding the LS method, since the equalities (2.4) and (2.5) are equivalent, either one of them may be used.

To this end, the models mentioned above will be evaluated with respect to both their CPU times and the determination coefficient between the variables $\widehat{y}$ and y, and accordingly we will point out the most advantageous model among them.

3. An Empirical Study

In this section, the regression coefficients of the three analyzed LAV regression methods are estimated. Using those estimations, the determination coefficient and the CPU time for each model are calculated. Additionally, the regression coefficients of those models are also estimated using the LS method in order to clearly see the difference.

In the comparison,

a uni-variable ($\beta_0 = 3, \beta_1 = 2$),

a 5-variable ($\beta_0 = 6, \beta_1 = 5, \beta_2 = 4, \beta_3 = 3, \beta_4 = 2, \beta_5 = 1$),

and a 10 variable ($\beta_0 = 6, \beta_1 = 4, \beta_2 = 3, \beta_3 = 2, \beta_4 = 1,\ \beta_5 = 1, \beta_6 = 1, \beta_7 = 2,$
$\beta_8 = 3, \beta_9 = 4, \beta_{10} = 5$)

regression model with 15, 100, and 1000 observations, respectively, were assigned using the MATLAB 7.0 simulation programme. The results are given in Tables 1, 2 and 3. The error distribution used in the three models was the standard normal distribution.

For this simulation study, a PC was used which has a Athlon XP 3200+ CPU, and 1 GB RAM.

Table 1. LAV and LS estimators, determination coefficient and CPU time for the uni-variate regression model

Model	n=15	n=100	n=1000
Model 1	CPU time=0.1094	CPU time=0.1094	CPU time=43.2656
	$b_0 = 3.3721$	$b_0 = 2.9317$	$b_0 = 3.0563$
	$b_1 = 1.8490$	$b_1 = 2.0307$	$b_1 = 1.9682$
	$R^2 = 0.7920$	$R^2 = 0.8430$	$R^2 = 0.8024$
Model 2	CPU time=0.1094	CPU time=0.5469	CPU time=86.1719
	$b_0 = 3.3721$	$b_0 = 2.9317$	$b_0 = 3.0563$
	$b_1 = 1.8490$	$b_1 = 2.0307$	$b_1 = 1.9682$
	$R^2 = 0.7920$	$R^2 = 0.8430$	$R^2 = 0.8024$
Model 3	CPU time=0.0938	CPU time=0.1250	CPU time=22.0781
	$b_0 = 2.3451$	$b_0 = 2.2088$	$b_0 = 2.4935$
	$b_1 = 2.1638$	$b_1 = 1.9904$	$b_1 = 1.9001$
	$R^2 = 0.7920$	$R^2 = 0.8430$	$R^2 = 0.8024$
LS	$b_0 = 3.6033$	$b_0 = 2.8029$	$b_0 = 3.0527$
	$b_1 = 1.7697$	$b_1 = 2.0528$	$b_1 = 1.9600$
	$R^2 = 0.7920$	$R^2 = 0.8430$	$R^2 = 0.8024$

Table 2. LAV and LS estimators, determination coefficient and CPU time for the 5-variate regression model

Model	n=15	n=100	n=1000
Model 1	CPU time=0.0313	CPU time=0.1250	CPU time=43.5000
	$b_0 = 2.5210$	$b_0 = 5.3935$	$b_0 = 5.9377$
	$b_1 = 2.2223$	$b_1 = 4.8609$	$b_1 = 4.9725$
	$b_2 = 3.8976$	$b_2 = 3.9519$	$b_2 = 4.0641$
	$b_3 = 3.0530$	$b_3 = 3.2549$	$b_3 = 3.0198$
	$b_4 = 2.4620$	$b_4 = 1.9515$	$b_4 = 1.9610$
	$b_5 = 1.1071$	$b_5 = 1.0578$	$b_5 = 1.0039$
	$R^2 = 0.9904$	$R^2 = 0.9758$	$R^2 = 0.9825$
Model 2	CPU time=0.0313	CPU time=0.6250	CPU time=87.2969
	$b_0 = 2.5210$	$b_0 = 5.3935$	$b_0 = 5.9377$
	$b_1 = 2.2223$	$b_1 = 4.8609$	$b_1 = 4.9725$
	$b_2 = 3.8976$	$b_2 = 3.9519$	$b_2 = 4.0641$
	$b_3 = 3.0530$	$b_3 = 3.2549$	$b_3 = 3.0198$
	$b_4 = 2.4620$	$b_4 = 1.9515$	$b_4 = 1.9610$
	$b_5 = 1.1071$	$b_5 = 1.0578$	$b_5 = 1.0039$
	$R^2 = 0.9904$	$R^2 = 0.9758$	$R^2 = 0.9825$
Model 3	CPU time=0.0310	CPU time=0.1875	CPU time=22.2500
	$b_0 = 2.2881$	$b_0 = 4.9002$	$b_0 = 5.4950$
	$b_1 = 5.2168$	$b_1 = 4.9405$	$b_1 = 4.9769$
	$b_2 = 3.8800$	$b_2 = 4.0426$	$b_2 = 4.0246$
	$b_3 = 3.2198$	$b_3 = 3.2886$	$b_3 = 3.0260$
	$b_4 = 2.4439$	$b_4 = 1.9253$	$b_4 = 1.9479$
	$b_5 = 0.9809$	$b_5 = 0.9688$	$b_5 = 0.9856$
	$R^2 = 0.9891$	$R^2 = 0.9756$	$R^2 = 0.9824$

Table 2 (Continued). LAV and LS estimators, determination coefficient and CPU time for the 5-variate regression model

Model	n=15	n=100	n=1000
LS	$b_0 = 2.8157$	$b_0 = 5.4497$	$b_0 = 6.0304$
	$b_1 = 4.9243$	$b_1 = 4.8832$	$b_1 = 4.9742$
	$b_2 = 4.0292$	$b_2 = 4.0819$	$b_2 = 4.0595$
	$b_3 = 3.0114$	$b_3 = 3.1320$	$b_3 = 2.9959$
	$b_4 = 2.5520$	$b_4 = 2.0432$	$b_4 = 1.9626$
	$b_5 = 1.0443$	$b_5 = 0.9823$	$b_5 = 1.0077$
	$R^2 = 0.9915$	$R^2 = 0.9768$	$R^2 = 0.9825$

Table 3. LAV and LS estimators, determination coefficient and CPU time for the 10-variate regression model

Model	n=15	n=100	n=1000
Model 1	CPU time=0.0469	CPU time=0.1406	CPU time=45.1094
	$b_0 = 16.2696$	$b_0 = 7.9841$	$b_0 = 6.7123$
	$b_1 = 5.0268$	$b_1 = 5.0952$	$b_1 = 5.0179$
	$b_2 = 3.6677$	$b_2 = 4.2496$	$b_2 = 3.9400$
	$b_3 = 2.4910$	$b_3 = 2.9671$	$b_3 = 3.0423$
	$b_4 = 1.8198$	$b_4 = 2.0000$	$b_4 = 1.9597$
	$b_5 = 1.0825$	$b_5 = 1.1824$	$b_5 = 1.0056$
	$b_6 = 05749$	$b_6 = 1.0097$	$b_6 = 0.9392$
	$b_7 = 1.8223$	$b_7 = 1.9107$	$b_7 = 2.0124$
	$b_8 = 2.6156$	$b_8 = 2.9122$	$b_8 = 3.0289$
	$b_9 = 3.9030$	$b_9 = 3.9603$	$b_9 = 3.9703$
	$b_{10} = 5.1235$	$b_{10} = 4.8153$	$b_{10} = 4.9809$
	$R^2 = 0.9966$	$R^2 = 0.9894$	$R^2 = 0.9913$
Model 2	CPU time=0.0469	CPU time=0.6875	CPU time=96.1563
	$b_0 = 16.2696$	$b_0 = 7.9841$	$b_0 = 6.7123$
	$b_1 = 5.0268$	$b_1 = 5.0952$	$b_1 = 5.0179$
	$b_2 = 3.6677$	$b_2 = 4.2496$	$b_2 = 3.9400$
	$b_3 = 2.4910$	$b_3 = 2.9671$	$b_3 = 3.0423$
	$b_4 = 1.8198$	$b_4 = 2.0000$	$b_4 = 1.9597$
	$b_5 = 1.0825$	$b_5 = 1.1824$	$b_5 = 1.0056$
	$b_6 = 05749$	$b_6 = 1.0097$	$b_6 = 0.9392$
	$b_7 = 1.8223$	$b_7 = 1.9107$	$b_7 = 2.0124$
	$b_8 = 2.6156$	$b_8 = 2.9122$	$b_8 = 3.0289$
	$b_9 = 3.9030$	$b_9 = 3.9603$	$b_9 = 3.9703$
	$b_{10} = 5.1235$	$b_{10} = 4.8153$	$b_{10} = 4.9809$
	$R^2 = 0.9966$	$R^2 = 0.9894$	$R^2 = 0.9913$

Table 3 (continued) LAV and LS estimators, determination coefficient and CPU time for the 10-variate regression model

Model	n=15	n=100	n=1000
Model 3	CPU time=0.0313	CPU time=0.2031	CPU time=23.8438
	$b_0 = 11.6454$	$b_0 = 5.4786$	$b_0 = 5.7983$
	$b_1 = 4.8151$	$b_1 = 5.1820$	$b_1 = 5.0459$
	$b_2 = 3.3773$	$b_2 = 4.3264$	$b_2 = 3.9378$
	$b_3 = 2.5964$	$b_3 = 2.9316$	$b_3 = 3.0599$
	$b_4 = 1.2866$	$b_4 = 1.9764$	$b_4 = 2.0064$
	$b_5 = 0.9002$	$b_5 = 1.0665$	$b_5 = 1.0011$
	$b_6 = 0.1587$	$b_6 = 1.1130$	$b_6 = 0.9727$
	$b_7 = 2.7779$	$b_7 = 1.9770$	$b_7 = 1.9611$
	$b_8 = 3.0930$	$b_8 = 2.8811$	$b_8 = 3.0030$
	$b_9 = 4.2475$	$b_9 = 4.0914$	$b_9 = 4.0029$
	$b_{10} = 4.7634$	$b_{10} = 4.8310$	$b_{10} = 4.9804$
	$R^2 = 0.9952$	$R^2 = 0.9890$	$R^2 = 0.9913$
LS	$b_0 = 8.8074$	$b_0 = 6.5142$	$b_0 = 6.6044$
	$b_1 = 4.9754$	$b_1 = 5.1098$	$b_1 = 5.0026$
	$b_2 = 3.7256$	$b_2 = 4.0950$	$b_2 = 3.9651$
	$b_3 = 2.6240$	$b_3 = 2.9785$	$b_3 = 3.0121$
	$b_4 = 1.7534$	$b_4 = 2.0897$	$b_4 = 1.9987$
	$b_5 = 1.0578$	$b_5 = 1.1140$	$b_5 = 1.0074$
	$b_6 = 0.8435$	$b_6 = 1.0442$	$b_6 = 0.9784$
	$b_7 = 2.1436$	$b_7 = 1.9151$	$b_7 = 1.9948$
	$b_8 = 2.9398$	$b_8 = 2.9441$	$b_8 = 3.0089$
	$b_9 = 3.9590$	$b_9 = 4.0555$	$b_9 = 3.9781$
	$b_{10} = 5.0658$	$b_{10} = 4.8485$	$b_{10} = 4.9777$
	$R^2 = 9980$	$R^2 = 0.9898$	$R^2 = 0.9913$

As shown in Tables 1, 2 and 3, the CPU times are not affected by the number of variables. Furthermore, the differences between the CPU times of Model 2 and Model 3 are about 0.02, and there is no difference between Model 2 and Model 1 when n=15. Differences between the CPU times of Model 2 and Model 3, and between Model 2 and Model 1 are about 0.52 and 0.48 second, respectively, when n=100. The CPU time of Model 2 is 2 and 4 times that of Models 1 and 2, respectively, when n=1000. On the other hand, differences between the CPU times of Model 2 and Model 3, and between Model 2 and Model 1 are about 73 and 41 seconds, respectively. This difference becomes clearly evident when the number of observations gets larger. Thus, for a larger number of observations, Model 3 is the fastest of these models. Additionally, all three models are giving close, but smaller values of the determination coefficients than the LS method.

In Table 1, the determination coefficients of all models are the same because the regression model is a univariate regression model. When comparing the three models with respect to their determination coefficients, it is clearly seen that the values given by Models 1 and 2 are much higher that that of Model 3. This is highly significant when the number of observations is small enough. And, when the number of observations gets larger, this difference diminishes. As a result, it is better to use Model 1 or Model 2 when the number of observations is small enough, and Model 3 when the number of observations is large enough to solve the LAV model.

References

[1] Charnes, A., Cooper, W. W., Henderson, A., *et. al. Introduction to Linear Programming*, (Wiley, 1953).

[2] Charnes, A., Cooper, W. W. and Ferguson, R. *Optimal estimation of executive compensation by linear programming*, Manage. Sci., **1**, 138–151, 1995.

[3] Ignizo, J. *Introduction to goal programming*, Sage, Newbury Park, California, 1985).

[4] Gonin R. and Money, A. H. *Nonlinear LP-Norm Estimation*, (Marcel Dekker, New York, 1989).

[5] Li, H. L. *Solve least absolute value regression problems using modified goal programming techniques*, Computers Op. Res. **25** (12), 1137–1143, 1998.

[6] Gentle, J. E., Narula, S. C. and Sposito, V. A. 1987,*Algorithms for unconstrained L_1 linear regression* in: Dodge, Y. (Ed.), *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (North-Holland, Amsterdam), 83–94, 1987.

[7] Weisberg, S. *Applied Linear Regression, 3rd edition*, (Wiley, New York, 2005).