International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 231-238

On Stability Properties of Multiple LAD-Regression

Yadolah Dodge and Giuseppe Melfi
Institut de Statistique
Espace de l'Europe, 4
Université de Neuchâtel
2002 Neuchâtel, Switzerland
E-mail: Yadolah.Dodge@unine.ch and Giuseppe.Melfi@unine.ch

[Received January 12, 2006; Revised June 2, 2007; Accepted September 7, 2007]

Abstract

We consider the stability of the linear multiple LAD-regression equation when two or more observations are added to the existing ones. In particular, we are interested to know whether the coefficients β's remain the same or, on the contrary, a change is produced.

Keywords and Phrases: L_1-norm, LAD regression, Robustness, Convex analysis.

AMS Classification: Primary 62J05; Secondary 62F35.

1 Introduction

We consider the following linear model:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_p X_{p,i} + \varepsilon_i,$$

where the coefficients β are determined by least absolute deviation (LAD) regression on the basis of n observations $(X_{1,i}, \ldots, X_{p,i}, Y_i)$, for $i = 1, \ldots, n$. We are interested in the changes in form or, on the contrary, the stability of the LAD-regression hyperplane, when a new observation is added to the n existing ones. Further, we investigate the case of the addition of several new observations.

This problem has been studied by Arthanari and Dodge (1993, p. 82-85). Another approach can be found in Dupačova (1992). In this paper we study this problem from a geometrical point of view and we generalize it to the case of several observations and several variables.

In order to find the equation of the LAD-regression hyperplane, we need to find the coefficients $\beta_0, \beta_1, \ldots, \beta_p$, which represent the solution of the optimization problem:

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} \mid Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \ldots - \beta_p X_{p,i} \mid . \tag{1}$$

If m new observations indexed by $n+1, \ldots, n+m$ are added, the minimization problem changes in principle. The aim of this paper is to determine some necessary and/or sufficient conditions to assure that a solution of the old optimization problem is also a solution for the new optimization problem.

2 Notations

We introduce the vector variables $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$, $\boldsymbol{X}_i = (1, X_{1,i}, \ldots, X_{p,i})$, and the functions $F_i(\boldsymbol{\beta}) = \mid Y_i - \boldsymbol{\beta}\boldsymbol{X}_i \mid$, where, as usual, the product of two vectors in $\mathbb{R}^{p+1}$ simply denotes the scalar product.
Let F be the function defined by

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{n} F_i(\boldsymbol{\beta}).$$

The optimization problem (1) is now:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} F(\boldsymbol{\beta}),$$

and F is a real-valued piecewise convex function, since made up of a sum of the n real-valued convex functions F_i on $\mathbb{R}^{p+1}$.
The functions considered in the previous section are not differentiable, but a somewhat weaker form of differentiability is suitable. In fact these functions are subdifferentiable. Consider a convex real-valued function φ:$\mathbb{R}^k \to \mathbb{R}$. We adopt the notation and the definition of subdifferentiability already used in Dodge and Roenko (1992), which is a natural generalization of the concept of differentiability.

Definition 2.1. A vector $\boldsymbol{v} \in \mathbb{R}^k$ is said to be a subgradient of φ at a point $\boldsymbol{x} \in \mathbb{R}^k$ if $\varphi(\boldsymbol{u}) \geq \varphi(\boldsymbol{x}) + \boldsymbol{v} \cdot (\boldsymbol{u} - \boldsymbol{x})$ for every $\boldsymbol{u} \in \mathbb{R}^k$.

Definition 2.2. The set of all subgradients of φ at $\boldsymbol{x}$ is called the subdifferential of φ at $\boldsymbol{x}$ and it will be denoted by $\partial\varphi(\boldsymbol{x})$.

So, for example, consider the absolute value function $\phi(x) = |x|$, $x \in \mathbb{R}$. We have

$$\partial\phi(x) = \begin{cases} -1 & \text{if } x < 0, \\ [-1, 1] & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Similarly, for the functions F_i introduced above,

$$\partial F_i(\boldsymbol{\beta}) = \begin{cases} -\boldsymbol{X}_i & \text{if } Y_i - \boldsymbol{\beta}\boldsymbol{X}_i > 0, \\ \text{Co }(-\boldsymbol{X}_i, \boldsymbol{X}_i) & \text{if } Y_i - \boldsymbol{\beta}\boldsymbol{X}_i = 0, \\ \boldsymbol{X}_i & \text{if } Y_i - \boldsymbol{\beta}\boldsymbol{X}_i < 0. \end{cases}$$

where by $\text{Co}(A)$ we mean the convex hull of A. Note that one can also define the preceding subdifferentials by the use of the sign function, so for example $\partial\phi(x) = \text{sign } x$ if $x \neq 0$, and $\partial F_i(\boldsymbol{\beta}) = -[\text{sign } (Y_i - \boldsymbol{\beta}\boldsymbol{X}_i)]\boldsymbol{X}_i$ if $Y_i - \boldsymbol{\beta}\boldsymbol{X}_i \neq 0$.
A useful property of subdifferentials is the linearity. So, for example, we have

$$\partial F(\boldsymbol{\beta}) = \sum_{i=1}^{n} \partial F_i(\boldsymbol{\beta}).$$

Through this paper we denote a vector by a bold character and an observation by a capital character, so for example a capital bold character means a vector of observations; a bold lowercase letter means a vector of variables, and so on.
Now consider the optimization problem

$$\inf_{\boldsymbol{x}\in\mathbb{R}^k} \varphi(\boldsymbol{x})$$

where φ is a convex function. If $\varphi > -\infty$ then φ reaches at least a minimum. In this case we have the following theorem (Rockafellar, 1996):

Theorem 9. *For a convex function $\varphi > -\infty$, $\boldsymbol{x}$ is a minimum if and only if $0 \in \partial\varphi(\boldsymbol{x})$.*

If m new observations indexed by $n+1, \ldots, n+m$ are added, the new optimization problem becomes

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} F^{\text{new}}(\boldsymbol{\beta}),$$

where

$$F^{\text{new}}(\boldsymbol{\beta}) = F(\boldsymbol{\beta}) + \sum_{i=1}^{m} F_{n+i}(\boldsymbol{\beta}).$$

The function F^{new} is again a convex function.
In the subsequent sections we give some necessary and sufficient conditions (or, simply, some sufficient conditions) to assure that a solution of the optimization problem for F is also a solution for the optimization problem for F^{new}.

3 Stability properties for the LAD-regression model

Suppose that $\boldsymbol{\beta}^*$ is a solution for the optimization problem $\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} F(\boldsymbol{\beta})$. Suppose a new observation $(X_{1,n+1}, \ldots, X_{p,n+1}, Y_{n+1})$ is taken into account. This corresponds to the case $m = 1$ described in Section 2. We have the following theorem, which generalizes a result of Dodge and Roenko (1992).

Theorem 10. *Let $\boldsymbol{\beta}^*$ be a solution for the optimization problem* $\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} F(\boldsymbol{\beta})$. *Let* $(X_{1,n+1},\ldots,X_{p,n+1},Y_{n+1})$ *be a new observation. Then $\boldsymbol{\beta}^*$ is a solution for the new optimization problem* $\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} F^{\text{new}}(\boldsymbol{\beta})$ *if and only if*

$$[\text{sign }(Y_{n+1}-\boldsymbol{\beta}^*\boldsymbol{X}_{n+1})]\boldsymbol{X}_{n+1}\in\partial F(\boldsymbol{\beta}^*).$$

Proof. If $Y_{n+1}=\boldsymbol{\beta}^*\boldsymbol{X}_{n+1}$, it is obvious that $\boldsymbol{\beta}^*$ remains a solution of the new minimization problem.
Suppose $Y_{n+1}>\boldsymbol{\beta}^*\boldsymbol{X}_{n+1}$. The subdifferential of F^{new} at the point $\boldsymbol{\beta}^*$ is

$$\partial F^{\text{new}}(\boldsymbol{\beta}^*)=\partial F(\boldsymbol{\beta}^*)-\boldsymbol{X}_{n+1}.$$

So by Theorem 9, the point $\boldsymbol{\beta}^*$ remains a solution of the new minimization problem if and only if $0\in\partial F(\boldsymbol{\beta}^*)-\boldsymbol{X}_{n+1}$, i.e. $\boldsymbol{X}_{n+1}\in\partial F(\boldsymbol{\beta}^*)$.
The case $Y_{n+1}<\boldsymbol{\beta}^*\boldsymbol{X}_{n+1}$ is similar.
For the case $p=1$, the above theorem gives Theorem 3 of Dodge and Roenko (1992).
Note that $\partial F(\boldsymbol{\beta}^*)$ is convex and it has dimension $p+1$, because there are at least $p+1$ points which determine the hyperplane $Y=\boldsymbol{\beta}^*\boldsymbol{X}$ (see for example Arthanari and Dodge, 1993).
Suppose now that several new observations are added, so now $m>1$. It is possible to give a necessary and sufficient condition in order that a solution $\boldsymbol{\beta}^*$ for the minimization problem for F remains a solution for F^{new}. We now generalize the previous theorem for the case of $n+m$ observations.

Theorem 11. *Let $\boldsymbol{\beta}^*$ be a solution for the optimization problem* $\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} F(\boldsymbol{\beta})$. *Let* $(X_{1,n+1},\ldots,X_{p,n+1},Y_{n+1})$, ..., $(X_{1,n+m},\ldots,X_{p,n+m},Y_{n+m})$ *be m new observations. Then $\boldsymbol{\beta}^*$ is a solution for the new optimization problem*

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} F^{\text{new}}(\boldsymbol{\beta})$$

if and only if

$$\sum_{i=1}^{m}[\text{sign }(Y_{n+i}-\boldsymbol{\beta}^*\boldsymbol{X}_{n+i})]\boldsymbol{X}_{n+i}\in\partial F(\boldsymbol{\beta}^*)+\sum_{Y_{n+i}=\boldsymbol{\beta}^*\boldsymbol{X}_{n+i}}\text{Co}(-\boldsymbol{X}_{n+i},\boldsymbol{X}_{n+i}). \tag{2}$$

Proof. This is an application of Theorem 9. In fact the subdifferential of F^{new} is

$$\partial F(\boldsymbol{\beta}^*)-\sum_{i=1}^{m}[\text{sign }(Y_{n+i}-\boldsymbol{\beta}^*\boldsymbol{X}_{n+i})]\boldsymbol{X}_{n+i}+\sum_{Y_{n+i}=\boldsymbol{\beta}^*\boldsymbol{X}_{n+i}}\text{Co}(-\boldsymbol{X}_{n+i},\boldsymbol{X}_{n+i}).$$

The above theorem is suitable for a massive computation, as for cases which one can find in real applications.
Note that ∂F^{new} is convex in $\mathbb{R}^{p+1}$, and actually is the convex hull of a set of points in $\mathbb{R}^{p+1}$. The left side of (2) represents a point in $\mathbb{R}^{p+1}$. On the right side, we have a convex set: the condition to assure that the model remains unchanged is equivalent to the membership of a certain point to a certain convex set.

4 How to determine the membership to a convex hull of a set of points

Let $\boldsymbol{W} \in \mathbb{R}^{p+1}$ and $S = \{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n\}$ be a set of points in $\mathbb{R}^{p+1}$.
The problem of determining whether $\boldsymbol{W}$ belongs to $\mathrm{Co}(S)$ is a classic problem in linear programming (see, for example, Yao, 1981).
A standard method to check whether $\boldsymbol{W}$ belongs to $\mathrm{Co}(S)$ uses a linear programming technique. The problem can be formulated as follows:

$$\begin{cases} \text{find} & \boldsymbol{W}, \\ \text{satisfying} & \boldsymbol{W} = \sum_{i=1}^{n} \lambda_i \boldsymbol{W}_i, \\ & \sum_{i=1}^{n} \lambda_i = 1, \\ & \lambda_i \geq 0. \end{cases}$$

This problem is often called a linear feasibility problem, and has a solution if and only if the following has no solution:

$$\begin{cases} \text{find} & z_0 \in \mathbb{R} \text{ and } \boldsymbol{z} \in \mathbb{R}^{p+1}, \\ \text{satisfying} & \boldsymbol{z}\boldsymbol{W}_i \leq z_0 \text{ for all } i = 1, \ldots, n, \\ & \boldsymbol{z}\boldsymbol{W} > z_0. \end{cases}$$

Geometrically, the meaning of this problem is simple. If it admits a solution $(z_0, \boldsymbol{z})$, then the set $\mathbb{H} = \{\boldsymbol{x} \in \mathbb{R}^{p+1}, \boldsymbol{z}\boldsymbol{x} = z_0\}$ is a hyperplane separating the polytope $\mathrm{Co}(S)$ from the inquiry point $\boldsymbol{W}$. Thus the existence of the separation means the nonredundancy. Now, to actually solve the problem, we set up the linear programming problem:

$$\begin{cases} \text{maximize} & \boldsymbol{z}\boldsymbol{W} - z_0, \\ \text{subject to} & \boldsymbol{z}\boldsymbol{W}_i - z_0 \leq 0 \text{ for all } i = 1, \ldots, n, \\ & \boldsymbol{z}\boldsymbol{W} - z_0 \leq 1. \end{cases}$$

The last inequality is artificially added to allow the linear programming problem to have a bounded solution. It is easy to see that the point is non-redundant if and only if the optimal value of the linear programming problem is strictly positive.
There is another approach to this problem. We have $\boldsymbol{W} \in \mathrm{Co}(S)$ if and only if $\mathrm{Co}(\boldsymbol{W}, S) = \mathrm{Co}(S)$. So every convex hull of a set of points is univocally determined by a minimal subset of points. For example, Mathematica command `ConvexHull` (of a set of points) gives the minimal set of points needed to determine the convex hull.

5 An example: Mayer's data

Consider the following data, derived from observations of the lunar crater Manilius (Mayer, 1750 and Stigler, 1986). Mayer studied the libration of the moon by observing the position of the crater Manilius as seen from the Earth. He found a linear relationship between certain mensurations and some location parameters of Manilius at moon's pole. He obtained a system of 27 linear equations in 3 unknown parameters (see Table 1), and proposed a method for determining these parameters, by selecting in a suitable manner some special observations in order to determine the best linear model. This method was subsequently improved by Boscovitch with the introduction of the least absolute deviation method.

Obs.	X_1	X_2	Y
1	−0.8836	0.4682	13.17
2	−0.9996	0.0282	13.13
3	−0.9899	−0.1421	13.20
4	−0.2221	−0.9750	14.25
5	−0.0006	−1.0000	14.70
6	−0.9308	0.3654	13.02
7	−0.0602	−0.9982	14.52
8	0.1570	−0.9876	14.95
9	−0.9097	0.4152	13.08
10	−1.0000	−0.0055	13.03
11	−0.9689	−0.2476	13.20
12	−0.8878	−0.4602	13.18
13	−0.7549	−0.6558	13.57
14	−0.5755	−0.8178	13.88
15	−0.3608	−0.9326	13.97
16	−0.1302	−0.9915	14.23
17	0.1068	−0.9943	14.93
18	0.3363	−0.9418	14.78
19	0.8560	−0.5170	15.93
20	−0.8002	−0.5997	13.48
21	0.9952	0.0982	15.92
22	0.8409	−0.5412	15.65
23	0.9429	−0.3330	16.15
24	0.9768	−0.2141	16.37
25	0.6262	0.7797	15.63
26	0.4091	0.9125	14.90
27	−0.9284	0.3716	13.12

Table 1: The libration of the moon. Mayer's data

The LAD-regression plane is

$$Y = 14.5676 + 1.51108X_1 - 0.133306X_2. \tag{3}$$

The plan of equation (3) passes through observations 1, 5 and 19.

In this case $\boldsymbol{\beta}^* = (14.5676, 1.51108, -0.133306)$ and since

$$\sum_{i=1}^{27}[\text{sign } (Y_{n+i} - \boldsymbol{\beta}^*\boldsymbol{X})]\boldsymbol{X}_{n+i} = (0, -1.2876, 0.2422)$$

we have:

$$\begin{aligned}\partial F(\boldsymbol{\beta}^*) = \quad & (0, -1.2876, 0.2422) \\ & +\text{Co}(-1, 0.8836, -0.4682), (1, -0.8836, 0.4682)) \\ & +\text{Co}(-1, 0.0006, 1), (1, -0.0006, -1)) \\ & +\text{Co}(-1, -0.8560, 0.5170), (1, 0.8560, -0.5170))\end{aligned}$$

This means that for Theorem 10 a new observation $(\boldsymbol{X}, Y)$ would not change the model, if sign $(Y - \boldsymbol{\beta}\boldsymbol{X})\boldsymbol{X}$ lies on $\partial F(\boldsymbol{\beta}^*)$, which represents a subset of $\mathbb{R}^3$ describing a parallelepiped spanned by the three vectors $(1, -0.8836, 0.4682)$, $(1, -0.0006, -1)$, $(1, 0.8560, -0.5170)$ and centered in $(0, -1.2876, 0.2422)$. Analogously, for Theorem 11, m further observations $(\boldsymbol{X}_1^*, Y_1^*)$, $\ldots$, $(\boldsymbol{X}_m^*, Y_m^*)$ would not change the model if a suitable algebraic sum of $\boldsymbol{X}_i$ (namely, $\boldsymbol{X}_i$ with the sign of $Y_i - \boldsymbol{\beta}\boldsymbol{X}_i$) lies on $\partial F(\boldsymbol{\beta}^*)$.

6 Conclusion

The interest of Theorem 11 resides on the fact that a set of new observations may be uninfluent for the determination of the LAD-regression model. Furthermore, its application shows a geometrical property. The convex region defined by (2) is a $(p+1)$ dimensional convex 'bag', containing some representative points, and whose meaning is the 'no change model region' in case of further adds of observations.

Finally we note that a further interpretation of the 'no change model region' is in terms of matrix determinant. Note that in its general form, i.e., when there are no redundant observations in the determination of the LAD-regression model, the region is determined by a translation of a parallelepiped spanned by $p+1$ vectors on $\mathbb{R}^{p+1}$ and the determinant of the corresponding matrix gives the volume, and then the extension, of the 'no change model region'.

References

[1] Arthanari, T.S.; Dodge, Y. (1993): Mathematical Programming in Statistics. Classic edition, *Wiley*, New York.

[2] Dodge, Y.; Roenko, N. (1992): Stability of L_1-norm regression under additional observations. *Computat. Statist. & Data Anal.* **14,** 385-390.

[3] Dupačova, J. (1992): Robustness of L_1 regression in the light of linear programming. In L_1-Statistical analysis and related methods, Y. Dodge Ed., North-Holland, 47-61.

[4] Mayer, J.T. (1750): Abhandlung über die Umwälzung des Monds um seine Axe. *Kosmographische Nachrichten und Sammlungen* **1,** 52-183.

[5] Rockafellar, R.T. (1996): Convex Analysis, *Princeton University Press*, USA.

[6] Stigler, S.M. (1986): The History of Statistics, *Belknap, Harvard*, USA.

[7] Yao, A.C.-C. (1981): A Lower Bound to Finding Convex Hulls. J. Assoc. Comput. Machin. 28, 780-787.