The initial location at the University of Vaasa, Finland


This paper is reproduced at the University of Vaasa in the electronic format with the permission of The Finnish Journal of Business Economics. Copyright © 1994 by The Finnish Journal of Business Economics and the authors.

http://lta.hse.fi/

http://lipas.uwasa.fi/~sjp/
http://www.uwasa.fi/laskentatoimi/english/personnel/salmitimo/

The LAD estimation program for Pynnönen and Salmi (1994) [online].
http://garbo.uwasa.fi/pub/pc/ts/ts5st17.zip
1-1994

43. VUOSIKERTA • ÅRGÅNG 43 • VOLUME 43.

Erkki K. Laitinen
Markov Process in Business Failure Prediction  p. 3

Annareetta Lumme – Ilkka Kauranen – Erkko Autio
The Growth and Funding Mechanisms of New, Technology-Based Companies
A Comparative Study between the United Kingdom and Finland  p. 20

Seppo Pynnönen – Timo Salmi
A Report on Least Absolute Deviation Regression with Ordinary Linear Programming  p. 36

Jouko Havas
Is it style or skill that makes a manager?  p. 50

Vesa Kanniainen – Hannu Hernesniemi
Verouudistuksen 1993 vaikutukset Helsingin Arvopaperipörssissä  s. 69
(Summary on page 81)

Petri Vehmanen
Laskentatoimi – ympäristönsä vanki ja muokkaaja  s. 82
(Summary on page 93)

Teija Laitinen
Eri voittokäsitteiden informaatiolisättö konkurssin ennustamisessa  s. 96
(Summary on page 101)

E. Olavi Salminen
Muutosjohtajan ulkoinen ja sisäinen menestyminen  s. 102
(Summary on page 106)

Toimituskunta
JAAKKO HONKO, PERTTI KETTUNEN, ALF-ERIK LERVIKS,
Redaktionsråd
CAJ-GUNNAR LINDBRÖM, VESA MÄKINEN, KALEVI PIHA,
Board of Editors
REUO RUUHELA, FEDI VAIVIO.
Toimitus
JAAKKO HONKO (påtoimittaja — huvudred. — editor)
Redaktion
MIKA KASKIMIES (toim.siht. — red.sekr. — secretary)
Editors
The Finnish Journal of Business Economics
Address: Runeberginkatu 14—16, 00100 Helsinki, Finland

ISSN 0024-3469
A Report on Least Absolute Deviation Regression with Ordinary Linear Programming

ABSTRACT

In this paper it is our purpose to revisit the least absolute deviation estimation in regression analysis, consider some of its theoretical properties, and consider its implementation from a computational mathematical programming point of view. We also consider goodness of fit statistics as well as approximate distributions of the associate test statistics for the parameters. Furthermore, we suggest a new goodness of fit statistic, called the LAD-coefficient of determination, that is adapted to the metrics used in LAD-estimation. Finally, some examples are provided to illustrate the behavior of the procedures in data that include outliers.

1. INTRODUCTION

Robust estimation refers to the ability of a procedure to produce highly insensitive estimates to model misspecifications. Hence, robust estimates should be »good» under wide range of possible data generating distributions. In the regression context, under normality with identically and independently distributed errors, the least squares is the most efficient among the unbiased estimation methods. However, when one gives up the normality assumption it is frequently possible to find estimation methods that are more efficient than the traditional least squares. Specifically this is true when the data generating process has fat tails resulting to several outliers compared to the normal distribution. In these cases the least squares becomes highly unstable and sample dependent because of the quadratic weighting, which makes the procedure very sensitive to outlying observations.

For example in finance and accounting research the idea is not at all uncommon that the underlying distributions could have infinite variances. For example, it has been for long observed that speculative price series tend to have volatility clusterings resulting to kurtic and thick tailed unconditional distributions [see e.g., Mandelbrot (1967), Fama and Roll (1968)]. If so then the least squares approach becomes
totally inappropriate, for it minimizes the squared deviations that heavily weight the outlying observations, typical to thick tailed distributions. A Finnish example of the application of the LAD estimation for growth estimation in long-run IRR assessment is Luoma (1983). Also Luoma and Pynnönen (1993) have found the LAD method useful in certain applications of firm’s steady state growth estimation.

The possibility of non-normal distributions, and infinite variance in particular, has led to development of alternative estimation methods to the least squares. Provided that one knows the generating distribution, a well established procedure is the method of maximum likelihood, which has several optimal properties. However, this method strongly relies on the knowledge of the distributional form, and hence by construction is not necessarily a robust method, except for the case where the underlying distribution itself is robust.

A large number of other estimation methods aimed at achieving robustness have been suggested and a considerable body of literature has developed. See for example, Gonin and Money (1989), Dodge (1987) and the references therein.

Generally the robust estimators in the literature can be classified as $M$-estimators, $L$-estimators, or $R$-estimators. Probably most attention has been paid to the $L$-estimators. For other type estimators, see e.g. Judge et al. (1985).

In order to briefly introduce to the general $L$-estimators, consider the following regression model

\[(1.1) \quad y_i = x_i^T \beta + \varepsilon_i,\]

where $y_i$ is the dependent variable, $x_i = (1, x_1, \ldots, x_p)'$ is a $(p + 1)$-vector of explanatory variables, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$ is the regression coefficient vector, $\varepsilon_i$ are uncorrelated constant variance disturbance terms, $i = 1, \ldots, n$, with $n$ being the sample size, and the prime denotes transpose. Then, given a sample of observations, the general minimization problem with $L$-estimators is

\[(1.2) \quad \min_{\beta} \left[ \sum_{\{i| y_i \geq x_i^T \beta\} \theta |y_i - x_i^T \beta| + \sum_{\{i| y_i < x_i^T \beta\} (1-\theta) |y_i - x_i^T \beta| \right], \]

where $\theta$ is called the $\theta$th regression quantile, $0 < \theta < 1$.

Selecting in (1.2) $\theta = 1/2$ gives perhaps the best known $L$-estimator, called the least absolute deviation (LAD). This method, as the name suggests, minimizes not the sum of squared deviations but the sum of absolute values of the deviations. Consequently it does not put excessive weight on highly deviating observations like the least squares does, and hence produces more robust estimators with respect to outliers.

In this paper it is our purpose to revisit the least absolute deviation estimation in regression analysis, consider some of its theoretical properties, and consider its implementation from a computational mathematical programming point of view. We also consider goodness of fit statistics as well as approximate distributions of
the associate test statistics for the parameters. Furthermore, we suggest a new goodness
of fit statistic, called the LAD-coefficient of determination, that is adapted to
the metrics used in LAD-estimation. Finally some examples are provided to illustrate the behavior of the procedures in data that include outliers.

2. LEAST ABSOLUTE DEVIATION REGRESSION

As noted in the previous chapter, the LAD estimator of the parameter vector \( \beta \)
in (1.1) is the one that minimizes

\[
(2.1) \quad \sum_{i=1}^{n} |y_i - x_i^T \beta|
\]

in contrast to the usual least squares, where the sum of squares is minimized.

The problem here is two-fold, on one hand computational and on the other statistical. The traditional solution on the computational side is to transform the problem into a linear programming (LP) context. Another popular approach is to turn the problem into a generalized (iterative) least squares estimation problem [see Sec. 4.2]. Recently Soliman and Christensen (1989) suggested another at least potentially very interesting solution that does not require a heavy LP-algorithm. The LP-algorithm is considered more closely in Chapter 3. First we shall, however, focus on the statistical aspects.

2.1 Goodness of fit

From the user’s view point an important question is how to evaluate the successfullness of the estimated model. In the least squares context the most used measure of goodness of fit is the coefficient of determination, \( R^2 \). It measures the closeness of the fit to the observed values. In the least squares sense the fit is perfect if \( R^2 = 1 \). The other end of the range is zero, meaning no explanatory power of the regressors. The coefficient of determination can be defined as the square of the correlation between observed and fitted values.

Using the correlation definition, \( R = \text{corr}(\hat{y}_{LAD}, y) \), where \( \hat{y}_{LAD} \) is the least absolute value fit, \( R^2 \) can be used also in the LAD estimation as a measure of coefficient of determination. However, in this context some arbitrariness is involved with \( R^2 \), at least for two partially interrelated reasons. First, if we look at the LAD estimation from the robustness point of view then using \( R^2 \) is against the logic because it is not robust, i.e., it is sensitive to outlying observations. The second reason is a direct consequence of this non-robustness. It is easy to construct examples of, say, two data sets otherwise similar except that one is contaminated with some outliers. The same model may still fit equally well in the LAD sense (using e.g. significance of regression coefficients as a measure), but does show almost nil \( R^2 \) with the outlier data. We shall demonstrate this drawback by an example in Chapter 5.
We suggest to alleviate the arbitrary nature of $R^2$ in the LAD estimation by defining the coefficient of determination rather as

$$G = 1 - \frac{\sum_{i=1}^{n} |y_i - x_i^T \hat{\beta}|}{\sum_{i=1}^{n} |y_i - \text{md}_y|}.$$  

(2.2)

In the definition $\hat{\beta}$ is the LAD estimator of the coefficient vector, $\beta$, and $\text{md}_y$ is the median of $y$. Like $R^2$, (2.2) assumes also values between zero and one, the former indicating no regression and the latter a perfect fit. These properties are due to the following relation,

$$\min_{\beta} \sum_{i=1}^{n} |y_i - x_i^T \beta| \leq \sum_{i=1}^{n} |y_i - \text{md}_y|,$$

(2.3)

where the equality holds if $x_i^T \beta = \text{md}_y$ for all $i = 1, \ldots, n$. It may be noted that (2.2) is an analog to the $R^2$, because

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - x_i^T \hat{\beta})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$  

(2.4)

2.2 Statistical inference on the regression coefficients

Let $X$ denote an $n \times (p + 1)$ matrix with row vectors $x_i = (1, x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$. Assume further that $n > p + 1$ and that $X$ is of full rank (i.e., rank $(X) = p + 1$), which guarantees the existence of the inverse $(X'X)^{-1}$. Then under fairly general conditions $\sqrt{n-1} (\hat{\beta} - \beta)$ is asymptotically normally distributed with zero mean and covariance matrix $\lambda^2 (X'X)^{-1}$, where $\lambda^2/n$ is the variance of the sample median of the residuals. Consequently individual coefficients, $\hat{\beta}_j$ are asymptotically normal with mean $\beta_j$ and variance $\lambda^2 (X'X)^{ij}$, where $(X'X)^{ij}$ denotes the $j^{th}$ diagonal element of the inverse of $X'X$, $j = 0, 1, \ldots, p$.

The problem is that the parameter $\lambda$ is unknown. However, already Cramér (1946; p. 369) has shown that asymptotically

$$\lambda = \frac{1}{2f(\text{md})},$$

(2.5)
where \( f(\text{md}) \) is the value of the error density function of the residuals, \( \epsilon \), at the median, \( \text{md} \). A consistent estimate for \( f(\text{md}) \) is

\[
\hat{f}(\text{md}) = \frac{2d}{n (e_{(m+d)} - e_{(m-d)})},
\]

where \( e_{(i)} \) denotes the ordered residuals, \( e_i = y_i - x_i' \beta \), \( m = n/2 \) is the median point, and \( d \) is an integer to be chosen. The best choice of \( d \) is not clear, for it depends on the smoothness of the empirical density function and the number of observation. Cox and Hinkley (1974), however, suggest that \( d \) should be kept small. Furthermore, since \( p + 1 \) of the residuals become zero by definition, it may be convenient to exclude these observations, so that the effective number of observations is \( \tilde{n} = n - p - 1 \). Consequently \( m \) is redefined as

\[
m = \frac{\tilde{n} + 1}{2} \quad \text{for } \tilde{n} \text{ odd, and}
\]

\[
= \frac{\tilde{n}}{2} + 1 \quad \text{for } \tilde{n} \text{ even},
\]

[cf. Gonin and Money (1989; p. 16)]. Integer \( d \) could be selected equal to one or two. For example, in the SHAZAM econometrics software package the default value of \( d \) is, however, set equal to integer part of \( n/6 \). In large samples this may become too big, hence, we suggest a compromise such that

\[
d = \max \left( 1, \min (4, \lfloor \tilde{n}/6 \rfloor) \right).
\]

where \( \lfloor a \rfloor \) denotes the integer part of a real number \( a \).

2.3 Properties of the LAD Estimators

It may be noted that the LAD estimator of \( \beta \) equals the maximum likelihood (ML) estimator if the disturbances follow a double exponential distribution with parameter \( \lambda > 0 \), such that

\[
f(\epsilon_i) = \frac{1}{2\lambda} \exp \left( -\frac{|\epsilon_i|}{\lambda} \right).
\]

Like the normal distribution, \( g(x) = (2\pi\sigma^2)^{-1/2} \exp \left\{ - \frac{(x - \mu)^2}{2\sigma^2} \right\} \), also distribution (2.9) is symmetric, but its kurtosis is 6, while it in the case of normality is 3. Hence, the double exponential distribution is more peaked than the normal distribution. Consequently the tails are fatter, resulting to outlier-prone samples.

Although model (2.9) is well defined, small sample properties of the LAD estimators are difficult to derive. However, general properties of the ML-estimators guarantee that the regression estimators are asymptotically efficient under the distribution (2.9). This means that in large samples the estimators are unbiased and
have smallest variances in the class of unbiased estimators. It may, however, be noted that unbiasedness does not necessarily imply the consistency of an estimator. A necessary condition for the consistency of the LAD regression estimators under general distribution of the error terms have been derived by Chen and Wu (1993).

Because of mathematical difficulties there are fairly few known finite sample properties of the LAD regression estimators. One result, however, is that, if the solution of the ensuing LP-problem (see Ch. 3) yields a unique solution, then the resulting estimator is unbiased [Taylor (1974)]. This holds also for more general error distributions than just (2.9), provided that the distribution is symmetrical about zero [see Taylor (1974)]. Moreover, in the case of multiple solutions, but symmetrical error distribution, unbiasedness can be gained by applying the algorithm suggested by Sielken and Hartley (1973). It may be noted that this result holds even when the variance of the error distribution is infinite. The variance of (2.9) is, however, $\lambda^2$, and hence finite.

Smith and Hall (1972) have confirmed by Monte Carlo simulation study that the LAD regression estimator is superior to OLS estimator in small samples under the error model (2.9). Hence superiority of LAD over OLS does not necessarily require an infinite variance, which theoretically makes OLS useless. An extensive review of Monte Carlo studies performed for investigating small sample properties of the LAD regression estimators are reported in Dielman and Pfaffenger (1982). The main results of the studies are that:

1. The LAD estimator have significantly smaller variances than the OLS estimator for a regression with high kurtosis.
2. The LAD estimators are closely normally distributed in the high kurtosis disturbances.

The second property suggests that the normal theory inference can be fairly safely applied in moderate sample sizes. In large samples the normal theory inference is fully applicable, because the important results due to Basset and Koenker (1978) show that the sampling distribution of the LAD regression estimators are asymptotically normal. These results hold under the assumptions that the error terms are independent and identically distributed, and hence is not necessary to assume the finiteness of the variance! As a consequence this result implies that for any error distribution the LAD estimators should be superior to OLS estimators if the median of the distribution is superior to the mean as an estimator of the location.

3. COMPUTATIONAL ASPECTS

3.1 Computer Aspects Outline

In this chapter we turn to the question of LAD estimation on computers. In particular we use »IBM compatible» personal computers (PCs) for this purpose. Here we shift our emphasis to the practitioner's point of view, because much of what fol-
ows is based on the authors’ computational experience with PCs, and with solving linear programming and linear goal programming problems.

The reason why we select PCs as our computational tools is prompted by their prevalence. There are no reliable figures in existence, but some estimates of the number of MsDos based systems go even as high as to a hundred million PCs worldwide. Even with the recent drastic developments in the speed in the PCs, with the advent of the 486 processors (and beyond), and fast capacity hard disks, our solutions will conform even to the 8086 XT standards. In technical terms we adhere to real mode MsDos programming with the 640Kb conventional memory limit.

We have developed a user-friendly computer program based on linear programming to solve LAD estimation problems with up to 25 explaining variables and 100 observations. This $25 \times 100$ limit is a consequence of the conventional memory limit. The capacity could be increased with a use of virtual techniques (arrays extending to the hard disk or RAM disk), but the current capacity is sufficient for the purposes of this paper.

Using linear programming and PCs is natural also because there has been a tendency to PC based linear programming packages from the mid-1980’s. See for example Jennergren (1985), Dowsland (1987), Stadler & Groeneveld & Hermannsen (1988), Ashford & Daniel (1988), Llewellyn & Sharda (1990), Orchard-Hays (1990). At the same time, according to our own experience, there has been steady development towards more user-friendly interfaces both in statistical and operations research (and more narrowly linear programming) computer packages.

In actual computer practice the LAD estimation procedures are prone to breakdowns because of round-offs aggravated by degeneracy problems characteristic in LAD estimation. This dilemma can occur even in well-established fully commercial statistical packages. This may be an indication of insufficient consideration for the numerical aspects in writing the computer code. To tackle this problem, we employ our own version of the round-off tolerance to reset small absolute values to zero. Per se, this is not a novel idea. It has been suggested e.g. in Orchard-Hays (1968).

In mathematical terms an accurate solution (barring the extremely rare cyclical problems) is always guaranteed for a correctly formulated LAD problem. But in a computer programming environment this need not be the case. Consequently, we also tackle the problem of the reliability of the results by incorporating our own version of a solution accuracy measure. In connection with linear programming the idea of the numerical accuracy confirmation of the solution is fairly common, and can occur in several alternative formats.

3.2 LAD Estimation Defined as a Linear Programming Problem

The LAD estimation problem was presented earlier in this paper in vector notation. To recount, in an ordinary index format better suitable for computer pro-
gramming, we can express (omitting the residual) the least absolute deviation (LAD) multiple regression analysis task as finding the estimates for the coefficients of

\begin{equation}
Y = a + b_1 X_1 + \ldots + b_p X_p.
\end{equation}

Let \( x(j,i) \) be the \( j \)th observation of variable \( X(i) \). Then with \( n \) observations the LAD task leads to the following linear programming (LP) problem

\begin{equation}
\min \sum_{j=1}^{n} (P(j) + N(j))
\end{equation}

subject to

\begin{equation}
a + \sum_{i=1}^{p} x(j,i) b(i) + P(j) - N(j) = y(j) \quad (j = 1, \ldots, n)
\end{equation}

\begin{equation}
P(j) \geq 0, N(j) \geq 0 \quad (j = 1, \ldots, n),
\end{equation}

In the above the \( P(j) \) denote positive deviations while the \( N(j) \) denote the absolute values of the negative deviations between the observed values and the fitted values (i.e. they are the residuals), \( x(j,i) \) denote the explaining variables, and \( y(j) \) dependent variables.

This problem is basically an ordinary linear programming. The parameters of the LAD estimation problem are, in turn, variables of the ensuing linear programming problem. These LP-variables, the intercept, \( a \), and the coefficients, \( b(i) \), are unrestricted in sign.

The customary substitution in a linear programming problem with variables unrestricted in sign is (c.f. Hadley (1972, p. 169) and Gass (1975, p. 317))

\begin{equation}
a = a' - a'', \quad a' \geq 0 \text{ and } a'' \geq 0
\end{equation}

\begin{equation}
b(i) = b'(i) - b''(i) \quad (i = 1, \ldots, p)
\end{equation}

where \( b'(i) \geq 0 \text{ and } b''(i) \geq 0 \).

As a sideline the following can be observed. The above problem is an ordinary linear programming task. The format, however, could also be regarded as a linear goal programming problem with a single priority level. (See, for example, Ignizio (1976; p. 31), and Schniederjans (1984; pp. 67-71).) The basic formulation of a goal programming problem involves minimization of the deviational variables, exactly as in (3.2)-(3.4). If the estimation task would warrant it, it would be relatively easy to assign different priorities to the observations.
3.3 Numerical Analysis Aspects

The substitution for the unrestricted variables (3.5)-(3.6), can cause crippling problems in solving the ensuing linear programming problem with a computer because of the resultant columnar linear dependence. Mathematically, it can be shown that the positive and negative parts of a redefined variable will not appear in the optimum solution simultaneously. However, in linear programming computer codes, which naturally use a finite accuracy (usually eight decimals for single precision or fifteen for double precision), this no longer holds.

It is possible that the computer solutions for simplex-algorithm will break down because of the round-off errors if measures like rounding small absolute values to zero are not taken. From practical experience we know that the likelihood of a break-down is drastically increased by columnar dependencies. This likelihood is even greater when the coefficients of the objective function are equal. For the LP methods of LAD estimation it is easily seen from the above definitions (3.1)-(3.6) that multiple columnar dependencies and the equality of the coefficients of the objective function are evident.

Our single-precision computer code is based on the ordinary simplex-algorithm (see for example Gass (1975), or almost any standard text-book of operations research with a linear programming chapter). We apply a tolerance factor of 0.000005. At each iteration all the elements of the simplex-tableau are checked. If an absolute value is below the tolerance it is rounded off to zero. The tolerance factor can be changed by the user. The initial tolerance value is based on our computational experience. This avoids with good probability the round-off break-down problems present even in some fully commercial software packages.

The second numerical aspect we tackle is the reliability of the optimum solution by applying our own version of the accuracy index. It is an obvious mathematical corollary of the simplex method for linear programming that the elements of the optimum (in fact any) simplex tableau can be also obtained by multiplying the respective column of the original tableau with the current inverse matrix of the base in the optimum (or any) tableau. In the software applications of LP this fact can be utilized to establish the numerical reliability of the solution obtained. We define and calculate the inaccuracy of the LP solution as follows. Denote

\[ b(i) = \text{The regression estimates from the optimum LP solution.} \]
\[ h(i) = \text{The recalculated regression estimates.} \]

The regression estimates are the optimum values of the \( b(i) \) as obtained from solving the LP task (3.2)-(3.4). The recalculated values, \( h(i) \), are obtained by multiplying the right-hand sides of the first simplex tableau by the inverse matrix of the base. A fairly common measure of the inaccuracy is

\[ (3.7) \quad d = \max_i |b(i) - h(i)|. \]
This definition, however, has the disadvantage of losing some of the information. To avoid a loss of information, we calculate the inaccuracy instead as the norm

\[
(3.8) \quad d = \sqrt{\sum_{i=1}^{p} (b(i) - h(i))^2}.
\]

If \(d\) exceeds one, we consider the solution unreliable. This criterium is based on our computational experience.

3.4 Of the Software

The detailed instructions for using the actual software package are outside the scope of the present paper. However, the instructions and the software are readily available as MsDos shareware. See Salmi \& Pynnönen (1992).

The main features of the software package are the following. The program is interactive, asking for the input in a logical order. Help can be obtained for each interactive question. The input can be taken from the keyboard or from a file. The data and the solutions can also be presented in a graphics (CGA and upwards) format. The capacity of the program is 25 variables and 100 observations, which leads to a linear programming problem with 250 variables and 100 constraints. On a 486/50 Mhz PC a typical \(25 \times 100\) sample program required 105 simplex iterations and took 37 seconds to solve. This demonstrates that a relatively heavy computational load is involved in LAD estimation. In general terms, if the LAD estimation involves \(p\) explaining variables and \(n\) observations, the resultant linear programming problem will contain \(2(1+p+n)\) variables and \(n\) constraints.

4. EXAMPLES

4.1 Robustness Considerations

In this chapter we shall demonstrate by a numerical example some of the problems addressed in our discussion. The data set contains certain financial data for a range of years for a Finnish company. We also use a second data that is a generated random sample with 100 observations and 25 variables. This is the maximum data matrix that can be handled by the STATLADR robust regression program by Salmi and Pynnönen (1992). The data set is available on request from the authors.

The first data set, listed in Table 4.1, consists of a logarithmic deflated growth rate \((Y)\) of a Finnish company, a time variable \((X_1)\), and a measure of the cyclical fluctuation in the aggregated funds from operations in logarithmic terms \((X_2)\). For a detailed description of the data, originally used for growth estimation, see Salmi, Dahlstedt and Luoma (1985).


There are no apparent outliers in the data set. Consequently the least squares is a liable method for estimating the growth model

\[(4.1) \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon.\]

It may also be expected that the LAD estimation should produce closely related results to the OLS estimators. The model has been estimated by OLS in Salmi, Dahlstedt, and Luoma (1985). It is re-estimated here, first by OLS, and then by LAD.

Running the OLS regression yields the estimates given in Table 4.2.

**Table 4.2. OLS estimates for model (4.1).**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>std.err</th>
<th>t-value</th>
<th>p-value</th>
<th>Goodness of fit ($R^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST</td>
<td>-142.08</td>
<td>51.90</td>
<td>-2.738</td>
<td>0.020</td>
<td>0.541</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.078</td>
<td>0.026</td>
<td>2.978</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>-1.490</td>
<td>1.124</td>
<td>-1.334</td>
<td>0.209</td>
<td></td>
</tr>
</tbody>
</table>

The LAD estimation [using the computer program by Salmi and Pynnönen (1992)] yields the results given in Table 4.3.

**Table 4.3. LAD estimates for model (4.1).**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>std.err</th>
<th>t-value</th>
<th>p-value</th>
<th>Goodness of fit (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST</td>
<td>-178.24</td>
<td>34.86</td>
<td>-5.113</td>
<td>0.000</td>
<td>0.417</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.097</td>
<td>0.018</td>
<td>5.470</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>-1.176</td>
<td>0.755</td>
<td>-1.557</td>
<td>0.150</td>
<td></td>
</tr>
</tbody>
</table>
Both estimation methods suggest the same general results. Variable $X_2$ is not statistically significant, and the time variable, $X_1$, is highly statistically significant. There are some differences in the numerical values of the estimates given by these two methods, but the general outlook is the same in the sense that the same variables are statistically significant and the signs of the estimates are the same.

Next contaminate the data set with a significant outlier by changing the ninth observation of $Y$ from its original value of 12.078 to 10.078. This roughly corresponds to a typing error which changes the original value by a factor of ten (more precisely $e^2$) before taking the logarithm.

As a consequence the OLS estimation results, reported in Table 4.4, change radically, suggesting that neither of the variables is significant.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>std.err</th>
<th>t-value</th>
<th>p-value</th>
<th>Goodness of fit ($R^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST</td>
<td>-170.89</td>
<td>123.4</td>
<td>-1.385</td>
<td>0.194</td>
<td>0.191</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.093</td>
<td>0.062</td>
<td>1.484</td>
<td>0.166</td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>-0.780</td>
<td>2.673</td>
<td>-0.292</td>
<td>0.676</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4. OLS estimates for model (4.1) with observation $Y_8$ contaminated.

The LAD estimates remain in this time unaltered (see Table 4.5). This is, however, not always the case. The contaminated observation in the example is already somewhat outlying downward. Thus strengthening the effect does not affect LAD estimates. Had we changed the outlier upwards from its original value would it also have affected the LAD estimates, nevertheless much less than the OLS estimates. In any case, these results are a clear indication of the robustness of the LAD method over OLS.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>std.err</th>
<th>t-value</th>
<th>p-value</th>
<th>Goodness of fit (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST</td>
<td>-194.66</td>
<td>45.39</td>
<td>-4.289</td>
<td>0.002</td>
<td>0.297</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.105</td>
<td>0.023</td>
<td>4.562</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$X_2$</td>
<td>-1.3186</td>
<td>0.983</td>
<td>-1.341</td>
<td>0.206</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5. LAD estimates for model (4.1) with observation $Y_8$ contaminated.

The example also demonstrates the arbitrariness of $R^2$ as a measure of goodness of fit in the LAD context, when there are outliers present. In the non-outlier data the squared correlation between fitted and observed values is 0.5330, whereas in the outlier data it collapses below one half the original value, being 0.190. The $G$ statistic behaves much less dramatically, although dropping significantly from 0.417 to 0.297, it drops relatively much less than the $R^2$. 
4.2 Aberrations

As noted in Chapter 2, compared to the least squares the computational load for the LAD-solution is much more demanding. Similarly the finite-sample statistical properties are complicated. This is probably one reason why the LAD-regression is not easily available in the most well known statistical packages, like SAS or BMBD. In some of the econometric packages, like SHAZAM, LAD (LAE in SHAZAM) is available. The LAE in SHAZAM is very fast, but it seems to run in troubles if the regression is close to a perfect fit.

In RATS the LAD-estimation is suggested to be performed by programming the statement

\[
\min_\beta \sum_{i=1}^{n} \left[ c^2 + (y_i - x_i \beta)^2 \right]^{1/2},
\]

where \( c \) is some constant. The solution of (4.2) approaches to LAD as \( c \to 0 \) [see, for further details, Doan (1992; pp. 5–10)]. Technically (4.2) leads to a weighted least squares solution. The problem is how to select the constant \( c \). The experience of the first author of this paper is that the result of the method are not as robust as the results of the proper LAD method.

REFERENCES


