Computational Statistics & Data Analysis 14 (1992) 207–211
North-Holland

A note on recent proposals for computing l_1 estimates

Gilbert W. Bassett
University of Illinois, Chigago, IL, USA

Roger W. Koenker
University of Illinois, Champaign, IL, USA

Received April 1991

Abstract: Recent methods for computing the least absolute value (l_1) estimate have been proposed. In contrast to the usual linear programming formulation of the l_1 problem, the new methods attempt to use least squares residuals to identify observations whose l_1 residuals are equal to zero. Examples are presented to show that the proposals do not produce estimates that are identical or even necessarily close to l_1, and hence the algorithms cannot be recommended as a method for computing the l_1 estimate.

Keywords: Least absolute value; Linear programming; Least squares.

1. Introduction

The standard method for computing the least absolute value (l_1) estimate derives from the linear programming formulation of the l_1 problem. The method is based on a version of the simplex algorithm that is customized for special features of the l_1 problem; see e.g. Barrodale and Roberts (1974), Bassett and Koenker (1978), and Bartels and Conn (1980).

Alternative methods for computing l_1 have been considered in a recent series of papers Soliman, Christensen, and Rouhi (1988) proposed using least squares residuals to identify the observations whose l_1 residuals are equal to zero. Examples by Herce (1990) showed that this did not necessarily produce the l_1 estimate and this led to a revised proposal (Christensen, Soliman, and Rouhi, 1990) in which the original method is implemented only after first discarding observations with large least squares residuals. A modified version of this latter algorithm was subsequently proposed in Soliman, Christensen, and Rouhi (1991) to handle l_1 estimation with nonlinear equality constraints.

Correspondence to: G.W. Bassett, University of Illinois at Chicago, Department of Economics (m/c144), Box 802451, Chicago, Il 60680-2451, USA. E-mail:u09006@uicvm.cc.uic.edu

The proposed computational methods are motivated by the l_1 estimate's relation to subsets of p observations. The l_1 estimate can be identified by the subset of p observations whose l_1 residuals are equal to zero; graphically, the l_1 fit goes through p observations, where p is the number of explanatory variables (including the intercept). This is the linear programming property in which an optimal solution is basic; the basic solutions to the l_1 linear program correspond to the set of hyperplanes that fit p observations. The best known instance of the property is the identification of the median estimate (vix., l_1 for $p = 1$) with an observation. The property means that computing l_1 is essentially like discovering which p subset of observations has zero residuals.

The new proposals are also motivated by a belief that residuals are not too sensitive to fitting criterion. If l_1 errors can be approximated by least squares errors (perhaps after deleting or reweighting observations) then least squares could be used to identify the observations that determine the l_1 estimate.

The proposed algorithms by Soliman, Christensen and Rouhi (1990, 1991) however do not produce estimates that are identical or even necessarily close to l_1. Least squares residuals cannot be used to decide which l_1 residuals are small or zero. When least squares and l_1 are very different from one another the algorithm fails. As shown in the examples below the discrepancy between l_1 and the estimate produced by the algorithm can be unbounded. Since the algorithm does not produce estimates that are necessarily close to l_1 it cannot be recommended as a method for computing l_1.

The next section presents examples. The first ones are for the simplest case of the location submodel where least squares is the sample mean and l_1 is the sample median. This is followed by a general technique for constructing counterexamples. Some concluding comments appear in the final section.

2. Examples

The original algorithm identified observations with zero l_1 residuals as those whose least squares residuals are smallest in absolute value. For the location submodel this means a median observation would be identified as the one closest to the sample mean. (In the location submodel the l_1 and least squares estimates are the median and mean, respectively). Examples where the median observation differs from the mean are readily constructed.

One case is $(0, 0, 0, 0, 6w, 23w, 132w)$, $w > 0$. [1] The sum of squared residuals is minimized at $23w$ and the sum of absolute residuals is minimized at zero. The sixth observation is closest—in this case equal—to the mean and therefore the initial algorithm by Soliman, Christensen and Rouhi (1988) would produce an estimate of $23w$. Since w is arbitrary the difference between this supposed median and the actual median is unbounded.

[1] The equality of the first four observations allows easy comparison of estimates; it plays no essential role in the counterexamples.

Table 1

	Original Data		OLS on Perturbed Data			OLS with (D) Deleted		
	Y	X	Y	X^{OLS}	Errors	Y	X^{OLS}	Errors
(A)	0	1	0	1	0.73	0	1	1.15
(B)	1	2	1	2	0.91	1	2	2.62
(C)	−1	2	−4	2	−4.09	−4	2	−2.38
(D)	2	3	12	3	11.09			
(E)	−2	3	−6	3	−6.91	−6	3	−3.92
(F)	0	4	0	4	−1.73	0	4	2.54
			Intercept	−1.5454545		Intercept	−0.6923077	
			Slope	0.81818182		Slope	−0.4615385	

This is also a counterexample to earlier proposals in which l_1 was presumed to be computed by reweighted least squares with weights determined by least squares residuals; Schlossmacher (1973). Other counterexamples may be found in Gallant and Gerig (1974).

The subsequent proposal modifies the original algorithm by rejecting observations that are outside one standard deviation of a preliminary least squares fit. [2] For the above example the standard deviation is $45.2w$ so the seventh observation is discarded. The mean of the truncated sample is $4.8w$ and the fifth observation is closest to the truncated least squares fit. The difference between this supposed estimate for the median and the actual median can again be made arbitrarily large.

This failure of the algorithm is not restricted to the location case. A general technique for constructing examples is as follows. Start with a pattern of data in which least squares and l_1 are equal. Perturb the data so that observations with positive residuals stay above, and negative residuals stay below the original fit. This transformation leaves l_1 unaltered [3], but shifts least squares to a new location that can be unboundedly far from the original fit. Values for the perturbed data can be selected so that the revised estimate differs from the l_1 estimate.

Consider the data in Table 1 and Figure 1. The least squares and l_1 estimates are equal; the slope and intercept are both zero. Let the observations labeled (D), (C), and (E) move as indicated by the arrows. This preserves the sign of the original residuals and hence does not change the l_1 estimate. But it does change the estimate produced by the original and modified versions of the algorithm. With the altered data (D) is discarded and the least squares estimate on the remaining data is shown by the dotted line. The smallest residuals are now at

[2] This modification leaves an open question regarding bimodal samples like (1, 1, 1, 9, 9, 9) where all of the observations are outside the cutoff point and would all be discarded.

[3] This is the counterpart of the median property in which the median stays fixed when any observations above the median are arbitrarily increased. It corresponds to the linear programming result in which an optimal solution is unaltered by relaxing a nonbinding constraint.

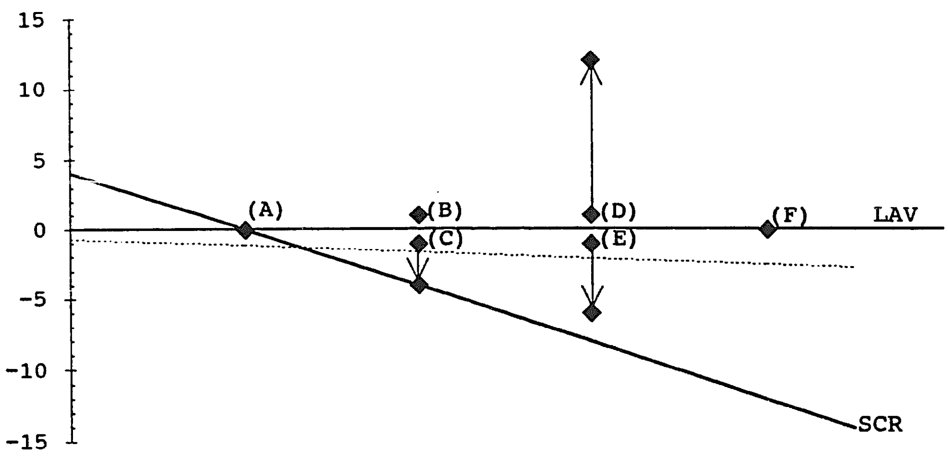

Fig. 1.

(A) and (C), which supposedly determine the l_1 estimate. This however is not the l_1 estimate; l_1 is still at its original position, along the X-axis.

3. Conclusion

The examples show that the proposed algorithm does not produce an estimate that is necessarily close to l_1. The method incorrectly presumed that least squares residuals are a reliable guide for computing l_1. Large least squares residuals can correspond to small l_1 residuals, and conversely. The method fails because it cannot purge itself of the initial dependence on the least squares estimates.

Finally, it should be noted that there are alternatives to the simplex algorithm that do successfully compute the l_1 estimate. Mekaton (1985) has suggested an approach based on successive nonlinear transformations. His interesting proposal implements the 'interior-based method' for linear programming due to Karmarker (1984).

References

Barrodale, I. and F.D.K. Roberts (1974). Solution of an Overdetermined System of Equations in the L_1-norm, *Communications Ass. Comput. Mach*, **17**, 319–320.

Bartels, R. and Conn, A.R. (1980) Linearly Constrained Discrete L_1 Problems, *Assoc. For Comp. C Mach Trans. on Math Software*, **6**, 594–608.

Bassett, G.W. and R.W. Koenker (1978). The Asymptotic Theory of Least Absolute Error Regression, *J. Amer. Statist. Ass.*, **73** 618–622.

Christensen, G.S., S.A. Soliman, and A. Rouhi (1990). Discussion of An Example Showing That a New Technique for lav Estimation Breaks Down in Certain Cases, *Computational Statistics & Data Analysis*, **9** 203–213.

Gallant, A.R. and T.M. Gerig (1974). Comments on Computing Minimum Absolute Deviations Regressions by Iterative Least Squares Regressions and by Linear Programming. (paper).

Herce, M.A. (1990). An Example Showing That a New Technique for lav Estimation Breaks Down in Certain Cases, *Computational Statistics & Data Analysis*, **9** 197–202.

Karmarker, N. (1984). A New Polynomial-Time Algorithm for Linear Programming, *Combinatorica*, **4** 373–395.

Meketon, M.S. (1985). Least Absolute Value Regression, AT&T Bell Laboratories, (unpublished).

Schlossmacher, E.J. (1973). An Iterative Technique for Absolute Deviations Curve Fitting. *J. Amer. Statist. Ass.* **68** 857–859.

Soliman, S.A., G.S. Christensen, and A.H. Rouhi (1988). A New Technique for Curve Fitting Based on Minimum Absolute Deviations, *Computational Statistics & Data Analysis*, **6** 341–351.

Soliman, S.A., G.S. Christensen, and A.H. Rouhi (1991). A New Algorithm for Nonlinear L_1-norm Minimization with Nonlinear Equality Constraints, *Computational Statistics & Data Analysis*, **11** 97–109.