A note on recent proposals for computing $l_1$ estimates

Gilbert W. Bassett
University of Illinois, Chicago, 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

0167-9473/92/$05.00 © 1992 – Elsevier Science Publishers B.V. All rights reserved
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 (viz., $l_1$ 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) the 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 $(0, 0, 0, 0, 6w, 23w, 132w)$, $w > 0$. 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

<table>
<thead>
<tr>
<th>Original Data</th>
<th>OLS on Perturbed Data</th>
<th>OLS with (D) Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X$</td>
<td>$X^{\text{OLS}}$</td>
</tr>
<tr>
<td>(A)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(B)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(C)</td>
<td>-1</td>
<td>-4</td>
</tr>
<tr>
<td>(D)</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>(E)</td>
<td>-2</td>
<td>-6</td>
</tr>
<tr>
<td>(F)</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Intercept: -1.5454545
Slope: 0.81818182

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. For the above example the standard deviation is 45.2 and so the seventh observation is discarded. The mean of the truncated sample is 4.8 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, 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.
(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


