ON $L_1$ AND CHEBYSHEV ESTIMATION

Gautam APBA

Middlesex Polytechnic at Enfield, Middlesex, England

and

Cyril SMITH

London School of Economics, London, England

Received 5 May 1972
Revised manuscript received 2 April 1973

The problem considered here is that of fitting a linear function to a set of points. The criterion normally used for this is least squares. We consider two alternatives, viz., least sum of absolute deviations (called the $L_1$ criterion) and the least maximum absolute deviation (called the Chebyshev criterion). Each of these criteria give rise to a linear program. We develop some theoretical properties of the solutions and in the light of these, examine the suitability of these criteria for linear estimation. Some of the estimates obtained by using them are shown to be counter-intuitive.

1. Introduction

The investigation of alternative criteria to least squares in fitting a linear equation to $N$ observations on $m$ independent and one dependent variable goes back at least to Fourier in 1820. He proposed an iterative procedure to minimise the sum of absolute deviations which is similar to the simplex method. Edgeworth [4] considered the same problem in 1887.

Many authors have noted the relationship of this problem, sometimes called $L_1$ approximation, to linear programming (see [2, 3, 5, 9]). The $L_1$ criterion has, in particular, interested econometricians who frequently wish to estimate parameters of linear models with relatively small numbers of observations [5].

A related criterion, whose minimisation is also reducible to a linear programme, is the maximum absolute deviation. This, the linear Chebyshev approximation problem, is sometimes called $L_\infty$ approximation. It has attracted a great deal of attention because of its applica-

tion to the computation of functions in terms of polynomials (see [1, 6, 7, 8]).
In this paper, we develop some properties of the linear equations fitted by these criteria and discuss their merit for econometric work.

2. Notations and preliminaries

Let \( x_1, ..., x_m \) be \( m \) independent variables and \( y \) be dependent on these. Suppose we have obtained \( N \) (\( N > m \)) observations on these variables, viz., \((y_1, x_{11}, x_{21}, ..., x_{m1}) \text{ to } (y_N, x_{1N}, x_{2N}, ..., x_{mN})\). Assume\(^1\) that no set of \( m + 2 \) observations lie on one hyperplane in \( m + 1 \) dimensions. Then:

(a) \[
\sum_{j=1}^{N} | y_j - b_0 - \sum_{i=1}^{m} b_i x_{ij} | \text{ is minimized,}
\]

(b) \[
\max_{j} | y_j - b_0 - \sum_{i=1}^{m} b_i x_{ij} | \text{ is minimized.}
\]

These problems have two interesting common features:
(i) the objective function in each case is a piecewise linear function of the parameters \( b_0, b_1, ..., b_m \);
(ii) both problems can be formulated and solved as standard linear programming problems.

As the properties of the solution developed in this paper rely heavily on some standard results in linear programming, it is necessary to develop further notation and give the linear programming formulation explicitly.

3.

3.1.1. Problem (a). Define variables \( d_j^+ \) and \( d_j^- \), representing positive and negative deviation, respectively, for the \( j^{th} \) observation. Then the linear programming problem corresponding to problem (a) is

\(^1\) This assumption allows us to eliminate the awkward cases of degenerate solutions which do not throw any light on the problem.