Journal of Modern Mathematics and Statistics 2 (3): 123-125, 2008
ISSN: 1994-5388

A Linear Goal Programming Model for the Linear Absolute Value Regression Problem

[1]U.N. Bassey and [2]E.O. Effanga
[1]Department of Mathematics, University of Ibadan, Ibadan, Nigeria
[2]Department of Mathematics/Statistics and Computer Science, University of Calabar, Calabar, Nigeria

Abstract: In this study, we reformulate the linear absolute regression problem as a linear goal-programming problem, whose associated dual problem is obtained and exploited computationally to estimate the regression parameters. It is shown that the goal programming method of estimating the regression parameters is better, in terms of the total absolute error, than the least squares method.

Key words: Linear absolute regression, linear goal programming, least squares, regression problem

INTRODUCTION

Multiple linear regression model has been widely used in showing relationships existing between a response variable and the explanatory variables (Moore and McCabe, 1989). A number of methods for the estimation of the regression parameters are available in the literature. These include methods of minimizing the sum of absolute residuals, minimizing the maximum of absolute residuals and minimizing the sum of squares of residuals (Weisberg, 1985). Of all these methods, the most commonly used is the one for minimizing the sum of squares of residuals, popularly known as the least squares method. It is, however, uncommon to use the method of minimizing the sum of absolute residuals because of certain difficulties inherent in it.

Goal programming is a tool for handling problems involving multiple but conflicting objectives. This tool often represents a substantial improvement in the modelling and analysis of multi-objective problems (Charnes and Cooper, 1977; Eiselt *et al.*, 1987; Ignizio, 1978).

In the present study, we attempt to overcome the difficulties associated with minimizing the sum of absolute residuals by reformulating the linear absolute regression problem as a linear goal programming problem and solving same by the usual simplex procedure.

A REVIEW OF MULTIPLE LINEAR REGRESSION MODEL

Let $x_{i0} = 1$ for i = 1, 2, …, n, let $X_1, X_2, \ldots, X_k$ be k independent random variables and let Y be a dependent random variable. Then a linear relationship of the form

$$y_i = \sum_{j=0}^{k} \beta_j x_{ij} + e_i, \quad i = 1, 2, \ldots, n \qquad (1)$$

is assumed, where $\beta_0, \beta_1, \ldots, \beta_k$ are the parameters to be estimated and $e_i (i = 1, 2, \ldots, n)$ are the error components which are assumed to be normally and independently distributed with zero mean and constant variance. The linear absolute residuals method requires us to estimate the values of the unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$ so as to minimize

$$\sum_{i=1}^{n} |y_i - \hat{y}_i|, \qquad (2)$$

Where:

$$\hat{y}_i = \sum_{j=0}^{k} \hat{\beta}_j x_{ij}, \quad i = 1, 2, \ldots, n \qquad (3)$$

and

$$\hat{\beta}_j, \ j = 0, 1, \ldots, k$$

are the estimated values of the unknown parameters β_j, j = 0, 1,....k.

The least squares principle requires us to choose $\beta_0, \beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

LINEAR GOAL PROGRAMMING FORMULATION OF THE LINEAR ABSOLUTE RESIDUALS METHOD

Let y_i be the ith goal, d_i^+ be positive deviation from the ith goal and d_i^- be the negative deviation from the ith goal. Then the problem of minimizing

Corresponding Author: U.N. Bassey, Department of Mathematics, University of Ibadan, Ibadan, Nigeria

$$\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

may be reformulated as

$$\left.\begin{array}{l} \min \sum_{i=1}^{n}\left(d_i^+ + d_i^-\right) \\ \text{Subject to:} \\ \sum_{j=0}^{k} x_{ij}\beta_j + d_i^+ - d_i^- = y_i, i = 1, 2, ..., n \\ d_i^+ \geq 0, \quad i = 1, 2, ..., n \\ d_i^- \geq 0, \quad i = 1, 2, ..., n \\ \beta_j \text{ unrestricted, } j = 0, 1, ..., k. \end{array}\right\} \quad (4)$$

Although, the primal linear goal programming problem Eq. 4 is solvable by the traditional simplex method, computations are simplified by solving its dual problem given below. Let u_i be the dual variable associated with the ith goal constraint. Then the dual of problem Eq. 4 is:

$$\left.\begin{array}{l} \max \sum_{i=1}^{n} y_i u_i \\ \text{Subject to:} \\ \sum_{i=1}^{n} x_{ij}u_i \leq 0, \quad j = 0, 1, ,...., k; \\ -1 \leq u_i \leq 1, \quad i = 1, 2, ..., n. \end{array}\right\} \quad (5)$$

The computational efficiency of problem Eq. 5 is enhanced by using the lower and upper bounding techniques. The lower bounding technique involves setting $u_i = w_i - 1$ in problem Eq. 5 to obtain problem Eq. 6:

$$\left.\begin{array}{l} \max\left\{\sum_{i=1}^{n} y_i w_i - \sum_{i=1}^{n} y_i\right\} \\ \text{Subject to:} \\ \sum_{i=1}^{n} x_{ij} w_i \leq \sum_{i=1}^{n} x_{ij}, \quad j = 0, 1, ..., k \\ 0 \leq w_i \leq 2, \quad i = 1, 2, ..., n. \end{array}\right\} \quad (6)$$

The upper bounding technique entails only a few variations of the simplex algorithm to obtain an optimal solution without additional rows or columns for the bounds (Eiselt *et al*., 1987) for a detailed description of the upper bounding technique).

RELATIONSHIP BETWEEN THE LEAST SQUARES METHOD AND THE LINEAR GOAL PROGRAMMING METHOD

Let $\hat{y}_{iG}$ be the estimate of the ith response using a goal programming technique and let $\hat{y}_{iL}$ be the estimate using the least square method. Then,

$$\min \sum_{i=1}^{n}|y_i - \hat{y}_{iG}| < \min \sum_{i=1}^{n}|y_i - \hat{y}_{iL}| \quad (7)$$

We do not know how this inequality may be proved analytically. However, an example is given below to demonstrate this relationship.

Example: We consider a regression equation of Y on X_1, X_2 and X_3. Our data for illustration is shown in the Table 1. The goal programming formulation is:

$$\text{minimize} \sum_{i=1}^{11}\left(d_i^+ + d_i^-\right)$$

subject to:

$$\begin{array}{l} a_0 + 0.45a_1 + 1.95a_2 + 0.34a_3 + d_1^+ - d_1^- = 27.1 \\ a_0 + 0.47a_1 + 5.13a_2 + 0.32a_3 + d_2^+ - d_2^- = 35.6 \\ a_0 + 0.44a_1 + 3.98a_2 + 0.29a_3 + d_3^+ - d_3^- = 31.4 \\ a_0 + 0.48a_1 + 6.25a_2 + 0.30a_3 + d_4^+ - d_4^- = 37.8 \\ a_0 + 0.48a_1 + 7.12a_2 + 0.25a_3 + d_5^+ - d_5^- = 40.2 \\ a_0 + 0.49a_1 + 6.50a_2 + 0.26a_3 + d_6^+ - d_6^- = 39.8 \\ a_0 + 0.53a_1 + 10.67a_2 + 0.10a_3 + d_7^+ - d_7^- = 55.5 \\ a_0 + 0.50a_1 + 7.08a_2 + 0.16a_3 + d_8^+ - d_8^- = 43.6 \\ a_0 + 0.55a_1 + 9.88a_2 + 0.19a_3 + d_9^+ - d_9^- = 52.1 \\ a_0 + 0.51a_1 + 8.72a_2 + 0.18a_3 + d_{10}^+ - d_{10}^- = 43.8 \\ a_0 + 0.48a_1 + 4.96a_2 + 0.28a_3 + d_{11}^+ - d_{11}^- = 35.7 \\ d_i^+ \geq 0, i = 1 \text{ to } 11 \\ d_i^- \geq 0, i = 1 \text{ to } 11 \\ a_i, i = 0, 1, 2, 3 \text{ are unrestricted.} \end{array}$$

The solution to the above problem by the simplex procedure using an optimizer in a Corel quarttro pro computer software package is:

$$a_0 = -2.82787, \quad a_1 = 77.80057,$$
$$a_2 = 1.541726, \quad a_3 = -21.174$$

Table 1: Regression equation of Y on X_1, X_2 and X_3

Y	27.1	35.6	31.4	37.8	40.2	39.8	55.5	43.6	52.1	43.8	35.7
X_1	0.45	0.47	0.44	0.48	0.48	0.49	0.53	0.50	0.55	0.51	0.48
X_2	1.95	5.13	3.98	6.25	7.12	6.50	10.67	7.08	9.88	8.72	4.96
X_3	0.34	0.32	0.29	0.30	0.25	0.26	0.10	0.16	0.19	0.18	0.28

Thus the estimated regression equation via the goal programming method is:

$$\hat{y}_{iG} = -2.82787 + 77.80057x_{i1} + 1.541726x_{i2} - 21.174x_{i3}, i = 1, 2, ..., 11$$

and

$$\min \sum_{i=1}^{11} |y_i - \hat{y}_{iG}| = 8.311739$$

The corresponding estimated regression equation via the least squares method is:

$$\hat{y}_{iL} = 23.35942 + 11.0438x_{i1} + 2.520824x_{i2} - 21.2573x_{i3}, i = 1, 2, ..., 11$$

and

$$\min \sum_{i=1}^{11} |y_i - \hat{y}_{iL}| = 11.965578$$

Thus

$$\min \sum_{i=1}^{n} |y_i - \hat{y}_{iG}| < \min \sum_{i=1}^{n} |y_i - \hat{y}_{iL}|$$

We have shown this relation to hold for several other examples not included in this study.

CONCLUSION

The goal programming technique has been shown to provide better estimates of the multiple linear regression parameters than the conventional least squares method.

REFERENCES

Charnes, A. and W.W. Cooper, 1977. Goal Programming and Multiple Objective Optimizations. Eur. J. Operat. Res., 1: 39-54.

Eiselt, H.A., G. Pederzoli and C.L. Sandblom, 1987. Continuous Optimization Models. W De G, New York.

Ignizio, J.P., 1978. A Review of Goal Programming-A Tool for Multiobjective Analysis. J. Opl. Res. Soc., 29 (11): 1109-1119.

Moore, D.S. and G.P. McCabe, 1989. Introduction to the practice of Statistics. W.H. Freeman and Company, New York.

Weisberg, S., 1985. Applied Linear Regression. 2nd Edn. John Wiley and Sons, Inc. New York.