LINEAR PROGRAMMING TECHNIQUES FOR REGRESSION ANALYSIS

HARVEY M. WAGNER*
Stanford University

In regression problems alternative criteria of "best fit" to least squares are least absolute deviations and least maximum deviations. In this paper it is noted that linear programming techniques may be employed to solve the latter two problems. In particular, if the linear regression relation contains p parameters, minimizing the sum of the absolute value of the "vertical" deviations from the regression line is shown to reduce to a p equation linear programming model with bounded variables; and fitting by the Chebyshev criterion is exhibited to lead to a standard-form $p+1$ equation linear programming model.

1. INTRODUCTION

KARST [7] has suggested recently an iterative procedure "for finding a straight line of best fit to a set of two dimensional points such that the sum of the absolute values of the vertical deviations of the points from the line is a minimum." Because linear programming is a relatively new tool to the statistician, we offer here an elementary presentation of the known results that the multi-dimensional version of Karst's problem and the model of multiple linear regression according to a Chebyshev criterion [6] may be solved directly by linear programming methods. Charnes, Cooper, and Ferguson [1] and other practitioners of linear programming have recognized that the problem of minimizing the sum of absolute deviations can be converted into a linear programming model consisting of k equations, where k is the number of observations, and an objective function which calls for minimizing the sum of non-negative variables.[1] First, by employing the fundamental dual theorem [2, 9, 11] in linear programming, we shall show how the problem can be handled by a p equation linear programming model with bounded variables [3, 4, 12], where p is the number of regression coefficients. Secondly, in a manner directly analogous to that of[2] Kelley [8], we shall demonstrate how a regular $p+1$ equation linear programming model can be utilized to find a line of best fit according to a Chebyshev criterion, *i.e.*, a line (or hyperplane) which minimizes the maximum of the vertical deviations from the sample points.

2. DUAL LINEAR PROGRAMMING PROBLEMS

For the purpose of subsequent reference throughout this paper, we give a version of Goldman and Tucker's canonical representation of "dual" linear programming problems [5]. The primal model consists of m linear relations in n unknowns x_l, in which the relations are partitioned into two mutually ex-

---

* This work was sponsored in part by the Office of Naval Research.

[1] In addition to exhibiting the model for the regression problem, Charnes, Cooper, and Ferguson present an interesting numerical illustration of the technique in which a nine parameter regression problem is solved by the utilization of linear inequalities among less than nine observations.

[2] The author is indebted to a referee for bringing this reference to his attention.

clusive and completely exhaustive classes M_1 and M_2, and similarly the variables into two such classes N_1 and N_2, as specified by (2).

$$\text{Maximize } c_1x_1 + \cdots + c_nx_n \tag{1}$$

subject to the constraints

$$a_{1h}x_1 + \cdots + a_{nh}x_n \begin{cases} \leqq b_h & \text{for each } h \text{ in } M_1 & \text{(2a)} \\ = b_h & \text{for each } h \text{ in } M_2 & \text{(2b)} \end{cases}$$

$$x_l \begin{cases} \text{non-negative for each } l \text{ in } N_1 & \text{(2c)} \\ \text{unrestricted in sign for each } l \text{ in } N_2. & \text{(2d)} \end{cases}$$

In particular cases, one or more of the classes M_1, M_2, N_1, and N_2 may be empty.

The corresponding dual model consists of n linear relations in m unknowns u_h with parallel partitioning

$$\text{Minimize } u_1b_1 + \cdots + u_mb_m \tag{3}$$

subject to the constraints

$$u_1a_{l1} + \cdots + u_ma_{lm} \begin{cases} \geqq c_l & \text{for each } l \text{ in } N_1 & \text{(4a)} \\ = c_l & \text{for each } l \text{ in } N_2 & \text{(4b)} \end{cases}$$

$$u_h \begin{cases} \text{non-negative for each } h \text{ in } M_1 & \text{(4c)} \\ \text{unrestricted in sign for each } h \text{ in } M_2. & \text{(4d)} \end{cases}$$

The fundamental dual theorem of linear programming [2, 5, 9, 11] states that a set of x_l^* satisfying (2) is optimal (1) if and only if there is a set of u_h^* satisfying (4) with

$$c_1x_1^* + \cdots + c_nx_n^* = u_1^*b_1 + \cdots + u_m^*b_m \quad . \tag{5}$$

Similarly a set of u_h^* satisfying (4) is optimal (3) if and only if there is a set of x_l^* satisfying (2) with (5) holding.

3. MINIMIZING THE SUM OF ABSOLUTE DEVIATIONS

Let x_{ij}, $i=1, 2, \cdots, k$, and $j=1, 2, \cdots, p$, denote a set of k observational measurements on p "independent" variables, and y_i, $i=1, 2, \cdots, k$, denote the associated measurement on the "dependent" variable. Note that in the case of curvilinear regression, we may have $x_{ij}=z_i^j$, or $x_{ij}=\log z_{ij}$, or $x_{ij}=w_{ij}v_{ij}$, etc. We wish to find regression coefficients b_j that

$$\underset{b_j}{\text{Minimize}} \sum_i \left| \sum_j x_{ij}b_j - y_i \right| \cdot \tag{6}$$

Using the reduction in Charnes, Cooper, and Ferguson [1], the problem (6) is transformed into

$$\text{Minimize } \sum_i \epsilon_{1i} + \sum_i \epsilon_{2i} \tag{7}$$

subject to the constraints

$$\sum_j x_{ij}b_j + \epsilon_{1i} - \epsilon_{2i} = y_i \qquad i = 1, 2, \cdots, k \tag{8a}$$

$$b_j \text{ unrestricted in sign} \tag{8b}$$

$$\epsilon_{1i}, \epsilon_{2i} \text{ non-negative.} \tag{8c}$$

We interpret ϵ_{1i} and ϵ_{2i} as vertical deviations "above" and "below" the fitted line for the i-th set of observations; *i.e.*, $\epsilon_{1i}+\epsilon_{2i}$ is the absolute deviation between the fit $\sum_j x_{ij}b_j$ and y_i. By the nature of the model (7) and (8), ϵ_{1i} and ϵ_{2i} cannot both be strictly positive in an optimal solution. Thus we have formulated the regression problem as a linear programming model of type (3) and (4).

The solution to (7) and (8) yields the regression equation

$$b_1x_1 + \cdots + b_px_p = y. \tag{9}$$

Note that if we wish the left hand side of (9) to include a coefficient for the intercept of the y axis to be determined by the linear fit, then we can let $x_p \equiv 1$, *i.e.*, let $x_{ip} = 1$ in (8). We may force the fitted line to pass through some point, the usual example being the set of sample means, either by adding to (8) the linear restriction

$$b_1\bar{x}_1 + \cdots + b_p\bar{x}_p = \bar{y} \tag{10}$$

or by the usual least squares approach of subtracting each coordinate of the point, in our example the sample mean for each variable, from all the corresponding observations (including y) and then by fitting (8) without a y-intercept coefficient; the latter approach simply consists of shifting the origin of the axes in a p-dimensional space to the selected point, and then of fitting the line (hyperplane) through the new origin.

According to (8b) we have not restricted the sign of b_j; but we may drop this unessential assumption if, in the context of the regression problem, we desire to permit only non-negative values for some or all b_j or to force the b_j to satisfy additional linear constraints of greater complexity [1].

It is noteworthy that collinearity in the x_{ij}, even to the degree that for some j' and j'', $x_{ij'} = x_{ij''}$, will not cause a failure in the linear programming algorithm for (7) and (8) or for any of the models below.

One drawback of the present model is evident: if the number of observations k is sizeable, (7) and (8) become computationally unwieldy. We shall now transform (7) and (8) into a more manageable dual problem which will yield the optimal b_j as a byproduct. To preserve generality in our treatment, assume the b_j are partitioned into the classes M_1 and M_2 according to (4c) and (4d). Then the dual relationship given in Section 2 implies that we can find a solution to (7) and (8) if and only if we can find a solution to

$$\text{Maximize } \sum_i y_i d_i \tag{11}$$

subject to the constraints

$$\sum_i x_{ij}d_i \begin{cases} \leqq 0 & \text{for each } j \text{ in } M_1 & \text{(12a)} \\ = 0 & \text{for each } j \text{ in } M_2 & \text{(12b)} \end{cases}$$

$$d_i \leqq 1 \qquad i = 1, 2, \cdots, k \tag{12c}$$

$$-d_i \leqq 1 \qquad i = 1, 2, \cdots, k \tag{12d}$$

$$d_i \text{ unrestricted in sign} \quad . \tag{12e}$$

Model (11) and (12) is even a larger problem than (7) and (8), since it consists of $p+2k$ relations. To reduce the problem to a model in p relations and k bounded variables we let

$$f_i \equiv d_i + 1 \qquad i = 1, 2, \cdots, k. \tag{13}$$

Then (11) and (12) are equivalent to

$$\text{Maximize } \sum_i y_i f_i - \sum_i y_i \tag{14}$$

subject to the constraints

$$\sum_i x_{ij} f_i \begin{cases} \leqq \sum_i x_{ij} & \text{for each } j \text{ in } M_1 \qquad (15a) \\ = \sum_i x_{ij} & \text{for each } j \text{ in } M_2 \qquad (15b) \end{cases}$$

$$0 \leqq f_i \leqq 2 \qquad i = 1, 2, \cdots, k \quad . \tag{15c}$$

The model now consists of p linear relations (15a) and (15b) in bounded non-negative variables (15c), and may be solved quite rapidly for small $p(<10)$ by special simplex algorithms for bounded variables problems [3, 4, 12]; such algorithms have been coded for nearly all medium and large scale electronic computers. Notice that if x_{ij} and y_i are deviations of sample values from their means, then the right hand side of (15a) and (15b) is zero and the constant in (14) is also zero. In the numerical solution of the model, it is customary to convert (15a) to equalities by appending *non-negative* "slack" variables [2] s_j

$$\sum_i x_{ij} f_i + s_j = \sum_i x_{ij} \qquad \text{for each } j \text{ in } M_1. \tag{16}$$

The coefficient of s_j in (14) is zero.

In a manner analogous to the techniques for solving *regular* linear programming problems [2, 8, 10], bounded variables algorithms produce an optimal "basic" set of variables, *i.e.*, p of the f_i and s_j (along with some of the f_i at their upper bound 2 (15c)) will enter the optimal solution, the values of the remaining variables being zero; we denote these optimal basic variables by v_i, $i=1, 2, \cdots, p$. Let γ_{ij} be the coefficients of the v_i in (15b) and (16), and λ_i the corresponding coefficients in (14). Then the regression coefficients b_j satisfy the relations[3]

$$\sum_j \gamma_{ij} b_j = \lambda_i \qquad i = 1, 2, \cdots, p. \tag{17}$$

No extra computations are needed to find b_j from (17). In the original simplex method [2, 10], b_j appears in the $(z_j - c_j)$ row of the final simplex tableau; in the revised simplex method [10], b_j is the "shadow price" for the optimal solution of (14) and (15). The optimal value of (14) is the minimized sum of absolute deviations.

If one desires to place additional constraints on the b_j, as in [1], the effect on (15) will be the addition of new variables; the number of relations remains

[3] Readers familiar with linear programming will recall that multiple optimal solutions to both the primal and dual model may exist. It is easy to verify that if the optimal values for b_j in (7) and (8) are unique, then (17) yields these values regardless of which alternative optimal solution is found in (14) and (15). If there are various values for b_j which produce the same optimal value for (7), then the solution to (15) will be degenerate, a condition which is present if some of the optimal basic variables in (15) are at their lower or upper bound, i.e., zero or two.

unchanged as does the dimension of (17). As suggested by Kelley [8], one may also wish to add new "independent" variables to improve the regression fit which has the effect of increasing the number of relations in (15). The latter problem may be handled by techniques dealing with "secondary constraints" [4, 11]. These advanced techniques permit one to add variables to (8), and consequently relations to (15), in a specified sequence and to determine the improvement without solving (14) and (15) from the beginning each time.

4. MINIMIZING THE MAXIMUM ABSOLUTE DEVIATION

We now consider a regression problem which in comparison to the model of the previous section contains one more equation but eliminates the "bounded variables" restriction. Employing a Chebyshev criterion of best fit, we seek b_j such that

$$\underset{b_j}{\text{Minimize}} \left\{ \underset{i}{\text{Maximum}} \left| \sum_j x_{ij}b_j - y_i \right| \right\}. \tag{18}$$

Paralleling Kelley's treatment [8], we transform (18) into the linear programming model

$$\text{Minimize } \epsilon \text{ (non-negative)} \tag{19}$$

subject to the constraints

$$-\epsilon \leqq \sum_j x_{ij}b_j - y_i \leqq \epsilon \qquad i = 1, 2, \cdots, k. \tag{10}$$

In (19) and (20) ϵ is the minimized value of the maximum absolute deviation. As before, to preserve generality we assume the b_j are partitioned according to (4c) and (4d). In preparation for the application of the dual theorem, we expand (19) and (20) into

$$\text{Minimize } \epsilon \tag{21}$$

subject to the constraints

$$-\sum_j x_{ij}b_j + \epsilon \geqq -y_i \qquad i = 1, 2, \cdots, k \tag{22a}$$

$$\sum_j x_{ij}b_j + \epsilon \geqq y_i \qquad i = 1, 2, \cdots, k \tag{22b}$$

$$b_j \begin{cases} \text{non-negative for each } j \text{ in } M_1 & (22c) \\ \text{unrestricted in sign for each } j \text{ in } M_2 & (22d) \end{cases}$$

$$\epsilon \text{ non-negative.} \tag{22e}$$

The dual formulation is then

$$\text{Maximize } -\sum_i y_i d_{1i} + \sum_i y_i d_{2i} \tag{23}$$

subject to the constraints

$$-\sum_i x_{ij}d_{1i} + \sum_i x_{ij}d_{2i} \begin{cases} \leqq 0 & \text{for each } j \text{ in } M_1 & (24a) \\ = 0 & \text{for each } j \text{ in } M_2 & (24b) \end{cases}$$

$$\sum_i d_{1i} + \sum_i d_{2i} \leqq 1 \tag{24c}$$

$$d_{1i}, d_{2i} \text{ non-negative.} \tag{24d}$$

Model (23) and (24) is a regular linear programming problem in $p+1$ relations and may be solved by any standard algorithm [2, 10]. If $d_{1i}(d_{2i})$ is positive in the optimal solution of (23) and (24), then the maximum deviation occurs for the i-th sample point, i.e., for the i-th relation in (22a), (22b), and this point will lie above (below) the fitted line.

As with the model in Section 3, we may derive from the optimal solution of (23) and (24) the values of the regression coefficients. Assume "slack variables" have been added to (24a) and (24c) to convert the relations to equalities, analogous to (16). The optimal basic solution consists of $p+1$ variables, which we denote by v_i, $i=1, 2, \cdots, p+1$. Let γ_{ij} be the coefficients of the v_i in (24a) and (24b), the former having been converted to equalities, and λ_i the corresponding coefficients in (23). Then the regression coefficients b_j and the error ϵ satisfy the relations[4]

$$\sum_j \gamma_{ij} b_j + \epsilon = \lambda_i \qquad i = 1, 2, \cdots, p + 1. \tag{25}$$

Once again, *no extra computations are needed to find b_j and ϵ from (25)*, since these values are automatically computed in the application of the simplex algorithms.

Finally the comments at the end of Section 3 concerning the addition of constraining relations and new "independent" variables are equally applicable here.[5]

5. A NUMERICAL EXAMPLE

Karst [7] examines the data in Table 211 which comprise deviations of the original data about their sample means.

TABLE 211

x_i	−12.5	−8.5	−6.5	−3.5	−2.5	−1.5	−0.5	2.5	4.5	8.5	8.5	11.5
y_i	− 8.4	−5.4	3.6	−2.4	−4.4	1.6	−0.4	−0.4	−2.4	3.6	5.6	9.6

He finds the least squares fit to be

$$y = .539x, \tag{26}$$

and the fit for the minimized sum of absolute deviations to be

$$y = .659x. \tag{27}$$

As we have suggested, (27) may be obtained by (14) and (15), where specifically we would find in (17)

$$8.5b = 5.6 \tag{28a}$$

$$b = .659. \tag{28b}$$

[4] We are assuming that we are dealing with the non-trivial case in which $\epsilon > 0$. Readers interested in the underlying mathematics of the model may verify that, given this assumption, the slack variable introduced in (24c) does not enter an optimal solution at a positive level and that an optimal basic set of variables v_i exists yielding (25) [5, 9].

[5] In lieu of a "secondary constraint" algorithm [4, 11] for handling the impact in (24) of additional variables in (22), Kelley [8] suggests the alternative of introducing into the enlarged basic solution an "artificial" variable with an arbitrarily high cost, and subsequently driving its value to zero.

The solution by model (21) and (22) yields for (25)

$$-6.5b + \epsilon = 3.6 \tag{29a}$$

$$11.5b + \epsilon = 9.6$$

$$b = .333 \qquad \epsilon = 5.767. \tag{29b}$$

That is, the Chebyshev fit is

$$y = .333x; \tag{30}$$

since the equations (29a) correspond to the optimal variables $d_{2,3}$ and $d_{2,12}$, the third and last sample point in Table 211 will lie above the fitted line and assume the maximum vertical deviation from it of 5.767.

REFERENCES

[1] Charnes, A., Cooper, W. W., and Ferguson, R. O., "Optimal Estimation of Executive Compensation by Linear Programming," *Management Science* 1 (1955), 138–51.

[2] Charnes, A., Cooper, W. W., and Henderson, A., *An Introduction to Linear Programming*, New York: Wiley, 1953.

[3] Charnes, A., and Lemke, C. E., *Computational Theory of Linear Programming: The Bounded Variables Problem*, Graduate School of Industrial Administration, Carnegie Institute of Technology, 1954.

[4] Dantzig, G. B., "Upper Bounds, Secondary Constraints, and Block Triangularity in Linear Programming," *Econometrica* 23 (1955), 174–83.

[5] Goldman, A. J., and Tucker, A. W., "Theory of Linear Programming," in H. W. Kuhn and A. W. Tucker, *Linear Inequalities and Related Systems*, Princeton: Princeton University Press, 1955, pp. 53–98.

[6] Hastings, C., *Approximations for Digital Computers*, Princeton: Princeton University Press, 1955.

[7] Karst, O. J., "Linear Curve Fitting Using Least Deviations," *Journal of the American Statistical Association* 53 (1958), 118–32.

[8] Kelley, J. E., "An Application of Linear Programming to Curve Fitting," *Journal of Industrial and Applied Mathematics* 6 (1958), 15–22.

[9] Tucker, A. W., "Dual Systems of Homogeneous Linear Relations," in *Linear Inequalities*, see [5], 3–18.

[10] Wagner, H. M., "A Comparison of the Original and Revised Simplex Methods," *Operations Research* 5 (1957), 361–9.

[11] Wagner, H. M., "A Practical Guide to the Dual Theorem," *Operations Research* 6 (1958), 364–84.

[12] Wagner, H. M., "The Dual Simplex Algorithm for Bounded Variables," *Naval Research Logistics Quarterly* 5, No. 3, 1958.