Linear programming

1978

UCD Department of Agricultural Economics

WORKING PAPER SERIES

University of California, Davis
Department of Agricultural Economics

A DUAL LINEAR PROGRAMMING ALGORITHM FOR LEAST SQUARES ESTIMATION OF LINEAR MODELS

by

Quirino Paris

Working Paper No. 78-18

A DUAL LINEAR PROGRAMMING ALGORITHM FOR LEAST SQUARES ESTIMATION OF LINEAR MODELS

BY Quirino Paris

Department of Agricultural Economics, University of California, Davis, California.

SUMMARY

The neglected subject of duality in statistical estimation is discussed in relation to linear least squares theory. It is found that although the minimization of the sum of squared residuals can be formulated as a quadratic programming problem, its dual corresponds to a linear program whose interpretation is that of maximizing the value of sample information. As a consequence, an unbiased estimate of the error variance can be obtained as a linear function of the sample data.

Some key words: Duality; Least squares estimation; Linear programming; Stepwise regression.

1. INTRODUCTION

Least squares estimation of linear models has always been regarded as a quadratic programming problem. In such a specification, the sum of squared residuals is minimized subject to the linear constraints defining the linear model. This formulation may be regarded as a primal problem. Until recently, it has gone entirely unnoticed that the corresponding dual problem is a linear programming problem. Thus--paradoxically as at first it may appear—it is possible to obtain least squares estimates of linear models by means of linear programming. A series of interesting corollaries immediately follows. Firstly, if the linear specification corresponds to a "true" model, the linear programming estimates possess the properties of unbiasedness and minimum variance among all linear unbiased estimators. Secondly, an unbiased estimate of the error variance can be obtained as a linear function of the sample information. Thirdly, the linear programming specification can easily be extended to include the generalized least squares and the least squares estimator restricted by linear inequalities. A new reinterpretation of the theory of estimable functions is also possible. Finally, experimental design models can be analyzed using the structure of network theory.

The fundamental linear duality results of least squares theory are given in §2. The structure of the LP model is analyzed vis-a-vis the problems multicollinearity, stepwise regression, and least squares restricted by linear inequalities in §3. A two way classification experimental design model is discussed in §4.

2. THE DUALITY OF LEAST SQUARES THEORY

Consider the linear model

$$y = X\beta + e \qquad (1)$$

where y is a vector of n observations, X is a (n x p) full rank matrix of fixed regressors, β is a vector of unknown parameters, and e is a vector of random errors with zero expectation and homoscedastic variance σ^2. The primal problem of finding least squares estimates of β, say b, can be stated as

$$\text{minimize } (1/2)u'u \qquad (2)$$

subject to $Xb + u = y,$

b and u unrestricted.

Using the familiar Lagrangean approach the dual of problem (2) can easily be derived. Thus, the necessary and sufficient conditions for solving problem (2) are obtained by differentiating the following Lagrangean function

$$L = (1/2)u'u + \pi'(y - Xb - u) \tag{3}$$

with respect to u, b, and π, where π is the vector of Lagrange multipliers (dual variables) associated with the constraints of (2):

$$\partial L/\partial u = u - \pi = 0 \tag{4}$$

$$\partial L/\partial b = -X'\pi = 0 \tag{5}$$

$$\partial L/\partial \pi = y - Xb - u = 0. \tag{6}$$

Substituting constraints (4) and (5) into the function (3) the dual to problem (2) can be formulated as

$$\text{maximize } L = \pi'y - (1/2)u'u \tag{7}$$

$$\text{subject to} \quad u - \pi = 0 \tag{8}$$

$$X'\pi = 0 \tag{9}$$

$$u, \pi \text{ unrestricted.}$$

The first set of dual constraints (8) indicates the peculiar property of a least squares problem: the dual variables, π, are identical to primal variables, u. The second set (9) explicitly exhibits the orthogonality property between residuals, u, and the matrix of regressors, X. Problem (7) is obviously a quadratic programming problem. The linear programming specification of the dual can be obtained by successive use of conditions (4), (5), and (6) in the objective function (7):

$$\max L = \pi'y - (1/2)\pi'(y - Xb) \quad \text{using (4) and (6)} \tag{10}$$

$$= (1/2)\pi'y \quad \text{using (5).}$$

Hence, the dual of problem (2) can finally be stated as

$$\max \pi' y \tag{11}$$

subject to

$$Xb + \pi = y \tag{12}$$

$$X'\pi = 0 \tag{13}$$

$$b, \pi \text{ unrestricted.}$$

Constraints (12) and (13) are equivalent to the least squares estimator. In fact, premultiplying (12) by X' one gets $X'Xb + X'\pi = X'y$ which, in view of (13), corresponds to the least squares normal equations. The interpretation of (11) as the dual problem of (2) is illustrated by Figure 1. In this figure, the traditional primal problem is represented by the minimization of noise defined by the convex function $u'u$ of squared residuals. The dual problem (11) is represented by the supporting hyperplane, $\pi'y$, to be maximized. Hence, the interpretation of the dual objective function is that of maximizing the value of sample information defined as a weighted combination of the sample observations, y, using weights π, representing the marginal valuation the primal constraints in problem (2). Notice that the dual objective function can also be interpreted as a Legendre transformation.

A most interesting feature of this LP approach is that an unbiased estimate of the error variance σ^2 is given by $\hat{\sigma}^2 = \pi'y/(n - p)$ which is a simple linear combination of the sample information. This result follows from condition (6) since $\pi'y = \pi'u$, and remembering that $u = \pi$.

3. LEAST SQUARES ESTIMATES RESTRICTED BY LINEAR INEQUALITIES

Suppose that the parameters β of model (1) are further restricted by the following set of linear inequalities $R\beta \leq r$, where R is a (k x p) full rank matrix of known coefficients, and r is a known vector of constraints. Then, the following linear programming problem can be regarded as the restricted dual least squares estimator

$$\max \{\pi'y - \psi'r\} \tag{14}$$

subject to

$$X\tilde{b} + \pi = y \tag{15}$$

$$X'\pi - R'\psi = 0 \tag{16}$$

$$-R\tilde{b} \geq - r \tag{17}$$

$$\psi \geq 0, \tilde{b}, \pi \quad \text{unrestricted,}$$

where ψ are dual variables associated with the additional restrictions and the other symbols retain the previous meaning. The objective function (14) can be interpreted as maximizing the value of the sample information minus the value of the exogenous information. To show that problem (14) corresponds indeed to the desired estimator, it is sufficient to derive the primal problem corresponding to it, which turns out to be the traditional specification of the restricted least squares problem. Hence, by differentiating the Lagrangean function

$$L = \pi'y - \psi'r - u'(X'\tilde{b} + \pi - y) - \gamma'(X'\pi - R'\psi) - \phi'(-R\tilde{b} + r) \tag{18}$$

with respect to all the primal and dual variables, one obtains the following necessary and sufficient conditions for optimality of problem (14):

$$\partial L/\partial\pi = y - u - X\gamma = 0 \tag{19}$$

$$\partial L/\partial u = -y + \pi + X\tilde{b} = 0 \tag{20}$$

$$\partial L/\partial\tilde{b} = -X'u + R'\phi = 0 \tag{21}$$

$$\partial L/\partial\gamma = -X'\pi + R'\psi = 0 \tag{22}$$

$$\partial L/\partial\psi = -r + R\gamma \leq 0, \qquad \psi'(\partial L/\partial\psi) = -\psi'r + \psi'R\gamma = 0 \tag{23}$$

$$\partial L/\partial\phi = R\tilde{b} - r \leq 0, \qquad \phi'(\partial L/\partial\phi) = \phi'R\tilde{b} - \phi'r = 0. \tag{24}$$

By using, as appropriate, conditions (19) through (24) into the Lagrangean function (18), the primal objective function corresponding to the dual problem (14) can be stated as

$$\begin{aligned}\min L &= (\pi'y - \pi'X\gamma) + (u'y - u'X\tilde{b}) - u'\pi \\ &\quad + (\gamma'R'\psi - r'\psi) + (\phi'R\tilde{b} - \phi'r) \qquad (25) \\ &= \pi'u + u'\pi - u'\pi \qquad \text{using (23) and (25)} \\ &= \pi'u \qquad \text{using (19) and (20).}\end{aligned}$$

Notice that the rank assumptions about matrices X and R guarantee that there exists a unique solution to both system {(19), (21), and (23)} and system {(20), (22), and (24)}. Hence, $u = \pi$, $\tilde{b} = \gamma$, and $\phi = \psi$.

Therefore, the primal specification to the dual problem (14) is

$$\min u'u \qquad (26)$$

$$\text{subject to} \qquad X\tilde{b} + u = y$$

$$R\tilde{b} \leq r$$

$$\tilde{b}, u \text{ unrestricted,}$$

which exactly corresponds to the traditional least squares problem restricted by linear inequalities. Notice that in the restricted least squares, the orthogonality between the residual vector u and the regressor matrix X no longer holds.

A few remarks are in order. Exact multicollinearity among regressors is not a computational problem for the LP estimator (14). The LP algorithm will naturally select those regressors according to the criterion of maximizing the net value of sample information. This choice will avoid any unnecessary loss of information that might be encountered when regressors are arbitrarily dropped in order to avoid the computational difficulties caused by multicollinearity.

The LP estimator (14) can naturally be regarded as a sort of stepwise regression, since at each iteration the variable which most increases the value of sample information is introduced into the basic solution. By duality, increasing the value of sample information means reducing the sum of squared residuals.

An unbiased estimate of the error variance of the restricted linear model is given by $\tilde{\sigma}^2 = (\pi' y - \psi' r)/(u - p + k)$.

4. LP ESTIMATION OF EXPERIMENTAL DESIGN MODELS

A two way classification model (without interaction) of analysis of variance can be stated as follows

$$y_{ij} = \mu + \tau_i + \alpha_j + e_{ij} \tag{27}$$

where $i = 1, \ldots, I$, $j = 1, \ldots, J$, y_{ij} is the response to the ith and jth treatments, μ is the mean effect, τ_i is the ith row effect, α_j is the jth column effect, and e_{ij} is the experimental error associated with the ith and jth treatments. We further assume that $E(e_{ij}) = 0$ for all i's and j's and $E(e_{ij}\, e_{i'j'}) = 0$, $E(e_{ij}^{\;2}) = \sigma^2$, where E denotes the expectation operator. Model (27) can be restated in matrix form using the notation of model (1) where the matrix X takes on the following structure

$$X = \begin{array}{c} \begin{array}{ccc} \mu & \tau_i & \alpha_j \end{array} \\ \left[\begin{array}{ccccc} 1 & 1 & & 1 & \\ \vdots & \vdots & & & \ddots \\ 1 & 1 & & & 1 \\ \vdots & \ddots & & & \vdots \\ 1 & & 1 & 1 & \\ \vdots & & \vdots & & \ddots \\ 1 & & 1 & & 1 \end{array}\right] \end{array}$$

The estimation of model (27) by the least squares method requires the same restrictions derived for model (1) and, in particular, the condition $X'\pi = 0$, where π (with typical element π_{ij}) is the vector of dual variables (alternatively, of estimated residuals) corresponding to constraint (27). Given the structure of the X matrix, such a condition implies the following series of restrictions upon the estimated residuals

$$\Sigma_i \Sigma_j \pi_{ij} = \Sigma_i \pi_{ij} = \Sigma_j \pi_{ij} = 0. \qquad (28)$$

Notice that to obtain estimates of μ, τ_i and α_j by means of the LP algorithm suggested in previous sections, it is not necessary to reparameterize the X matrix according to the dictates of the theory of estimable functions. This theory requires that the parameters τ_i and α_j be restricted, for example, as

$$\Sigma_i \tau_i = \Sigma_j \alpha_j = 0. \qquad (29)$$

Of course, the projection of the y vector onto the X space is the same whether obtained via the impositions of either restrictions (28) or (29).

5. CONCLUSIONS

The novel result of this paper is that there exists a class of quadratic programming problems whose duals are linear programs. The least squares estimator of linear models belongs to this class, if the specified model is a "true" model. In this case, the least squares estimates obtained by linear programming are unbiased and of minimum variance among all linear unbiased estimates. It follows that if the linear model is misspecified, the least squares estimates of the unknown parameters are biased, and the dual problem of minimizing the sum of squared residual cannot be stated as a linear program. Hence, a one to one correspondence exists between unbiasedness of the least squares estimates and linearity of the dual objective function.

I wish to acknowledge helpful discussions with R. Pope, R. Green and D. Halimi without burdening them with the responsibility of any remaining error.

jma 12/12/78

FIGURE 1 Illustration of Duality in Least-Squares Theory