Linear programming
C

1978

UCD Department of Agricultural Economics

[redacted]

WORKING PAPER SERIES

University of California, Davis
Department of Agricultural Economics

Giannini Foundation Paper No. ___.

RESTRICTED AND GENERALIZED LEAST SQUARES BY LINEAR PROGRAMMING: DUALITY IN LEAST-SQUARES THEORY

by

Quirino Paris

Working Paper No. 78-14

Although any error is mine, I greatly benefited from discussing the subject of this paper with G. King, R. Pope, R. Green and D. Halimi.

Restricted and Generalized Least Squares by Linear Programming: Duality in Least-Squares Theory

1. INTRODUCTION

In 1973, Sielken and Hartley proposed two very interesting linear programming (LP) algorithms for unbiased estimation of linear models. The two criterion functions chosen by those authors were the minimization of the sum of absolute residuals and the minimization of the maximum absolute residual, respectively. The use of LP techniques for econometric purposes offers the appealing aspects of relative computational ease and, more importantly, of interpretative potential. Yet, it appears that the use of linear programming for deriving other more commonly-used estimators has gone unnoticed.[1/] In particular, the least-squares estimator can be obtained by a linear programming algorithm. The extension of the LP algorithm for computing generalized least-squares estimates and least-squares estimates subject to linear inequalities is straight-forward.

The main interest in the new LP algorithm lies in the dual interpretation of the least-squares problem rather than in computational aspects. Hence, the new development is relevant because it brings forward the almost completely neglected duality side of statistical estimation. The dual problem of minimizing the familiar sum of squared errors is given an elegant interpretation in terms of the supporting hyperplane to the convex function represented by the sum of squared residuals. On a more intuitive basis, the dual problem of minimizing noise defined by the sum of squared residuals can be interpreted as the problem of maximizing the value of information, defined as a weighted combination of the sample information. The weights are the dual variables of the problem of minimizing the sum of squared residuals subject to the linear regression system. Because of the self-duality property of the least-squares problem (clearly demonstrated by the proposed LP algorithm), the dual variables correspond identically to the primal variables. With hindsight knowledge, one might conjecture that the self-duality property of the least-square procedure constitutes the reason for the disregard of the duality side in statistical and econometric estimation, although one can hardly find such a statement in the published literature. It would also appear that self-duality has a great deal to do with unbiasedness of the estimator. Hence, the question of duality in biased estimators springs up naturally. This topic might be the subject of further investigation.

2. THE DUAL OF THE LEAST-SQUARES PROBLEM

Consider the linear model

$$y = X\beta + e \tag{2.1}$$

where y is a vector of $\underline{n}$ observations, X is an $\underline{n}$ x $\underline{p}$ matrix of rank $\underline{p}$ of fixed regressors, β is a vector of unknown parameters, and e is a vector of random variables with mean zero and homoscedastic variance σ^2. The problem is to find estimates b of the vector β of unknown parameters such that $y = Xb + u$, and b minimizes the quadratic form $u'u$ of residuals, where $u = y - Xb$. Equivalently, the primal least-squares problem can be formulated as the following quadratic programming specification:

$$\min\ (1/2)(u_1 - u_2)'(u_1 - u_2) \tag{2.2}$$

subject to
$$X(b_1 - b_2) + (u_1 - u_2) \le y$$
$$-X(b_1 - b_2) - (u_1 - u_2) \le -y$$
$$u_1 \ge 0,\ u_2 \ge 0,\ b_1 \ge 0,\ b_2 \ge 0,$$

where $u = u_1 - u_2$, and $b = b_1 - b_2$. In order to simplify the notation, the following equivalent formulation of (2.2) will be used throughout:

$$\min\ (1/2)u'u \tag{2.3}$$

subject to
$$Xb + u = y$$
$$b \text{ and } u \text{ unrestricted.}$$

The dual of problem (2.3) can be derived according to the usual Lagrangean procedure. Hence, choosing π as the vector of Lagrange multipliers, the necessary and sufficient conditions for maximizing (2.3) are obtained by differentiating the following Lagrangean function with respect to u, b and π

$$L = (1/2)u'u + \pi'[y - Xb - u] \tag{2.4}$$

and setting the corresponding derivatives equal to zero; that is:

$$\partial L/\partial u = u - \pi = 0, \tag{2.5}$$

$$\partial L/\partial b = -X'\pi = 0, \tag{2.6}$$

$$\partial L/\partial \pi = y - Xb - u = 0, \tag{2.7}$$

$$u'(\partial L/\partial u) = u'u - u'\pi = 0, \tag{2.5.i}$$

$$b'(\partial L/\partial b) = -b'X\pi = 0, \tag{2.6.i}$$

$$\pi'(\partial L/\partial \pi) = \pi'y - u'u = 0. \tag{2.7.i}$$

By substituting the dual constraints (2.5) and (2.6) into the Lagrangean function, the dual to problem (2.3) can be stated as

$$\max L = \pi' y - (1/2)u'u \tag{2.8}$$

subject to
$$u - \pi = 0 \tag{2.9}$$
$$X'\pi = 0 \tag{2.10}$$

u and π unrestricted.

Two important remarks are in order. Constraint (2.9) establishes that the dual variables π are identical to the primal variables u and, therefore, hints at the self-duality property of least squares. Constraint (2.10) makes explicit the characteristic orthogonality between the regressor matrix X and the residuals u.

As written above, the objective function (2.8) of the dual problem is, obviously, a quadratic (concave) function. Notice, however, that by successive use of conditions (2.5), (2.7) and (2.6), it can be transformed into a linear function as follows:

$$\begin{aligned} \max L &= \pi' y - (1/2)\pi'(y - Xb) && \text{using (2.5) and (2.7)} \quad (2.11) \\ &= \pi' y - (1/2)\pi' y + (1/2)\pi' Xb && \\ &= (1/2)\pi' y. && \text{using (2.6)} \end{aligned}$$

Thus, the dual of problem (2.3) can be restated as the linear program

$$\max \pi' y \tag{2.12}$$

subject to
$$\left.\begin{aligned} Xb + \pi &= y \quad \text{n of these} && (2.13) \\ X'\pi &= 0 \quad \text{m of these} && (2.14) \end{aligned}\right\}$$

b, π unrestricted.

have m+n indep. linear constraints and m+n variables (u, b), so constraints define the solution - obj. func. is irrelevant since feasible region is a single point

Several remarks come to mind. Clearly, the constraints of (2.12) are equivalent to the least-squares estimator. In fact, pre-multiplying (2.13) by X' one gets $X'Xb + X'\pi = X'y$, which in view of (2.14) corresponds to the system of least-squares normal equations. Secondly, the interpretation of (2.12) as the dual problem of (2.3) can be illustrated by Figure A.

In this figure, the traditional primal problem is represented by the convex function $u'u$ to be minimized. The dual problem (2.12) is represented by the supporting hyperplane, $\pi'y$, to be maximized. The objective function (2.12) can also be restated in terms of the Legendre transformation $\phi(g) = u'u - g'b$, where $g = \partial(u'u)/\partial b$. The maximization of $\phi(g)$ subject to the condition $g = 0$, corresponds to minimizing $u'u$. Obviously, $-(\partial\phi/\partial g) = b$, the least-squares estimator. Another appealing interpretation of the objective function (2.12) is that of maximizing the value of sample information defined as a weighted combination of the sample observations, y, with weights π representing the marginal valuation of the primal constraint in problem (2.3).

Perhaps, one of the most striking features of this LP approach is that an unbiased estimate of the variance σ^2 can be obtained as a simple linear combination of the sample information. In fact, from (2.7.i) such an unbiased estimate is $\hat{\sigma}^2 = u'u/(n - p) = \pi'y/(n - p)$.

Of course, if X is a matrix of full rank, problem (2.12) is a LP problem with a unique optimal solution.[2/] That solution must be the least-squares solution, as implied by the orthogonality condition (2.14). Finally, the self-duality of the least-squares problem can be made graphically explicit by considering (2.12), (2.13), and (2.14) as the primal problem. Choosing vectors u and d as the dual variables of constraints (2.13) and (2.14), respectively, one obtains the following dual problem

$$\min u'y \tag{2.15}$$

$$\text{subject to} \qquad X'u = 0$$

$$Xd + u = y$$

$$d, u \text{ unrestricted}$$

which is obviously identical to (2.12).

3. GENERALIZED LEAST-SQUARES ESTIMATES BY LINEAR PROGRAMMING

When the error term e of (2.1) is heteroscedastic with full rank covariance matrix Σ, the appropriate least-square procedure is to minimize a weighted sum of squared errors defined as $(1/2)u'\Sigma^{-1}u$, where Σ^{-1} is the inverse of the Σ matrix. If Σ is not known it can be estimated by $\hat{\Sigma}$ using the solution vector π of problem (2.12). The second stage of the procedure can thus be formulated as the following linear programming problem

$$\max \pi'\hat{\Sigma}^{-1}y \tag{3.1}$$

$$\text{subject to} \qquad \hat{\Sigma}^{-1}X\hat{b} + \hat{\Sigma}^{-1}\pi = \hat{\Sigma}^{-1}y$$

$$X'\hat{\Sigma}^{-1}\pi = 0$$

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

It can be easily verified that the above problem corresponds to the generalized least-squares estimator $\hat{b} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y$. The interpretation of (3.1) is entirely analogous to that of (2.12) with the only difference that the axes of the regressors X have been rotated by the matrix $\hat{\Sigma}^{-1}$.

4. RESTRICTED LEAST-SQUARES ESTIMATES BY LINEAR PROGRAMMING

Another interesting problem arises when the parameters of the linear model $y = X\beta + e$ (now taken as homoscedastic without loss of generality) are subject to linear inequality restrictions such as $R\beta \leq r$, where R is a $\underline{k}$ x $\underline{p}$ full rank matrix of known coefficients, and r is a known vector of constraints. The primal least-squares problem can thus be stated as follows:

$$\min \ (1/2)u'u \tag{4.1}$$

subject to

$$X\tilde{b} + u = y \tag{4.2}$$

$$R\tilde{b} \leq r \tag{4.3}$$

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

The derivation of the corresponding dual problem requires to differentiate the following Lagrangean function

$$L = (1/2)u'u + \pi'[y - X\tilde{b} - u] + \psi'[R\tilde{b} - r] \tag{4.4}$$

with respect to u, π, $\tilde{b}$ and ψ, where ψ is a nonnegative vector of Lagrange multipliers associated with constraint (4.3). The desired derivatives are as follows:

$$\partial L/\partial u = u - \pi = 0 \tag{4.5}$$

$$\partial L/\partial \tilde{b} = -X'\pi + R'\psi = 0 \tag{4.6}$$

$$\partial L/\partial \pi = y - X\tilde{b} - u = 0 \tag{4.7}$$

$$\partial L/\partial \psi = R\tilde{b} - r \leq 0, \qquad \psi'(\partial L/\partial \psi) = \psi' R\tilde{b} - \psi' r = 0. \tag{4.8}$$

Using conditions (4.5) through (4.8) in (4.4) as appropriate, the dual objective function of problem (4.1) can be stated as

$$\max \ (1/2)[\pi' y - \psi' r]. \tag{4.9}$$

Hence, the following linear programming problem can be regarded as the dual of (4.1):

$$\max \{\pi' y - \psi' r\} \tag{4.10}$$

subject to

$$X\tilde{b} + \pi = y \tag{4.11}$$

$$X'\pi - R'\psi = 0 \tag{4.12}$$

$$-R\tilde{b} \geq -r \tag{4.13}$$

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

The objective function (4.10) can now be interpreted as maximizing the value of sample information, $\pi' y$, minus the value of the exogenous information, $\psi' d$, or, in other words, maximizing the net value of sample information. Judge and Takayama, in their famous paper on regression analysis with inequality restrictions, formulated the same problem as a quadratic program and stated (p. 169): ". . . since the objective function (of the least-squares problem) is a quadratic form in β, use of the linear programming approach is precluded." To verify the correspondence between problem (4.10) and the least-squares estimator restricted by linear inequalities, it is sufficient to notice that a solution $\tilde{b}$, π, and ψ satisfying constraints (4.11), (4.12), and (4.13) corresponds to the desired least-squares estimator. In fact, premultiplying (4.11) by X' one gets

$$X'y = X'X\tilde{b} + X'\pi = X'X\tilde{b} + R'\psi$$

which can be solved for $\tilde{b}$ to obtain

$$\begin{aligned} \tilde{b} &= (X'X)^{-1}X'y - (X'X)^{-1}R'\psi \\ &= b - (X'X)^{-1}R'\psi \end{aligned} \tag{4.14}$$

easily recognizable as the least-squares estimator restricted by linear inequalities. In this case, an unbiased estimate of σ^2 is given by the linear function

$$\tilde{\sigma}^2 = (\pi' y - \psi' r)/(n + k - p).$$

5. CONCLUDING REMARKS

All the estimators in the above specifications are of the least-squares type. Except for the restricted case, when the error term, e, possesses a homoscedastic variance the associated estimator is best linear unbiased, in the sense that it is characterized by minimum variance among the class of linear unbiased estimators. In this case, the estimator possesses all the properties of a maximum likelihood estimator. If the variance of e is heteroscedastic, the corresponding estimator is only asymptotically unbiased. These properties, of course, are well known. What does not appear to be equally known is the fact that, although least-squares procedures can be formulated as a self-dual mathematical programming problem, the duality of minimizing the sum of squared errors (with or without inequality restrictions on the parameters) corresponds to a LP problem whose geometric interpretation fits the notion of duality in the sense of the Legendre transformation of minimizing the sum of squared error. The orthogonality between regressors and the residual errors, characteristic of least-squares estimators, allows the conversion of the Legendre transformation to a linear function interpreted as the supporting hyperplane to the convex set inscribed by the sums of the squared errors function. It is worth noticing again that this approach suggests the novel feature of obtaining an unbiased estimate of the variance σ^2 as a linear combination of the sample observations. Finally, the linear programming algorithm naturally implies a step-wise regression procedure, as at each iteration, that variable which most increases the value of sample information is introduced into the basic solution. By duality, increasing the value of information corresponds to reducing the sum of squared residuals.

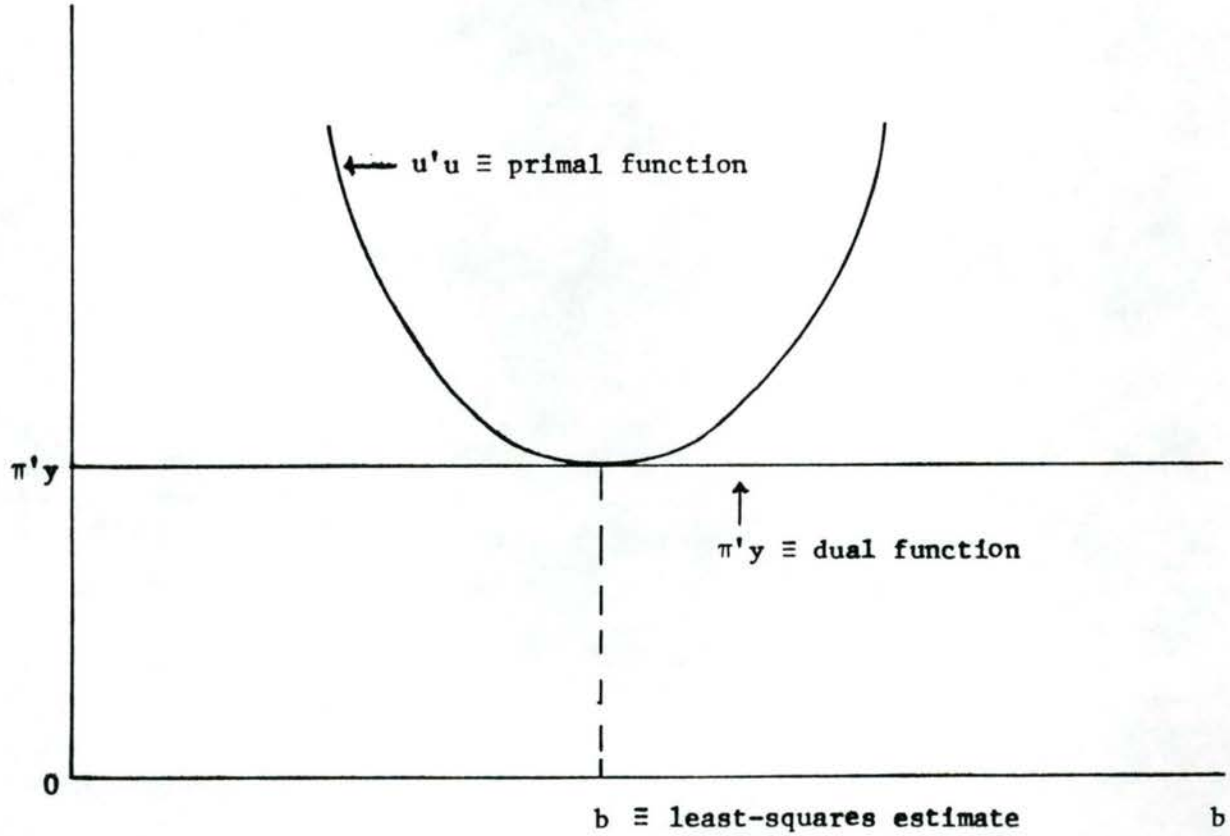
jma 12/8/78

FOOTNOTES

1/ Sielken and Hartley (see reference, p. 639) state that, to their knowledge, their two algorithms "are the only such algorithms which have been proposed and proven to be unbiased." (Emphasis is not mine.)

2/ If X is of rank $r < p$ one can either use the reparametrization approach traditionally used in least-squares theory, or allow the LP program to select a suitable basis. Usually, the reparametrization involves a subjective choice of the regressor(s) to be eliminated from the X matrix. This arbitrary choice may often imply elimination of unnecessary sample information. On the contrary, the LP formulation does not involve any a priori arbitrary choice and will eliminate regressor(s) according to the objective of retaining the largest amount of information.

FIGURE A. Illustration of Duality in Least-Squares Theory



REFERENCES

Judge, G. G., and T. Takayama (1966), "Inequality Restrictions in Regression Analysis," Journal of the American Statistical Association, 61, 166-181.

Sielken, R. L., and H. O. Hartley (1973), "Two Linear Programming Algorithms for Unbiased Estimation of Linear Models," Journal of the American Statistical Association, 68, 639-641.