Linear programming

1979

UCD Department of Agricultural Economics

WORKING PAPER SERIES

University of California, Davis
Department of Agricultural Economics

INEQUALITY RESTRICTED LEAST SQUARES BY LINEAR PROGRAMMING: DUALITY IN LEAST-SQUARES THEORY

by

Quirino Paris*

Working Paper No. 79-9

ABSTRACT

Inequality Restricted Least Squares by Linear Programming: Duality in Least-Squares Theory

Duality in statistical estimation is discussed within the context of linear least squares theory. The main result is that the quadratic programming problem of minimizing the residual sum of squared residuals subject to the linear model possesses a dual that is a linear program. The interpretation of this dual is that of maximizing the value of sample information defined as a linear combination of the sample data. Least-squares estimates restricted by linear (deterministic and stochastic) inequalities can easily be obtained by linear programming.

KEY WORDS: Linear programming; inequality restricted least squares; duality of statistical estimation, stochastic inequalities.

*Quirino Paris is Associate Professor, Department of Agricultural Economics, University of California, Davis, CA 95616. I wish to acknowledge the helpful discussions with G. King, R. Pope, R. Green, D. Halimi, Art Havenner, the students of my mathematical programming course, and the useful comments of the Journal's Editor as well as those of an anonymous Associate Editor, all of whom should not be responsible for any remaining error. Giannini Foundation Paper No. ___.

INEQUALITY RESTRICTED LEAST SQUARES BY LINEAR PROGRAMMING: DUALITY IN LEAST-SQUARES THEORY

1. INTRODUCTION

During the past two decades many statisticians have recognized that the specification of some relevant empirical problems involves the imposition of inequality restrictions upon the parameters of linear models. Accordingly, the traditional normal equations associated with unrestricted least-squares do not offer a feasible procedure for estimating the desired parameters. The problem of estimating by least-squares methods those linear models whose parameters are subject to inequality restrictions has always been considered only a quadratic programming (QP) problem. A clear expression of this sentiment can be found in a rather famous paper by Judge and Takayama who wrote (p. 169): ". . .since the objective function (of the inequality restricted least-squares problem) is a quadratic form in β (the parameter vector), use of the linear programming approach is precluded (emphasis added). Thus, unable to use conventional techniques in obtaining a solution to this constrained minimization problem, we reformulate the problem so that a solution can be obtained by quadratic programming procedures."

We conjecture that this statement--although incorrect--represents the belief of the totality of econometricians and statisticians. In fact, the perusal of the relevant literature has revealed a series of important studies which support the viewpoint of Judge and Takayama.

Theil and Rey, for example, were among the first to propose a quadratic programming algorithm for estimating the transition probabilities of a first-order Markov chain model. Malinvaud (pp. 317-319) reemphasizes the exclusive quadratic programming nature of the problem. A similar position is reasserted by Lovell and Prescott (p. 914). Liew confirms the quadratic programming specification very explicitly and goes as far as to propose the solution of the QP problem by either Lemke's or Dantzig-Cottle's algorithms for solving the corresponding linear complementarity problem. More recently, Lee, Judge and Zellner published a book entirely devoted to the subject of estimating alternative specifications of transition probability models by means of quadratic programming. But, perhaps, the most revealing statement about econometricians' opinion of the exclusive quadratic programming nature of the problem is provided by Arrow and Hoffenberg. The two authors were concerned with the problem of estimating interindustry input-output coefficients which, obviously, ought to be nonnegative. In a methodological chapter they wrote (pp. 55-56): "However, it did seem worth while to make some attempt to use inequalities in fitting. One possibility is to disregard the simultaneous equation nature of the problem and fit $r_i(t)$ to the remaining variables . . .by the method of least-squares, where the minimization is to be constrained by inequalities. This becomes a problem in quadratic programming. . .; the computational problems are large but by no means completely impractical. An alternative to minimizing the sum of squares is minimizing the sum of absolute deviations. From a theoretical point of view, it would be difficult to argue the superiority of one method over the other, while the latter has the advantage of being capable of expression as a linear programming problem." (Emphasis added.) The two authors thanked G. B. Dantzig for suggesting the alternative approach.

If Arrow and Hoffenberg, as well as Judge and Takayama, desired to use linear programming (LP) techniques for dealing with the estimation of inequality restricted linear models, they did not need either to abandon the least-squares approach or to rely exclusively upon quadratic programming. In fact, and contrary to a widespread belief, the least-squares estimation of inequality restricted linear models can be performed by linear programming.

This admittedly surprising result may be regarded--at least in some instances--as offering computational advantages especially when and where quadratic programming computer subroutines are either inaccessible or unreliable.

The same result, however, has a second and perhaps more interesting aspect. Indeed, one may conjecture that this unsuspected result went unnoticed because statisticians have not studied with any degree of fervor the structure of duality of the least-squares estimator. The linear programming specification presented in this paper is, in fact, dual to the problem of minimizing the sum of squared residuals. The corresponding geometric interpretation is that the objective function of the proposed LP problem constitutes a supporting hyperplane to the convex set defined by the function of squared residuals. Intuitively, the objective function of the new linear program can be interpreted as maximizing the net value of sample information. An interesting by-product of this new way of looking at least-squares estimation is that the variance of the error terms can be computed as a linear combination of the sample observations.

To facilitate the introduction and the discussion of duality in least-squares theory, Section 2 is devoted to the explanation of terminology and procedures connected with the duality of unrestricted least squares. Section 3 deals with the duality results associated with least-squares estimates restricted by exact linear inequality. A further analysis of the structure of least-squares duality is undertaken in Section 4 by means of the Legendre transformation of the sum of squared residuals. In Section 5, it is shown that the LP algorithm proposed in this paper can handle also the least-squares mixed estimation of linear models subject to stochastic inequalities. Some computational aspects of the new LP approach to least-squares estimation are reviewed in Section 6. Finally, in Section 7, a numerical application--requiring the estimation of a Cobb-Douglas production function--is used to illustrate the LP approach in terms of both unrestricted and inequality restricted least squares.

2. THE DUAL OF THE LEAST-SQUARES PROBLEM

Consider the linear model

$$y = X\beta + e \qquad (2.1)$$

where y is a vector of n observations, X is an n x p matrix of rank 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:

$$\min u'u/2 \qquad (2.2)$$

subject to

$$Xb + u = y$$

b and u unrestricted.

The specification of (2.2) is a typical constrained minimization problem whose solution (if it exists) is obtained by means of the Lagrangean method.[1/] Hence, choosing π as the n-dimensional vector of Lagrange multipliers, associated with the constraints of (2.2), the necessary and sufficient conditions for solving (2.2) are obtained by differentiating the following Lagrangean function with respect to u, b and π

$$L = u'u/2 + \pi'[y - Xb - u] \qquad (2.3)$$

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

$$\partial L/\partial u = u - \pi = 0, \qquad (2.4)$$

$$\partial L/\partial b = -X'\pi = 0, \tag{2.5}$$

$$\partial L/\partial \pi = y - Xb - u = 0. \tag{2.6}$$

In the terminology of mathematical programming, the relation (2.3) constitutes a saddle function whose saddlepoint (if it exists) provides a solution to the original problem (2.2). The objective function of the dual problem associated to the primal (2.2) is obtained by minimizing the Lagrangean function with respect to the Lagrange multipliers, π, for given values of u and b. In general terms, the derivatives of the Lagrangean function with respect to the primal variables constitute the dual constraints. Similarly, the derivatives of the same function with respect to the Lagrange multipliers (heretoforth called also dual variables) represent the primal constraints. Hence, in our case, relations (2.4) and (2.5) are dual constraints for the unrestricted least-squares problem, while relation (2.6) is, obviously, the primal constraint.

Therefore, by substituting the dual constraints (2.4) and (2.5) into the Lagrangean function, the dual to problem (2.2) can be stated as

$$\max L = \pi' y - u'u/2 \tag{2.7}$$

subject to

$$u - \pi = 0, \tag{2.8}$$

$$X'\pi = 0, \tag{2.9}$$

u and π unrestricted.

Two important remarks are in order. First of all, constraint (2.8) establishes that the dual variables π are identical to the primal variables u, that is, the Lagrange multipliers are identically equal to the residuals. Secondly, constraint (2.9) makes explicit the characteristic least-squares orthogonality between the regressor matrix X and the residuals u.

As written above, the objective function (2.7) of the dual problem is, obviously, a quadratic (concave) function. Notice, however, that by successive use of conditions (2.4), (2.6), and (2.5), it can be transformed into a linear function as follows:

$$\begin{aligned} \max L &= \pi'y - \pi'(y - Xb)/2 \qquad \text{(using (2.4) and (2.6))} \qquad (2.10) \\ &= \pi'y - \pi'y/2 + \pi'Xb/2 \\ &= \pi'y/2 \qquad \text{(using (2.5))}. \end{aligned}$$

Thus, the dual of problem (2.2) can be restated as the linear program

$$\max \pi'y/2 \qquad (2.11)$$

subject to

$$Xb + \pi = y, \qquad (2.12)$$

$$X'\pi = 0, \qquad (2.13)$$

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

Constraint (2.12) replaces (2.8) since $u = y - Xb$. Notice that problems [(2.7) - (2.9)] and [(2.11) - (2.13)] are equivalent in the sense that they provide exactly the same solutions. It is very easy, in fact, to reconstruct the specification (2.7) - (2.9) beginning from (2.11) - (2.13), by further use of relations (2.4) - (2.6). Hence, both formulations can be regarded as representing the dual of the unrestricted least-squares problem. Of course, the linear programming version is of greater interest because it was previously unsuspected and, furthermore, it may be easier to implement computationally than the QP version (2.7) - (2.9).

Several remarks come to mind. Clearly, the constraints of (2.11) are equivalent to the least-squares estimator. In fact, premultiplying (2.12) by X' one gets $X'Xb + X'\pi = X'y$, which in view of (2.13) corresponds to the system of least-squares normal equations. Secondly, the interpretation of (2.11) as the dual problem of (2.2) 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 (2.11) is represented by the supporting hyperplane, $\pi'y$, to be maximized. Another appealing interpretation of the objective function (2.11) 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 (2.2). This interpretation is suggested by the fact that the sample information is contained in the quantities constituting the sample observations and by adopting the traditional economic interpretation of the dual variables, π, as "shadow prices". Hence, if the vector y represents the "quantity" of information contained in the sample and the vector π represents the associated "prices," the linear combination $\pi'y$ may be regarded as the value of sample information.

Perhaps, the most striking feature of this LP approach is the discovery that an unbiased estimate of the variance σ^2 can be obtained as a linear combination of the sample observations, y. In fact, premultiplying (2.6) by π' and using (2.4) and (2.5), such an unbiased estimate is $\hat{\sigma}^2 = u'u/(n - p) = \pi'y/(n - p)$. This appears to be an entirely novel procedure for computing the variance.[2/] Notice also that, although the dual variables, π, depend upon the sample observations, y, the term $\pi'y$ is still to be regarded as a linear combination exactly as in the conventional linear programming specifications.

3. A LINEAR PROGRAM FOR INEQUALITY RESTRICTED LEAST-SQUARES

We can now tackle the problem which concerned Arrow and Hoffenberg, Theil and Rey, Judge and Takayama, and other researchers. The problem is to determine whether linear programming techniques can be used to obtain least-squares estimates of parameters in linear models restricted by linear inequalities. In view of the results obtained in Section 2, the answer, of course, is positive and can be articulated as follows.

Let us suppose that the parameters of the linear model $y = X\beta + e$, whose elements were defined in Section 2, are subject to linear inequality restrictions such as $R\beta \leq r$, where R is a $k \times p$ full rank matrix of known coefficients, and r is a known vector of constraints. In this case, the primal least-squares problem can be stated as follows:

$$\min u'u/2 \tag{3.1}$$

subject to

$$X\tilde{b} + u = y, \tag{3.2}$$

$$R\tilde{b} \leq r, \tag{3.3}$$

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

The derivation of the corresponding dual problem follows the outline developed in Section 2 and requires to differentiate the following Langrangean function

$$L = u'u/2 + \pi'[y - X\tilde{b} - u] + \psi'[R\tilde{b} - r] \tag{3.4}$$

with respect to u, π, $\tilde{b}$ and ψ, where ψ is a nonnegative vector of Lagrange multipliers associated with constraint (3.3). The constrained minimization problem (3.1) - (3.3) is of a slightly different structure than that in Section 2 because of the presence of the inequality constraints (3.3). For this reason the classical Lagrangean method must be modified according to the theory of nonlinear programming developed by Kuhn and Tucker.[3/] Hence, the desired derivatives (called also Kuhn-Tucker conditions) are as follows:

$$\partial L/\partial u = u - \pi = 0, \tag{3.5}$$

$$\partial L/\partial \tilde{b} = -X'\pi + R'\psi = 0, \tag{3.6}$$

$$\partial L/\partial \pi = y - X\tilde{b} - u = 0, \tag{3.7}$$

$$\partial L/\partial \psi = R\tilde{b} - r \leq 0, \quad \psi'(\partial L/\partial \psi) = \psi' R\tilde{b} - \psi' r = 0. \tag{3.8}$$

Using conditions (3.5) through (3.8) in (3.4), as appropriate, the objective function of the dual problem can be transformed into a linear function as follows:

$$\max L = u'u/2 + \pi'y - \psi'r - u'u + (\psi'R - \pi'X)\tilde{b} \qquad (3.9)$$

$$= \pi'y - \psi'r - u'u/2 \qquad \text{by (3.5) and (3.6)}$$

$$= \pi'y - \psi'r - \pi'(y - X\tilde{b})/2 \qquad \text{by (3.5) and (2.7)}$$

$$= \pi'y/2 - \psi'r + \psi'R\tilde{b}/2 \qquad \text{by (3.6)}$$

$$= \{\pi'y - \psi'r\}/2 \qquad \text{by (3.8)}$$

In its last equivalent specification this function is linear since both the y and r vectors are known.

Hence, the following linear programming problem can be used to obtain least-squares estimates of linear models subject to inequality restrictions:

$$\max \{\pi'y - \psi'r\} \qquad (3.10)$$

subject to

$$X\tilde{b} + \pi = y \qquad (3.11)$$

$$X'\pi - R'\psi = 0 \qquad (3.12)$$

$$-R\tilde{b} \geq - r \qquad (3.13)$$

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

The objective function (3.10) 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.[4/] To verify the correspondence between problem (3.10) - (3.13) and the least-squares estimator restricted by linear inequalities, it is sufficient to notice that a solution $\tilde{b}$, π, and ψ satisfying constraints (3.11), (3.12), and (3.13) corresponds to the desired least-squares estimator. In fact, premultiplying (3.11) by X' one gets

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

which--in view of the full rank property of X--can be solved for $\tilde{b}$ to obtain

$$\begin{aligned}\tilde{b} &= (X'X)^{-1}X'y - (X'X)^{-1}R'\psi \\ &= b - (X'X)^{-1}R'\psi \qquad (3.14)\end{aligned}$$

easily recognizable as the least-squares estimator restricted by linear inequalities.[5/] It is, thus demonstrated that also the least-squares method (just like the least absolute deviation method) for estimation of linear models is capable of expression as a linear programming problem.

4. A FURTHER ANALYSIS OF LEAST-SQUARES DUALITY: THE LEGENDRE TRANSFORMATION

An alternative and perhaps deeper analysis of duality in least-squares theory can be developed in terms of the Legendre transformation. For simplicity we refer the discussion of this section to the unrestricted least-squares problem of Section 2.

By definition, the Legendre transformation of the differentiable function f(z), where z is a vector variable, is given by the function $\phi(d) = f(z) - d'z$, where $d \equiv \partial f/\partial z$ is the gradient vector of f(z). By applying this notion to the least-squares problem, the function

$f(\cdot)$ is given by the quadratic form $f(b) = u'u$. Furthermore, $d = X'Xb - X'y$. Hence, the corresponding Legendre transformation is:

$$\phi(d) = u'u - d'b \qquad (4.1)$$

$$= (y - Xb)'(y - Xb) - (X'Xb - X'y)'b$$

$$= y'(y - Xb)$$

$$= y'u.$$

Since the function $f(b)$ is convex, the maximization of the Legendre transformation, $\phi(d)$, subject to the condition $d = 0$, corresponds to the minimization of $f(b)$. Therefore, $\phi(0) = u'u = u'y$. In Figure 1, the function $\phi(d)$ can be interpreted as the intercept on the residual sum of squares (RSS) axis of the tangent to the convex function $u'u$. The maximization of the intercept, subject to the condition $d = 0$, is equivalent to the minimization of $u'u$.

In other words, the Legendre transformation (4.1) represents the family of tangents to the convex function $u'u$. Since tangents are, obviously, linear functions, the optimal tangent for a least-squares problem is that which corresponds to the minimum of the function $u'u$. Such a tangent has intercept $\phi(d = 0)$ and corresponds to the hyperplane $y'u$ or, equivalently, $y'\pi$.

5. LEAST-SQUARES MIXED ESTIMATION BY LINEAR PROGRAMMING

Oftentimes, theoretical conclusions or personal belief about events suggest the use of this prior information in conjunction with the sample information for estimating the parameters of a model.

While, in Section 3, the exogenous information was assumed to exhibit the structure of exact (or deterministic) inequalities ($Rb \leq r$), a degree of uncertainty about the exact nature of this information can be easily specified by introducing a random element, v, into the inequalities.

Hence, suppose that the parameters of the linear model ($Xb + u = y$) obey also the following stochastic inequality system

$$Gb \leq g + v \tag{5.1}$$

where G is a $\ell \times p$ matrix and g is a $\ell \times 1$ vector of known coefficients. Each component of the $\ell \times 1$ random vector is supposed to have zero expectation and variance σ_v^2. The intuitive interpretation of relation (5.1) is that the researcher does not possess a 100 percent confidence (or knowledge) about the structure of the inequalities that the parameter vector b ought to satisfy. He, therefore, allows a margin of error, v, in the specification of the vector g. It is desirable, however, to make such an error as small as possible. Hence, the appropriate primal least-squares problem is specified as follows

$$\min \{u'u/2 + v'v/2\} \tag{5.2}$$

subject to

$$Xb + u = y \tag{5.3}$$

$$Gb - v \leq g \tag{5.4}$$

b, u, v unrestricted.

By following, step by step, the procedure developed in Section 3, it is possible to arrive at the specification of the following linear program:

$$\max \{\pi' y - \psi_s' g\}/2 \quad (5.5)$$

subject to

$$Xb + \pi = y \quad (5.6)$$

$$X'\pi - G'\psi_s = 0 \quad (5.7)$$

$$Gb - \psi_s \leq g \quad (5.8)$$

ψ, b, π unrestricted.

The dual variables ψ_{si}, $i = 1 \ldots \ell$, represent the cost of satisfying the stochastic inequalities (5.1). Although we use the same symbol, ψ, as in Section 3, they are obviously of a different nature than those associated with the exact inequalities $Rb \leq r$. In this case, the specification contains two sources of error: the sample and the a priori information. The formulation in problem (5.2) - (5.4) or, alternatively, problem (5.5) - (5.8), corresponds to a slightly more general version of the mixed estimation specification discussed, for example, by Goldberg and Theil.

6. COMPUTATIONAL ASPECTS OF THE LP ALGORITHM

No particular problem exists for implementing the linear programming formulations developed in previous sections using the LP subroutines conventionally available at computer centers. The only requirements are that they be reliable and, secondly, that all the unrestricted variables appearing in the LP formulations be redefined as the difference of two nonnegative variables. Hence, by letting $b = b_1 - b_2$ and $\pi = \pi_1 - \pi_2$, where b_1 and b_2 are p x 1 and π_1 and π_2 are n x 1 nonnegative vectors, the more explicit version of the unrestricted least-squares problem (2.1) - (2.13) becomes

$$\max y'(\pi_1 - \pi_2)/2 \tag{6.1}$$

subject to

$$X(b_1 - b_2) + (\pi_1 - \pi_2) = y,$$

$$X'(\pi_1 - \pi_2) = 0,$$

$$\pi_1 \geq 0,\ \pi_2 \geq 0,\ b_1 \geq 0,\ b_2 \geq 0.$$

Similarly, the inequality restricted least-squares problem (3.10) - (3.13) must be converted to the following computer-ready specification

$$\max \{y'(\pi_1 - \pi_2) - \psi' r\} \tag{6.2}$$

subject to

$$X(\tilde{b}_1 - \tilde{b}_2) + (\pi_1 - \pi_2) = y \tag{6.3}$$

$$X'(\pi_1 - \pi_2) - R'\psi = 0 \tag{6.4}$$

$$-R(\tilde{b}_1 - \tilde{b}_2) \geq -r \tag{6.5}$$

$$\psi \geq 0,\ \tilde{b}_1 \geq 0,\ \tilde{b}_2 \geq 0,\ \pi_1 \geq 0,\ \pi_2 \geq 0.$$

It is well known that the available and efficient LP subroutines based upon the simplex algorithm can handle several hundred constraints and several thousands of variables. Therefore, the dimensionality of the least-squares problem cannot possibly present any inconvenience. In principle, it would clearly seem that linear programming is a simpler and more efficient computational technique than either quadratic programming or any other algorithm for the linear complementarity problem. This belief must have been shared also by Arrow and Hoffenberg, Judge and Takayama while carrying out their original studies. In practice, however, the advantages and disadvantages of any algorithm constitute an empirical question. They depend, to a large extent, upon the sophistication of the programming language and diagnostic techniques used for implementing the desired algorithm on the computer.

Consider first the linear programming problem (6.2). With n observations, p parameters and k constraints, its dimensionality (that is, the dimensionality of the associated simplex tableau) is of $(n + p + k)$ rows and $(2n + 2p + k)$ columns. The corresponding input requirements are the matrix of regressors X, the vector of observations y, the matrix R and vector r describing the constraints. No computations or transformations of the sample data are necessary prior to inputting the sample information.

Consider now the quadratic programming algorithm suggested by Judge and Takayama, as well as by Lee, Judge, and Zellner for solving inequality restricted least squares. Using as a reference the simplex tableau of Judge and Takayama (p. 172) it would appear that the dimensionality of their quadratic programming formulation is of $(p + k)$ rows and $(2p + 2k)$ columns.

Notice, however, that the reduced dimensions of the QP specification is associated with more complex input requirements and lack of final information about residuals as compared to the LP method. In fact, the QP approach of Judge and Takayama requires that the matrix $X'X$ as well as the vector $X'y$ be inputted necessitating, therefore, nonnegligible preliminary data manipulations. Furthermore, their QP approach does not directly provide either the series of residual terms or the sum of their squares. These estimates will have to be computed separately, after the estimates of the coefficients are obtained.

The LP formulation, on the contrary, does not require any prior transformation of the sample information and provides the estimates of the coefficients, error terms and their sum of squares directly in the solution of the LP problem. The inverse of the optimal basis associated with the optimal solution gives the variance-covariance matrix of the coefficients.

The comparison between the LP algorithm proposed in this paper and the linear complementarity problem suggested by Liew follows more or less the lines indicated above for the LP versus QP algorithms, with one additional advantage in favor of the LP specification. In general, LP algorithms based upon the simplex method are articulated in two phases: Phase I, which utilizes the dual-simplex algorithm, establishes the feasibility of the solution, while Phase II--using the primal-simplex algorithm--attempts to achieve optimality. If, during computations performed in Phase II, feasibility is lost (due to rounding off errors, for example), the subroutine reverts to Phase I and so on. The algorithms for solving the linear complementarity problem of Lemke as well as that of Dantzig and Cottle, cited by Liew, do not possess this double phase articulation, and, in large problems, may encounter difficulties in successfully completing the computations, due principally to rounding-off errors.

So far, the matrix of fixed regressors, X, was assumed to be of full rank. The linear programming formulation proposed in this paper, however, is not restricted to this assumption. 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 unneccessary 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.

In the case of a non-full-rank matrix X, it is admissible that the LP formulation possesses alternative optimal solutions. But also this aspect does not constitute any problem whatsoever. To see this, it is sufficient to recall the theory of estimable functions. Under this theory, the alternative schemes of reparametrization which might exist give rise to the same projection of the vector of observations, y, onto the space of regressors, X.[6/] Of course, the sets of individual coefficient estimates associated with the alternative reparametrizations would differ from each other. Analogously, in the non-full-rank LP specification all the alternative optimal solutions produce the same minimization of the residual sum of squares and, therefore, the same projection of y onto X. Of course, the individual elements of the alternative optimal solutions are not the same but this is no different from the conventional approach toward reparametrization.

Furthermore, the LP formulation naturally provides a criterion for stepwise regression. In fact, at each iteration the variable that most increases the value of the objective function, $\pi'y$ (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.

Finally, notice that the LP algorithm proposed in this paper is capable to provide generalized least-squares estimates of the parameters.

When the error term e of (2.1) is heteroscedastic with full rank covariance matrix Σ, the appropriate least-squares procedure is to minimize a weighted sum of squared errors defined as $u'\Sigma^{-1}u/2$ 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 (2.11). The second stage of the procedure can thus be formulated as the following linear programming problem

$$\max \pi'\hat{\Sigma}^{-1}y \tag{6.6}$$

subject to

$$\hat{\Sigma}^{-1}X\hat{b} + \hat{\Sigma}^{-1}\pi = \hat{\Sigma}^{-1}y,$$

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

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

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 (6.6) is entirely analogous to that of (2.11) with the only difference that the axes of the regressors X have been rotated by the matrix $\hat{\Sigma}^{-1}$. A linear programming specification can easily be formulated for the case of generalized least-squares restricted by linear inequalities.

7. A NUMERICAL ILLUSTRATION

To illustrate the theoretical development outlined in previous sections we will now discuss three numerical examples using the same sample data presented in Table 1. The information refers to a ten-year period of the U.S. private nonfarm GNP which is postulated to have been generated by a Cobb-Douglas technology specified by capital and labor:

$$Y_i = BK_i^{b_1}L_i^{b_2}E_i, \quad i = 1, \ldots, 10 \tag{7.1}$$

where E is a multiplicative random disturbance and B, b_1, and b_2 are the coefficients to be estimated. When using the least-squares approach it is customary to take the natural log-transform of (7.1) which can be stated as

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i \quad i = 1, \ldots, 10 \qquad (7.2)$$

where $y_i \equiv \log Y_i$, $X_{1i} \equiv \log K_i$, $X_{2i} = \log L_i$, $b_0 \equiv \log A$, $e_i \equiv \log E_i$. The transformed random error e_i is assumed to possess zero expectation and constant variance σ^2 and, furthermore, it is independent of all other observations.

Unrestricted Least-Squares Estimation

The first example consists in the computation of the unrestricted least-squares estimates of the parameters β_0, β_1, β_2. For this purpose, we adopt the specification (6.1), where y is a (10 x 1) vector and X is a 10 x 3 matrix of regressors.

The estimated coefficients, the associated residuals and the corresponding sum of squared residuals (RSS) are presented in Table 2. The computations were carried out on a 1130 IBM computer using the 1130LP-MOSS software for linear programs. Exactly the same figures were obtained using a conventional regression package based upon normal equations.

Least-Squares Estimation with Deterministic Inequalities

The second numerical example specifies the estimation of the Cobb-Douglas coefficients in (7.2) under the restrictions that $b_1 \leq 3b_2$ and $b_1 \geq 0$, $b_2 \geq 0$. In this case, to obtain the desired least-squares estimates by linear programming, one must use the specification given in (6.2) - (6.5) with the following modifications: since b_1 and b_2 are supposed to be nonnegative, the rows of (6.4) which (by duality) correspond to these coefficients are now inequalities. Furthermore, R = (0, 0, 1, -3) and r = 0. A configuration of the structure of the simplex tableau corresponding to this example is given in Table 3. The type of constraints characterizing this problem is indicated in the second column where E = equality, L = less than or equal and G = greater than or equal. The letter N indicates the objective function's row, while the letter R represents a real number between 10 and 100. The least-squares estimates of the coefficients, residuals, their sum of squares, and of the dual variable ψ are presented in Table 2. The computations were carried out with the same LP software and hardware indicated previously. Exactly the same solution was obtained by quadratic programming using a computer subroutine prepared for the Rand Corporation by Cutler and Pass and implemented on a Burroughs 7700/6700 system.

Least-Squares Estimation with Stochastic Inequalities

The third numerical illustration postulates that knowledge of the ratio between the coefficients b_1 and b_2 is not certain. Furthermore, theory suggests that an economic equilibrium is achieved only when the production function exhibits either decreasing or, at most, constant returns to scale. Given the nature of the sample data, also this second inequality is not applicable with full confidence. Hence, the new specification introduces stochastic inequalities as follows

$$b_1 - 3b_2 \leq v_1$$

$$b_1 + b_2 \leq 1 + v_2$$

where v_1 and v_2 are random errors with zero expectation, homoscedastic variance σ_v^2 and zero covariance.[7/] In this case, the appropriate LP specification of the problem is that given in (5.5) - (5.8), modified to account for the nonnegativity of the coefficients b_1, b_2. The corresponding estimates are presented in Table 2. Exactly the same estimates were obtained by quadratic programming.

In the course of developing the least-squares estimates for the numerical examples presented here, it was realized that the linear programming formulation of the least-squares problem possesses also a practical advantage as a method of checking the reliability of LP computer subroutines. Indeed, it is in the theoretical nature of the LP specification of the least-squares problem (either restricted or unrestricted) that the primal solution must be exactly equal to the dual solution. Hence, the LP least-squares formulation contains in itself an absolute consistency check. If one inputs the data correctly but does not obtain the verification that the primal solution is equal to the dual solution, he can conclude that the LP software is unreliable. In fact, using this check, it was discovered that the LP subroutine copyrighted by Burroughs Corporation for its 7700/6700 system since 1971 and called "TEMPO," does not give the correct primal optimal solution and the corresponding value of the objective function.

8. CONCLUSIONS

The study of the structure of duality in least-squares theory has uncovered unsuspected results. The most interesting is that least-squares estimates can be obtained by formulating an appropriate linear programming problem. The advantages of this discovery are both computational and interpretative. I conjecture that several other useful insights and uses of this novel LP approach to least squares will be added shortly to those presented in this paper.

ke
8/7/79

FOOTNOTES

1/ For a clear exposition of the Lagrangean method as it applies to constrained minima see Hadley (pp. 60-75).

2/ A rather exhaustive search of the statistical literature has failed to reveal any hint that the error variance in least-squares may be computed as a linear combination of the sample observations using the residuals as weights.

3/ A very intelligible discussion of Kuhn-Tucker theory of non-linear programming can be found in Hadley (pp. 185-205).

4/ The factor (1/2), appearing in (3.9), can be dropped in the LP problem (3.10) without modifying the solution.

5/ See, for example, Theil, p. 44, equation (8.7).

6/ See, for example, Graybill, p. 229.

7/ The homoscedasticity of the variance of v_1 and v_2 as well as the zero covariance between the two terms is assumed for simplicity. A more complex specification of the error terms can be easily handled by the LP algorithm.

REFERENCES

Arrow, K. J. and M. Hoffenberg (1959), A Time Series Analysis of Interindustry Demands, Amsterdam: North-Holland.

Cutler, L. and D. S. Pass (1971), A Computer Program for Quadratic Mathematical Models to be Used for Aircraft Design and Other Applications Involving Linear Constraints, R-516-PR, Santa Monica, Rand Corporation.

Goldberg, A. S. (1964), Econometric Theory, New York: John Wiley.

Graybill, F. A. (1961), An Introduction to Linear Statistical Models, New York: McGraw-Hill.

Hadley, G. (1964), Nonlinear and Dynamic Programming, Reading, MA: Addison-Wesley.

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

Kuhn, H. W. and A. W. Tucker (1951), "Nonlinear Programming," in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Neyman, J., Editor, Berkeley: University of California Press, 481-492.

Lee, T. C., G. G. Judge and A. Zellner (1977), Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, second revised edition, Amsterdam: North-Holland.

Liew, C. K. (1976), "Inequality Constrained Least-Squares Estimation," Journal of the American Statistical Association, 71, 746-751.

Lovell, M. C. and E. Prescott (1970), "Multiple Regression with Inequality Constraints: Pretesting Bias, Hypothesis Testing and Efficiency," Journal of the American Statistical Association, 65, 913-925.

Malinvaud, E. (1966), Statistical Methods of Econometrics, Chicago: Rand McNally.

Sato, R. (1970), "The Estimation of Bias Technical Progress in the Production Function," International Economic Review, 11, 179-208.

Theil, H. and G. Rey (1966), "A Quadratic Programming Approach to the Estimation of Transition Probabilities," Management Science, 12, 714-721.

Theil, H. (1971), Principles of Econometrics, New York: John Wiley.

TABLE 1

The Sample Data

Year	Private nonfarm GNP Y	Capital K	Man hours L
1951	173,398	317,629	104,801
1952	178,864	332,480	106,168
1953	186,264	343,207	109,195
1954	184,482	344,371	103,523
1955	200,993	362,673	107,954
1956	205,730	376,986	124,341
1957	210,125	389,533	111,104
1958	205,700	383,892	106,250
1959	221,385	402,047	110,732
1960	227,593	412,304	111,881

Source: R. Sato

TABLE 2

LP Estimates of Alternative Least-Squares Specifications

Estimates	Unrestricted Cobb-Douglas	Cobb-Douglas with deterministic inequalities	Cobb-Douglas with stochastic inequalities
RSS	.00118	.00240	.00511
b_o	-1.66522	-3.39608	-.44618
b_1	1.03118	.93520	.75911
b_2	.05662	.31173	.25189
$u_1 = \pi_1$	.01038	.00806	-.01916
$u_2 = \pi_2$	- .00644	- .00768	-.02608
$u_3 = \pi_3$	- .00024	- .00560	-.01673
$u_4 = \pi_4$	- .01032	- .00175	-.01547
$u_5 = \pi_5$	.01963	.02247	.02038
$u_6 = \pi_6$	- .00499	- .03449	-.02131
$u_7 = \pi_7$	- .01124	- .00888	.00333
$u_8 = \pi_8$	- .01495	- .00259	.00437
$u_9 = \pi_9$	.00854	.01480	.03237
$u_{10} = \pi_{10}$	.00964	.01567	.03830
ψ_1	--	.00142	.00344
ψ_2	--	--	.01100

TABLE 3

Structure of the Simplex Tableau for Example 2

	Constraint Type	COL01	COL02	COL03	COL04	COL05	COL06	COL07	COL08	COL09	COL10	COL11	COL12	COL13	COL14	COL15	COL16	COL17	COL18	COL19	COL20	COL21	COL22	COL23	COL24	COL25	RHS
ROWOBJ	N					B	B	B	B	B	B	B	B	B	B	-B	-B	-B	-B	-B	-B	-B	-B	-B	-B		
ROW 01	E	1	-1	B	B	1										-1											B
ROW 02	E	1	-1	B	B		1										-1										B
ROW 03	E	1	-1	B	B			1										-1									B
ROW 04	E	1	-1	B	B				1										-1								B
ROW 05	E	1	-1	B	B					1										-1							B
ROW 06	E	1	-1	B	B						1										-1						B
ROW 07	E	1	-1	B	B							1										-1					B
ROW 08	E	1	-1	B	B								1										-1				B
ROW 09	E	1	-1	B	B									1										-1			B
ROW 10	E	1	-1	B	B										1										-1		B
ROW 11	E					1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1		
ROW 12	E					-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1		
ROW 13	L					B	B	B	B	B	B	B	B	B	B	-B	-B	-B	-B	-B	-B	-B	-B	-B	-B	-1	
ROW 14	L					B	B	B	B	B	B	B	B	B	B	-B	-B	-B	-B	-B	-B	-B	-B	-B	-B	A	
ROW 15	G			-1	A																						

FIGURE A. Illustration of Duality in Least-Squares Theory