The Constant Elasticity of Substitution Production Function and Its Application in Research

Marlen F. Miller *Norman K. Whittlesey*
Terry N. Barr

COLLEGE OF AGRICULTURAL RESEARCH CENTER
Washington State University Technical Bulletin 80

Published by the College of Agriculture Research Center, Washington State University August, 1975

CONTENTS

THE AUTHORS AND THIS BULLETIN

Marlen Miller is a former assistant agricultural economist in the department of agricultural economics, Washington State University. He is now chairman of the department of economics at Pacific Lutheran University.

Norman K. Whittlesey is an agricultural economist, department of agricultural economics, Washington State University.

Terry Barr was a graduate student in the department of agricultural economics, WSU. He is now an economist with the Economic Research Service, U.S. Department of Agriculture.

The work reported here was done under projects 0201 and 1386.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of those who reviewed the manuscript and suggested many improvements, particularly Leroy Blakeslee, Dave Price, and Joseph Roop. Grateful appreciation is also extended to Gayle Beckstrom for checking the notation and mathematics throughout the manuscript. Nevertheless, the authors accept full responsibility for any errors that may still exist.

THE CONSTANT ELASTICITY OF SUBSTITUTION PRODUCTION FUNCTION AND ITS APPLICATION IN RESEARCH

Marlen F. Miller, Norman K. Whittlesey and Terry N. Barr

SUMMARY AND CONCLUSIONS

The major objective of this study was to shed some light on the controversy over which form of production function, Cobb-Douglas or Constant Elasticity of Substitution (C-D and CES), should be preferred in applied research. Toward this end, the theoretical superiority of the CES function was outlined. It became clear that no definitive statement about statistical superiority was possible on the basis of one application. Indeed, the CES function must meet the same test that the C-D function has undergone, namely the test of the robustness of the application to various data and problems. To test the CES function, the problem of estimation was considered and a relatively simple estimation technique used.

In the application of the CES function, the elasticity of substitution was approximately equal to one, implying a reduction of the CES to the C-D. However, the estimating technique performed relatively well near an indeterminate region with regard to the CES. Achievement of convergence from both sets of starting values would have been very desirable; however, the computer program would not permit a better approximation near an indeterminate form of the CES. The approximately equal marginal value products under two different algebraic specifications (CES and C-D) were particularly encouraging with respect to policy questions. If the marginal value products had differed substantially with g near zero, it would be hard to support policy recommendations in future applications where g differed from zero.

The CES specification, while not yielding substantially different marginal value products, did provide parameter estimates of greater precision for machinery investment, family labor, and land investment. Based on the t-test for the C-D formulation, the coefficients on these inputs were not significantly different from zero. Under the CES specification, the coefficients were more precise based on the support plane confidence intervals. The hypothesis that the restriction of $g = 0$ under the C-D form might affect the significance of some inputs in the production process is at least suggested.

One conclusion does stand out clearly from the foregoing analysis. Where there is doubt as to which single-equation production model to use in applied research, both the CES and the C-D must be fitted. It is no longer enough to argue that the CES has additional estimation difficulties that are not compensated for in terms of additional useful information. These difficulties are now relatively easy to surmount and the potential for misspecification is clearly suggested by the above application. With the improved feasibility of estimating the CES production function, the dominance of the C-D production function in applied research will certainly be challenged.

THE PRODUCTION FUNCTION

Introduction

This bulletin discusses the general properties of production functions. Its primary focus is on the controversy surrounding the elasticity of factor substitution to be assumed in production function models. Should the elasticity of substitution be restricted to unity as in the Cobb-Douglas function, which might bias the results? Or is there something to be gained by using the constant elasticity of substitution (CES) function, which allows the elasticity of substitution to be other than unity? Finally, this report investigates estimation procedures for the CES function and shows the results of estimates obtained from farm survey data in Washington.

Definitions

Ragnar Frisch was concerned that something essential could be lost from the theory of production by the increased concentration on mathematical relations rather than discussions proceeding directly from "the fundamental logic of the laws of production" (32). The early writers on production and distribution did concentrate on the logic of production and the fundamental concepts of product, production factors, productivity, etc., which are derivable from any production formulation (98). There can be no question that the analysis of production can never be independent of the logic of production. This dependence need not, however, deter consideration of alternative mathematical forms. Indeed, the question whether the Cobb-Douglas or the CES function is the superior theoretical form turns on the fundamental concept of substitutability of factors and must be resolved before any consideration of statistical superiority. If the theoretical form of a production function is such that it can yield more useful and correct information on the production process, it warrants statistical investigation.

The most familiar production relationships are described by amounts of inputs associated with specified amounts of output. That is, in a single input production process, a large number of input and output combinations are possible. To form a production function, each level of input in combination with the producer's technology must be evaluated in terms of the output achieved. The producer's technology consists of all the technical information about the physical combinations of inputs necessary for the production of the output. The object is to find the most efficient use of resources for a given output level or the point of maximum output for a given input level.

Given the assumption that an efficient point exists for every fractional input, the set of efficient points can be approximated by a continuous function. This function is the production function of traditional theory, i.e., the theoretical production function. It is defined for a given state of technical knowledge (96). In this most abstract sense, there are no restrictions on the production function except the technical restrictions of transforming inputs into products. The function becomes a frontier for values of the technologically efficient process (12,98).

The *statistically* fitted aggregate production function is a relationship between a set of outputs and a set of inputs of individual factors per unit of time. Since the function must be single valued to provide an efficient set of production possibilities, it is assumed to express the maximum product obtainable from the input combination at the existing state of technical knowledge. The statistical production function, which "assumes" that the input-output combinations solve the technical maximization problem, is something different from the theoretical production function, which is "defined" as the solution of the technical maximization problem.

Experimental production functions

The engineering, or experimental, type of production function provides a basis for normative answers to questions about which processes one should select to produce a particular product under given economic conditions. The function is derived from a designed experiment, generally with a specific, well-defined and homogeneous product. This is clearly a technical problem and is to be contrasted to the selection of the best input combination for the production of a *particular output level.* This last is an economic problem dependent upon input and output prices (51).

The fact that previous industry production studies had generally used statistically determined cost curves rather than statistically determined production functions led to H. B. Chenery's pioneering work on engineering production functions (13). Previous studies had to base the cost curves upon input combinations that had proven feasible for the entrepreneurs. They could not describe the broader range of productive possibilities that had been developed experimentally but not used commercially. Technical feasibility and not commercial feasibility should be the only restriction on possible techniques in the estimation of production functions.

Chenery's paper suggests a method by which engineering data can be used to approximate a production function in the industrial field, which corresponds more closely to the production function of economic theory. If the major input into production is a machine, the engineering production function would consider the characteristics of the machine such as speed, size, ease of operation, etc., and not its cost. This procedure makes it feasible to examine the impact of new characteristics of experimental machines (14, p. 297-325).

The engineering production function studies are not, however, free of difficulties (48). Kurz and Manne (67) used data showing the capabilities of 115 different machine tools to perform 129 alternative metal-working tasks. Productivity coefficients were established that showed the number of pieces per day that could be produced by one worker using a particular type of machine to do a given job. A censoring rule was then applied to eliminate the inefficient machines. "If in the performance of that task, one machine tool had a higher investment cost and not a higher output than a second machine tool, the first was said to be 'inefficient' and was deleted from our analysis" (67, p. 665-666). Next they used characteristics of the tasks, e.g., smaller or larger pieces or operations with wider or narrower tolerance, and not the tasks themselves as independent variables in the production function.

E. G. Furubotn has observed, however, that the censoring rule seriously undermines the entire study (34). The production function is not independent of input prices. Thus it is very unstable, since a new collection of efficient processes may appear with any price change in factors. Previously inefficient machine tools suddenly become efficient if their price falls relative to the other tools, since the censoring rule will change the composition of the set of efficient tools. If an *economic* solution is wanted, both price and technological data have to be used. But if a production function is to be established, information on prices is not needed. The efficient technical alternatives can be separated from the total array of processes on the basis of objective physical criteria (34).

It is not entirely clear how the separation between technical feasibility and commercial feasibility is to be achieved. Since factor prices partly determine the character of the production function by influencing its design, an array of techniques may show a labor-saving bias simply because labor has become more expensive than other factors. J. R. Hicks suggested that when labor becomes dearer than capital, the search for labor-saving techniques is stimulated. This is a corollary of Hicks' theory of induced inventions (52, p. 121-127). However, one might argue that the profit maximizer is interested in reducing costs in total and not particular labor costs or capital costs. Therefore, he has no bias toward labor-saving knowledge unless it is easier to get (95, p. 147). The concept of the engineering production function properly applied can clearly lead to an improved approximation of the true production function that is based on technical feasibility.

Nonexperimental production functions

The second type of production function of major concern to economists is the nonexperimental production function, often called a statistical production function. The use of data on inputs and outputs generated in the real world in a nonexperimental environment is much cheaper than data derived from precise experimental studies. However, one must recognize the cost of defining the nonexperimental production function for more aggregated products and categories of inputs. The nonexperimental types of functions are more descriptive than experimental types but are criticized for vagueness of identification and meaning (largely due to aggregation). Even so, many attempts have been made to fit the various forms (117).

Since this report is mainly concerned with nonexperimental production functions, it greatly simplifies exposition to use "statistical" and "nonexperimental" interchangeably. The experimental type of production function provides valuable answers to many of the criticisms of statistical production functions in general, but has narrow applicability and will not be considered further in this report.

Statistical and theoretical production functions

At least two major differences exist between the statistical and the theoretical functions. First, the statistical function can only approach the ex ante concept of a pure theoretical production function, since it is based on observations of resource combinations that were expected to be economically efficient at a particular time. The lag in capital expenditures for more technically efficient processes makes the statistically-fitted production function based on historical data only approximately correct. Thus, to be properly interpreted, production functions apply only for a particular point in time.

The second major difference between the pure theoretical and statistically fitted production function is the degree of aggregation. The theoretical function evolved for a single firm, while much of the data used in statistically fitted productions is of multi-firm or industry-wide origin (63). Hildebrand and Liu provide an excellent, critical review of some past research on nonexperimental production functions and discuss the aggregation problem (54). They show that the degree of aggregation must be clearly defined if one is to interpret the results of any production analysis correctly.

The basic problem in interpretation of a nonexperimental production function study is in the specification of the economic model used to analyze the data. The lack of homogeneity of individual resource units must be explained to derive conclusions from a single production function that is based on an aggregation of resource units. Do they face different factor prices, possess different expectations, or is the assumption of profit maximization or cost minimization naive? These factors must be analyzed if one is to rationalize the different positions of the individual units on the same production function.

There will always be a gap between the "pure" theoretical production function and the statistically-fitted production function. The economist should strive to close the gap as much as possible by recognizing the limitations placed on results by the data and the procedure. Even the most severe critics of statistical production functions conclude that they are useful, even though uncertain (85, 102).

Production function controversies

The state of the art of production economics has been critically analyzed (76). Two points of substantial controversy arose. The first related to engineering.

The second controversy was over the wide and unreconciled disparities in estimates of the elasticity of substitution from the CES (Constant Elasticity of Substitution) function using cross-section or time-series data. Marc Nerlove concluded, in reference to cross-section studies by Minasian, Solow, and Hildebrand and Liu, that results were very sensitive to small changes in model specification or data use (87). For example, all three studies used *Survey of Manufacturers* cross-section data for 1956 and 1957 and found the elasticity of substitution in the electrical machinery industry to be 1.26, 0.37, and 1.0996, respectively. The second elasticity value is based on regional aggregates; the others are based on state aggregates.

Time series studies reveal the same type of diversity. Biases and technical difficulties in various studies make it impossible to completely reconcile the diversity of results within each type. However, a pattern is established between cross-section and time series. The estimate of the elasticity of substitution from cross-section data is consistently larger than the time series estimates, leaving doubt as to which is better. For arguments in favor of cross-section studies, see (41). For a discussion favoring time series, see (5).

More recently, Paul Zarembka added to the controversy by showing that for most empirical purposes, the elasticity of substitution should be assumed equal to unity and the Cobb-Douglas function used rather than the CES (122,40). Maddala and Kadane, however, concluded that under some circumstances, a restriction of the elasticity of substitution to one can bias the estimates of returns to scale substantially (68).

The rest of this paper compares the Cobb-Douglas and CES functions for empirical research.

THE FORM OF THE PRODUCTION FUNCTION

The two production forms in question are now briefly examined—the Cobb-Douglas (C-D) and Constant Elasticity of Substitution (CES) production functions.

One algebraic form of the k-input C-D function is

$$Y = A X_1^{b_1} X_2^{b_2} \ldots X_k^{b_k}$$

where Y is output and the X values are the resource inputs (15). The parameter represents the efficiency of the technology, and the b_i are factor productivities. All can be estimated from data on output and inputs. In the original C-D formulation, the sum of the b_i equaled unity, but this restriction is often relaxed. The parameters and their economic implications are treated in detail later.

Again letting Y represent output and the X_i inputs, the CES function can be written for the k-input case as (1,115):

$$Y = \gamma [b_1X_1^{-g} + b_2X_2^{-g} + \ldots + b_kX_k^{-g}]^{v/-g}$$

$$\sum_{i=1}^{k} b_i = 1$$

the parameters of the function are γ, v, g, and the b_i.

The parameter γ is a scale parameter denoting the efficiency of the technology. It is a relationship between all inputs and output and does not affect the relations among inputs.

The b's, of course, measure the degree to which the production process relies on each input.

The parameter v represents the degree of homogeneity of the function or its degree of returns to scale. Although in the original formulation by Arrow et al. (1) this parameter was restricted to one, the above function can represent any degree of returns to scale of an empirical situation.

The parameter g determines the elasticity of substitution, which measures the ease of input substitution. The elasticity of substitution, σ, for the CES function is:

$$\sigma = 1/1+g$$

The formula for the elasticity of substitution is derived for the C-D and CES in appendix I.

Therefore, g is considered the substitution parameter and has an admissible range of values from -1 to infinity. See appendix II for a demonstration of the generality of the CES function for alternative values of g.

Elasticity of substitution

The principle of substitution of factors is quantified by the elasticity of substitution. It is the ratio of the proportional change in one variable resulting from a proportional change in another, holding all other variables constant.

If the variables are the ratio of the amount of capital, K, and labor, L; and the ratio of the price of capital, r, and the price of labor, w, the elasticity of substitution, as defined above is:

$$\sigma = \frac{\Delta\ (K/L)\ /\ (K/L)}{\Delta\ (w/r)\ /\ (w/r)}$$

Profits are maximized when labor and capital are used to the point where the marginal productivity of labor (MP_L) equals the wage rate, and the marginal productivity of capital (MP_k) equals the price of each unit of capital. Therefore, the following is equivalent to the previous form.

$$\frac{\Delta\ (K/L)\ /\ (K/L)}{\Delta\ (MP_L/MP_k)\ /\ (MP_L/MP_k)}$$

In the two-input production relationship, the relative shares of labor and capital are wL/PQ and rK/PQ where r, K, w, and L are as previously defined and PQ is the total value of output.

The ratio of the relative shares then becomes wL/rK. This ratio increases and decreases as the prices of the inputs vary. The extent of substitution of capital or labor for the other is quantified by the elasticity of substitution.

$\sigma > 1$ implies capital will be substituted for labor in greater proportion than the wage rate has risen relative to capital's return, so labor's relative share will fall.

$\sigma = 1$ implies the relative shares will remain constant as capital is substituted for labor in the same proportion as the wage increase.

$\sigma < 1$ implies the relative share of labor will rise because capital cannot be substituted for labor in the same proportion as the wage rate has risen relative to the price of capital.

Profit is maximized when the price-marginal productivity ratios of all inputs are equal to each other and to the output's unit price.

The demand curve of a firm for a productive resource under the assumption of profit maximization must therefore depend upon these variables:

1. The price of other productive resources (w, r)
2. the price of the output
3. the technology as embodied in the production function.

For one of the earlier treatments of derived demand, see (78), and for a more recent one, see (31).

Elasticity of substitution: C-D and CES production functions compared

The C-D function restricts the elasticity of substitution to be unity irrespective of the data used (see appendix I). This property guarantees that the relative income shares of inputs are constant for any changes in the relative supplies of inputs. Thus it provides a rationale for the observed relative constancy of factor shares in developed countries (66,107). However, it is doubtful that such a rationale is generally supportable.

Murray Brown has commented on the property of assumed unitary elasticity of substitution, ". . . the use of a production function which assumes the elasticity of substitution is always on the knife edge of unity encourages the rejection of the results for anything but a very crude approximation to production function estimation" (8, p. 37-50). However, the burden of proof of other than unitary elasticity lies with the estimation of alternative production forms. Jora D. Minasian has suggested that more caution should be exercised in making specific assumptions about the value of the elasticity of substitution (76).

The variability of the substitution parameter is largely responsible for the generality of the CES function (see appendix II). If $g = 0$, the elasticity of substitution is unity as in the C-D function. Indeed, if $g = 0$, the CES function reduces to the C-D form. If $g = -1$, the CES function reduces to a linear production function. The important realization is that the empirical data can dictate a value for the elasticity of substitution through the CES production function. Mathematical proofs are in appendix II.

The most desirable feature of the CES function—the elasticity of substitution is constant but not restricted a priori to any value—is not free of criticism. When the elasticity of substitution is specified as constant, it only is assumed that changes in relative factor inputs and prices do not alter the elasticity. Hence the constancy refers to the invariance with respect to changes in relative factor supplies and not to changes in technology. Since there is no way to be sure that all the change in the elasticity is solely due to technological change, there exists a potential specification error.

Nagesh S. Revankar (92) has sought to overcome the problem of the constancy of the elasticity of substitution by advancing the Variable Elasticity of Substitution (VES) production function.

Though both the CES and VES cover the special cases of the C-D and Leontief forms, the VES is more general in one important sense. The CES requires that the elasticity of substitution be the same at all possible input combinations whereas the VES allows the value to vary along any isoquant. The VES can be written

$$Y = \gamma K^{\alpha(1-\delta p)} [L + (p-1)K]^{\alpha\delta p}$$

with the elasticity of substitution being

$$\sigma = 1 + \frac{(p-1)}{1-\delta p} \cdot K/L .$$

A significant amount of variation in the K/L ratio must occur for effective discrimination between the CES and VES. However, the generalization to more than two inputs does present difficulties in structure and estimation.

Given the criticism regarding constancy of the elasticity of substitution, which applies to both the C-D and the CES, the generality of this constancy must also be examined. Irrespective of the number of inputs, the partial elasticities of substitution are independent of factor prices and are identical for all pairs of production factors (115). This property for all generalized CES functions has given rise to an alternative CES form. For other attempts to generalize the CES function to k inputs, see (71,79).

R. Sato has developed a two-level CES production function, closely related to the generalized CES, that takes special recognition of the fact that elasticities of substitution between pairs of inputs may not all be the same (83,97). In the two-level CES, the resources are divided into mutually exclusive groups; the direct partial elasticity of substitution is constant between resources within a group but not between resources in different groups. Each group is defined as a single CES form with constant elasticity as

$$Z_r = \left(\sum_{t=1}^{k} \theta_t X_t^{-g_r} \right)^{1/-g_r} , \quad \sum_{t=1}^{k} \theta_t = 1.$$

The two-level CES is then defined for the group r = 1, . . . , m as

$$Y' = a_0 \left[\sum_{r=1}^{m} a_r Z_r^{-g} \right]^{-1/g} , \quad \sum_{r=1}^{m} a_r = 1.$$

Note that an intra-class elasticity, g_r, exists within the r^{th} group and an inter-class elasticity, g, exists among input groups. The form presents difficulties in that a priori grouping of the factors is necsssary and direct estimation by the Taylor series expansion must be used. Sato used indirect estimation techniques, making the equality of the ratio of marginal products to the price ratio as a side condition.

There is no question that the CES is *not* theoretically superior to all other production function forms. The other functions, although superior theoretically with regard to particular parameters, present additional difficulties for estimation that are not in the realm of this study. The question to be answered here is whether the CES is superior in applied research where the C-D form has been dominant for so many years. A theoretical specification such as the CES, which permits the empirical environment to determine the degree of substitutability between inputs, would justifiably be considered superior to more restrictive specifications.

Technology and the Production Function

One determinant of the demand for factors depends heavily upon the technology of production. The relations between outputs and inputs and between the inputs themselves are determined by the ruling technology.

Murray Brown has identified four characteristics of a production function, that taken together form an abstract technology (8). These characteristics are the efficiency of technology, the degree of economies of scale, the degree of input intensity of a technology, and the ease of input substitution.

Efficiency of technology

The efficiency characteristic can be thought of as a scale transformation of inputs into output. The more efficient the technology, the larger the output for any given level of inputs. Thus if input levels of K_1 and L_1 produced an output of Y_1 at one technology and a greater output of Y_2 at another technology, the latter technology is more efficient.

Since an increase in the efficiency of a technology augments output but does not alter input relations, it is represented in the C-D form by the parameter A.

Increases or decreases in the value of A directly increase or decrease the output for the same input levels. The value of the parameter, A, is determined by the empirical data.

The CES production function is interpreted in the same manner with the scale parameter γ denoting the efficiency of the technology. Again, increases or decreases in the value of γ directly increase or decrease the output for the same input levels, with the value of γ determined by the empirical data.

Degree of economies of scale

A firm enjoys increasing, constant or decreasing returns to scale if for a given percentage increase in all inputs, output is increased at all. Any economies of scale measured in an aggregated manner, say for an industry, are a combination of internal and external economies (116). Internal economies depend on the resources within an individual firm. External economies are those arising from the general development and growth of an industry.

In the C-D production function, the sum of the input coefficients indicates the degree of returns to scale. The marginal products of the two-input case where the C-D function is used are:

$$\frac{\partial Y}{\partial X_1} = A\, b_1 X_2^{b_2}\, X_1^{b_1 - 1} = b_1 \frac{Y}{X_1}$$

$$\frac{\partial Y}{\partial X_2} = A\, b_2 X_1^{b_1}\, X_2^{b_2 - 1} = b_2 \frac{Y}{X_2}.$$

The factor intensities can then be written:

$$b_1 = \frac{X_1}{Y}\frac{\partial Y}{\partial X_1} \qquad b_2 = \frac{X_2}{Y}\frac{\partial Y}{\partial X_2}$$

From these equations it is inferred that b_i is a measure of the percentage change in output attributable to a percentage change in input N_i. The sum of b_i $i=1, \ldots,$ n then measures the total percentage change in output for a given percentage change in all inputs, i.e., the returns to scale. Since each b_i is determined empirically, if not otherwise restricted, the degree of returns to scale depends upon the data. Remember, however, that Maddala and Kadane found a substantial bias in estimates of returns to scale when the elasticity of substitution was constrained to unity as in the C-D form (68).

For the CES function, the scale parameter, v, determines the degree of return to scale. Given

$$Y = \gamma[b_1 X_1^{-g} + b_2 X_2^{-g}]^{v/-g} \text{ where } b_1 + b_2 = 1$$

The marginal products can be written

$$\frac{\partial Y}{\partial X_1} = \gamma(-v/g)[b_1 X_1^{-g} + b_2 X_2^{-g}]^{(v/g)^{-1}} b_1(-g)X_1^{-g-1}$$

$$= b_1 v \frac{Y}{[b_1 X_1^{-g} + b_2 X_2^{-g}]} \cdot \frac{X_1^{-g}}{X_1}$$

$$\frac{\partial Y}{\partial X_2} = \gamma(-v/g)[b_1 X_1^{-g} + b_2 X_2^{-g}]^{(-v/g)-1} b_2(-g)X_2^{-g-1}$$

$$= b_2 v \frac{Y}{[b_1 X_1^{-g} + b_2 X_2^{-g}]} \cdot \frac{X_2^{-g}}{X_2}$$

The input elasticities could then be written as above

$$\frac{\partial Y}{\partial X_1} \cdot \frac{X_1}{Y} = b_1 v \frac{Y}{[b_1 X_1^{-g} + b_2 X_2^{-g}]} \cdot \frac{X_1^{-g}}{X_1} \cdot \frac{X_1}{Y}$$

$$= v \frac{b_1 X_1^{-g}}{[b_1 X_1^{-g} + b_2 X_2^{-g}]}$$

and

$$\frac{\partial Y}{\partial X_2} \cdot \frac{X_2}{Y} = b_2 v \frac{Y}{[b_1 X_1^{-g} + b_2 X_2^{-g}]} \cdot \frac{X_2^{-g}}{X_2} \cdot \frac{X_2}{Y}$$

$$= v \frac{b_2 X_2^{-g}}{[b_1 X_1^{-g} + b_2 X_2^{-g}]}.$$

The sum of the input elasticities would again measure the total percentage change in output for a given percentage change in inputs.

$$\frac{\partial Y}{\partial X_1} \cdot \frac{X_1}{Y} + \frac{\partial Y}{\partial X_2} \cdot \frac{X_2}{Y} = v \frac{b_1 X_1^{-g} + b_2 X_2^{-g}}{b_1 X_1^{-g} + b_2 X_2^{-g}} = v.$$

The parameter v is determined by the empirical data. (R. K. Diwan seeks evidence of economies of scale while examining alternative forms of the C-D and CES functions (21).)

Input intensity of a technology

The concern here is with the technological requirements of the production process but not the current changes in output or relative factor supplies. The structure requirements of the technology determine the intensity of one input relative to some other input. The relative intensity of the inputs will be reflected in the data on inputs and outputs and not established a priori in the production function form. This intensity is measured by the relative values of the b_t coefficients in both the C-D and CES functions.

The ease of input substitution

The technology of the production function, and thus factor demand, depends on the form of the production function, at least with regard to input substitution. Neither the C-D nor the CES form specifies the first three characteristics, efficiency, return to scale, and intensity, independent of the data. With regard to factor subsituation, however, the C-D function provides less information on factor demand.

Functional Form and Technological Change

Definitions of technological change

Technological change here means any change in the production process that permits the same level of output to be produced with less input or enables the former levels of inputs to increase output. For production function analysis, the phenomenon is often regarded as going on at some externally given rate and increasing, over time, the output possible by any combination of factors.

It is appealing to define technological change in terms of changes in the "abstract technology" of the production function developed in the last section. The input-input and input-output relations will be determined by the technology that exists at a given time. Changes in the underlying technology of a production function will be reflected in changes in the characteristics that form the abstract technology.

Technological change is classified by its impact on the marginal products of the inputs on each other. A "neutral" technological change improves the productivity of inputs for each input combination but leaves unchanged the relative marginal products of the inputs. If relative marginal products are unchanged, there is no change in *relative* factor demand as the result of technological change, although absolute demand may change (52). Neutral technological change will thus affect input-output relationships but not the input-input relationship.

An excellent comparison of the Hicksian definition of technological change used here and other formulations is in Salter (95); see also Hahn and Matthews (45).

A reexamination of the characteristics of the abstract technology shows that changes in two of the characteristics can be considered neutral technological changes. These characteristics are changes in the efficiency of technology and changes in the economies of scale.

Nonneutral technological change includes changes in input intensity or the degree of input substitution that alters input-input or input-output relationships. Nonneutral technological change is classified according to its impact on the *relative* marginal products. For instance, if the marginal product of capital rises relative to the marginal products of other inputs, we have a capital-using nonneutral technological change. A rise in the marginal product of labor relative to the marginal products of the other inputs, is a labor-using nonneutral technological change.

The necessary distinction between neutral and nonneutral technological change becomes apparent when considering the impact of technological progress on factor shares, factor-demand, or output (29). Isolation of the types of technological change that promote growth and the conditions necessary for such growth are the goals of many production studies (46,84). One of the most important results of any aggregate production function will be the linking of factor supplies and output, in the aggregate, over time, and thus a clearer understanding of the engines of economic growth.

Technological changes and production function form

The C-D production function can show changes in in three of the four characteristics of the underlying technology. Neutral technological changes will be reflected by changes in the parameter A or in the sum of the exponents of the inputs. Changes in these parameter values can be determined empirically and reflect changes in the efficiency of the technology and the degree of returns to scale.

The C-D function cannot reflect changes in the degree of input substitution. Therefore, nonneutral technological change can be reflected only by changes in the relative factor intensities. These changes in factor intensities will be reflected in changes in the ratios of the inputs to each other. Clearly, changes in intensities due to a nonneutral technological change could lead to changes in the sum of these intensities, indicative of a neutral technological change. It becomes very hard to isolate neutral technological change in the summation of the parameters from nonneutral technological change reflected in the change in relative intensities.

The CES function that allows measurement of changes in all four of the characteristics of the abstract technology is not without criticism. Neutral technological change is reflected by changes in the efficiency parameter γ and in the scale parameter, v. However, two forces affect the parameter v and it would be difficult to determine their relative influence. Economies of scale can result from an expansion of the scale of operations with a given technology, or given the scale, a technological change could occur. The same limitation applies to the sum of the exponents in the C-D function.

The CES function does separate neutral and nonneutral technological change by permitting changes in factor intensities and changes in the degree of input substitution to be reflected in parameters other than γ or v. Relative factor intensities are determined by the relative values of the coefficients b_1 and b_2, while the substitution parameter will reflect changes in input substitution.

Summary of the Evidence

The impact of the functional form on the results has been examined in three major areas: input substitution, factor demand and returns to scale, and technological change. A brief review of the conclusion in each area should suggest the theoretical superiority of the CES production function.

Input substitution. The CES function assumes that the elasticity of substitution is constant but not restricted to an a priori value. The C-D function compels the elasticity of substitution to be constant at a value of unity.

Factor demand. The impact of the production function upon factor demand is largely by way of the underlying technology. The CES function was found superior in allowing the data to dictate the technology, particularly with regard to factor substitution.

Technological change. The CES function can represent technological change better than the C-D.

In the areas where the CES function was found superior, it was not without criticism. But the generality of the CES function is the key to its theoretical superiority over the C-D production function. This generality is, of course, not without costs. The problem of statistically fitting the CES function is the challenge when using it for applied research.

STATISTICAL ESTIMATION PROBLEMS

Estimation Techniques

C-D production function

The two-input C-D function can be written as

$$Y = A\, x_1^{b_1}\, x_2^{b_2}$$

Deviations between theory and observation create problems when estimating production parameters. The statistical specification of the production function must accept the fact that most of the disturbance or discrepancies between the model and observation are due to factors excluded from the equation and not entirely rationalized by economic theory. Such things as errors in measurement, the randomness of human behavior, etc., are not always quantifiable with specific variables. The simplest approach is to assume the disturbances enter into the equation multiplicatively as a *random* variable u.

$$Y = Ax_1^{b_1}\, x_2^{b_2}\, u \qquad u = e^{v} .$$

Under a logarithmic transformation, this stochastic C-D function becomes linear in the parameters

$$\ln Y = \ln A + b_1 \ln x_1 + b_2 \ln x_2 + \ln u$$

or

$$\ln Y = \ln A + b_1 \ln x_1 + b_2 \ln x_2 + v .$$

Assuming the proper distribution of the disturbance term v, straightforward application of least squares will provide estimates of the parameters of the equation. The ease with which this function can be fitted with empirical data virtually explains its popularity.

The production function equation may be only one equation of a system of equations describing the production relations causing concern about single equation estimation of the C-D function. Single equation estimates of the C-D production function are biased when the independent variables, in this case inputs, are correlated with the disturbance term. Under such conditions, the independent variables are functions of the disturbance in the given equation. The result is a violation of the underlying assumptions of single equation regression regarding truly independent variables. If the disturbance in the production equation affects only the output and not the independent variables, then there is no simultaneous equation (57).

Others have said that the disturbance term in a given behavioral equation for an input is composed additively of two terms:

1. the disturbance in the production function equation,
2. a disturbance specific to the behavioral equation, and uncorrelated with the disturbance term in the production function (82, 123).

Indeed, one would expect an input to respond in some manner to fluctuations in output, particularly if the resource is used solely in one industry, and has restricted mobility. Thus the production function generally specified for inputs x_i as

$$Y = F(x_i, u)$$

is considered directly in the factor demand (behavioral) equation for input x_i as

$$x_i = f(z_j, Y, v)$$

where v is the disturbance term for the behavioral equation of the first input and z_j represents other input market factors. Thus

$$x_i = f[z_j, f(x_i, u), v]$$

And the total disturbance term for the behavioral equation is composed of u and v; the result is potential simultaneous equation bias (93).

The traditional, or classical, approach to specification of the C-D production function assumes that firms operate on a nonstochastic production function and maximize profits. The specification is nonstochastic in the sense that an exact functional relationship between variables is postulated. The smallest amount of contact with economic data, however, will demonstrate that not all data points lie exactly on the lines representing the functions. The econometrician makes the equation stochastic, or random, with the introduction of the random disturbance term (87).

An alternative technique is the nontransmission approach. The technique requires that each input be uncorrelated with the disturbance term in the production function to eliminate simultaneous equation bias. This requirement can be met by assuming that the production function is stochastic, or random, and that the firm maximizes the mathematical expectation of profit (93, p. 1-16; 123).

This assumption is particularly applicable to agricultural firms where practically all inputs are determined before full information on actual output has been received. Under such conditions, the use of inputs is at least more independent of actual output and correspondingly less correlated with the disturbance term in the production function. Of course, it would be absurd to argue that inputs could always be wholly independent of output decisions. But by assuming a stochastic production function, entrepreneurs are seen as aware of the stochastic nature of production. As Zellner, Kmenta and Dreze observed, ". . . one-period maximization of expected returns is just a step in the direction of a proper treatment of stochastic elements in a firm's sequential decision-making process under uncertainty" (123, p. 794).

The proper strategy for the error term in the estimation of the C-D function deserves comment. Clark Edwards hypothesized that the multiplicative error is assumed for the stochastic C-D function not because of economic theory, but because it permits the logarithmic transformation. He therefore fit several alternative C-D functions with a multiplicative error and with an additive error (25) such as

$$Y = A\, x_1^{b_1}\, x_2^{b_2}\, u$$

and

$$Y = A\, x_1^{b_1}\, x_2^{b_2} + v.$$

Although the results were substantially different, indicative of error misspecification, it is hard to see how much of the difference was due to the alternative estimation techniques used. The first equation can be estimated by normal least squares after a logarithmic transformation. But the second equation remains nonlinear and requires a nonlinear estimating technique.

The argument may be academic. The error term in a production function is designed to measure the effects of missing variables. It could be argued that under the C-D formulation, the misspecification would have a multiplicative effect upon output as do the specified inputs. The multiplicative error term in the function may thus satisfy the least squares assumption by appealing to the multiplicative central limit theorem (35).

However, when a multiplicative error term is transformed logarithmically into a multiplicative lognormal disturbance for linear regression, attention is shifted from the conditional mean to the conditional median. Since the conditional mean is the prime target of most studies, Goldberger has advised that researchers report the value of the adjusted residual variance when linear logarithmic regressions are run so that results can be adjusted to avoid potential bias (36). Goldberger did concede that in practice, the minimum variance unbiased estimators after adjustment may not differ substantially from those achieved by normal least squares with no adjustment.

The possibility of simultaneous multiplicative and additive errors has also been examined, but no statistical grounds have been presented for a choice between the alternative specifications (37).

CES production function

For the CES function, a logarithmic transformation will not resolve the nonlinear property. The two-input CES can be written

$$Y = \gamma[b_1 x_1^{-g} + (1-b_1)\, x_2^{-g}]^{-v/g}\, u$$

with u being a multiplicative random disturbance. A logarithmic transformation results in

$$\ln Y = \ln \gamma - v/g \ln [b_1 x_1^{-g} + (1-b_1) x_2^{-g}] + \ln u\ .$$

The function remains nonlinear, as no monotonic transformation can separate the parts of the second term:

$$\ln [b_1 x_1^{-g} + (1-b_1)\, x_2^{-g}]\ .$$

The diversities in methods of estimation and in empirical results are a direct result of the need to find estimates of γ, b_1, g, and v in an equation that is strongly nonlinear. Initial attempts at estimating the CES function were designed to estimate the nonlinear equation

$$Z = [b_1 x_1^{-g} + (1-b_1)\, x_2^{-g}]\ .$$

To accomplish this task, estimates of g and b_1 must be derived from the data. One approach was suggested by Arrow et al. in their article that introduced the CES function (1,88). If the CES function is restricted to the case of constant returns to scale, i.e., v = 1, one can estimate the elasticity of substitution from the marginal productivity condition under cost minimization by regressing the value of output per worker on the wage rate. If, however, the CES function is generalized to allow empirical determination of the return to scale, this method is no longer feasible.

Assuming cost minimization and using marginal productivity conditions, one could also equate the input price ratio to the ratio of the marginal products

$$\frac{w}{r} = \frac{MPx_1}{MPx_2}\ .$$

Including a multiplicative error term and transforming by use of logarithms we obtain

$$\ln \frac{w}{r} = \ln \frac{MPx_1}{MPx_2} + \ln u.$$

In Appendix I, it is shown that

$$\frac{\partial y}{\partial x_1} = b_1 v \frac{y}{[b_1\, x_1^{-g} + b_2\, x_2^{-g}]} \cdot \frac{x_1^{-g}}{x_1}$$

and

$$\frac{\partial y}{\partial x_2} = b_2 v \frac{y}{[b_2\, x_2^{-g} + b_1\, x_1^{-g}]} \cdot \frac{x_2^{-g}}{x_2}$$

since

$\frac{\partial y}{\partial x_2}$ is equivalent to MPX_i

$$\ln \left[\frac{MP_{x_1}}{MP_{x_2}}\right] = \ln \left[\frac{\frac{\partial y}{\partial x_1}}{\frac{\partial y}{\partial x_2}}\right] = \ln \frac{b_1}{1-b_1} - (g+1) \ln \frac{x_1}{x_2}.$$

$$\text{Then } \ln \frac{w}{r} = \ln\left[\frac{b_1}{1-b_1}\right] - (g+1) \ln \frac{x_1}{x_2} + \ln \mu.$$

Fitting this equation with data on input prices and quantities will result in estimates of b_1 and g, denoted $\hat{b}_1$ and $\hat{g}$.

Substituting these estimates into the nonlinear term, Z, results in

$$\ln Y = \ln \hat{\gamma} - v/\hat{g} \ln [\hat{b}_1 x_1^{-\hat{g}} + (1-\hat{b}_1) x_2^{-\hat{g}}] + \ln \mu .$$

This form is now linear in the remaining unknown parameters and can be estimated by simple least squares.

Although alternative side conditions can be used, one must be careful of single equation bias and serial correlations with respect to specification errors, as well as the questionable validity of the initial assumption of cost minimization. The feasibility of side conditions under non-constant returns to scale remains tenuous (21).

Estimation of the CES function directly has the advantage of requiring only the specification of the production function form and not side conditions based on the assumption of cost minimization and marginal productivity conditions.

The Taylor Series Approximation uses initial values possibly derived from one of the previous indirect estimation techniques, to a first-order Taylor Series expansion.[1] Kmenta has suggested that the technique be applied to the logarithmic form of the CES, i.e.

$$\ln Y = \ln \gamma - v/g \ln [b_1 x_1{}^{-g} + (1-b_1) x_2{}^{-g}] + \ln u.$$

This form can be written more concisely as

$$\ln Y = \ln \gamma - v/g\ f(g) + \ln u.$$

The term f (g) can now be expanded around the value g = 0 in a Taylor Series to approximate its value (64,72). Then, disregarding the terms of third and higher orders, the expansion becomes[2]

$$f(g) = f(0) + f'(0)\cdot(g-0) + \frac{f''(0)}{2!}(g-0)^2$$

$$= 0 + [-b_1 \ln x_1 + (1-b_1) \ln x_2]\ (g)$$

$$+ 1/2\ [b(1-b)[\ln x_1 - \ln x_2]^2](g^2)$$

[1]Taylor's Theorem: Let f be a function that is continuous together with its first n-1 derivatives on an interval containing a and x. Then the value of the function at x is given by

$$f(x) = f(a) + f'(a)\cdot(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^n(a)}{n!}(x-a)^n + R_n(x,a)$$

where

$$R_n(x,a) = \int_a^x \frac{(x-t)^n}{n!} f^{n+1}(t)\ dt$$

See George B. Thomas, Jr. (112, p. 790-791).

[2] J. Kmenta found that provided the second order term is included, the error from neglect of the higher order terms is not serious unless both the capital-labor ratio and the elasticity of substitution are either very high or very low (65, p. 186).

$$f(g) = -g\ [b_1 \ln x_1 + (1-b_1) \ln x_2] + 1/2g^2\ [b(1-b)\ (\ln x_1 - \ln x_2)^2] .$$

The logarithmic form of the CES function may then be approximated by

$$\ln Y = \ln \gamma - v/g\ [-g[b_1 \ln x_1 + (1-b_1) \ln x_2] + 1/2\ g^2\ [b_1(1-b_1)(\ln x_1 - \ln x_2)^2]] + z$$

$$= \ln \gamma + vb_1 \ln x_1 + v(1-b_1) \ln x_2 - 1/2\ gvb_1\ (1-b_1)\ (\ln x_1 - \ln x_2)^2 + z$$

where z = [u — (v/g) — neglected higher order terms of the expansion of f (g)]. This form is linear in the unknown parameters and allows direct estimation of the parameters.

The approach must be used with some reservations. First, the undue amount of multicollinearity that could be introduced by expansion around g = 0 might only be prevented if one had some a priori estimate of the distribution parameter, b_1 (1). Second, some doubt is created by this method, since the expansion occurs around such a crucial variable as g. If g = 0, the CES becomes equivalent to the C-D and one has assumed away the crucial question of which form is proper and introduces a bias in favor of the latter. Third, and most important, when more than two variable inputs are used, the equation obtained from the above expansion is highly overidentified and consistent estimates are not possible (50).

The traditional method for estimating the coefficients of a linear function such as

$$Y_t = A + b_1 x_{1t} + b_2 x_{2t} + u_t \qquad t=1,n$$

is to minimize the expression

$$S = \sum_{t=1}^{n} (u_t)^2 = \sum_{t=1}^{n} (Y_t - A - b_1 x_{1t} - b_2 x_{2t})^2 .$$

If one approaches the CES function in a similar manner, the expression becomes

$$S = \sum_{t=1}^{n} (Y_t - \gamma[b_1 x_{1t}{}^{-g} + (1-b_1)\ x_{2t}{}^{-g}]^{-v/g})^2 .$$

One technique for minimizing this expression was used by Clark Edwards in estimating C-D production functions with additive errors. Using a Taylor series expansion of the CES function, initial guesses, labeled γ^o, $b^o{}_1$, g^o, and v^o, were used to form the following expression,

$$Y_t^o = \gamma^o[b_1^o x_{1t}^{-g^o} + (1-b_1^o)\ x_{2t}^{-g^o}]\ v^o/g^o + \left(\frac{\partial Y_t}{\partial \gamma}\right)_o (\gamma-\gamma^o)$$

$$+\left(\frac{\partial Y_t}{\partial b_1}\right)_o (b_1-b_1^o) + \left(\frac{\partial Y_t}{\partial g}\right)_o (g-g^o) + \left(\frac{\partial Y_t}{\partial v}\right)_o (v-v^o)$$

where the first derivatives are evaluated at the initial values designated by subscript zeros (25). Note that this is a first-order Taylor series expansion of the untransformed CES function as opposed to Kmenta's approach of expanding only one term of the logarithm of the CES function.

The next step is to find values for γ, b_1, g, and v, which minimize

$$S_1 = \sum_{t=1}^{n} (Y_t - Y_t^o)^2.$$

These values are then labeled γ^1, b_1^1, g^1, and v^1 and used to form the above expression at new values.

At the second stage, then,

$$Y_t^1 = \gamma^1\ [b_1^1\ x_{1t}^{-g^1} + (1-b_1)\ x_{2t}^{-g^1}]^{-v^1/g^1} + \left(\frac{\partial Y_t}{\partial \gamma}\right)_1 (\gamma-\gamma^1)$$

$$+ \left(\frac{\partial Y_t}{\partial b_1}\right)_1 (b_1-b_1^1) + \left(\frac{\partial Y_t}{\partial g}\right)_1 (g-g^1) + \left(\frac{\partial Y_t}{\partial v}\right)_1 (v-v^1)$$

and the expression to be minimized is

$$S_2 = \sum_{t=1}^{n} (Y_t - Y_t^1)^2$$

which gives new values for γ, b_1, g, and v. This procedure continues until the parameter values converge to a particular set of values. The number of iterations necessary depends on the quality of initial guesses and, the procedure may only succeed in finding a local minimum.

A simultaneous equation approach to estimation of the CES production function used by Bodkin and Klein is a modification of a technique developed by Eisenpress and Greenstadt (4,26). Beginning with a similar expression to be minimized as above with the addition of the term $10\lambda^t$ to allow for disembodied neutral technological change at the rate of 10λ-1, the expression to be minimized would be written

$$S = \sum_{t=1}^{n} (Y_t - \gamma 10^{\lambda t}\ [b_1 x_{1t}^{-g} + (1-b_1)\ x_{2t}^{-g}]^{-v/g})^2$$

a system of equations necessary for minimization is formed. The system becomes

$$\frac{\partial S}{\partial \lambda} = 0 \qquad \frac{\partial S}{\partial \gamma} = 0 \qquad \frac{\partial S}{\partial b_1} = 0$$

$$\frac{\partial S}{\partial g} = 0 \qquad \frac{\partial S}{\partial v} = 0.$$

Each of these equations can then be expanded about assumed initial values of the parameters possibly derived from the Taylor Series approach. The system of equations after expansion is illustrated by

$$\frac{\partial S}{\partial \gamma} = \left[\frac{\partial S}{\partial \gamma}\right]_o + \left[\frac{\partial^2 S}{\partial \gamma^2}\right]_o (\gamma-\gamma^o) + \left[\frac{\partial^2 S}{\partial \gamma \partial b_1}\right]_o (b-b_1^o)$$

$$+\left[\frac{\partial^2 S}{\partial \gamma \partial\ g}\right]_o (g-g^o) + \left[\frac{\partial^2 S}{\partial \gamma \partial v}\right]_o (v-v^o)$$

$$\frac{\partial S}{\partial b_1} = \left[\frac{\partial S}{\partial b_1}\right]_o + \left[\frac{\partial^2 S}{\partial b_1^2}\right]_o (b_1-b_1^o) + \left[\frac{\partial^2 S}{\partial b_1 \partial \gamma}\right]_o (\gamma-\gamma^o)$$

$$+ \left[\frac{\partial^2 S}{\partial b_1 \partial g}\right]_o (g-g^o) + \left[\frac{\partial^2 S}{\partial b_1 \partial v}\right]_o (v-v^o)$$

and so forth.

These equations form a system of live linear equations in the five unknown parameters. The procedure then becomes iterative as the system is solved for parameter values, λ^1, γ^1, b_1^1, g^1, and v^1, which minimize S, and new equations are formed. The procedure followed by Eisenpress and Greenstadt was modified to approach the minimum S more efficiently.

A more recent estimation technique developed by Tsurumi uses a procedure based on Marquardt's maximum neighborhood algorithm (114). Employing initial parameter estimates based on the approach of Arrow et al., the following first-order Taylor Series expansion.

$$Y = \gamma e^{\lambda}\ [b_1 x_1^{-g} + (1-b_1)x_2^{-g}]^{-v/g} = \left[\frac{\partial Y}{\partial \gamma}\right]_o (\gamma-\gamma^o)$$

$$+ \left[\frac{\partial Y}{\partial b_1}\right]_o (b_1-b_1^o) + \left[\frac{\partial Y}{\partial g}\right]_o (g-g^o) + \left[\frac{\partial Y}{\partial v}\right]_o (v-v^o)$$

$$+ \left[\frac{\partial Y}{\partial \lambda}\right]_o (\lambda-\lambda^o) + r$$

was fitted to obtain estimates of λ, γ, b_1, g, and v. r is a remainder term for the expansion. As in the previous procedures, this iterative scheme can be carried on until it satisfies some convergence criterion established for the model. Tsurumi then computes the five partials,

$$f_1 = \frac{\partial Y}{\partial \gamma},\ f_2 = \frac{\partial Y}{\partial b_1},\ f_3 = \frac{\partial Y}{\partial g},\ f_4 = \frac{\partial Y}{\partial v},\ f_5 = \frac{\partial Y}{\partial \lambda},$$

at the parameter values obtained at convergence. The second step is to regress the dependent variable

$$Y - \gamma e^{\lambda}[b_1 x_1^{-g} + (1-b_1)x_2^{-g}]^{-v/g}$$

on these estimated partials. An alternative dependent variable actually used by Tsurumi is obtained by dividing the original expansion by one of the partials such that

$$\frac{Y - \gamma e^{\lambda}[b_1 x_1^{-g} + (1-b_1)x_2^{-g}]^{-v/g}}{f_1}$$

$$= (\gamma-\gamma^o) + \frac{f_2}{f_1}(b_1-b^o) + \frac{f_3}{f_1}(g-g^o)$$

$$+ \frac{f_4}{f_1}(v-v^o) + \frac{f_5}{f_1}(\lambda-\lambda^o) + r.$$

Tsurumi concluded that the choice of the exogenous variable and data may dictate the validity of Eisenpress and Greenstadt's conclusion that the two-stage algorithm applied to a nonlinear system tended to yield poor results (114).

Comparing Estimating Problems of the C-D and CES Functions

The estimation of the C-D production function has its difficulties. But in comparison to the difficulties of estimating the CES production function, its statistical simplicity is obvious. However, simplicity is not enough for statistical superiority. For applied research, statistical superiority really turns on the question of robustness of the estimating technique with respect to different types of data and the reliability of the results. Any alternative to the C-D function that is to be considered superior must match its statistical robustness and provide more useful and correct information. A brief re-examination of the shortcomings of the current techniques for estimating the CES may provide an insight as to the areas that future work on its estimation should focus.

The principal disadvantage of the indirect estimation techniques is that they require an assumption of equilibrium. A side relation, which is derived from marginal productivity theory under cost minimization, and a production function, derived from theory, must be posited. The direct estimation techniques require acceptance of only the latter theory.

The major criticisms of all the direct estimation techniques lie in the heavy reliance upon Taylor Series approximations; these require increasingly better starting values to obtain convergence as the number of inputs is expanded. Inasmuch as the C-D is easily generalized to more than two inputs, the above shortcomings are very serious. The development of a consistent technique for obtaining starting values for a k-input CES function would greatly reduce estimation difficulties if the Taylor Series approach must be used. The techniques for achieving convergence could likewise be improved along the lines of Tsurumi's approach by using the Marquardt algorithm, which will be discussed later. Despite the above criticisms, the direct estimation approaches seem most promising. The following chapter will develop quite carefully a technique for obtaining starting values for the Marquardt algorithm.

AN ALTERNATIVE APPROACH TO DIRECT ESTIMATION OF THE CES PRODUCTION FUNCTION

Next, a method of estimating the CES function by maximum likelihood functions and the Marquardt algorithm will be discussed. Since the Marquardt algorithm is the method used in the example of this study and since it is based on both the Taylor Series method and the method of steepest descent, we shall outline these concepts first.

Least Squares Estimation

Before examining the techniques individually, a review of the solution of a linear model would be informative for comparison. Assume a *linear* model defined by

$$Y = X\theta + \varepsilon$$

where

$$Y = \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} X_{11} & \cdots & X_{t1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ X_{1n} & & X_{tn} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_t \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

The sum of squares for this model is defined as

$$SS(\theta) = \sum_{i=1}^{n} (Y_i - \sum_{t=1}^{k} \theta_t X_{ti})^2 \qquad [1.0]$$

or in matrix form

$$SS(\theta) = Y'Y - 2\theta'X'Y + \theta'X'X\theta \ . \qquad [2.0]$$

The linear model is solved by minimizing the sum of squares. It is accomplished when all the partial derivatives with respect to the θ_t are equal to zero.

$$\frac{\partial SS(\theta)}{\partial \theta} = -2X'Y + 2X'X\theta = 0$$

$$= -X'Y + X'X\theta = 0 \ . \qquad [3.0]$$

The solution set $\hat{\theta}$ thus becomes

$$\hat{\theta} = (X'X)^{-1}X'Y \qquad [4.0]$$

Note that in this linear model the solution for θ is dependent only upon X and Y and no θ appears on the right side of [4.0]. The normal equations derived directly from [1.0] are of the form

$$\frac{\partial SS(\theta)}{\partial \theta_t} = \sum_{t=1}^{k}(Y_i - \sum_{t=1}^{k} \theta_t X_{ti}) \frac{\partial(\sum_{t=1}^{k} \theta_t X_{ti})}{\partial \theta_t} = 0 \qquad [4.1]$$

$$= \sum_{t=1}^{k} (Y_i - \sum_{t=1}^{k} \theta_t X_{ti}) X_{ti} = 0 \qquad [4.2]$$

since $\sum_{t=1}^{k} \theta_t X_{ti}$ is linear in the θ_ts, the partial derivative is independent of θ_t.

Substituting $\hat{\theta}$ into equation [2.0] yields the sum of squares of the solution set

$$SS(\hat{\theta}) = Y'Y - 2\hat{\theta}'X'Y + \hat{\theta}'X'X\hat{\theta}$$

$$= Y'Y - \hat{\theta}'X'Y - \hat{\theta}'X'Y + \hat{\theta}'X'X\hat{\theta}$$

$$= Y'Y - \hat{\theta}'X'Y - \hat{\theta}'(X'Y - X'X\hat{\theta})$$

since $\hat{\theta}$ satisfies [3.0] by minimizing $SS(\theta)$ this reduces to

$$SS(\hat{\theta}) = Y'Y - \hat{\theta}'X'Y.$$

This constitutes the residual sum of squares. It represents the smallest amount of error that cannot be explained by the regression. The amount of explained variation attributable to the regression must therefore be

$$SS(\theta) - SS(\hat{\theta})$$

$$= Y'Y - 2\theta'X'Y + \theta'X'X\theta - Y'Y + \hat{\theta}'X'X\hat{\theta}$$

$$= \theta'X'X\theta - 2\theta'X'Y + \hat{\theta}'X'X\theta$$

$$= (\theta - \hat{\theta})'X'X(\theta - \hat{\theta}) .$$

Thus, all values of θ that satisfy $SS(\theta)$ = constant value K are given by

$$K = SS(\hat{\theta}) + (\theta - \hat{\theta})'X'X(\theta - \hat{\theta}).$$

This equation can be identified as the equation for a closed ellipsoidal contour with center at $\hat{\theta}$. As K increases, the implication is that the parameter sets of θ that satisfy the equation are further away from $\hat{\theta}$ and define a sum of squares contour of larger circumference. A $100(1-\alpha)\%$ confidence region for the true value of θ that yields a particular value K is given by

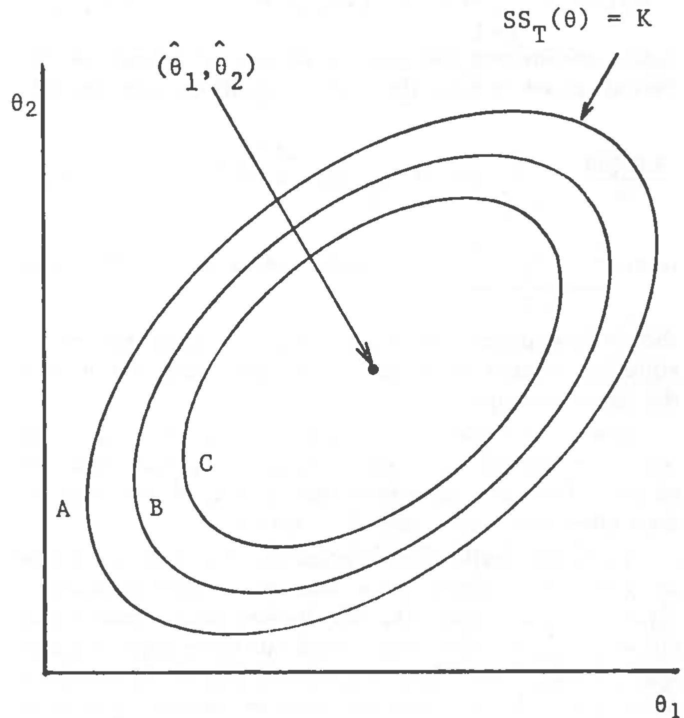
$$K = SS(\theta) = SS(\hat{\theta}) \{1 + \frac{t}{n-t} F(t,n-t, 1-\alpha)\}$$

where n = sample size and t = number of parameters.

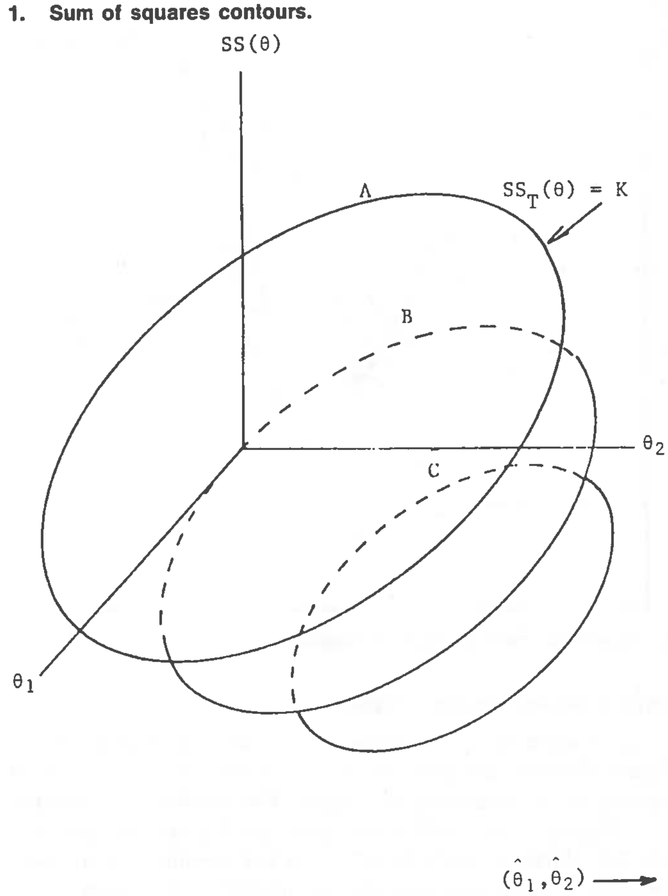
The ellipsoidal contour concept can be seen more clearly in figure 1. A simple two-parameter linear model is depicted in terms of the sum of squares contours. In this two-dimensional space of θ_1 and θ_2, the contours are concentric ellipses about the point where $SS(\theta)$ is minimized, denoted $(\hat{\theta}_1, \hat{\theta}_2)$. Figure 2 shows the third dimension of the contours.

When the model becomes nonlinear, the absence of an X matrix in the linear sense changes the problem drastically. To illustrate the complications assume a general *nonlinear* model of the form

$$Y = f(X,\theta) + \varepsilon$$



1. Sum of squares contours.



2. Three-dimensional sum of squares contours.

alternatively

$$Y_i = f(X_i,\theta) + \varepsilon_i \quad i=1,\ldots,n.$$

The sum of squares in this case becomes

$$SS(\theta) = \sum_{i=1}^{n} \{Y_i - f(X_i,\theta)\}^2 . \qquad [5.0]$$

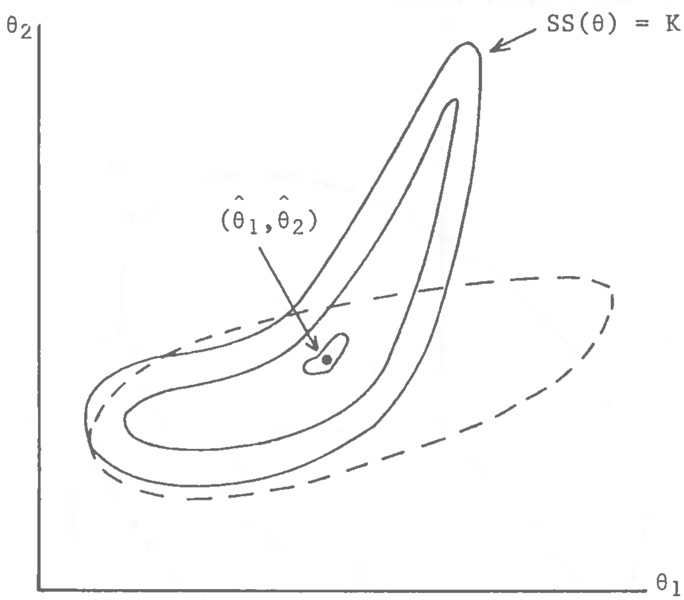
Again minimizing the sum of squares by setting all first partials equal to zero, the normal equations take the form

$$\frac{\partial SS(\theta)}{\partial \theta} = \sum_{i=1}^{n} \{(Y_i - f(X_i,\theta)) \frac{\partial f(X_i,\theta)}{\partial \theta}\} = 0 \qquad [6.0]$$

where $\frac{\partial f(X_i,\theta)}{\partial \theta}$ is not independent of θ. This form should be compared to that in [4.2]. Clearly, the normal equations cannot be linear in the parameter set θ as in the linear example.

When the model is nonlinear, the sum of squares contours are not of any regular shape as in the linear example. They are sometimes banana-shaped and may not even close, but rather stretch to infinity.

A possible well-behaved nonlinear function is depicted in terms of constant value sum of squares contours in figure 3. In comparison, the broken line represents an elliptical contour that would exist for some approximating linear model. The figure also shows that at the minimum point $(\hat{\theta}_1, \hat{\theta}_2)$ the nonlinear contour becomes nearly elliptical.



3. Nonlinear sum of squares contours.

Taylor series approximation

This technique, sometimes referred to as the Gauss or Gauss-Newton method, uses the results of linear least squares in a succession of stages. The model is expanded as a Taylor series with corrections to the several parameters calculated at each iteration on the assumption of *local linearity*. Assuming a nonlinear model of the form

$$Y_i = f(X_i,\theta) \qquad i=1, \ldots, n$$

the linear approximation of $f(X_i, \theta)$ at some given value of θ, denoted θ^o, is

$$Y_i = f(X_i,\theta) = f(X_i, \theta^o) + \sum_{t=1}^{k} \left[\frac{\partial f(X_i,\theta)}{\partial \theta_t}\right]_{\theta=\theta^o} (\theta_t - \theta_t^o) + \varepsilon_i$$

where the partial derivatives are evaluated at $\theta = \theta^o$. Then

$$Y_i = f(X_i,\theta^o) + \sum_{t=1}^{k} \beta_t A_{ti}^o \qquad i=1, \ldots, n \qquad [7.0]$$

where

$$\beta_t = (\theta_t - \theta_t^o) \text{ and } A_{ti}^o = \left[\frac{\partial f(X_i,\theta)}{\partial \theta_t}\right]_{\theta=\theta^o} \quad t=1\ldots k .$$

The expression (7.0) is linear in the β_t and least squares estimates of β_t can be obtained such that

$$\theta_t^1 = \hat{\beta}_t + \theta_t^o$$

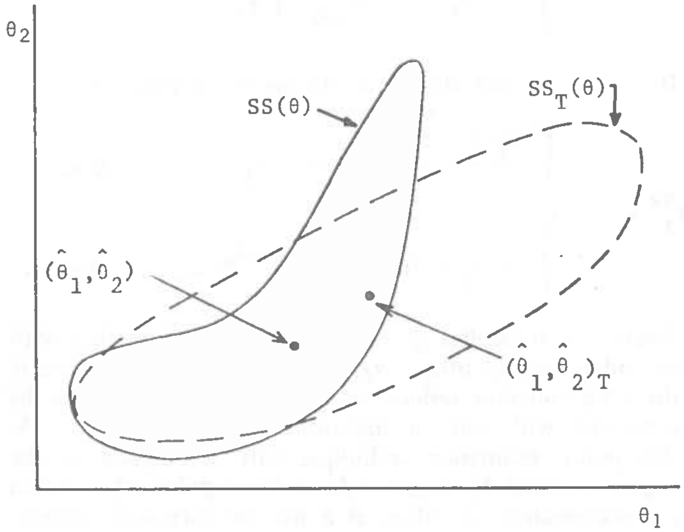
since θ_t^o are the known starting values. The new values of θ_t, denoted θ_t^1, replace the original starting values, θ_t^o, and the procedure is repeated until some convergence criterion is satisfied, e.g., $\left[\frac{\theta_t^{(j+1)} - \theta_t^j}{\theta_t^j}\right] < C$, where C is a predetermined small number, e.g., 10^{-6} and the parameter set has converged to some set of values. Additionally, at each iteration the sum of squares is calculated to see if a reduction in its value is achieved. If a reduction in value is not achieved at each successive iteration, the estimation procedure is diverging and a closer starting point must be picked.

$$SS_T(\theta) = \sum_{i=1}^{n} \{Y_i - f(X_i,\theta^o) - \sum_{t=1}^{n} \beta_t A_{ti}^o\}^2$$

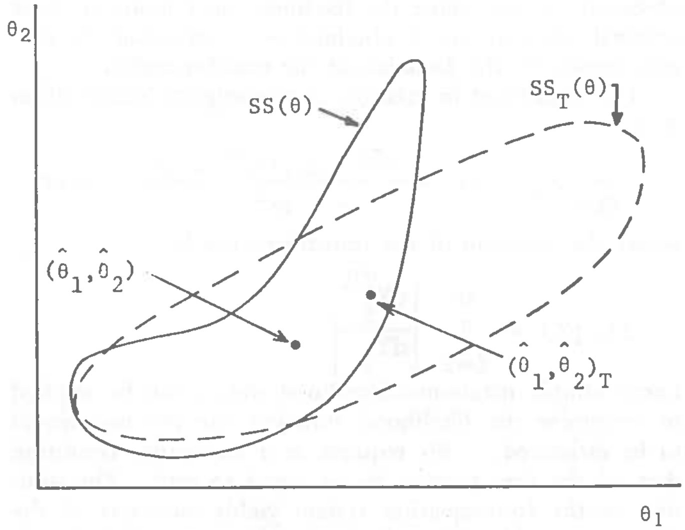
as compared to the actual sum of squares contour

$$SS(\theta) = \sum_{i=1}^{n} \{Y_i - f(X_i,\theta)\}^2 .$$

Clearly as $\theta^o \rightarrow \theta$, the contours become identical and the solution sets likewise become identical. The assumption of local linearity implies that θ^o is very close to θ. In figure 4, the local linearity assumption is valid and will lead to a correct solution since the parameter set $(\hat{\theta}_1, \hat{\theta}_2)$ is approximately the same for both the actual nonlinear sum of squares contour and the Taylor series approximation. In figure 5, however, the local linearity assumption breaks down and the solution is on a higher sum of squares contour than the solution contour and likely to diverge, rather than converge. The general direction of the Taylor series iterations is toward the center of the contour but the step size can be incremented to detect divergence properties (47). The path to convergence is not always smooth nor successful.



4. A successful Taylor series approximation.



5. An unsuccessful Taylor series approximation.

Method of steepest descent (gradient method)

The steepest descent method focuses on the sum of squares defined by

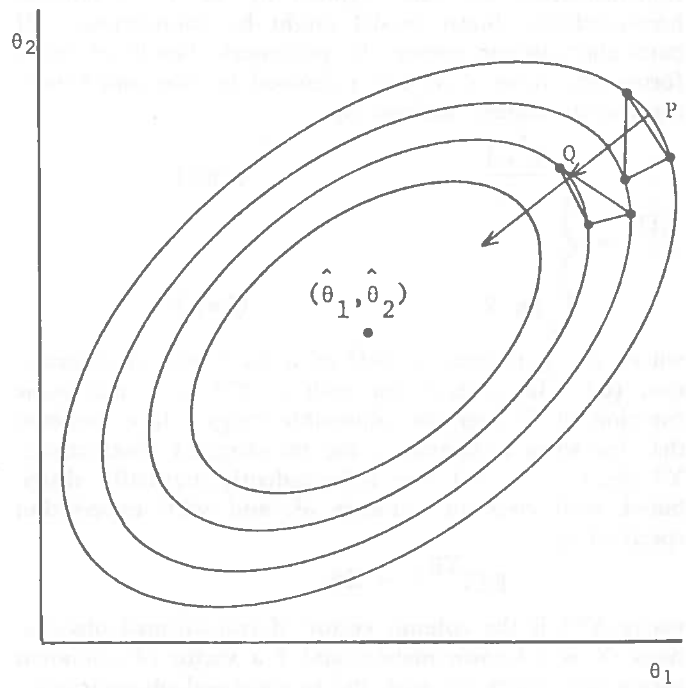
$$SS(\theta) = \sum_{i=1}^{n} \{Y_i - f(X_i, \theta)\}^2 .$$

to search for the minimum. The technique is analogous to noting the altitude of a number of points on the side of a mountain, fitting a plane to these points, and noting the direction of steepest descent in that plane. More specifically, the step is taken in the direction of the negative gradient of $SS(\theta)$, i.e.,

$$-\left[\frac{\partial SS(\theta)}{\theta_1}, \frac{\partial SS(\theta)}{\partial \theta_2}, \ldots, \frac{\partial SS(\theta)}{\partial \theta_k}\right]^T .$$

Starting in a particular region of θ, make several runs by selecting alternatives values for $\theta_1, \ldots, \theta_k$ and evaluate $SS(\theta)$ at these alternative combinations. Determine levels of $SS(\theta)$. Then determine the relative magnitudes and signs of the slopes between contours and find a direction to give the maximum reduction in $SS(\theta)$.

In figure 6, the procedure begins at point P with a search in this local area. The direction, which is at right angles to the contours, leads to point Q. There, repeat the procedure until convergence to (θ_1, θ_2) is achieved. The procedure eventually defeats itself, for as one approaches the minimum, the slopes become more and gradual and thus it becomes harder to estimate the proper plane with further sets of trials (16). Thus after some rapid initial progress, convergence may be achieved quite slowly if at all. As with the Taylor series, the method of steepest descent has a number of modifications that are used to control the step size once the direction is found.



6. The method of steepest descent.

The combination of these techniques would have obvious advantages. The method of steepest descent makes very rapid progress until an area of approximate local linearity is reached; that is the best area for the application of the Taylor series expansion. The Marquardt algorithm is composed of precisely this combination.

Transformation and Concentration

Transformation of the function

Hood and Koopmans suggested an approach for production function analysis that is a feasible alternative to the Taylor Series and simultaneous equation techniques (58). The technique for determining starting values for a nonlinear algorithm will be described for a k-input CES production function. The k-input CES function can be written

$$Y = \gamma[b_1 x_{1i}^{-g} + b_2 x_{2i}^{-g} + \dots + b_k x_{ki}^{-g}]^{-v/g}$$

$$i = 1,\dots,n \qquad \sum_{t=1}^{k} b_t = 1.$$

In a more concise form it becomes

$$Y_i = \gamma[\sum_{t=1}^{k} b_t x_{ti}^{-g}]^{-v/g} \quad \text{where } i = 1,\dots,n$$

Traditionally in the application of least squares regression, it is assumed that the observations $Y_1, \dots, Y_n$ are independently, normally distributed with constant variance and expectations specified by a model *linear* in the set of parameters. The k-input CES clearly is not linear in the parameters.

Box and Cox (6) have suggested that after a suitable transformation has been applied to the Y_i, a normal, homoscedastic, linear model might be appropriate. Of particular interest among the parametric family of transformations from Y to Y^{TR} examined by Box and Cox is the transformation defined by

$$Y^{TR} = \begin{cases} \dfrac{Y^{\lambda}-1}{\lambda} & (\lambda \neq 0) \\ \ln Y & (\lambda = 0) \end{cases}$$

where the parameter λ defines a particular transformation (6). In general, for each λ, Y^{TR} is a monotonic function of Y over the admissible range. It is assumed that for some unknown λ, the transformed observations, $Y^{TR}(i=1, \dots, n)$ are independently normally distributed with constant variance σ^2, and with expectation specified by

$$E(Y^{TR}) = X\theta$$

where Y^{TR} is the column vector of transformed observations, X is a known matrix, and θ a vector of unknown parameters associated with the transformed observations.

Rewriting the k-input CES function under consideration

$$Y_i = [\sum_{t=1}^{k} \theta_t x_{ti}^{-g}]^{-v/g} \quad \text{where } \theta_t = \gamma^{-g/v} b_t$$

and using the above transformation in addition to an error term, μ the optimal value of λ for estimation, can be found:

$$Y_i^{TR} = \begin{cases} \dfrac{Y_i^{\lambda}-1}{\lambda} = \dfrac{\{[\sum_{t=1}^{k} \theta_t x_{ti}^{-g}]^{-v/g}\}^{\lambda} - 1}{\lambda} + \mu_i & (\lambda \neq 0) \\ \ln Y_i = \ln \{[\sum_{t=1}^{k} \theta_t x_{ti}^{-g}]^{-g/v}\} + \mu_i & (\lambda \neq 0) \end{cases}$$

If $\lambda = -g/v$ the above transformation appears as

$$Y_i^{TR} = \begin{cases} \dfrac{Y_i^{\lambda}-1}{\lambda} = \dfrac{\sum^{k} \theta_t x_{ti}^{-g} - 1}{\lambda} + \mu_i & (\lambda \neq 0) \\ \ln Y_i = \ln \{[\sum^{k} \theta_t x_{ti}^{-g}]^{-v/g}\} + \mu_i & (\lambda = 0) . \end{cases}$$

Since $\lambda = 0$ implies $g = 0$ the problem of dealing with an indeterminate form, $-v/g$, is avoided. For if $g = 0$, the CES function reduces to the C-D form and can be estimated with only a logarithmic transformation. At this point, estimation technique will be applied to the region defined by $\lambda \neq 0$. At a later point, the region for examination of values of λ will be narrowed substantially.

The purpose of this transformation was to allow estimation of the parameter vector θ. To achieve this estimation, the likelihood function associated with the transformed function must be examined. Appendix III demonstrates that the probability density for the untransformed observations, and hence the likelihood in relation to these original observations, is obtained by multiplying the normal density by the Jacobian of the transformation.

The likelihood in relation to the original observations Y is

$$\ell = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp - \frac{(Y^{TR} - X\theta)'(Y^{TR} - X\theta)}{2\sigma^2} \cdot J(\lambda;Y)$$

where the Jacobian of the transformation is

$$J(\lambda;Y) = \prod_{i=1}^{n} \left| \frac{dY_i^{TR}}{dY_i} \right| .$$

Large sample maximum likelihood theory can be applied to maximize the likelihood function for the parameters to be estimated. This requires as a first-order condition that all the first partials be set equal to zero. The solution to the four-equation system yields estimates of the parameters. However, these equations are nonlinear in the parameters and it seems we have circled back to our original problem of nonlinearity.

The logarithm of the likelihood function, although its first partials are also nonlinear, provides a solution.

$$L = \ln \ell = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \{(Y^{TR}-X\theta)'(Y^{TR}-X\theta)\} + \ln J(\lambda;Y).$$

The function, L, is composed of two parameter sets, λ and θ. This makes it possible to assign temporarily fixed values to the parameters of one set in order to carry out a *provisional* maximization for those of the other set (58, p. 157). Given λ, the function L is, except for a constant factor, the likelihood for a standard least squares problem. The specification of λ implies that g must also be specified. Hence the maximum-likelihood estimates of θs are the least squares estimates for the dependent variable Y^{TR} and the estimate of σ^2, denoted $\hat{\sigma}^2(\lambda)$, is the residual sum of squares RSS(λ) divided by the degrees of freedom, i.e., $\hat{\sigma}^2(\lambda) = RSS(\lambda)/n{-}k{-}l$.

Since an analysis of variance is unchanged by a linear transformation, the initial transformation is equivalent to

$$Y^{TR*} = \begin{cases} Y^{\lambda} & (\lambda \neq 0) \\ \ln Y & (\lambda = 0). \end{cases}$$

Thus least squares estimates of the θs and $\hat{\sigma}^2(\lambda)$ obtained from

$$Y^{TR} = \frac{Y^{\lambda}-1}{\lambda} = \frac{\sum_{t=1}^{k} \theta_t x_{ti}^{-g} - 1}{\lambda}$$

are equivalent to those obtained from

$$Y^{TR*} = Y^{\lambda} = \sum_{t=1}^{k} \theta_t X_{ti}^{-g}$$

when g and v are fixed at specific values. The likelihood function for Y^{λ} is not, however, the same as the likelihood function for Y^{TR}. This can be easily seen if one compares the alternative Jacobians. The Jacobin for Y^{λ} is

$$\prod_{i=1}^{n} \frac{dY_i^{\lambda}}{dY_i} = \prod_{i=1}^{n} |\lambda Y_i^{\lambda-1}|$$

whereas the Jacobian of $Y^{TR} = \frac{Y^{\lambda}-1}{\lambda}$ is

$$\prod_{i=1}^{n} \frac{dY_i^{TR}}{dY_i} = \prod_{i=1}^{n} |Y_i^{\lambda-1}| .$$

The likelihood function that is being maximized is that of Y^{TR} not Y^{λ}. However, the fixing of λ allows the use of the least squares results employing Y^{TR} or Y^{λ} as equivalent transformations to maximize L.

Concentration of the likelihood function

The logarithm of the likelihood function of the transformed CES is a function of two parameter sets and the inputs, i.e.,

$$L = \ln \ell = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\{(Y^{TR} - X\theta)'(Y^{TR} - X\theta)\} + (\lambda-1)\sum_{i=1}^{n} \ln Y_i$$

where X is a matrix of inputs of the form

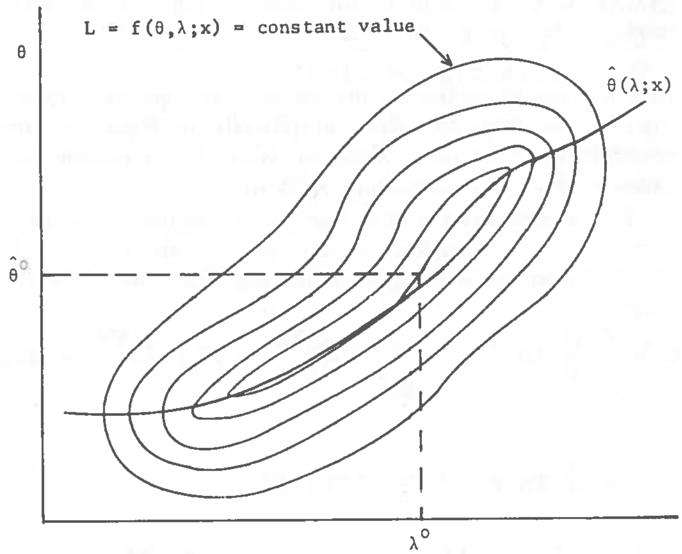
$$X = \begin{bmatrix} X_{11}^{-g} & X_{21}^{-g} & \cdots & X_{k1}^{-g} \\ X_{12}^{-g} & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ X_{1n}^{-g} & \cdots & \cdots & X_{kn}^{-g} \end{bmatrix} .$$

Thus $\quad L = f(\theta, \lambda; x)$

If values are assigned to g and v, thus determining the value of λ, then least squares estimates of the transformed model can be used to find the maximum value for L, given each λ. Under these conditions, the parameter set θ used to evaluate L becomes a function of the parameter set λ and x, i.e.,

$$\hat{\theta} = \hat{\theta}(\lambda; x) .$$

This can be illustrated graphically (figure 7). If we consider $\hat{\theta}$ and λ as scalars, the likelihood function forms contours related to different combinations of parameter values. Any point on a given contour represents the same value for the likelihood function.



7. The parameter set as a function of λ.

The function $\hat{\theta}(\lambda;x)$ is a line on the graph connecting the values of $\hat{\theta}$ that maximize L for a given λ. Thus $\hat{\theta}^o$ is the set of values of the parameters that maximizes L for $\lambda = \lambda^o$. The task of finding the maximizing values of θ for each λ is solved and the next step is to find the maximizing values of λ, denoted $\hat{\lambda}(x)$.

This step is made easier by recognizing the concept of a concentrated likelihood function (58). The above procedure has concentrated the likelihood function

$$L = f(\Theta, \lambda; x)$$

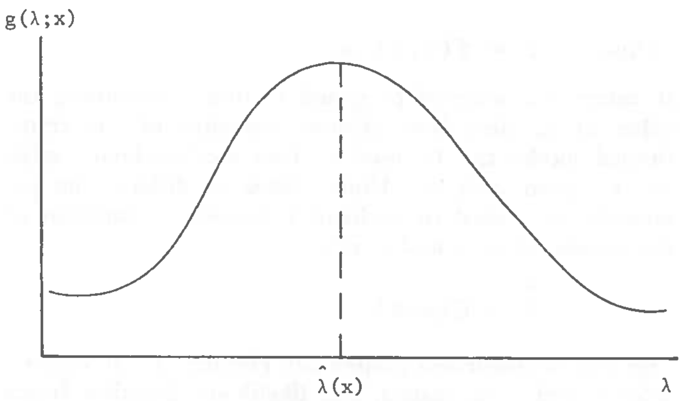
on λ by making the maximizing value for θ, denoted $\hat{\theta}$, a function of λ, i.e.,

$$\hat{\Theta} = \hat{\Theta}(\lambda; x) \quad .$$

Thus

$$L^* = f(\hat{\Theta}(\lambda; x), \lambda; x) = g(\lambda; x) \quad .$$

This concentrated likelihood function, $g(\lambda; x)$, contains only one parameter set and can be maximized. This is graphically demonstrated in figure 8 treating λ as a scalar.



8. Maximization of the concentrated likelihood function.

The value denoted $\hat{\lambda}(x)$ would be the value of λ that maximizes the concentrated likelihood function, $g(\lambda;x)$. Corresponding to the value of $\hat{\lambda}(x)$ (the maximizing values of g and v) is a parameter set.

$$\hat{\Theta} = \hat{\Theta}(\hat{\lambda}(x); x) = \hat{\Theta}(x)$$

that was found earlier by the use of least squares. Viewing the two-step procedure graphically in figure 9, the maximizing value of λ, denoted $\hat{\lambda}(x)$, has a unique parameter set, $\theta(x)$, associated with it.

The mathematics of the likelihood function concentration is relatively simple. Given that we can find $\hat{\theta} = \hat{\theta}(\lambda; x)$ by simple least squares, the log likelihood function

$$L = -\frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\{(Y^{TR} - x\Theta)'(Y^{TR} - x\Theta)\}$$

$$- \frac{n}{2}\ln \sigma^2 + \ln J(\lambda; Y)$$

with a fixed λ reduces to a concentrated likelihood function which, except for a constant, is

$$L^* = \frac{n}{2}\ln \hat{\sigma}^2(\lambda) + \ln J(\lambda; Y) = g(\lambda; x) \quad .$$

Evaluating the Jacobian for fixed λ

$$\ln J(\lambda; Y) = \ln \prod^{n} \left|\frac{dY_i^{\lambda}}{dY_i}\right| = \ln \prod^{n} |Y_i^{\lambda-1}|$$

$$= (\lambda-1)\sum_{i=1}^{n} \ln Y_i \quad .$$

Substituting, the concentrated likelihood function, except for a constant, becomes

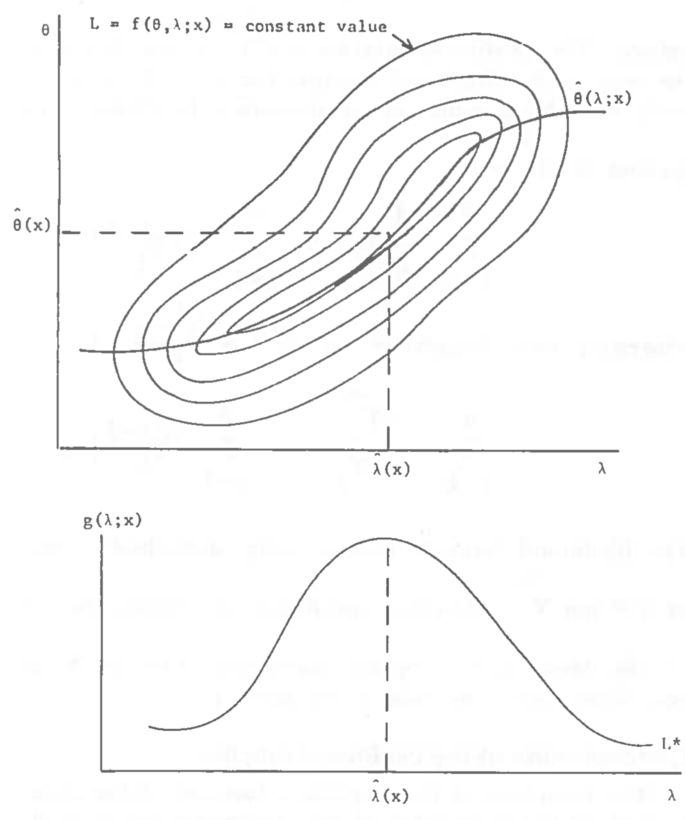
$$L^* = -\frac{n}{2}\ln \hat{\sigma}^2(\lambda) + (\lambda-1)\sum_{i=1}^{n} \log Y_i$$

$$= g(\lambda; x) \quad .$$

The estimation procedure is complete once one finds the value of λ, denoted $\hat{\lambda}(x)$, that maximizes L* and concurrently the value of $\hat{\theta}$ based on this λ value.

In summary, a four-step procedure is used:

1. Perform the transformation upon the CES function and apply least squares to the dependent variable y^{λ} for given λ, to get estimates of θ_t.
2. Use the residual sum of squares from the above to estimate σ^2, denoted $\hat{\sigma}^2(\lambda)$, for each value of λ.
3. Evaluate L* and graph to select the value $\hat{\lambda}(x)$ that maximizes L*.



9. The two-step maximization procedure.

4. Given the value $\hat{\lambda}(x)$, return to the regressions performed in step (1) to find the maximizing value for the θ_ts corresponding to $\hat{\lambda}(x)$ and these become $\hat{\theta}(x)$.

The obvious difficulty, of course, is that λ is composed of two unknown parameters, v and g. The maximizing value of λ, $\hat{\lambda}(x)$, can be formed with any number of combinations of v and g. Obviously, if we select values for g, we are biasing our results on a crucial parameter. Therefore, it would seem better to try to find some a priori evidence regarding the returns to scale parameter (v). Roy Black concluded that the relative bias was 5% or less if v was constrained to the C-D value (3). Maddala and Kadane, however, found that using the C-D function to estimate v in a CES function biased the value downward when the elasticity of substitution was greater than unity and biased it upward when the elasticity was less than unity (68).

The degree of bias was not evaluated. It would appear best to conduct the search over both parameters. Initially it seems best to set v at the C-D value and search over values of g to form λ. Once $\hat{\lambda}(x)$ has been found, the amount of change in $\hat{\lambda}(x)$ resulting from changes in v can be analyzed to get an approximate value for $\hat{\lambda}(x)$. Some flexibility exists, since the technique is used solely to obtain starting guesses for the Marquardt algorithm.

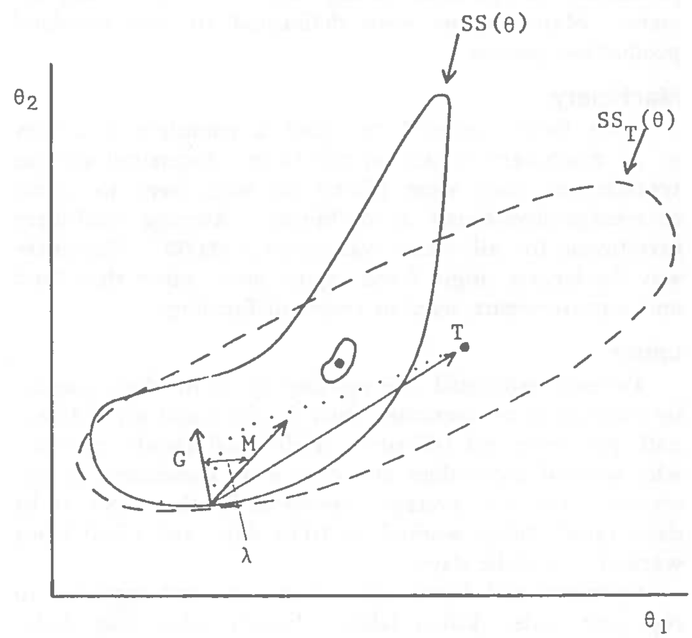
The Marquardt Algorithm

Most algorithms for the least squares estimation of nonlinear parameters have used either of two approaches. Either the model is expanded as a Taylor series with corrections to the several parameters calculated at each iteration on the assumption of local linearity, or various modifications of the method of steepest descent have been used. Each of these approaches has a serious shortcoming. The Taylor series often diverges on successive iterations instead of converging. The steepest-descent (or gradient) methods are agonizingly slow to converge after the first few iterations (69).

The Marquardt algorithm is based on the maximum neighborhood method which, in effect, performs an optimum interpolation between the Taylor series method and the gradient method (69). The algorithm combines the desirable property of the gradient method of converging from an initial guess that may be outside the region of convergence of other methods, with the desirable property of the Taylor series of rapid convergence once the vicinity of the converged values has been reached. The best features of these algorithms are combined while avoiding their most serious limitations. An examination of these alternative techniques will provide a better grasp of the problems that are overcome and the reliability of any solution.

Marquardt initially examined the behavior of the two techniques from a common point on some $SS(\theta)$ contour. Monitoring the angle γ between the alternative directions that the techniques dictated, he found that γ usually falls between 80° and 90°. Any improvement in these methods must in some sense interpolate between the alternative solution vectors (69).

Figure 10 illustrates the behavior of the Marquardt algorithm. The vector computed by the Marquardt algorithm will be somewhere inside the angle γ, depending on the selection of some constant. The algorithm can be constrained to either technique or allowed to interpolate between them as shown by the vector M. The vector G represents the gradient vector and the vector T the Taylor series solution. The angle γ is formed by these two vectors. In this example, the Taylor series goes to a higher sum of squares contour rather than a lower contour; thus it is diverging. The gradient vector is going across the elongated base of the nonlinear contour.



10. The Marquardt algorithm.

Impact of the TCA approach

Generally, a problem can be constructed to defeat any solution procedure. Also, most pure methods can be improved through modification. Thus no method can be called "best" for all nonlinear problems, but the TCA (Transformation-Concentration-Algorithm) approach certainly insures a greater versatility with regard to modifications of the algorithm or of the production function form. The CES, like the C-D, can now be estimated with only data on inputs and outputs, without regard to factor prices, expansion paths, etc.

The question of model superiority cannot be resolved by merely advancing a new technique that simplifies estimation. The C-D function has remained popular in applied research because of its roubustness with all types of data. The CES function and TCA approach must also stand this test before a final judgment can be made.

APPLICATION OF THE NONLINEAR ESTIMATION TECHNIQUE

Preliminary Evidence

The Data

The cross-section data used for applying the estimation technique were obtained from a study of the current financial structure of farms in the Columbia Basin Project of central Washington (119). Financial and farm management data for 1966 were obtained by personal interviews on 92 randomly selected farms of average or above size. Farms of this size, averaging 370 acres, reduced the probability of operators having full-time nonfarm employment. Major inputs were delineated for the cropland production process.

Machinery

Each farm surveyed provided a complete inventory of all machinery in use on the farm. Estimated current replacement costs were placed on each item to arrive at average investment in machinery. Average machinery investment for all farms was about $20,000. Machinery was the largest single fixed capital asset, other than land and improvements, used in cropland farming.

Labor

Farmers estimated the number of 10-hr days worked by themselves as operators, their families and hired labor, both part-time and full-time. Individual family members who worked more than 200 days were considered as operators. On the average, operators worked 283 10-hr days, family labor worked 30 10-hr days, and hired labor worked 150 10-hr days.

Operator and hired labor were grouped together to represent male, skilled labor. Family labor was designated as unskilled. The limited amount of information on the characteristics of the labor input dictated this classification. Only labor directly connected with cropland production was included.

Cash expenditures

Cash production expenditures include cash farm business expenses, such as custom labor and machinery for very specialized jobs, fertilizer, seed, taxes, and interest paid on mortgages and debts. Depreciation expenses on machinery and improvements are not included. Since many farms also raise livestock, it was necessary to exclude expenses not directly attributable to cropland production.

Land

All land used in cropland production was classified by type and expressed in terms of current value. The farms surveyed ranged from 75 to 1,150 acres of cropland and averaged 370 acres. Average investment in land was $165,000. The estimated value of land provided by the farmer was corrected for land that was fallowed, in the soil bank, or used for other than crop production.

The above data provide crude information on two types of land input, two types of fixed capital, and one type of variable capital. The corresponding measure of output was that part of gross farm income directly attributable to actual cropland operations.

Using these definitions, 69 of the original 92 farm surveys were considered as the sample set. Those eliminated lacked some information.

Fitting the Cobb-Douglas function

Initially, the C-D function was fitted to obtain an estimate of the returns to scale. Remember that the C-D estimate of returns to scale is relatively good if the actual elasticity of substitution approaches unity. The C-D form, fitted after a logarithmic transformation, was

$$Y = Ax_1^{b_1}x_2^{b_2}x_3^{b_3}x_4^{b_4}x_5^{b_5}u$$

where Y = cropland revenues
x_1 = machinery evaluated at replacement costs
x_2 = operator labor and hired labor expressed in 10-hr days
x_3 = family labor expressed in 10-hr days
x_4 = cash expenditures
x_5 = land expressed in dollars of current value.

The results of the regression using the logarithmic form of the C-D function are summarized in the correlation matrix in table 1.

TABLE 1. Correlation matrix for C-D function.

	x_1	x_2	x_3	x_4	x_5	Y
x_1	1.0000	.6393	- .0295	.6528	.6511	.6917
x_2		1.0000	- .3562	.7217	.6427	.7347
x_3			1.0000	-.1888	-.0737	-.1370
x_4				1.0000	.6340	.8813
x_5					1.0000	.6346

The following relationships among inputs are suggested by the correlation matrix:

1. Family labor is inversely related to all other inputs. Therefore, as the farm becomes larger and output increases, the reliance on family labor declines. The smaller farms generally use family labor to a larger extent to reduce labor costs.
2. The high correlation between cash expenditures and farm income indicates the importance of variable capital in this area.
3. The validity of farmers' estimates of land value may explain part of the low correlation between land and gross farm income, but two characteristics of the sample may also provide some insights. First, the largest farms were almost completely devoted to forage and small grain crops while the higher valued land composing the smaller farms was more intensively cropped (119). Since the land input is expressed in dollar values, size in acres is lost. Thus, smaller acreages of high value are considered identical in land input to larger farms of less per acre value. The gross farm income from these supposedly identical land inputs depends

heavily upon the relative prices for the respective crops. Second, the composition of payments for rented land is important. If the rent is paid on a crop share basis, that portion for rent does not appear in gross farm income. If the rent is paid in cash, it is included in gross farm income, but also added as a cash farm expense. Thus, the investment for cash-rented land is probably overestimated, while the return to crop-share rented land is underestimated.

4. The correlation between machinery investment and farm income reflects the diversity of machinery requirements on farms that are intensively cropped and farms that are largely forage and small grain. Equal yields in terms of gross farm income may have very different machinery requirements, depending on type of operation.
5. The major labor input is apparently operator labor and hired labor. It is inversely related to family labor, as expected, and the results imply that family labor is a relatively insignificant input.

The dominant characteristic of the data at this point is the key role played by the variable factors, cash farm expenditures and nonfamily labor. The regression results in table 2 support this conclusion.

TABLE 2. Regression results for the Cobb-Douglas production function.

Parameters		Estimated coefficients	t-Values[a]
$\hat{A}$		.1460[b]	-1.6113
Machinery	$\hat{\beta}_1$	.2107	1.6446
Op. and hired labor	$\hat{\beta}_2$	.3766	1.9173
Family labor	$\hat{\beta}_3$	.0208	.9439
Cash expend	$\hat{\beta}_4$	.7510	7.7209
Land	$\hat{\beta}_5$	.0234	.1635

Sum of $\hat{\beta}_t$ = 1.383

Corrected R^2 = .8129

F = 54.7334

[a]The t-value is the ratio of the estimated coefficient to the estimated standard error of the coefficient.

[b]This is the anti-logarithm of the estimate of log A.

Two statistics in particular will be used to examine the results of the C-D function. The F-value is used to test the overall relation, that is, whether $x_1, x_2, \ldots, x_5$ influence Y. The value for the F-test with probability of .01 of rejecting a true hypothesis and degrees of freedom $(t + 1)-1$ and $n - (t + 1)$ is

$$F_{5,63} = 3.32$$

The F-value based on the regression is 54.7334 and the hypothesis that the x_t *do not* influence Y is therefore rejected. The correct R^2, or corrected coefficient of determination, implies that about 81% of the variance of the dependent variable can be explained by the independent variables.

The t-values test whether each coefficient, β_t, is equal to zero. The critical t-value for testing the hypothesis of $\beta_t = 0$ at different levels of significance with $n-(t+1)$ degrees of freedom are, for this study,

$$t_{63,\ .01} = 2.66$$
$$t_{63,\ .05} = 2.00$$
$$t_{63,\ .10} = 1.67$$
$$t_{63,\ .50} = .678.$$

The cash expenditure input is significantly different from zero at the .01 level. The coefficient for operator and hired labor is significant at the .10 level, while the coefficient for machinery is significant at some level greater than .10. Although family labor is significant at the .50 level the probability of rejecting the hypothesis that $\beta_3 = 0$ when it is true, namely .50, is so large that it cannot be considered much better than the coefficient for land. The level of significance is clearly higher for β_3 than for β_5, but neither appears to explain much of the movements of gross farm income. The efficiency parameter A is significant at some level slightly more than .10.

A complete reliance upon the t-test for judging the relative significance of the coefficients must be avoided. Draper and Smith have observed,

> The effect of an x-variable (x_q say) in determining a response may be large when the regression equation includes only x_q. However, when the same variable is entered into the equation after other variables, it may affect the response very little, due to the fact that x_q is highly correlated with variables already in the regression equation (23).

The t-test of the coefficient for machinery may be an example of such a case.

Using an F test for a subset of coefficients based on the sum of squares attributable to the subset, it may be possible to explain the relatively low significance level of the coefficient for machinery. The results of the regression imply that at least two variables, cash farm expenses and operator and hired labor, have some explanatory powers. Further, the high correlation between machinery and each of these variables may be masking some of the influence of the machinery input. But of greater concern is whether less significant variables are affecting the degree of certainty one should have in including machinery as a significant input. The tests would certainly be more important if one were trying to select variables to include in the model, but the test does provide useful insights, even though the intention is to retain all inputs and ascertain relative productivities.

Beginning with a regression containing cash farm expenditures and operator and hired labor, the remaining variables can be added to determine the addition to the explained sum of squares from each set relative to the residual sum of squares for the regression with all the variables included. The F test for the relative subsets can then be constructed. The hypotheses and the respective conclusions based on the F test at a .01 level of significance are listed below:

Machinery	$H_0 : \beta_1 = 0$	Reject
Family labor	$H_0 : \beta_3 = 0$	Accept
Land	$H_0 : \beta_5 = 0$	Accept
Machinery and land	$H_0 : \beta_1 = \beta_5 = 0$	Accept
Machinery and family labor	$H_0 : \beta_1 = \beta_3 = 0$	Accept
Family labor and land	$H_0 : \beta_3 = \beta_5 = 0$	Accept
Machinery, family, labor and land	$H_0 : \beta_1 = \beta_3 = \beta_5 = 0$	Accept

The test of the first hypothesis indicates that the machinery input is significant at the .01 level when included in a regession with only cash expenditures and operator and hired labor. The tests of the other hypotheses imply that the inclusion of family labor or land investment, or both, is masking the total influence that the machinery input may exert upon gross farm income.

The returns to scale under the three-input function, namely cash expenditures, operator and hired labor, and machinery investment, was 1.3166 as compared to returns to scale of 1.383 under the original formulation. This is particularly important, since the objective of this preliminary investigation of the C-D function is to arrive at an approximate starting value for the return to scale parameter in the CES function.

Concentrated Likelihood Function

With an estimated value of v = 1.383, the 4-step procedure outlined above can be applied to the CES function of the form

$$Y = \gamma[b_1x_1^{-g} + b_2x_2^{-g} + b_3x_3^{-g} + b_4x_4^{-g} + b_5x_5^{-g}]^{-v/g} \qquad \sum_{t=1}^{5} b = 1$$

where Y and the x_t are defined as in the C-D case.

Step 1. Perform the transformation upon the CES function and apply least squares to the dependent variable Y^{λ} for a given λ, to get estimates of b_t. The constant λ equals $-g/v$. Inasmuch as v has been set at v = 1.383 it is only necessary to search over values for g.

The CES form under examination can be written more concisely after transformation as

$$\frac{Y^{\lambda}-1}{\lambda} = \frac{\sum_{t=1}^{5} \theta_t X_t^{-g} - 1}{\lambda}$$

where

$$\sum_{t=1}^{5} \theta_t = \sum_{t=1}^{5} \gamma^{-g/v} b_t$$

The range of acceptable values for g is $-1 \leq g \leq \infty$, assuming the proper curvature for isoquants (1). The choice as to the effective range of g over which to examine the likelihood must be tempered with some judgment as to possible values for the elasticity of substitution. For this demonstration, values of g ranged from –.9 to +.9, which implies an elasticity of substitution ranging from 10 to .53.

For an analysis of variance, the equivalent form

$$Y^{\lambda} = \sum_{t=1}^{5} \theta_t x_t^{-g}$$

can be used. Since v is constrained to the C-D estimate, specifying values for g will determine the values of λ and the above equation becomes linear in the parameters with estimates of θ_ts, denoted $\hat{\theta}_t$s, derivable from the equation

$$Y^{\lambda} = \theta_1x_1^{-g} + \theta_2x_2^{-g} + \theta_3x_3^{-g} + \theta_4x_4^{-g} + \theta_5x_5^{-g} + \varepsilon.$$

Step 2. Use the residual sum of squares from the regression above to estimate $\hat{\sigma}^2(\lambda)$ for each λ. The interest at this point centers on the estimate of the residual sum of squares. This statistic divided by the degrees of freedom yields $\hat{\sigma}^2(\lambda)$, the adjusted residual variance. The estimated adjusted residual variance for each value of λ is in table 3.

Step 3. Evaluate and graph L* to find the value of λ, denoted $\hat{\lambda}(x)$, that maximizes the concentrated likelihood function. As demonstrated before, the concentrated likelihood function is, except for a constant, C,

$$L^* = -\frac{n}{2} \ln \hat{\sigma}^2(\lambda) + (\lambda-1) \sum_{i=1}^{69} \ln Y_i$$

Table 3 shows an evaluation L* over the range of $-.9 \leq g \leq .9$. The graph of the likelihood function in fig. 11 indicates that its maximum value falls in the range $-.1 < g < .1$. The elasticity of substitution must therefore lie in the range $1.11 < \sigma < .91$. Since the maximization value for g would appear to lie close to zero, remember that if g = 0, the CES production function reduces to the C-D form.

Since the nonlinear algorithm will perform most efficiently when the starting guesses are reasonably close to the true values, it is desirable to examine the range $-.1 \leq g \leq .1$ more closely. Table 4 and fig. 12 show the results of this examination.

Figure 12 implies that the data are consistent with a set of starting values which could reduce the CES function to the C-D form. Some crucial questions must be asked at this point. By constraining v to the C-D estimate has a bias been introduced toward the C-D form, i.e., g = 0? Is it necessary to search the concentrated likelihood function over both parameters g and v? The answers to these questions, at least for this example, are in table 5 and fig. 12. The maximizing value of g appears to be invariant with respect to the value of the returns to scale parameter, at least over the range $1.00 \leq v \leq 1.766$. These results will be discussed in more detail later.

The concentrated likelihood function appears to yield two equally good starting values for g. These values are implied when the maximizing values $\hat{\lambda}(x)$ are found with $v = 1.38$. The maximizing values are $\hat{\lambda}(x) = +.006$ and $\hat{\lambda}(x) = -.006$.

Step 4. Given the value $\hat{\lambda}(x)$, it is possible to return to the regressions performed in step 1 to find the $\hat{\theta}_t$ corresponding to the value of $\hat{\lambda}(x)$ and denote these as $\hat{\theta}_t(x)$ In this case, two values of $\hat{\lambda}(x)$ are being examined and from table 4 the following information can be obtained:

$\hat{\lambda}(x) =$	+.006	−.006
$g =$	−.008	+.008
$\hat{\theta}_1(x) =$	.1549	.1612
$\hat{\theta}_2(x) =$	.2858	.2791
$\hat{\theta}_3(x) =$	.0161	.0150

TABLE 3. Maximization of the concentrated likelihood function $-.9 \le g \le .9$

g	λ=-g/v	$\hat{\sigma}^1(\lambda)$	L*	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$
+.9000	-.6508	.0000001	-686.370	1.5066	.0885	.0001	4.0090	6.5989
.8000	-.5785	.0000004	-654.380	1.2110	.1011	.0003	3.1982	3.9529
.7000	-.5061	.0000014	-644.277	.9625	.1154	.0004	2.5595	2.3559
.6000	-.4338	.0000043	-629.737	.7558	.1312	.0008	2.0515	1.3873
.5000	-.3615	.0000127	-613.849	.5873	.1488	.0013	1.6474	.8025
.4000	-.2890	.0000352	-600.031	.4529	.1685	.0022	1.3261	.4523
.3000	-.2170	.0000853	-574.977	.3458	.1897	.0036	1.0637	.2427
.2000	-.1446	.0001664	-543.593	.2634	.2138	.0059	.8539	.1217
.1000	-.0723	.0001846	-492.313	.2002	.2410	.0095	.6827	.0535
-.1000	.0723	.0037	-490.479	.1163	.3088	.0231	.4290	-.0012
-.2000	.1446	.0686	-529.010	.0895	.3520	.0351	.3364	-.0088
-.3000	.2170	.7177	-565.281	.0694	.4042	.0517	.2622	-.0110
-.4000	.2890	5.9364	-585.148	.0541	.4641	.0738	.2015	-.0104
-.5000	.3615	44.4052	-601.153	.0428	.5382	.1026	.1549	-.0089
-.6000	.4338	309.0442	-614.828	.0341	.6241	.1374	.1179	-.0071
-.7000	.5061	2067.9533	-627.198	.0275	.7239	.1767	.0892	-.0054
-.8000	.5785	13566.5852	-638.815	.0224	.8394	.2177	.0672	-.0041
-.9000	.6508	87812.5081	-649.906	.0184	.9694	.2541	.0504	-.0030

TABLE 4. Maximization of the concentrated likelihood function $-.1 \le g \le .1$

g	λ=-g/v	$\sigma^1(\lambda)$	L*α	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$
.100	-.0723	.00018	-492.313	.2002	.2410	.0095	.6827	.0535
.075	-.0542	.00015	-472.692	.1869	.2483	.0107	.6449	.0420
.050	-.0361	.000097	-444.428	.1744	.2558	.0120	.6088	.0322
.025	-.0181	.000035	-396.012	.1625	.2631	.0134	.5734	.0238
.020	-.0145	.000025	-381.079	.1612	.2663	.0137	.5700	.0225
.018	-.0130	.000020	-373.150	.1598	.2660	.0138	.5655	.0218
.016	-.0116	.000017	-365.580	.1595	.2676	.0140	.5648	.0213
.014	-.0101	.000013	-356.099	.1579	.2671	.0141	.5598	.0206
.012	-.0087	.000010	-345.682	.1578	.2689	.0143	.5596	.0202
.010	-.0070	.0000067	-330.683	.1518	.2609	.0139	.5394	.0190
.008	-.0060	.0000050	-319.988	.1612	.2791	.0150	.5724	.0196
-.008	.0060	.0000064	-319.800	.1549	.2858	.0161	.5542	.0155
-.010	.0070	.0000090	-330.400	.1432	.2663	.0152	.5123	.0139
-.012	.0087	.000014	-345.275	.1479	.2772	.0159	.5296	.0140
-.014	.0101	.000020	-355.819	.1463	.2764	.0159	.5243	.0135
-.016	.0116	.000027	-365.067	.1463	.2786	.0162	.5247	.0131
-.018	.0130	.000035	-372.988	.1449	.2781	.0162	.5200	.0126
-.020	.0145	.000045	-380.478	.1447	.2800	.0164	.5199	.0122
-.025	.0181	.000075	-395.376	.1416	.2795	.0167	.5097	.0110
-.050	.0361	.00044	-443.073	.1328	.2892	.0187	.4821	.0061
-.075	.0542	.0014	-470.882	.1248	.2989	.0208	.4550	.0021
-.100	.0723	.0037	-490.479	.1164	.3088	.0232	.4290	-.0012

$\hat{\theta}_4(x) = \quad .5542 \quad .5724$

$\hat{\theta}_5(x) = \quad .0155 \quad .0196.$

From each of these parameter sets, a set of β^*_t can be obtained by using the constraints

$$\Sigma\ \hat{\theta}_t(X) = \gamma^{-g/v} \text{ and } \theta_t(X) = \gamma^{-g/v}\ \hat{\beta}^*_t\ .$$

The starting values for g, v, and the $\hat{\beta}_t$ are in table 6. The impact of changing the scale parameter can also be clearly seen in table 6. Not only is the maximizing value of g invariant with respect to changes in the value of v, but the factor intensities β_t also appear invariant. This preliminary evidence suggests that if the elasticity of substitution exceeds unity, higher returns to scale dictate a lower efficiency parameter. But if the elasticity of substitution is less than unity, higher returns to scale dictate a higher efficiency parameter.

Marquardt Algorithm in Estimation Estimates at Convergence

Initially, the starting values corresponding to v = 1.383 and g = –.008 were used to estimate the CES form

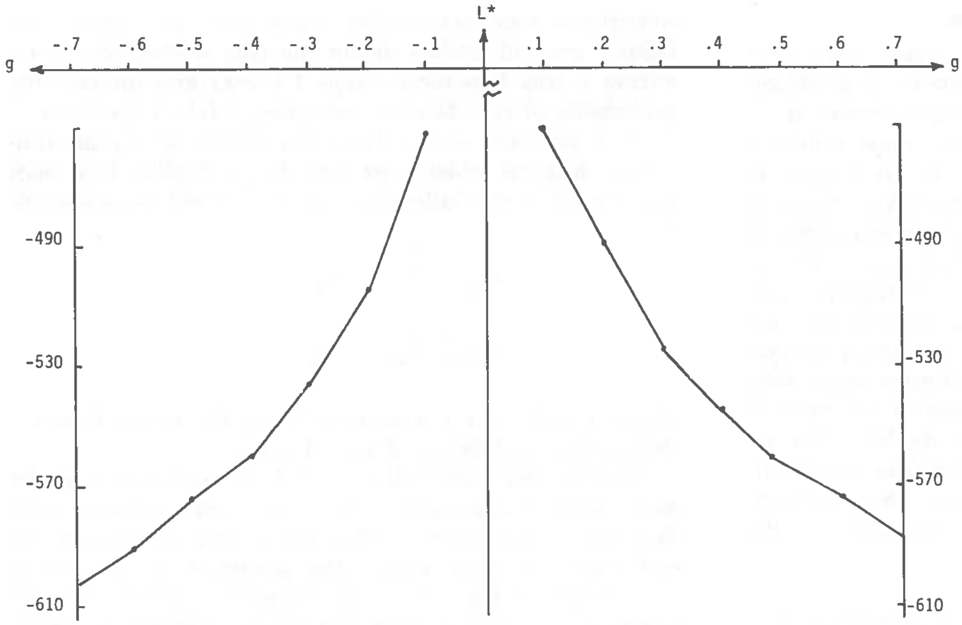
$$Y = \gamma[b_1x_1^{-g} + b_2x_2^{-g} + b_3x_3^{-g} + b_4x_4^{-g} + b_5x_5^{-g}]^{-v/g}$$

TABLE 5. Maximization of concentrated likelihood function over two variables.

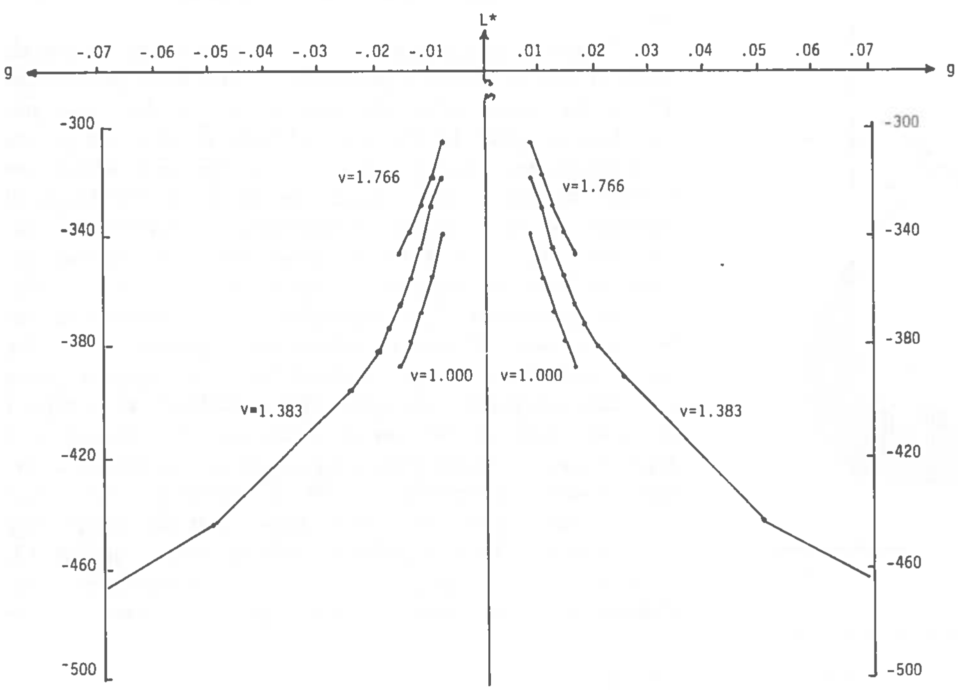
g	=-g/v	$\hat{\sigma}^2$ (λ)	L*	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$
				v = 1.000				
.016	-.016	.000029	-387.742	.2115	.3537	.0187	.7419	.0269
.014	-.014	.000023	-378.540	.2114	.3565	.0190	.7430	.0265
.012	-.012	.000018	-367.881	.2113	.3593	.0192	.7441	.0260
.010	-.010	.000013	-355.226	.2112	.3621	.0195	.7452	.0256
.008	-.008	.0000086	-339.832	.2111	.3650	.0197	.7464	.0251
-.008	.008	.000012	-339.742	.2102	.3885	.0218	.7556	.0216
-.010	.010	.000020	-355.019	.2101	.3916	.0221	.7568	.0212
-.012	.012	.000029	-367.593	.2100	.3946	.0224	.7579	.0208
-.014	.014	.000042	-378.178	.2099	.3977	.0227	.7591	.0203
-.016	.016	.000057	-388.052	.2098	.4008	.0230	.7603	.0199
				v = 1.766				
.016	-.009	.000011	-347.983	.1266	.2129	.0111	.4511	.0174
.014	-.008	.0000086	-339.953	.1274	.2159	.0113	.4539	.0170
.012	-.007	.0000067	-330.683	.1289	.2200	.0116	.4589	.0168
.010	-.006	.0000050	-320.056	.1313	.2258	.0120	.4675	.0166
.008	-.005	.0000036	-307.394	.1355	.2348	.0126	.4823	.0167
-.008	.005	.0000044	-307.181	.1279	.2359	.0134	.4567	.0127
-.010	.006	.0000064	-319.743	.1217	.2261	.0129	.4342	.0117
-.012	.007	.0000090	-330.360	.1172	.2194	.0126	.4181	.0109
-.014	.008	.000012	-339.453	.1137	.2145	.0124	.4056	.0102
-.016	.009	.000015	-347.612	.1109	.2108	.0123	.3955	.0096

TABLE 6. Starting values for Marquardt algorithm over two parameters.

	g=-.008			g=+.008		
v	1.000	1.383	1.766	1.000	1.383	1.766
$\Sigma\Theta_\tau$	1.3977	1.0265	.8466	1.3673	1.0473	.8819
$\hat{\gamma}^*$	1.84×10^{18}	138.00	9.68×10^{-10}	8.13×10^{-18}	.0003	6560000
$\hat{\beta}^*_1$	.1504	.1509	.1511	.1544	.1539	.1536
$\hat{\beta}^*_2$	.2779	.2784	.2786	.2669	.2665	.2662
$\hat{\beta}^*_3$	.0156	.0157	.0158	.0144	.0143	.0143
$\hat{\beta}^*_4$	.5406	.5399	.5395	.5459	.5465	.5469
$\hat{\beta}^*_5$	.0154	.0151	.0150	.0184	.0187	.0189



11. The concentrated likelihood over the interval $-.7 \leq g \leq .7$



12. The concentrated likelihood function over the interval $-.075 \leq g \leq .075$ with v variable.

directly using the Marquardt algorithm. After 17 iterations, the algorithm converged to the values given in table 7.

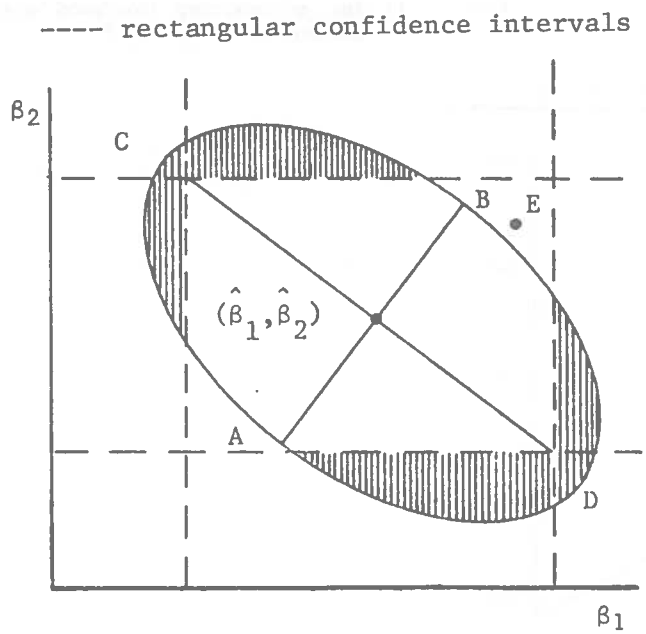
However, when the algorithm was started at the values corresponding to $v = 1.383$ and $g = +.008$, a serious shortcoming in the procedure appeared. If the *true value* of g at convergence is $-.0083$ (as was discovered in the first solution), the program must iterate very closely at $g = 0$. An examination of the CES form shows the trouble. At the point $g = 0$, the CES reduces to the C-D form, but in the algorithm, the term $-v/g$ becomes indeterminate, since division by zero is being attempted. The initial output from the algorithm with the second set of values showed that the iterations were converging toward $g = 0$ before the program was ended by division by zero.

TABLE 7. Convergence values for nonlinear algorithm.

Parameters	Starting values	Values at convergence
$\hat{v}$	1.383	1.4532
$\hat{g}$	-.008	-.0083
$\hat{\gamma}$	138.000	2.6723
$\hat{\beta}_1$	.1509	.1519
$\hat{\beta}_2$	.2784	.2706
$\hat{\beta}_3$	.0157	.0151
$\hat{\beta}_4$	.5399	.5280
$\hat{\beta}_5$	.0151	.0148

Confidence Intervals

The usual tests appropriate in the linear model are generally not appropriate when the model is nonlinear (23). When the error ε of the nonlinear model is assumed to be normally distributed, $\hat{\theta}$ is no longer normally distributed, $\hat{\sigma}^2 = RSS(\theta)/(N-(t+1)$ is no longer an unbiased estimate of σ^2, and there is no variance-covariance matrix of the linear form. Thus the examination of the results from the Marquardt algorithm cannot be carried out using the usual regression tests. Marquardt concluded that the support-plane confidence intervals for each parameter individually were the most realistic portrayal of the precision of the parameter estimates separately. Figure 13 shows the rationale for the use of this type of confidence interval for a two-parameter model. The actual joint 95% confidence region for the true parameter, β_1 and β_2, will be an ellipse that encloses the parameter set that the data regarded as jointly reasonable for the parameters $\hat{\beta}_1$ and $\hat{\beta}_2$.



13. Rectangular confidence intervals over an ellipsoid.

The conventional 1-parameter 95% confidence intervals give a minimum length interval for each parameter, on the assumption that the *remaining simultaneous parameter estimates are the same as their corresponding population values.* These intervals are represented as the distance between the dashed lines perpendicular to the β_t axis in question. If one tries to interpret these intervals simultaneously, i.e., regard the rectangle that is defined by the individual confidence intervals as a joint confidence region, then coordinates of points like E would be considered reasonable values. Of greater concern is the fact that the coordinates of points in the shaded region of the ellipse will be rejected as reasonable values. The rectangular confidence intervals clearly underestimate the true interval within which a parameter set may lie and still remain in the actual confidence ellipsoid. Thus 95% rectangular confidence intervals (a probability of .05 of rejecting a true hypothesis) when used as a joint confidence interval underestimate the true probability of rejecting a true hypothesis (type I error) and increase the probability of type II error, accepting a false hypothesis.

The problem derives from the results of a joint confidence interval which state that the probability is at least 1-α that *all* of the following statements hold simultaneously:

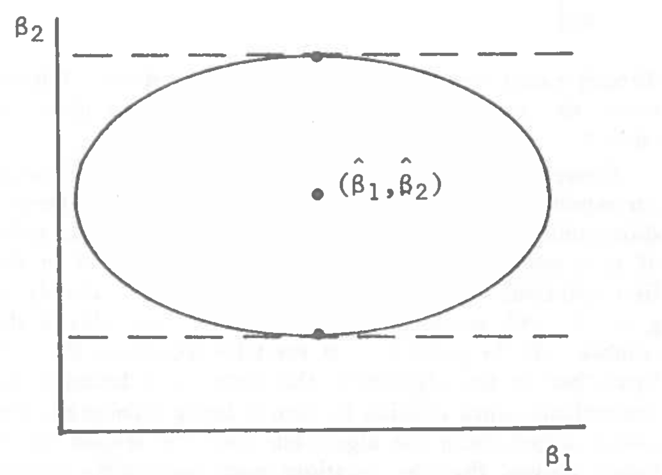
$$a_1 \leq \hat{\beta}_1 \leq b_1$$

$$a_2 \leq \hat{\beta}_2 \leq b_2$$

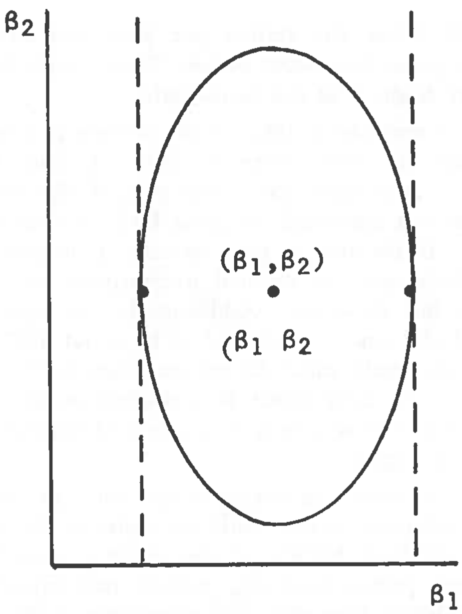
where a and b are constants defining the intervals and α defines the probability of type I error.

Ideally, one would like to find the end points of the major axes of the region. In other words, construct axes that bisect each other within the ellipse and locate the end points of these axes. The points A, B, C, and D are located in fig. 13. This, however, involves the rotation of the original axes and raises problems of definition.

Marquardt has suggested that the maximum symmetric interval within which a parameter β_t lies with probability (1-α), no matter what the true values of the other parameters, is given by the interval between the two planes of support (tangent planes) of the ellipsoid, which are normal to the β_t-axis. Since the slope of the ellipsoid depends on the degree of correlation between the parameters, fig. 14 and 15 are cases where the parameters seem to have no dependence upon one another. In fig. 14, the parameter β_2 is determined in a narrow interval but β_1 is not. Figure 15 shows the opposite case. The joint confidence interval defined by these support plane intervals adequately identifies the probability of a type I error but only at the cost of increasing the chances of a type II error. So the 95% support plane confidence intervals denote a probability of .05 of rejecting a parameter set that lies within the actual ellipse; but the probability of accepting a false hypothesis, such as point E in fig. 13, is increased. In figs. 14 and 15, the rectangular confidence intervals *may* be good approximations but as



14. A support plane confidence interval, β_2, well determined.



15. A support plane confidence interval, β_1, well determined.

the correlation between parameters increases, they become less applicable.

With regard to the CES function, the convenience of portraying the parameter space in two dimensions is lost. But insofar as the model can be represented adequately in the vicinity of the least squares estimates by a linearized Taylor series expansion, the foregoing confidence interval can be applied directly. To ascertain the extent of the deviations from linearity, one can assume that the nonlinear model is correct and that the deviations of the $\hat{\beta}_t$ from their true values are due to random errors.

$$\frac{(\Phi-\hat{\Phi})/n-t-3}{\hat{\Phi}/n-t-3} \leq F_{1-\alpha(t+3,n-t-3)} \ .$$

After selecting a confidence probability (1-α), one can use the above to determine the critical value of Φ, denoted $\tilde{\Phi}$. Then the parameters are varied one at a time by trial and error to determine the upper and lower limits where Φ assumes the critical value $\tilde{\Phi}$. If the deviations from linearity are negligible within the confidence region, then the conventional one-parameter limits will about equal the nonlinear confidence limits based on critical values of the residual sum of squares for alternative parameter sets. As is clear in table 8, the existence of parameter values near zero greatly complicates the computation of the nonlinear confidence limits. The deviations from linearity appear most severe with regard to γ and v. For the remaining parameters for which nonlinear limits were found, the assumption of linearity is supported. The previously discussed support plane confidence intervals are also in table 8.

The substitution parameter g and the factor intensities, β^*_t appear to be contained in relatively small intervals. The efficiency parameter γ^* and the scale parameter v do not appear to be estimated very precisely. This might be expected, given the invariance of g and the β^*_t with changes in $\hat{v}$ and $\hat{\gamma}^*$. The only evidence on the overall fit is that convergence was actually achieved at a parameter set.

CES and C-D estimates and their implications

With the CES form yielding a value for g of –.0083, the elasticity of substitution becomes 1.008. Since the CES reduces to the C-D form when g = 0, the policy implications under each model should be very similar if the CES is properly estimated. The alternative marginal value products for each input are given in table 9.

TABLE 9. Marginal value products of inputs under alternative production function forms.

Input	Cobb-Douglas	CES
Machinery	.4133	.4470
Op and Hired Labor	32.3495	33.2826
Farm Labor	60.6966	61.7596
Cash Expenditures	.9610	.9322
Land	.0068	.0061

TABLE 8. Support plane confidence intervals and deviations from linearity.

Parameter Value	Confidence limits: One-parameter		Nonlinear		Support plane	
$\hat{\gamma}$ = 2.6723	-8.2628	13.6074	1.8137	3.5331	-28.2568	33.6014
$\hat{v}$ = 1.4532	1.1152	1.7912	1.3669	1.4818	.4973	2.4092
$\hat{g}$ = - .0083	- .0080	- .0086	None found		- .0075	- .0090
$\hat{\beta}_1$ = .1519	.1475	.1562	.1497	.1534	.1397	.1640
$\hat{\beta}_2$ = .2706	.2597	.2816	.2676	.2722	.2397	.3016
$\hat{\beta}_3$ = .0151	.0153	.0157	None found		.0135	.0167
$\hat{\beta}_4$ = .5280	.5109	.5451	.5239	.5293	.4796	.5764
$\hat{\beta}_5$ = .0148	.0145	.0152	None found		.0139	.0158

Apparently, the data are dictating a production function of the C-D form. The technique has fitted the CES function at a value of g that corresponds closely to the C-D value. The results support the work of Maddala and Kadane (68) and Zarembka (122). Maddala and Kadane's analytical findings that the C-D estimator of returns to scale is biased downward when the elasticity of substitution is greater than unity is supported by these results, although the confidence interval on v is very wide in the CES function. Zarembka's conclusion that the C-D form was adequate was also supported, but only after both forms were fitted. Both forms must be used to verify such findings.

The marginal value products represent the addition to total revenue brought about by an additional unit of an input. On the basis of the results, we conclude that:

1. An additional $1.00 of machinery investment yields about 40¢ in additional revenue.
2. An additional 10-hr day of operation or hired labor yields approximately $32 of additional revenue.
3. An additional 10-hr day of family labor yields approximately $60 in additional revenue. This variable has come under question in both forms as to its validity in the analysis.
4. An additional $1.00 of cash farm expenditures yields 95¢ in additional revenue.
5. An additional $1.00 of land investment yields about 0.5¢ in additional revenue. This low marginal value product is partly the result of inadequate data.

Since the C-D and CES forms yield almost equivalent results, the policy implications are the same. Generally, the return to the variable factors, labor and cash expenditures, is adequate, but the return to the fixed factor, land, appears too low.

Disregarding the return to family labor, which is at this point a questionable input, each of the inputs needs some elaboration. The return to machinery investment implies that generally machinery will pay for itself in 2½ years. However, no account is taken of depreciation, which will lower the return per year and extend the break-even point to a later period. Thus, while the figure is probably high, it is not unrealistic.

The fact that about 28% of the sample is rented land may explain the low return to land. If the land was rented on a crop share basis, this part of the gross farm income was not measured, so gross farm income is underestimated. If the rent is paid in cash, it is part of cash farm expenses and an implied overestimate of the land investment has occurred. Additionally, the appreciation in value of the land is estimated to be about 5.8% (119). These factors could raise the return from 0.5% to about 8%. Also, the land input is measured at its value in 1966 and thus the return is to a stock of capital and not to a flow of capital.

A profit-maximizing entrepreneur will use units of a variable productive service until the value of the marginal product (VMP or MVP) of the input is exactly equal to the input prices, assuming perfect competition. The results in table 9 show that cash expenditures, the variable capital measure, are adding about 93¢ to total revenue for every $1.00 they add to total cost. Since this input is an aggregate of such things as seed, fertilizer, etc., one cannot attribute a return to each item separately, but the overall return appears to approximate a profit-maximizing rate of use.

The return to operator and hired labor is composed not only of a wage to hired labor but a return to the operator. Although one can't measure the productivity of each type of labor separately, it is clear that a great deal of the short-run returns over and above the costs of production are due to the productivity of this skilled labor input.

Both production function forms suggest that the farming in this region is subject to increasing returns to scale with inputs highly substitutable for one another. The returns to factors are relatively consistent with what one would expect within the limitations of the data.

REFERENCES AND LITERATURE CITED

1 Arrow, Kenneth, Hollis B. Chenery, Bagicha Minhas & Robert M. Solow. 1961. Capital-labor substitution and economic efficiency. Rev. Econ & Stat. XIII(3):225-250.

2 Allen, R. G. D. 1960. Mathematical analysis for economists. Macmillan & Co., Ltd., London.

3 Black, Roy S. 1970. Alternative production function forms and their properties. Unpubl. Ph.D. dissertation, Univ. Minn.

4 Bodkin, Ronald G. & Lawrence D. Klein. 1967. Nonlinear estimation of aggregate production functions. Rev. Econ. & Stat. XLIX(1):28-44.

5 Boddy, Raford. 1967. Comment on Marc Nerlove's paper. In Studies in income and wealth. Vol. 31: The theory and empirical analysis of production. Murray Brown [ed.] Columbia Univ. Press, Columbia.

6 Box, G. E. P. & D. R. Cox. 1964. An analysis of transformations. J. Royal Stat. Soc., Series B XXVI(2):211-43.

7 Brown, Byron. 1970. Tests for Cobb-Douglas and CES production functions. Intl. Econ. Rev. XI(1):77-83.

8 Brown, Murray. 1966. On the theory and measurement of technological change. Cambridge Univ. Press, Cambridge, England.

9 ———— [ed.] 1967. Studies in income and wealth. Vol. 31: The theory and empirical analysis of production. Natl. Bur. Econ. Res., New York.

10 ———— & J. S. de Cani. 1963. A measure of technological employment. Rev. Econ. & Stat. XLV(4):386-394.

11 ———— & ————. 1963. Technological change and the disrtibution of income. Intl. Econ. Rev. IV(3):289-309.

12 Carlson, Sune. 1939. A study on the pure theory of production. P. S. King & Sons, Ltd., London.

13 Chenery, H. B. 1949. Engineering production functions. Quart. J. Econ. LXIII(4):507-531.

14 ————. 1953. Process and production functions from engineering data. In W. W. Leontief et al. [Eds.] Studies in the structure of the American economy. Oxford Univ. Press, New York.

15 Cobb, Charles W. & Paul H. Douglas. 1928. A theory of production. Amer. Econ. Rev. XVIII(1) supplement p. 138-166.

16 Curry, H. B. 1944. The method of steepest descent in non-linear minimization problems. Quart. J. Applied Math. II(3):258-261.

17 Davies, Owen L. 1954. The design and analysis of industrial experiments. Hafner Publ. Co., New York.

18 Dhrymes, Phoebus J. 1967. Adjustment dynamics and the estimation of the CES class of production functions. Intl. Econ. Rev. VIII(2):209-216.

19 ——————. 1962. On devising unbiased estimators for the parameters of the Cobb-Douglas production function. Econometrica XXX(2):297-304.

20 —————— & Paul Zarembka. 1970. Elasticities of substitution for two-digit manufacturing industries: a correction. Rev. of Econ. & Stat. LII(1):115-117.

21 Diwan, R. K. 1966. Alternative specifications of economies of scale. Economica XXXIII(132):442-453.

22 ——————. 1963. An empirical estimate of the CES production function. Paper given at the European meeting of the Econometric Society, Copenhagen, Denmark.

23 Draper, N. R. & H. Smith. 1966. Applied regression analysis. John Wiley & Sons, Inc. New York.

24 Durand, David. 1937. Some thoughts on marginal productivity with special reference to Professor Douglas' analysis. J. Political Econ. XLV(6):740-758.

25 Edwards, Clark. 1962. Nonlinear programming and nonlinear regression procedures. J. Farm Econ. LXIV(1):100-114.

26 Eisenpress, H. & J. L. Greenstadt. 1964. Nonlinear full-information estimation. Paper given at meeting of Econometric Society, Chicago, Ill. Dec. 28-30.

27 —————— & ——————. 1966. The estimation of nonlinear econometric systems. Econometrica XXXIV (4):851-861.

28 Ferguson, C. E. 1963. Cross-section production functions and the elasticity of substitution in American manufacturing industry. Rev. of Econ. & Stat. LXV(3):305-313.

29 ——————. 1968. Neoclassical theory of technical progress and relative factor shares. Southern Econ. J. XXXIV(4):490-504.

30 ——————. 1969. The neoclassical theory of production and distribution. Cambridge Univ. Press, London.

31 Friedman, Milton. 1962. Price theory, a provisional text. Aldine Publ. Co., Chicago.

32 Frisch, Ragnar. 1965. Theory of production. Rand McNally & Co., Chicago.

33 Fuchs, Victor, R. 1963. Capital-labor substitution: a note. Rev. of Econ. & Stat. XLV(4):436-438.

34 Furubotn, E. G. 1965. Engineering data and the production function. Amer. Econ. Rev. LV(3):512-516.

35 Goldberger, Arthur S. 1964. Econometric theory. John Wiley & Sons, Inc., New York.

36 ——————. 1968. The interpretation and estimation of Cobb-Douglas functions. Econometrica XXXVI(3-4):464-472.

37 Goldfeld, Stephen M. & Richard E. Quandt. 1970. The estimation of Cobb-Douglas type functions with multiplicative and additive errors. Intl. Econ. Rev. XI(2):251-257.

38 Green, H. A. J. 1964. Aggregation in economic analysis. Princeton Univ. Press, Princeton.

39 Griliches, Zvi. 1963. Estimates of the aggregate agricultural production function from cross-sectional data. J. Farm Econ. XLV(2).

40 ——————. 1966. Production functions in manufacturing: some preliminary results. Econometrica XXXIV(5):65-68 (supplementary issue).

41 ——————. 1967. Production functions in manufacturing: some prelimary results. Studies in income and wealth. Vol. 31: The theory and empirical analysis of production. Murray Brown [Ed.] Columbia Univ. Press, New York.

42 ——————. 1964. Research expenditures, education and the aggregate agricultural production function. Amer. Econ. Rev. LIV(6):961-974.

43 ——————. 1962. Agricultural production functions: a review. Amer. Econ. Rev. LII(1):282-285.

44 Guttman, L & Duane A. Meeter. 1964. Use of transformations on parameters in nonlinear theory; I. Transformations to accelerate convergence in nonlinear least squares. Dept. of Statistics Tech. Rept. 37, Univ. Wisconsin, Madison.

45 Hahn, F. H. & R. C. O. Matthews. 1964. The theory of economic growth. Econ. J. LXXIV(296):825-850.

46 Hamburg, D. 1959. Production functions, innovations, and economic growth. J. Political Econ. LXVIII(3):238-245.

47 Hartley, H. O. 1961. The modified Gauss-Newton method for fitting of nonlinear regression functions by least squares. Technometrics III(2):269-280.

48 Heady, Earl O. & John L. Dillon. 1961. Agricultural productions functions. Iowa State Univ. Press, Ames.

49 —————— & Luther G. Tweeten. 1963. Resource demand and structure of the agricultural industry. Iowa State Univ. Press, Ames.

50 Hedley, Douglas D. 1969. Estimation of an agricultural production function. Mimeo. Michigan State Univ. East Lansing. P. 6-7.

51 Henderson, James M. & Richard E. Quandt. 1958. Microeconomic theory, a mathematical approach.—McGraw-Hill Book Co., New York.

52 Hicks, J. R. 1963. The theory of wages. 2nd Amer. ed. St. Martin's Press, New York.

53 ——————. 1946. Value and capital. 2nd ed. Clarendon Press, Oxford.

54 Hildebrand, George H. & Ta-Chung Liu. 1965. Manufacturing production functions in the United States. W. W. Humphrey Press, Inc., Geneva, N.Y.

55 Hildebrand, John R. 1960. Some difficulties with empirical results from whole-farm Cobb-Douglas-type production functions. J. Farm Econ. XLII(4):897-904.

56 Hoch, Irving. 1962. Estimation of production function parameters combining time-series and cross-section data. Econometrica XXX(1):34-53.

57 ——————. 1958. Simultaneous equation bias in the context of the Cobb-Douglas production function. Econometrica XXVI(4):566-578.

58 Hood, William C. & Tjalling C. Koopmans [Eds.]. 1953. Studies in econometric method. John Wiley & Sons, New York.

59 Jenrich, R. I. & P. F. Sampson. 1968. Application of stepwise regression to nonlinear estimation. Technometrics X(1):63-72.

60 Johnston, J. 1963. Econometric methods. McGraw Hill Book Co., Inc., New York.

61 Kamien, M. 1963. A comment on alternative derivations of the two input production function with constant elasticity of substitution. Mimeo, Carnegie Inst. of Tech.

62 Kendrick, J. W. & Ryuzo Sato. 1963. Factor prices, productivity and growth. Amer. Econ. Rev. LIII(2):974-1003.

63 Klein, L. R. 1946. Macroeconomics and the theory of rational behavior. Econometrica XIV(2):93-108.

64 Kmenta, J. 1967. On estimation of the CES production function. Intl. Econ. Rev. VIII(2):180-189.

65 ——————. 1964. Some properties of alternative estimates of the Cobb-Douglas production function. Econometrica XXXII(1-2):183-188.

66 Kravis, I. 1959. Relative income shares in fact and theory. Amer. Econ. Rev. LXIX(5):917-949.

67 Kurz, M. & A. S. Manne. 1963. Engineering estimates of capital-labor substitution in metal machinery. Amer. Econ. Rev. LIII(4):662-681.

68 Maddala, G. S. & J. B. Kadane. 1967. Estimation of returns to scale and the elasticity of substitution. Econometrica XXXV(3-4):419-423.

69 Marquardt, Donald W. 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Industrial Applied Math. XI(2):431-444.

70 Marshall, Alfred. 1920. Principles of economics. 8th ed. St. Martin's Press, Inc., New York.

71 McFadden, D. 1963. Constant elasticity of substitution production functions. Rev. of Econ. Studies XXX(83):73-83.

72 McKinnon, R. I. 1962. Wages, capital costs and employment in manufacturing: a model applied to 1947-58 U. S. data. Econometrica XXX(3):510-514.

73 Meade, James E. 1962. A neoclassical theory of economic growth. Allen & Unwin, New York.

74 Mendershausen, Horst. 1938. On the significance of Professor Douglas' production function. Econometrica VI(2): 143-153.
75 Miller, Marlen F. 1967. A measurement of factor productivities in Minnesota agriculture. Unpubl. Ph.D. dissertation, Univ. Minnesota.
76 Minasian, Jora D. 1961. Elasticities of substitution and constant-output demand curves for labor. J. Political Econ. LXIX(3):261-270.
77 Minhas, Bagicha S. 1962. The homohypallagic production function, factor-intensity reversals and the Heckscher-Ohlin theorem. J. Political Econ. LXX(2):138-156.
78 Mosak, Jacob L. 1938. Interrelations of production, price and derived demand. J. Political Econ. XLVI(6):761-787.
79 Mukerji, V. 1963. A generalized SMAC function with constant ratios of elasticities of substitution. Rev. Econ. Studies XXX(84):233-236.
80 Mundlak, Y. 1968. Elasticities of substitution and the theory of derived demand. Rev. Econ. Studies XXXV(102): 225-239.
81 ————. 1962. On estimation of production and behavioral functions. In Carl F. Christ [Ed.]. Measurement in economics. Stanford Univ. Press, Stanford.
82 ———— & Irving Hoch. 1965. Consequences of alternative specifications in estimation of Cobb-Douglas production functions. Econometrica XXXIII(4):814-828.
83 ———— & A. Rasin. 1966. Multistage multiproduct elasticity of substitution production function. Unpubl. manuscript, Univ. Wisconsin.
84 Nelson, Richard R. 1967. Aggregate production functions and economic growth policy. Studies in Income and Wealth. Vol. 31: The theory and Empirical Analysis of Production. Natl. Bur. Econ. Res., New York.
85 ————. 1964. Aggregate production functions and medium-range growth projections. Amer. Econ. Rev. LIV(5): 574-606.
86 ————. 1967. Recent empirical studies of the CES and related production functions. Studies in Income and Wealth. Vol. 31: The Theory and Analysis of Production. Murray Brown [Ed.]. Columbia Univ. Press, Columbia.
87 Nerlove, Marc. 1965. Estimation and identification of Cobb-Douglas production functions. Rand McNally, Chicago.
88 Philpot, Gordon. 1970. Labor quality, returns to scale and the elasticity of factor substitution. Rev. Econ. and Stat. LII(2):194-199.
89 Ramsey, James B. 1968. A comment on the marginal physical product curves for the CES and VES production functions. Amer. Econ. Rev. LVIII(3):482-485.
90 Reder, M. W. 1959. Alternative theories of labor's share. In Moses Abramovitz [Ed.] Allocation of economic resources. Stanford Univ. Press, Stanford.
91 ————. 1943. An alternative interpretation of the Cobb-Douglas function. Econometrica XI(3-4):259-264.
92 Revankar, Nagesh S. 1966. The constant and variable elasticity of substitution production functions: A comparative study in U. S. manufacturing industries. Systems Formulation, Methodology and Policy Workshop paper 6616, Social Systems Res. Inst., Madison, Wis.
93 ————. 1966. The non-transmission approach to complete identification of the Cobb-Douglas equation system. paper 6612—see reference 92.
94 Robinson, Joan. 1955. The production function. Econ. J. LXV(257):67-71.
95 Salter, W. E. G. 1966. Productivity and technical change. Cambridge Univ. Press, Cambridge, England.
96 Samuelson, Paul A. 1970. Economics, 8th ed. McGraw-Hill, Inc., New York.
97 Sato, R. 1967. A two-level constant elasticity of substitution production function. Rev. of Econ. Studies XXXIV(98): 201-218.
98 Schumpeter, Joseph A. 1954. History of economic analysis. Oxford Univ. Press, New York.
99 Scitovsky, T. 1964. A survey of some theories of income distribution. Studies in income and wealth. Vol. 27: The behavior of income shares. Louis Weiner & Charles L. Schultze [Eds.] Princeton Univ. Press, Princeton.
100 Simkin, C. G. F. 1955. Aggregate production functions. Econ. Record XXXI(60):32-39.
101 Smith, C. A. 1955. Survey of the empirical evidence on economies of scale. In Business concentration and price policy. Princeton Univ. Press, Princeton.
102 Smith, Victor E. 1945. Pure theoretical and statistically fitted production functions. Quart. J. Econ. LIX(4):543-562.
103 Solow, Robert M. 1956. A contribution to the theory of economic growth. Quart. J. Econ. LXX(1):65-94.
104 ————. 1964. Capital, labor and income in manufacturing. Studies in income and wealth. Vol. 27: The behavior of income shares. Princeton Univ. Press, Princeton.
105 ————. 1967. Some recent developments in theory of production. Studies in income and wealth. Vol. 31: The theory and empirical analysis of production. Murray Brown [Ed.] Columbia Univ. Press, Columbia.
106 ————. 1957. Technical change and the aggregate production function. Rev. Econ. and Stat. XXXIX(3):312-320.
107 ————. 1958. The constancy of relative shares. Amer. Econ. Rev. XLVIII(4):618-631.
108 Spang, H. A., III. 1962. A review of minimization techniques for nonlinear functions. J. Soc. Industrial Applied Math. IV(4):343-365.
109 Szakolczai, Gyorgy & Janos Stahl. 1967. Generalized CES production function from Hungarian time-series data for middle-and-long-range planning purposes. Econometrica XXXV (5):21-23.
110 ———— & ————. 1969. Increasing or decreasing returns to scale in the constant elasticity of substitution production function. Rev. Econ. & Stat. LI(1):84-90.
111 Theil, H. 1958. Economic forecasts and policy. North-Holland Publ. Co., Amsterdam.
112 Thomas, George B. Jr. 1966. Calculus and analytical geometry. 3rd ed. Addison-Wesley Publ. Co., Reading, Mass.
113 Thompson, Earl A. 1970. Nonpecuniary rewards and the aggregate production function. Rev. Econ. & Stat. LII(4): 395-404.
114 Tsurumi, Hiroki. 1970. Nonlinear two-stage least squares estimation of CES production function applied to the Canadian manufacturing industries, 1926-1939, 1946-1967. Rev. Econ. & Stat. LII(2):200-207.
115 Uzawa, H. 1962. Production functions with constant elasticities of substitution. Rev. Econ. Studies XXIX(81):291-299.
116 Walters, A. A. 1962. Economies of scale in the aggregate production function. Discussion paper, series A, no. 29. Univ. Birmingham, Birmingham, Ala.
117 ————. 1963. Production and cost functions: an econometric survey. Econometrica XXXI(1-2):1-66.
118 Whitaker, J. K. 1964. A note on the CES production function. Rev. Econ. Studies XXXI(86):166-167.
119 Whittlesey, Norman K. & Dwaine E. Umberger. 1970. Financial structure of large farms in the Columbia Basin Project. Wash. Agr. Exp. Sta. Bull. 719.
120 Wicksell, Knut. 1934. Lectures on political economy. George Routledge, London.
121 Zarembka, Paul. 1968. Functional form in the demand for money. Amer. Stat. Assn. J. XLIII(322):502-511.
122 ————. 1970. On the empirical relevance of the CES production function. Rev. Econ. & Stat. LII(1):47-53.
123 Zellner, A., S. Kmenta, and S. Dreze. 1966. Specification and estimation of Cobb-Douglas production function models. Econometricia XXXIV(4):784-795.
124 ———— & Nagesh S. Revankar. 1969. Generalized production functions. Rev. Econ. Studies XXXVI(106):241-250.

APPENDIX I

Elasticity of substitution for the two-point production function C-D and CES

If the production function is of the form

$$Y = f(x_1, x_2)$$

then the elasticity of substitution, σ, is by definition

$$\sigma = \frac{d(x_2/x_1) \ / \ (x_2/x_1)}{d(f_1/f_2) \ / \ (f_1/f_2)}$$

where the ratio of the marginal products, (f_1/f_2), has been substituted for the marginal rate of substitution between inputs.

Evaluating the above expression

$$d(x_2/x_1) = -x_2/x_1^2 \, dx_1 + 1/x_1 \, dx_2$$

letting

$$z = f_1/f_2 = -dx_2/dx_1$$

then

$$dx_2 = -zdx_1$$

and

$$d(x_2/x_1) = (-x_2/x_1^2 - z/x_1)dx_1 \, .$$

Additionally

$$d(f_1/f_2) = d(z) = \partial z/\partial x_1 \, dx_1 + \partial z/\partial x_2 \, dx_2$$

$$= \partial z/\partial x_1 \, dx_1 - \partial z/\partial x_2 \, zdx_1$$

where

$$\frac{\partial z}{\partial x_1} = \frac{f_2 f_{11} - f_1 f_{21}}{f_2^2}$$

and

$$\frac{\partial z}{\partial x_2} = \frac{f_2 f_{12} - f_1 f_{22}}{f_2^2}$$

f_i and f_{ii} being first and second partial derivatives of y with respect to x_i.

Substituting these expressions into our original formula

$$\sigma = \frac{dx_1\,(-x_2/x_1^{\,2} - z/x_1)}{dx_1\left[\left(\frac{f_2f_{11}-f_1f_{21}}{f_2^{\,2}}\right) - z\left(\frac{f_2f_{12}-f_1f_{22}}{f_2^{\,2}}\right)\right]} \cdot \frac{f_1/f_2}{x_2/x_1}$$

which reduces to

$$\sigma \doteq \frac{f_1f_2[x_1f_1 + x_2f_2]}{-\,x_1x_2[f_{11}f_2^{\,2} - 2f_1f_2f_{12} + f_{22}f_1^{\,2}]} \;.$$

Utilizing this formula one can examine both the Cobb-Douglas and CES input production functions.

The Cobb-Douglas may be written in unrestricted form as

$$Y = Ax_1^{\,b_1}\,x_2^{\,b_2} \;.$$

The derivatives of the function are

$$f_1 = b_1\,Y/x_1 \qquad f_{12} = b_1b_2\,\frac{Y}{x_1x_2} \qquad f_{11} = \frac{b_1Y(b_1-1)}{x_1^{\,2}}$$

$$f_2 = b_2\,Y/x_2 \qquad f_{21} = b_1b_2\,\frac{Y}{x_1x_2} \qquad f_{22} = \frac{b_2Y(b_2-1)}{x_2^{\,2}} \;.$$

The elasticity of substitution then becomes

$$\sigma = \frac{(b_1\,\frac{Y}{x_1})(b_2\,\frac{Y}{x_2})\left[x_1(b_1\,\frac{Y}{x_1}) + x_2(b_2\,\frac{Y}{x_2})\right]}{-x_1x_2\left[\frac{b_1Y(b_1-1)}{x_1^{\,2}}\,b_2^{\,2}\,\frac{Y^2}{x_2^{\,2}} - 2(b_1\,\frac{Y}{x_1})(b_2\,\frac{Y}{x_2})(b_1b_2\,\frac{Y}{x_1x_2}) + \frac{b_2Y(b_2-1)}{x_2^{\,2}}\,(b_1^{\,2}\,\frac{Y^2}{x_1^{\,2}})\right]}$$

which reduces to

$$\sigma = \frac{b_1b_2\frac{Y^2}{x_1x_2}\left[b_1Y + b_2Y\right]}{-x_1x_2\left[\frac{b_1^{\,2}b_2^{\,2}Y^3}{x_1^{\,2}x_2^{\,2}} - \frac{b_1b_2^{\,2}Y^3}{x_1^{\,2}x_2^{\,2}} - 2\,\frac{b_1^{\,2}b_2^{\,2}Y^3}{x_1^{\,2}x_2^{\,2}} + \frac{b_1^{\,2}b_2^{\,2}Y^3}{x_1^{\,2}x_2^{\,2}} - \frac{b_1^{\,2}b_2Y^3}{x_1^{\,2}x_2^{\,2}}\right]}$$

$$= \frac{b_1{}^2 b_2 Y^3 + b_1 b_2{}^2 Y^3}{b_1{}^2 b_2 Y^3 + b_1 b_2{}^2 Y^3}$$

$$= 1.$$

The unitary elasticity of substitution is, of course, a famous property of the Cobb-Douglas form.

Beginning again with the generalized form of the elasticity of substitution

$$\sigma = \frac{f_1 f_2 \left| x_1 f_1 + x_2 f_2 \right|}{-x_1 x_2 \left[f_{11} f_2{}^2 - 2 f_1 f_2 f_{12} + f_{22} f_1{}^2 \right]} .$$

The CES function is written

$$y = \gamma [b_1 x_1{}^{-g} + (1-b_1) x_2{}^{-g}]^{-v/g}$$

let $T = b_1 x_1{}^{-g} + (1-b_1) x_2{}^{-g}$ then

$$y = \gamma T^{-v/g}$$

$$T^{-v/g} = y/\gamma$$

$$T = y^{-g/v} \gamma^{g/v}$$

$$\frac{\partial y}{\partial x_1} = \gamma(-v/g) T^{-v/g-1} . b_1 (-g) x_1{}^{-g-1}$$

$$= \gamma b_1 v (y^{-g/v} \gamma^{g/v})^{-v/g-1} x_1{}^{-g-1}$$

$$\frac{\partial y}{\partial x_1} = \gamma^{-g/v} b_1 v\, y^{1+g/v} x_1{}^{-g-1} = f_1$$

$$\frac{\partial y}{x_2} = \gamma^{-g/v} (1-b_1) v\, y^{1+g/v}\, x_2{}^{-g-1} = f_2$$

$$\frac{\partial^2 y}{\partial x_1{}^2} = (\gamma^{-g/v} b_1 v) [(g/v + 1) (y^{g/v} \frac{\partial y}{\partial x_1} . x_1{}^{-g-1}) + y^{1+g/v} (-g-1) x_1{}^{-g-2}]$$

$$= \gamma^{-g/v} b_1 v [y^{g/v+1} x_1{}^{-g-1}] y^{-1} (g/v+1) \frac{\partial y}{\partial x_1} + (\gamma^{-g/v} b_1 v\, y^{1+g/v} x_1^{-g-1}) x_1^{-1} (-g-1)$$

$$= \quad \frac{\partial y}{\partial x_1} y^{-1}(g/v + 1) \frac{\partial y}{\partial x_1} \quad + \quad \frac{\partial y}{x_1} x_1^{-1}(-g-1)$$

$$\frac{\partial^2 y}{\partial x_1^2} = y^{-1}(g/v+1)\left(\frac{\partial y}{\partial x_1}\right)^2 - x_1^{-1}(g+1)\frac{\partial y}{\partial x_1} = f_{11}$$

$$\frac{\partial^2 y}{\partial x_2^2} = y^{-1}(g/v+1)\left(\frac{\partial v}{\partial x_2}\right)^2 - x_2^{-1}(g+1)\frac{\partial y}{\partial x_2} = f_{22}$$

$$\frac{\partial^2 y}{\partial x_1 x_2} = y^{-1}(g/v+1)\frac{\partial y}{\partial x_1}\frac{\partial y}{\partial x_2} \quad = f_{12} .$$

Examining the denominator of the elasticity of substitution formula first

$$D = -x_1 x_2 [f_{11} f_2^2 - 2f_1 f_2 f_{12} + f_{22} f_1^2]$$

$$D = -x_1 x_2 [f_2^2 (y^{-1}(g/v+1) f_1^2 - x_1^{-1}(g+1) f_1) - 2f_1 f_2 (y^{-1}(g/v+1) f_1 f_2$$

$$+ f_1^2 (y^{-1}(g/v+1) f_2^2 - x_2^{1}(g+1) f_2)]$$

$$= -x_1 x_2 [f_1^2 f_2^2 (y^{-1}(g/v+1) - x_1^{-1}(g+1) f_1 f_2^2 - 2f_1^2 f_2^2 y^{-1}(g/v+1)$$

$$+ f_1^2 f_2^2 y^{-1}(g/v+1) - x_2^{-1}(g+1) f_2 f_1^2]$$

$$= -x_1 x_2 [-x_1^{-1}(g+1) f_1 f_2^2 - x_2^{-1}(g+1) f_2 f_1^2]$$

$$= x_2 (g+1) f_1 f_2^2 + x_1 (g+1) f_2 f_1^2$$

$$D = (g+1) f_1 f_2 (x_1 f_1 + x_2 f_2)$$

$$\sigma = \frac{f_1 f_2 [x_1 f_1 + x_2 f_2]}{(g+1) f_1 f_2 [x_1 f_1 + x_2 f_2]}$$

$$\sigma = \frac{1}{g+1} .$$

Thus, if one is able to estimate the parameter g, one can determine the elasticity of substitution between inputs.

APPENDIX II

Generality of the CES production function

The range of values of g gives the CES function its generality. The data will determine the value of g from negative one to plus infinity. The value of $g < -1$ is ruled out largely because economic theory requires some rationality with respect to production. If $g < -1$ the isoquants are concave from below, implying a negative marginal product for one input. If $g = -1$ the isoquants are straight lines and if $g > -1$ the isoquants are convex from below and all marginal products are positive.

If $g = 0$ then $\sigma = 1$ which, of course, is the Cobb-Douglas assumption. This reduction is simply accomplished for the two-input case since

$$Y = \gamma[b_1 x_1^{-g} + (1-b_1)x_2^{-g}]^{-v/g}$$

given $g = 0$ reduces to

$$Y = \gamma[b_1 + (1-b_1)]^{\infty} \quad .$$

Since $\gamma[1]^{\infty}$ is not defined, l'Hospital's rule may be applied in this case to the logarithm of both sides.

$$\log \frac{Y}{\gamma} = -\frac{v}{g} \log [b_1 x_1^{-g} + (1-b_1)x_2^{-g}] = \frac{f(g)}{h(g)}$$

where $f(g) = -v \log [b_1 x_1^{-g} + (1-b_1)x_2^{-g}]$

and $h(g) = g$

then

$$\frac{\frac{df}{dy}}{\frac{dh}{dg}} = \frac{-v \dfrac{[-b_1 x_1^{-g} \log x_1 - (1-b_1)x_2^{-g} \log x_2]}{b_1 x_1^{-g} + (1-b_1)x_2^{-g}}}{1}$$

and

$$= \frac{\lim\limits_{g \to o} \frac{df}{dg}}{\lim\limits_{g \to o} \frac{dh}{dg}} = \frac{v \cdot \lim\limits_{g \to o} \left\{ \dfrac{b_1 x_1^{-g} \log x_1 + (1-b_1)x_2^{-g} \log x_2}{b_1 x_1^{-g} + (1-b_1)x_2^{-g}} \right\}}{\lim\limits_{g \to o} 1}$$

Thus, as g goes to zero we have,

$$\log Y - \log \gamma = \frac{v\ [b_1 \log x_1 + (1-b_1) \log x_2]}{b_1 + (1-b_1)}$$

$$= v\ [b_1 \log x_1 + (1-b_1) \log x_2]$$

$$\log Y = \log \gamma + vb_1 \log x_1 + v(1-b_1) \log x_2$$

$$Y = \gamma x_1^{vb_1} x_2^{v(1-b_1)}$$

which is the Cobb-Douglas function.

If $g = -1$ $\sigma = \infty$ then

$$Y = \gamma[b_1 x_1^{-g} + (1-b_1)x_2^{-g}]^{-v/g}$$

reduces to

$$Y = \gamma[b_1 x_1 + (1-b_1)x_2]^{v}$$

assuming constant returns to scale, i.e., $v = 1$, the CES function reduces to a linear production

$$Y = \gamma b_1 x_1 + \gamma(1-b_1)x_2$$

If $g = \infty$ $\sigma = 0$ then

$$Y = \gamma[b_1 x_1^{-g} + (1-b_1)x_2^{-g}]^{-v/g}$$

reduces to

$$Y = \gamma[b_1 x_1^{-\infty} + (1-b_1)x_2^{-\infty}]^{v/\infty}$$

$= 0°$ which is indeterminate.

Rewriting the CES function, and taking logs of both sides,

$$Y = \gamma\left[x_1^{-g}[b_1 + (1-b_1)\ \frac{x_2}{x_1}^{-g}]\right]^{-v/g}$$

$$= \gamma x_1^{v}\left[b_1 + (1-b_1)\left[\frac{x_2}{x_1}\right]^{-g}\right]^{-v/g}$$

$$Y/\gamma x_1^{\,v} = \left[b_1 + (1-b_1)\left[\frac{x_2}{x_1}\right]^{-g} \right]^{-v/g}$$

$$\log Y/\gamma x_1^{\,v} = -\frac{v}{g} \log \left[b_1 + (1-b_1)\left[\frac{x_2}{x_1}\right]^{-g} \right]$$

Again, applying L-Hospital's rule to evaluate $\lim_{g\to\infty}$ on the right hand side,

$$\frac{-v \lim\limits_{g\to\infty}\left[\dfrac{1}{b_1 + (1-b_1)\left|\frac{x_2}{x_1}\right|^{-g}}\right] \cdot \left[- (1-b_1)\left[\frac{x_2}{x_1}\right]^{-g} \log\left[\frac{x_2}{x_1}\right]\right]}{\lim\limits_{g\to\infty} 1} = v\,\frac{1}{b_1}\cdot 0$$

Thus as $g\to\infty$ we have

$$\log \frac{Y}{\gamma x_1^{\,v}} = v\,\frac{1}{b_1} \cdot 0 = 0 = \log 1$$

$$Y = \gamma x_1^{\,v}$$

Assuming constant return to scale $Y = \gamma x_1$. One could alternatively have found $Y = \gamma x_2$. This is the Leontief production function $Y = \gamma(\min x_i)$.

APPENDIX III

The likelihood function of transformation

The deterministic version of the CES function

$$Y_i = [\sum^{k} \Theta_t X_{ti}^{-g}]^{-v/g} \qquad i=1, \ldots, n \qquad \Theta_t = \gamma^{-g/v} b_t$$

is obtained by applying the transformation of the form

$$Y_i^{TR} = \frac{Y_i^{\lambda}-1}{\lambda} \quad (\lambda \neq 0) \qquad \text{where } \lambda = -g/v$$

and adding a disturbance term. The version under this transformation becomes

$$Y_i^{TR} = \frac{Y_i^{\lambda}-1}{\lambda} = \frac{\left[\sum_{t=1}^{k} \Theta_t X_{ti}^{-g}\right] - 1}{\lambda} + u_i \,.$$

Let X be a random variable of the continuous type having a probability density function, $f(x)$. Let A be the 1-dimension space where $f(x) > 0$. Next consider the random variable Z = u(x) where Z = u(x) defines a one-to-one transformation which maps the set A onto the set B. Let the inverse of Z=u(x) be denoted by $x = u^{-1}(z)$ and let the derivative dx/dz = v'(z) be continuous and not vanish for all points in the set B. Then the probability density function (p.d.f.) of the random variable z = u(x) is given by

$$f(z) = f(u^{-1}(z))|v'(z)|$$

for all z contained in the set B.

In the stochastic version of the CES consider the random variable Y^{TR} which has a p.d.f. of

$$f(Y_i^{TR}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(Y_i^{TR} - \mu Y_i^{TR})^2}{2\sigma^2}$$

where σ^2 is the variance and the mean of the transformed variable is

$$\mu Y_i^{TR} = \frac{1}{\lambda} \left[\sum_{t=1}^{k} \Theta_t X_{ti}^{-g} - 1 \right] .$$

The random variable Y_i will be defined as one-to-one transformation of Y_i^{TR}. The transformation $Y_i = u\ (Y_i^{TR})$ will map the set A composed of Y^{TR} into the set B composed of Y. Corresponding to Z = u(x) above, the specific transformation in this is,

$$Y = u\ (Y^{TR}) = (\lambda\ Y_i^{TR} + 1)^{Y\lambda} \qquad \lambda \neq 0.$$

The inverse, denoted above as $Y_i^{TR} = u^{-1}(Y_i)$, must thus be defined as

$$Y_i^{TR} = \frac{Y_i^{\lambda} - 1}{\lambda} .$$

The derivative of the inverse becomes

$$v'(Y_i) = \frac{d(Y_i^{TR})}{dY_i} = Y_i^{\lambda-1} .$$

Then the p.d.f. of Y_i becomes

$$f(Y_i) = \frac{1}{\sqrt{\partial \pi \sigma^2}} \exp - \frac{(Y_i^{TR} - \mu Y_i^{TR})^2}{2\sigma^2} \, |Y_i^{\lambda-1}| \, .$$

The likelihood of a random sample $Y_1, \ldots, Y_n$ is a product of the individual probability density functions:

$$\ell(Y_1, \ldots, Y_n) = f(Y_1, \ldots, Y_n) = f(Y_1).f(Y_2), \ldots, f(Y_n) = \prod_{i=1}^{n} f(Y_i).$$

Thus

$$\ell(Y_1, \ldots, Y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(Y_i^{TR} - \mu Y_i^{TR})^2}{2\sigma^2} |Y_i^{\lambda-1}|$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp \sum_{i=1}^{n} \frac{n(Y_i^{TR} - \mu Y_i^{TR})^2}{2\sigma^2} \prod_{i=1}^{n} |Y_i^{\lambda-1}| \, .$$

The logarithm of the likelihood function is

$$L = \ln \ell = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \sum_{i=1}^{n} \frac{(Y_i^{TR} - \mu Y_i^{TR})^2}{2\sigma^2} + (\lambda - 1) \sum_{i=1}^{n} \ln Y_i .$$

The procedure at this point would be to find values of λ and θ (recall μY_i^{TR} is composed of θs) which will maximize L, since the parameter values which maximize ln ℓ will maximize ℓ. The first-order conditions for a maximum require that all the first partial derivatives of the log likelihood function be equal to zero. The first partials are nonlinear in the parameters, and a two-step maximization procedure is used.

STATE AGRICULTURAL EXPERIMENT STATIONS

1875 1975