Research Bulletin 0962 1985

Duality Theory and Applied Production Economics Research: A Pedagogical Treatise

Agriculture Research Center
College of Agriculture and Home Economics
Washington State University

TABLE OF CONTENTS

Page

GLOSSARY OF MATHEMATICAL NOTATION UTILIZED IN THIS BULLETIN

q = q(·); q is a function of all arguments within the parentheses (the general function notation)

$\prod_{i=1}^{n}$; the product of elements i = 1,...n

Max (or Min); $X_1,...,X_n$ — maximize or minimize over the choice variables $X_1,...,X_n$

$\sum_{i=1}^{n}$; the sum of elements i = 1,...,n

$\geq, \leq$; greater (less) than or equal to

$X_{i|c}$; the value of the variable X_i conditional upon the value c

A = {X:·} ; A is the set of all X's defined by the implicit arguments after the colon

ε ; is an element of

$\cup$; union of sets

$\cap$; intersection of sets

$>, <$; strictly greater (less) than

$\subset$; is contained in

$\notin$; not an element of

$\partial c/\partial X_i$; the partial derivative of c with respect to X_i

ln A ; the natural logarithm of A

$\forall$; for all

s.t.; subject to

w/r/t; with respect to

$\rightarrow$; approaches

$\neq$; not equal to

Duality Theory and Applied Production Economics Research: A Pedagogical Treatise

Douglas L. Young, Ron C. Mittelhammer, Ahmad Rostamizadeh and David W. Holland[1]

INTRODUCTION

What is Duality?

While "duality" has become perhaps the most fashionable new development in neoclassical microeconomics during the past decade, few textbook writers or researchers have attempted a concise prose definition of the term. We offer the following as such a definition:

> Duality in neoclassical microeconomics refers to the existence, under appropriate regularity conditions, of "dual functions" which embody the same essential information on preferences or technology as familiar primal functions such as production and utility functions. Dual functions describe the results of optimizing responses to input and output prices and constraints rather than global responses to input and output quantities as in the corresponding primal functions.

To illustrate this definition, consider a familiar pair of primal and dual functions—the single-product firm production and cost functions,

(1) $q = q(X_1, X_2, \ldots, X_n)$

and

(2) $c = c(r_1, r_2, \ldots, r_n, q)$,

where q represents output, c is total cost, and X_i's and r_i's are input quantities and input prices, respectively. The production function in (1), referred to as the "primal", describes output response globally to all possible combinations of input quantities. The cost function in (2), which is a "dual" of the production function in (1), describes the optimal or **minimum** cost of producing any level of output given a set of input prices and the production technology. An example from consumer theory would be the indirect utility function (dual), which shows the **maximum** value of utility associated with given commodity prices and level of money income. The familiar (primal) utility function, on the other hand, describes the level of utility associated with all possible combinations of commodity quantities.

We will see that the dual functions contain information about both optimal behavior and the structure of the underlying technology or preferences, whereas the primal functions describe only the latter.

Historical Development

The first application of duality appears to have been made by Hotelling in a 1932 article. However, the first comprehensive development of duality in production economics, including the explicit derivation of many fundamental theories and lemmas, appeared in Shephard's path-breaking 1953 book, *Cost and Production Functions*. Shephard's theoretical contribution initially received relatively little attention. In the early 1970's, further theoretical work by McFadden, Diewert, Berndt and Christensen, and Lau, among others, opened the way for empirical applications. Some of the earliest empirical applications of dual functions were in the areas of agricultural production (Lau and Yotopoulos; Binswanger) and electric power generation (Christensen and Green; Fuss). Lau and Yotopoulos employed a firm-level dual profit function to examine returns to scale, output supply and factor demand elasticities, and interfirm efficiency comparisons for farms in the Indian Punjab. Binswanger estimated an aggregate dual cost function for the U. S. agricultural sector to examine the nature of technical change, factor substitution possibilities and elasticities of factor demand and output supply. Following these pioneering applications in the 1970's, the recent "boom" in the use of dual approaches by agricultural economists did not occur until the early 1980's. As evidence of this surge of interest, there were at least six articles in 1982 issues of the *American Journal of Agricultural Economics* which used duality theory (Babin, Willis, and Allen; Chambers; Heien; Ray; Lopez; and Ball and Chambers). Accompanying this surge of empirical applications, duality appeared as part of the standard treatments in most intermediate and advanced microeconomics textbooks during the late 1970's (e.g., Varian; Silberberg).

[1]Young and Mittelhammer are Associate Professors, Holland is a Professor, and Rostamizadeh is a former graduate student in the Department of Agricultural Economics at Washington State University, Pullman, WA.

The authors gratefully acknowledge the editorial assistance of Don Blayney and Derek Santos in preparing this publication. We also thank Jennifer Brathvode, Nancy Rowe, Deidre Ryan, and Sue Osterback who valiantly struggled through successive versions of this manuscript from 1981 through 1984. Without their expert word processing skills and patience, this manuscript would never have seen the light of day.

Why Study Duality?

Practicing economists trained during earlier eras may feel an understandable reluctance to learn what appears to be a redundant new overlay on their familiar neoclassical theory. This reluctance is often reinforced by the use of unfamiliar mathematical language in modern treatments of duality theory.

Dual approaches permit estimating the same information of practical value to policy makers and managers—supply and demand elasticities, returns to scale coefficients and technical progress parameters, for example—that applied economists have supplied by traditional primal approaches for years. Many may ask why the familiar traditional approaches are not adequate? For some applications, we maintain they will be. Indeed, they will be superior to dual approaches for certain problems. However, data, econometric, or theoretical considerations will permit more accurate or less difficult and costly analysis of many problems with dual approaches. Later in this bulletin, we discuss in detail some of the advantages—and limitations—of dual approaches and provide guidance to readers on how to identify problems where dual approaches are appropriate.

We also argue that understanding duality reinforces a general understanding of microeconomic theory. Certain fundamental relationships such as the slopes of factor demand and output supply curves, homogeneity of demand and supply curves and symmetry of cross-price effects of demand follow immediately from dual approaches. Duality also facilitates a complete systems approach to examining interrelated demand and supply structures in which all theoretical restrictions across equations are enforced (Theil, 1980).

Why the Need for a Pedagogical Treatise?

In spite of the recent surge of duality applications and a voluminous theoretical literature, many economists trained in earlier eras probably have an incomplete understanding of duality theory. Furthermore, we believe the purely theoretical and mathematical treatments in modern texts leave many current students unaware of the full potential and limitations of duality theory in applied research. Most modern treatments of duality build few bridges between dual and familiar primal approaches. Consequently, the full potential of duality as an integrating influence in a student's understanding of microeconomic theory is not realized. Consequently, a pedagogical treatise on duality which assumes a relatively low level of mathematical sophistication—and which stresses the relationship of dual to traditional primal approaches—can serve the needs of both current students and professionals interested in updating their repertoire of theoretical tools.[2]

We have attempted to write this bulletin at a level accessible to the typical first or second year graduate student who has a working knowledge of differential calculus and linear algebra and some exposure to elementary set theory. For the convenience of readers, a glossary of mathematical notation utilized in the bulletin follows the Table of Contents.

There are risks associated with every newly fashionable theoretical development and duality has not been immune to these. The enthusiasm to rush to print with the "new tool" has led some researchers to sacrifice statistical/theoretical rigor or to force inappropriate problems to fit the methodology. We hope this bulletin will help reduce these risks for potential dualists.

Objectives

The four major objectives for this bulletin are:

1. To review and illustrate in a pedagogically effective manner fundamental duality relationships in the neoclassical theory of the firm.
2. To present, as an example, a dual aggregate cost function analysis of U. S. agricultural sector relationships with special attention to examining empirical problems of the dual approach.
3. To identify and describe major theoretical and empirical advantages and limitations of dual approaches to applied production economics research problems.
4. To offer recommendations to applied researchers on improved problem identification when considering dual approaches to production economics research problems.

NECESSARY VOCABULARY AND CONCEPTS

Terminology and the Nature of Response

The first obstacle encountered by the traditionally trained economist in considering dual approaches to familiar problems is the new verbal and mathematical vocabulary. Those familiar with production, profit and cost functions depicted in input or output space must shift gears mentally to conceptualize dual functions in price space. Unfortunately, the same unmodified term—profit function—is typically used to describe the familiar primal profit function (a function of output or input quantities) as is used to describe

[2]We caution readers that empirical research applications of duality also require substantial econometric background to select and use appropriate estimating procedures for alternative functional forms, to impose and test theoretical restrictions and for testing substantive hypotheses (see Fuss and McFadden, and Deaton and Muellbauer). The emphasis in this bulletin will be primarily on theory rather than econometric procedures. A sound understanding of the theory is an essential prerequisite to appropriate empirical applications.

the dual profit function (a function of output and input prices). Less potential for confusion exists with the term cost function, which is dual to the production function, because this term applies uniquely to cost as a function of input prices and of output. However, those familiar with graphical expositions of total or average cost in output space may temporarily forget that the cost functions are functions of input prices as well. Fortunately, the modifier "indirect" is usually used in referring to the dual indirect production function and indirect utility function to distinguish them from their direct counterparts.

Pope (p. 347) observed that some people will be "more comfortable with their knowledge of technology than with economic response" and will prefer primal functions for this reason. For example, agricultural production economists are accustomed to "eyeballing" the estimated parameters of production functions to ensure that they make sense in terms of the known realities of the production process. However, estimated parameters of dual functions must be evaluated by whether they make sense as rational behavioral responses to price or constraint changes.

Identifying Constraints

Because dual functions often represent constrained optimizing responses, new students of duality must pay very close attention to embedded constraints. For example, factor demand functions emerging from dual functions can be either output-constrained, cost-constrained, or ordinary (unconstrained profit maximizing) responses to factor price variations. Of course, constant-output, constant-cost, or ordinary demand functions can also be obtained by appropriate constrained or unconstrained optimization of primal cost, production, or profit functions defined in input space (Ferguson). The restrictive nature of constant-output or constant-cost factor demands derived from familiar primal functions is generally readily apparent from their self-descriptive names and from the familiar procedures by which they are derived. In contrast, the properties of factor demands derived from dual functions are usually less readily apparent. For example, it may slip by a casual reader that **partial** differentiation of a dual cost function (which includes output as an argument) with respect to an input price in accordance with Shephard's Lemma necessarily yields a **constant-output** input demand function. Unfortunately, while of great importance to practitioners, the underlying behavioral assumptions and constraints associated with factor demands, output supplies and other policy-relevant relationships emerging from dual approaches receive scanty emphasis in most textbook treatments of duality.

Three Important Dual Functions in Production Theory

We will build our overview of dual approaches to the neoclassical theory of the firm around three fundamental dual functions: the profit function, the indirect production function and the cost function.

Profit Function

In the simple case of a single-product firm, the dual profit function can be written as:

(3) $$\pi^* = \pi^*(P, r_1, \ldots, r_n)$$

where P and the r_i's denote output price and input prices respectively. In duality theory notation, asterisks are typically used to denote that the dependent variable is the outcome of an optimization process. In (3), π^* is the maximum level of profit associated with the exogenous competitive prices P, and $r_1, \ldots, r_n$. π^* can be derived by maximizing the primal profit function with respect to choices of input levels, $X_1, \ldots, X_n$, as in (4):

(4) $$\underset{X_i\text{'s}}{\text{Max}}\ \pi = \text{Max}\ [P \cdot q(X_1, \ldots, X_n) - \Sigma r_i X_i]$$

Simultaneous solution of the n first-order conditions for a maximum from (4) yields the n ordinary factor demands:[3]

(5) $$X_i^* = X_i^*(P, r_1, \ldots, r_n),\ i = 1, \ldots, n$$

Substituting (5) into the right hand side bracketed expression in (4) yields (3), the dual profit function. Observe that π^* is a function only of prices and that it shows the maximum profit a rational producer with the given technology can obtain given the specified price vector. To distinguish π^* from π use of a different name like "dual maximum profit function" would be helpful but this practice has not been widely adopted in the literature.

The primal profit function in (4) is referred to as the "direct objective function"; whereas, the dual associated with optimizing (4), π^*, is referred to as the "indirect objective function."

Assuming the production function in (4) is quasiconcave and continuous, we are assured that any π^* derived from (4) meets the necessary regularity conditions for a neoclassical profit function. However, duality would have little empirical appeal if one always needed to first know the associated primal function to obtain a "theoretically valid" dual. The empirical payoff is in being able to estimate the dual function directly

[3]We assume, for the most general case, that the production function is at least quasiconcave, with a region of strict concavity in which unique profit maximum solutions exist. Henceforth, when quasiconcavity is assumed, we also assume the existence of a strictly concave region which will contain the profit-maximizing solutions associated with π^*.

from economic data thus deriving the desired information about technology parameters, supply and demand elasticities and other policy-relevant knowledge directly from the dual function, bypassing the primal entirely.

A crucial question is: What properties must a function of input and output prices possess for it to be interpretable as a theoretically valid dual profit function? These properties, referred to as "regularity conditions" in duality theory, are:

A. continuous with respect to input and output prices;
B. linearly homogeneous (homogeneity of degree one) in input and output prices;
C. nondecreasing in output price and nonincreasing in input prices (monotonicity in output and input prices); and
D. convex in input and output prices.

It can be shown by formal "existence proofs" that any function of output and input prices satisfying these four properties is a theoretically valid representation of profit maximizing responses for some well-behaved neoclassical production technology (McFadden). Well-behaved implies that a unique maximum to the primal profit maximization problem exists.

We will not duplicate formal existence proofs here, but we will attempt to provide readers with an intuitive appreciation of the reasonableness of certain conditions.

The requirement of linear homogeneity of the profit function can be easily confirmed by multiplying all prices in the primal profit function by a constant scale factor, $\lambda>0$;

$$(6) \qquad \lambda P \cdot q(X_1,\ldots,X_n) - \Sigma\lambda r_i X_i = \lambda\pi,$$

and recognizing that the first-order conditions associated with the profit maximization problem imply that the same optimum levels of inputs are used regardless of the value of λ. The latter fact is easily demonstrated by dividing both sides of all first-order conditions of (4) by λ and recognizing that the solution values for the X_i's are unchanged.

Examination of the primal profit function also indicates the reasonableness of the proposition that increasing output price, *ceteris paribus,* cannot decrease profit. Similarly, increasing input prices, *ceteris paribus,* cannot increase profit.

The convexity property of π^* is less obvious, but it can be intuitively motivated as follows. The definition of convexity is that

$$(7) \qquad \lambda\pi^*(P_1,r_{11},\ldots,r_{n1}) + (1-\lambda)\pi^*(P_2,r_{12},\ldots,r_{n2}) \geqslant \pi^*((\lambda P_1 + (1-\lambda)P_2),(\lambda r_{11} + (1-\lambda)r_{12}),\ldots,(\lambda r_{n1} + (1-\lambda)r_{n2}))$$

for $0\leqslant\lambda\leqslant1$. Since $\pi^*(\cdot)$ is linearly homogeneous, and given the previous explanation of the derivation of $\pi^*(\cdot)$ via (4), the left-hand side of the above inequality can be represented as

$$(8) \qquad \underset{\substack{X_{11},\ldots,X_{n1}\\X_{12},\ldots,X_{n2}}}{\text{Max}} \quad \lambda P_1 q(X_{11},\ldots,X_{n1}) - \Sigma\lambda r_{i1}X_{i1} + (1-\lambda)P_2 q(X_{12},\ldots,X_{n2}) - \Sigma(1-\lambda)r_{i2}X_{i2}$$

However, the right-hand side of the inequality in (7) can be found by solving problem (8) subject to the additional constraints that $X_{i1} = X_{i2}$ for $i=1,\ldots,n$. Since the latter maximization problem has additional constraints compared to the former maximization problem, its optimized value must be $\leqslant$the optimized value of the lesser constrained problem. Thus $\pi^*(\cdot)$ must be convex.

Applied researchers generally select functional forms for π^* which impose the homogeneity and continuity requirements. As discussed later, it is possible to test *ex post* whether estimated equations meet the convexity and monotonicity requirements, although this is not always done in practice.

Before leaving the dual profit function, note that the concept can also be extended to yield the multiple-product profit function,

$$(9) \qquad \pi^* = \pi^*(P_1,\ldots,P_m, r_1,\ldots,r_n)$$

where the P_i's and r_i's refer to output prices and input prices, respectively. The regularity conditions A–D above extend directly to (9).

Indirect Production Function

The dual indirect production function, presented in (10) below, shows the maximum output available from a given technology, given input prices, and a cost constraint, c.

$$(10) \qquad q^* = q^*(r_1,\ldots,r_n, c)$$

The associated direct objective function that is maximized with respect to the X_i's to establish the functional relationship between maximum output, input prices, and cost level is:

$$(11) \qquad \underset{X_i\text{'s}}{\text{Max}}\ L = \text{Max}\ q(X_1,\ldots,X_n) + \lambda(c-\Sigma r_i X_i)$$

Simultaneous solution of the (n + 1) first-order conditions from (11) yields the solution value for λ and the n constant-cost input demand functions,

(12) $X^*_{i|c} = X^*_{i|c}(r_1,\ldots,r_n,c),\ i = 1,\ldots,n$

where $X^*_{i|c}$ refers to the demand for input i conditional on the cost level c. Substituting (12) into $q(X_1,\ldots,X_n)$, the traditional production function expression, yields q^*.

To be interpretable as a theoretically valid representation of cost-constrained maximization of a neoclassical production technology, a function of input prices and the cost level must satisfy the following regularity conditions:

A. continuous with respect to input prices and cost,
B. homogeneous of degree zero in input prices and cost,
C. nonincreasing in input prices and nondecreasing in cost (monotonicity in input prices and cost level), and
D. quasiconvex in input prices.

Again, one can confirm the reasonableness of the homogeneity and monotonicity conditions by examining the direct objective function in (11). Homogeneity of degree zero is a direct implication of the fact that multiplying c and the r_i's by a factor $k>0$ does not change the feasible set of X_i's from which to choose the maximum output level, and thus the maximum output level itself is left unchanged. Maximum production is nondecreasing in cost, since increasing c enlarges the set of feasible X_i's, and is nonincreasing in r_i's, since increasing r_i's decreases the set of feasible X_i's.

Quasiconvexity of $q^*(\cdot)$ in input prices is less obvious. By definition, the function $q^*(\cdot)$ will be quasiconvex in input prices if the set $A = \{r: q^*(r,c) \leq a\}$ is a convex set for any given c and a, where r is a vector of input prices (Varian, p. 254). We will show that convex combinations of any arbitrary points r_1 and $r_2 \in A$ define points that also belong to A, and thus A is a convex set.

Let r_1 and $r_2 \in A$, and thus $q^*(r_1,c) \leq a$ and $q^*(r_2,c) \leq a$. Now define a convex combination of r_1 and r_2 as $r^* = \lambda r_1 + (1-\lambda) r_2$ where $\lambda \in (0,1)$.

Examine the cost-constrained sets of admissible input combinations at the three input price levels:

(13) $S_1 = \{X: r_1' X \leq c\}$
(14) $S_2 = \{X: r_2' X \leq c\}$
(15) $S^* = \{X: r^{*\prime} X \leq c\}$.

If $X^* \in S^*$, then it must be the case that $X^* \in S_1 \cup S_2$, for assume the contrary that $X^* \notin S_1 \cup S_2$. Then X^* satisfies $r^{*\prime}X^* = \lambda r_1' X^* + (1-\lambda) r_2' X^* \leq c$ and also $r_1' X^* > c$ and $r_2' X^* > c$. But these latter two conditions can be transformed to

(16) $\lambda r_1' X^* > \lambda c$ and $(1-\lambda) r_2' X^* > (1-\lambda)c$

which when added together, imply

(17) $\lambda r_1' X^* + (1-\lambda) r_2' X^* > c + (1-\lambda)c$

which since $\lambda + (1-\lambda) = 1$, and given the definition of r^*, implies

(18) $r^{*\prime} X^* > c$,

a contradiction. Thus, if $X^* \in S^*$, then $X^* \in S_1 \cup S_2$, which implies $S^* \subset S_1 \cup S_2$.

Finally, since, by definition,

(19) $q^*(r^*,c) = \text{Max } q(X) \text{ for } X \in S^*$
$\leq \text{Max } q(X) \text{ for } X \in S_1 \cup S_2$

because $S^* \subset S_1 \cup S_2$ implies there are potentially more X vectors from which to choose in solving the latter maximization problem, then

(20) $q^*(r^*,c) \leq a$

because both $q^*(r_1,c)$ and $q^*(r_2,c) \leq a$. Thus $r^* \in A$, and A is thus demonstrated to be a convex set and $g^*(\cdot)$ is quasiconvex in input prices.

Cost Function

The procedure for deriving a cost function,

(21) $c^* = c^*(r_1,\ldots,r_n,q)$,

from the constrained minimization of total factor cost subject to an output constraint,

(22) $\underset{X_i\text{'s}}{\text{Min}}\ L = \underset{X_i\text{'s}}{\text{Min}}\ \Sigma r_i X_i + \lambda(q - q(X_1,\ldots,X_n))$

is well known. Solving (22) yields the n constant-output factor demands

(23) $X^*_{i|q} = X^*_{i|q}(r_1,\ldots,r_n,q),\ i = 1,\ldots,n$

where $X^*_{i|q}$ refers to the demand for input i conditional on output level q. Substituting (23) into $\Sigma r_i X_i$ provides an expression for the **minimum** level of cost in terms of input prices and output level, expression (21).

To be interpretable as representing a theoretically valid output-constrained minimization of cost given a well-behaved production technology, a function of input prices and level of output must satisfy the following regularity conditions;

A. continuous with respect to input prices,
B. linearly homogeneous in input prices,
C. nondecreasing in input prices (monotonicity in input prices), and
D. concave in input prices.

The homogeneity and monotonicity requirements are both intuitively plausible. Linear homogeneity can be intuitively motivated by recalling the familiar graphical exposition in input space of the minimum cost of producing q at the tangency of the isoquant and an isocost line. If all input prices double, isocost lines will retain the same slopes. The points of tangency with isoquants will not change. Consequently, the minimum cost of producing a given output will also double because the input quantities utilized remain the same but the input prices have doubled. Also, one would not expect the minimum cost of producing a given output to decrease as any input price increased since this would imply that there existed an input combination that could have produced q while lowering cost, a contradiction of the fact that q was being produced at minimum cost prior to the input price increase.

The concavity property is less intuitively apparent, but Varian provides an ingenious graphical explanation which we repeat here:

> Suppose we graph cost as a function of the price of a single input, with all other prices held constant. If the price of a factor rises, costs will never go down (property C.), but they will go up at a decreasing rate (property D.). Why? Because as this one factor becomes more expensive and other prices stay the same, the cost-minimizing firm will shift away from it to use other inputs.
>
> This is made more clear by considering Figure 1...
>
> Let x^* be a cost-minimizing bundle at prices w^*. Suppose the price of factor 1 changes from w_1^* to w_1. If we just behave passively and continue to use x^*, our costs will be $c = w_1X_1^* + \sum_{i=2}^{n} w_i^*X_i^*$. The minimal cost of production c(w,y) must be less than this "passive" cost function; thus, the graph of c(w,y) must lie below the graph of the passive cost function, with both curves coinciding at w_1^*. It is not hard to see that this implies c(w,y) is concave with respect to w_1 (Varian, pp. 29–30).

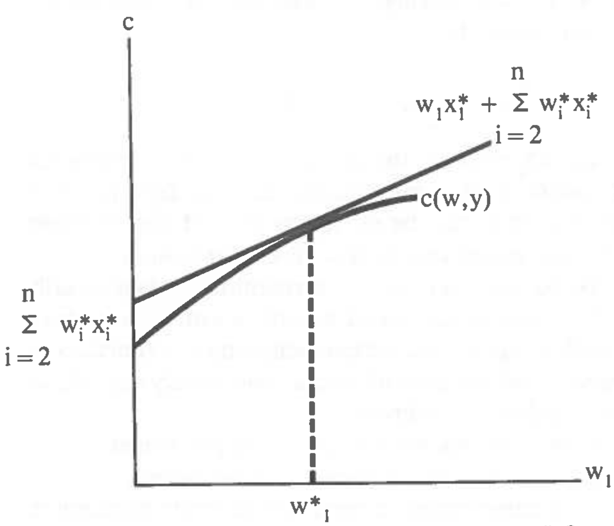

Figure 1. The Cost Function and the "Passive" Cost Function

A Simple Example: Duals for the Cobb-Douglas Production Technology

In this section we present the dual functions introduced above for the familiar Cobb-Douglas (C-D) production technology. The example dual functions for this simple technology will be used to illustrate the meaning of the regularity conditions in each case.

For simplicity, consider the simple two-input C-D production function with decreasing returns to scale:

$$(24)\qquad q = AX_1^a X_2^b$$

where (A, a, b) >0 and (a + b) <1. This continuous and strictly concave production function will generate a unique profit maximum for fixed input and output prices.

C-D Cost Function

We will begin with the cost function, as it is probably the most familiar dual function to most readers. The C-D cost function which emerges from the optimization problem described in (22) is:

$$(25)\qquad C = D\, r_1^{a/E}\, r_2^{b/E}\, q^{1/E}$$

where a and b are the C-D production function elasticities, E = (a + b), and $D = E(Aa^ab^b)^{-1/E}$ is a positive constant. We observe from (25) that the C-D cost function is "self dual", that is, it has the same functional form in prices as the C-D production function has in inputs.

Recall that regularity conditions require that a function of input prices and level of output be continuous, linearly homogeneous, non-decreasing and concave in input prices if it is to be considered a theoretically valid dual cost function. First it is obvious from differentiability of C that (25) is continuous in its input price arguments. Next, we verify linear homogeneity by multiplying all input prices by λ:

$$(26)\qquad \begin{aligned} C(\lambda r_1,\lambda r_2,q) &= D(\lambda r_1)^{a/E}(\lambda r_2)^{b/E}q^{1/E} \\ &= \lambda^{(a+b)/E}D r_1{}^{a/E} r_2{}^{b/E} q^{1/E} \\ &= \lambda C(r_1,r_2,q). \end{aligned}$$

Third, we verify that C is nondecreasing in input prices given the imposed parameter restrictions:

$$(27)\qquad \partial C/\partial r_1 = (a/E)\, D r_1{}^{(a/E)-1} r_2{}^{b/E} q^{1/E} > 0$$

and

$$(28)\qquad \partial C/\partial r_2 = (b/E)\, D r_1{}^{a/E} r_2{}^{(b/E)-1} q^{1/E} > 0.$$

Last, we can show that C is concave in input prices by confirming that its Hessian matrix, H, is negative semidefinite. C has the C-D functional form, $C = Zr_1^s r_2^t$, where Z, s, and t have been substituted for $Dq^{1/E}$, a/E, and b/E, respectively, for notational simplicity. Note that $Z>0$ and $0<s<1$ and $0<t<1$ due to the parameter restrictions applied to the production function. For H to be negative semidefinite the first principal minor, $C_{r_1r_1}$, must be less than or equal to zero and the second, $C_{r_1r_1}C_{r_2r_2} - C_{r_1r_2}C_{r_2r_1}$, must be greater than or equal to zero. Evaluating the necessary derivatives yields:

(29) $$C_{r_1r_1} = s(s-1)Zr_1^{s-2}r_2^t < 0$$

(30) $$C_{r_2r_2} = t(t-1)Zr_1^s r_2^{t-2} < 0$$

(31) $$C_{r_1r_2} = C_{r_2r_1} = s\,t\,Zr_1^{s-1}r_2^{t-1} > 0$$

Based on these derivatives, $C_{r_1r_1}C_{r_2r_2} - C_{r_1r_2}C_{r_2r_1}$ is greater than zero given the aforementioned parameter restrictions and we can conclude that H is negative definite and C is concave in input prices.

For estimated empirical functions, one could numerically check for concavity by evaluating the characteristic roots of H at each observation point. The Hessian will be negative semidefinite and the cost function concave if and only if all the characteristic roots are nonpositive.

C-D Profit Function

Solving the profit maximization problem in (4) for the C-D technology yields the following dual profit function:

(32) $$\pi^* = A^{1/g}P^{1/g}(a/r_1)^{a/g}(b/r_2)^{b/g} - VP^{1/g}r_1^{-a/g}r_2^{-b/g} - WP^{1/g}r_1^{-a/g}r_2^{-b/g}$$

where: A, a, and b are C-D production function parameters as previously defined,
P, r_1, and r_2 are output and input prices,
$g = 1-a-b > 0,$
$V = a^{(1-b)/g}b^{b/g}A^{1/g} > 0,$
$W = a^{a/g}b^{(1-a)/g}A^{1/g} > 0.$

Regarding the regularity conditions, first note that (32) is clearly continuous in r_1, r_2 and P since it is differentiable in these arguments. Next we can confirm that the profit function is linearly homogeneous in output and input prices by multiplying all prices by the positive value λ to obtain;

(33) $$\begin{aligned}\pi^*(\lambda P, \lambda r_1, \lambda r_2) &= A^{1/g}P^{1/g}\lambda^{1/g}(a/r_1)^{a/g}(1/\lambda)^{a/g}(b/r_2)^{b/g}(1/\lambda)^{b/g}\\ &\quad - VP^{1/g}\lambda^{1/g}r_2^{-b/g}\lambda^{-b/g}r_1^{-a/g}\lambda^{-a/g}\\ &\quad - WP^{1/g}\lambda^{1/g}r_2^{-b/g}\lambda^{-b/g}r_1^{-a/g}\lambda^{-a/g}\\ &= \lambda^{(1-a-b)/g}(AP)^{1/g}(a/r_1)^{a/g}(b/r_2)^{b/g}\\ &\quad - \lambda^{(1-a-b)/g}VP^{1/g}r_2^{-b/g}r_1^{-a/g}\\ &\quad - \lambda^{(1-a-b)/g}WP^{1/g}r_2^{-b/g}r_1^{-a/g}\\ &= \lambda\pi^*(P, r_1, r_2)\end{aligned}$$

To verify the monotonicity requirement, examine the derivatives of π^*. First, the derivative with respect to output price is given by

(34) $$\begin{aligned}\partial\pi^*/\partial P &= (1/g)P^{(1/g)-1}[A^{1/g}(a/r_1)^{a/g}(b/r_2)^{b/g} - Vr_1^{-a/g}r_2^{-b/g} - Wr_1^{-a/g}r_2^{-b/g}]\\ &= (1/g)P^{-1}\,\pi^* > 0 \text{ for positive } \pi^*.\end{aligned}$$

As long as π^* is positive, $\partial\pi^*/\partial P$ will be positive. Consequently π^* is nondecreasing with respect to P for all positive π^*.

For the assumed decreasing returns to scale C-D production technology, requiring a positive π^* is not a limiting assumption because a positive profit will always exist given any positive output price. To see this, note that differentiation of (25) with respect to q yields the marginal cost (MC) function $MC = (E^{-1}Dr_1^{a/E}r_2^{b/E})q^{(1-E)/E}$, so that MC emanates from the origin and is monotonically increasing in output space. The area under this MC curve until its intersection with the horizontal output-price line represents total cost at the optimum output level defined by MC = P. This area is necessarily less than the corresponding area under the horizontal output-price line which represents total revenue; hence, profit is positive.

Next we verify that π^* for this C-D technology is nonincreasing with respect to input prices. The derivative of π^* with respect to r_1 is;

(35) $$\begin{aligned}\partial\pi^*/\partial r_1 &= (AP)^{1/g}(b/r_2)^{b/g}a^{a/g}(-a/g)r_1^{-(a/g)-1}\\ &\quad - VP^{1/g}r_2^{-b/g}(-a/g)r_1^{(-a/g)-1}\\ &\quad - WP^{1/g}r_2^{-b/g}(-a/g)r_1^{-(a/g)-1}\\ &= (-a/g)(1/r_1)\pi^* < 0 \text{ for positive } \pi^*\end{aligned}$$

One can similarly show $\partial\pi^*/\partial r_2 < 0$. Thus π^* is monotonically decreasing in input prices for positive π^* and, as described above, the assumption of positive π^* is not restrictive.

Technically, we could verify the convexity in output price and input prices of this C-D profit function by showing that the principal minors of the 3x3 Hessian matrix of (32) were all greater than or equal to zero; that is, the Hessian matrix is positive semidefinite. However, the algebra of this demonstration becomes very tedious. We appeal instead to the general proof of the convexity of the profit function presented earlier. Because the C-D production technology used in this example is continuous, differentiable, quasiconcave and can generate a unique profit maximum, its profit function satisfies the requirements of the earlier proof.

With estimated empirical functions, we could numerically check for convexity by evaluating the characteristic roots of the Hessian matrix of the profit function at each observation point. The Hessian is positive semidefinite and the profit function convex if and only if all the characteristic roots are nonnegative.

C-D Indirect Production Function

Solving the output maximization problem in (11) for the C-D technology yields the following indirect production function:

$$(36) \qquad q^* = s\, c^{a+b} r_1^{-a} r_2^{-b}$$

where $s = A(a/(a+b))^a\,(b/(a+b))^b$ and c, a and b, and r_1 and r_2 are cost level, production function elasticities and input prices, respectively, as before.

As before, differentiability of q^* implies continuity in input prices and cost. To show that this C-D indirect production function is homogeneous of degree zero in input prices and cost, multiply input prices and cost by a positive value λ, to obtain

$$(37) \qquad q^*(\lambda r_1, \lambda r_2, \lambda c) = s\lambda^{a+b} c^{a+b} \lambda^{-a} r_1^{-a} \lambda^{-b} r_2^{-b} = \lambda^0 q^*(r_1, r_2, c) = q^*(r_1, r_2, c)$$

Next we confirm that q^* is nondecreasing in cost:

$$(38) \qquad \partial q^*/\partial c = (a+b)\, s\, c^{a+b-1} r_1^{-a} r_2^{-b} > 0$$

We also verify that q^* is nonincreasing in input prices:

$$(39) \qquad \partial q^*/\partial r_1 = -a\, s\, c^{a+b} r_1^{-a-1} r_2^{-b} < 0$$

A similar result holds for $\partial q^*/\partial r_2$.

To verify the quasiconvexity in input prices of the C-D indirect production function, we again refer the reader to the earlier general proof of this condition which holds for any proper indirect production function.

INPUT DEMAND AND OUTPUT SUPPLY ANALYSIS THROUGH DUAL FUNCTIONS

Quantitative estimates of factor demand and output supply functions and their elasticities are among the most popular forms of economic intelligence applied production economists offer managers and policy makers. The capacity to derive complete systems of factor demand and output supply relationships from directly estimated dual functions, with all theoretical requirements enforced, accounts for a substantial part of duality's appeal to many empirical researchers.

Three Fundamental Lemmas

Derivations of input demand and output supply functions from dual profit, indirect production, and cost functions are described by the following three fundamental lemmas of duality theory:

I. HOTELLING'S LEMMA:

The negative partial derivative of the *profit* function with respect to the i'th input price yields the ordinary demand function for input i;

$$(40) \qquad -\partial \pi^*/\partial r_i = X_i^*$$

The partial derivative of the *profit* function with respect to the output price yields the output supply function;

$$(41) \qquad \partial \pi^*/\partial P = S_q^*$$

II. ROY'S IDENTITY:

The negative ratio of the partial derivative of the *indirect production function* with respect to the i'th input price to the partial derivative of the *indirect production function* with respect to cost yields the constant-cost demand for input i;

$$(42) \qquad -(\partial q^*/\partial r_i)/(\partial q^*/\partial c) = X_{i|c}^*$$

III. SHEPHARD'S LEMMA:

The partial derivative of the *cost* function with respect to the i'th input price yields the constant-output demand function for input i.

$$(43) \qquad \partial c^*/\partial r_i = X_{i|q}^*$$

Technical Development of the Envelope Theorem

Each of these propositions can be derived by applying a result known as the Envelope Theorem, which establishes equality between the partial derivatives of a direct objective function and its associated indirect objective function. Given its pivotal importance in proving the above three lemmas, we state and prove this theorem here. Before stating the theorem, it will be useful to reexamine the "indirect objective function" concept to demonstrate how the concept generalizes beyond the examples of indirect objective functions presented heretofore.

Definition: Indirect Objective Function

Consider the problem of maximizing or minimizing the direct objective function

(44) $Z = f(w_1,\ldots,w_n;\ a_1,\ldots,a_m)$

in the case of an unconstrained optimization problem; or the direct objective function

(45) $L = f(w_1,\ldots,w_n;\ a_1,\ldots,a_m) - \lambda g(w_1,\ldots,w_n;\ a_1,\ldots,a_m)$

in the case where Z is being optimized subject to the constraint

(46) $g(w_1,\ldots,w_n;\ a_1,\ldots,a_m) = 0,$

where $w_1,\ldots,w_n$ are the choice variables and $a_1,\ldots,a_m$ are the parameters of the problem. The first-order conditions for the optimization problem are given by

(47) $\partial f(w_1,\ldots,w_n;\ a_1,\ldots,a_m)/\partial w_i = 0$ for $i = 1,\ldots,n$

if the optimization problem is unconstrained. On the other hand, if the optimization is subject to the constraint (46) then the first-order conditions from the Lagrangian form of the maximization problem are (46) and

(47a) $\partial f(w_1,\ldots,w_n;\ a_1,\ldots,a_m)/\partial w_i - \lambda \partial g(w_1,\ldots,w_n;\ a_1,\ldots,a_m)/\partial w_i = 0$ for $i = 1,\ldots,n.$

Assuming that the appropriate second-order conditions hold, conditions (47), or (46) and (47a) can be solved for the optimum levels of the choice variables as

$$w_i = w_i^*(a_1,\ldots,a_m) \text{ for } i = 1,\ldots,n.$$

Then the indirect objective function associated with (44) is given by

(48) $Z^* = f(w_1^*(a_1,\ldots,a_m),\ldots,w_n^*(a_1,\ldots,a_m);\ a_1,\ldots,a_m) = \Psi(a_1,\ldots,a_m).$

Note that the indirect objective function simply represents the maximum value of Z for any values of the parameters $(a_1,\ldots,a_m)$ of the optimization problem. Recalling the previous discussions of the profit function, indirect production function and cost function, it is recognized that each of these dual functions are "indirect objective functions" in the sense of the above definition. In all three cases, the choice vector$(w_1,\ldots,w_n)$ in the definition is the vector of inputs $(X_1,\ldots,X_n)$ in the development of the three dual functions. Regarding the vector of parameters referred to in (44), the parameters of the profit function are the output and input prices $(P,r_1,\ldots,r_n)$, the parameters of the indirect production function are the input prices and cost level $(r_1,\ldots,r_n,c)$, and the parameters of the cost function are the input prices and output level $(r_1,\ldots,r_n,q)$.

We are now in a position to state and prove the Envelope Theorem.[4]

Envelope Theorem and Proof

The partial derivative of an indirect objective function with respect to a parameter is equal to the partial derivative of the associated direct objective function with respect to the same parameter evaluated at the optimal point $(w_1^*,\ldots,w_n^*)$. Mathematically,

(49) $$\partial \Psi(a_1,\ldots,a_m)/\partial a_j = \partial Z^*(w_1^*(a_1,\ldots,a_m),\ldots,w_n^*(a_1,\ldots,a_m);\ a_1,\ldots,a_m)/\partial a_j = \partial f(w_1^*,\ldots,w_n^*;\ a_1,\ldots,a_m)/\partial a_j = \partial Z/\partial a_j$$

in the case where (44) is optimized without constraint, and

(50) $$\partial \Psi(a_1,\ldots,a_m)/\partial a_j = \partial Z^*(w_1^*(a_1,\ldots,a_m),\ldots,w_n^*(a_1,\ldots,a_m);\ a_1,\ldots,a_m)/\partial a_j = \partial f(w_1^*,\ldots,w_n^*;\ a_1,\ldots,a_m)/\partial a_j - \lambda \partial g(w_1^*,\ldots,w_n^*;\ a_1,\ldots,a_m)/\partial a_j = \partial L/\partial a_j$$

in the case where (44) is optimized subject to the constraint (46).

Proofs

Unconstrained Case: Differentiating the indirect objective function (48) with respect to a_j yields

(51) $$\partial Z^*/\partial a_j = \sum_{i=1}^{n} (\partial f/\partial w_i^*)(\partial w_i^*/\partial a_j) + (\partial f/\partial a_j).$$

However, since $(w_1^*,\ldots,w_n^*)$ represents the optimal levels of the choice variables, and since at the optimum the first-order conditions (47) must necessarily hold, $\partial f/\partial w_i^* = 0$ for all i, and thus

[4]We only deal with the case of one constraint. The definition extends in an obvious way to the case of multiple constraints. See Silberberg, p. 171.

$$(52)\qquad \partial\psi/\partial a_j = \partial Z^*/\partial a_j = \partial f/\partial a_j = \partial Z/\partial a_j.$$

Constrained Case: Differentiating the indirect objective function (48) with respect to a_i yields (51). At the optimal levels of the choice variables $(w_1^*,\ldots,w_n^*)$, the first-order conditions (47a) will necessarily hold, and also

$$(53)\qquad g(w_1^*(a_1,\ldots,a_m),\ldots,w_n^*(a_1,\ldots,a_m)) \equiv 0$$

since the constraint (46) must be satisfied at all optimum points. Differentiating (53) with respect to a_j yields

$$(54)\qquad \sum_{i=1}^{n} (\partial g/\partial w_i^*)(\partial w_i^*/\partial a_j) + (\partial g/\partial a_j) \equiv 0$$

Multiplying (54) by the Lagrangian multiplier , and then subtracting the result from (51) (which subtracts zero and thus does not change the value of the expression) yields

$$(55)\qquad \partial Z^*/\partial a_j = \sum_{i=1}^{n} [(\partial f/\partial w_i^*) - \lambda(\partial g/\partial w_i^*)](\partial w_i^*/\partial a_j) + (\partial f/\partial a_j) - \lambda(\partial g/\partial a_j).$$

Then since the bracketed expression must be zero because of the first-order conditions (47a), we have that

$$(56)\qquad \partial\psi/\partial a_j = \partial Z^*/\partial a_j = (\partial f/\partial a_j) - \lambda(\partial g/\partial a_j) = \partial L/\partial a_j.$$

Q.E.D.

Proofs of the Three Fundamental Lemmas

Using the Envelope Theorem, each of the fundamental lemmas above can be proved straightforwardly. For the dual profit function, the direct objective function (Z) is described by (4). Consequently, using r_i in the role of the parameter (a_j), letting $Z^* = \pi^*$ and using the Envelope Theorem, we obtain

$$(57)\qquad \partial\pi^*/\partial r_i = \partial Z/\partial r_i = \partial[P\cdot q(X_1^*,\ldots,X_n^*) - \Sigma r_i X_i^*]/\partial r_i = -X_i^*$$

Note that the asterisk is attached to the solution of the derivative to emphasize that it is the optimum value of this variable associated with the optimization problem described by (4). Multiplying through by negative one yields Hotelling's Lemma as expressed in (40). We should emphasize that (57) is valid when evaluated at any relevant values of the parameters $(P,r_1,\ldots,r_n)$ and at all associated input levels $(X_1^*,\ldots,X_n^*)$ maximizing profit (4). It is therefore legitimate to interpret (57) as a functional relationship having the parameter values as arguments of the function. To emphasize this fact, Hotelling's Lemma could be written as

$$(58)\qquad -\partial\pi^*(P,r_1,\ldots,r_n)/\partial r_i = X_i^*(P,r_1,\ldots,r_n).$$

To derive result (41) of Hotelling's Lemma, we differentiate the profit function and the direct objective function with respect to the parameter P, and then apply the Envelope theorem to obtain

$$(59)\qquad \partial\pi^*/\partial P = \partial Z/\partial P = \partial[P\cdot q(X_1^*,\ldots,X_n^*) - \Sigma r_i X_i^*]/\partial P = q(X_1^*,\ldots,X_n^*),$$

and by definition $q(X_1^*,\ldots,X_n^*) = S_q^*$ is quantity supplied at the prices $(P,r_1,\ldots,r_n)$. Again note that (59) is valid when evaluated at any relevant values of the parameters $(P,r_1,\ldots,r_n)$ resulting in all associated input levels $(X_1^*,\ldots,X_n^*)$ that maximize profit (4). Therefore, it is legitimate to interpret (59) as a functional relationship having the parameter values as arguments, and thus we may write

$$(60)\qquad \partial\pi^*(P,r_1,\ldots,r_n)/\partial P = q(X_1^*(P,r_1,\ldots,r_n),\ldots,X_n^*(P,r_1,\ldots,r_n)) = S_q^*(P,r_1,\ldots,r_n).$$

To derive Roy's identity, note that the direct objective function (11) evaluated at the input levels $(X_{1|c}^*,\ldots,X_{n|c}^*)$ that maximize production subject to the cost level c is given by

$$(61)\qquad L = q(X_{1|c}^*,\ldots,X_{n|c}^*) - \lambda(c - \Sigma r_i X_{i|c}^*).$$

Next, differentiate the indirect production function (10) and the direct objective function with respect to the parameters r_i and c, and use the result (50) or (56) of the Envelope Theorem to obtain

$$(62)\qquad \partial q^*/\partial r_i = \partial L/\partial r_i = \lambda X_{i|c}^*$$

$$(63)\qquad \partial q^*/\partial c = \partial L/\partial c = -\lambda$$

hence it is clear that

$$(64)\qquad -(\partial q^*/\partial r_i)/(\partial q^*/\partial c) = X_{i|c}^*,$$

where $X_{i|c}^*$ is the optimal level of input i used in maximizing production at input prices $r_1,\ldots,r_n$ and given cost level c. Similar to the arguments presented above, the validity of (64) at all relevant values of the parameters $(r_1,\ldots,r_n,c)$ legitimizes a functional representation of (64) as

(65) $\quad (-\partial q^*(r_1,\ldots,r_n,c)/\partial r_i)/(\partial q^*(r_1,\ldots,r_n,c)/\partial c) = X^*_{i|c}(r_1,\ldots,r_n,c)$

which defines the constant-cost input demand function for input i.

Shephard's Lemma can be derived by first noting that the direct objective function for the problem of minimizing cost subject to the output level q, when evaluated at the optimal input levels $(X^*_{1|q},\ldots,X^*_{n|q})$, is given by (recall (22))

(66) $\quad L = \Sigma r_i X^*_{i|q} + \lambda(q - q(X^*_{1|q},\ldots,X^*_{n|q}))$.

Differentiating the indirect objective function (21), i.e. the cost function, with respect to the parameter r_i, and using the result (50) or (56) of the Envelope Theorem obtains

(67) $\quad \partial c^*/\partial r_i = \partial L/\partial r_i = X^*_{i|q}$

where $X^*_{i|q}$ is the optimal level of input i used in minimizing the cost of producing the output level q at input prices $(r_1,\ldots,r_n)$. The validity of (67) at all relevant values of the parameters $(r_1,\ldots,r_n,q)$ legitimizes a functional representation of (67) as

(68) $\quad \partial c^*(r_1,\ldots r_n,q)/\partial r_i = X^*_{i|q}(r_1,\ldots,r_n,q)$

which defines the constant-output input demand function for input i.

Properties of Input Demand and Output Supply Functions as Derived from Duality Theory

The regularity conditions on the parent dual functions imply the restrictions on factor demand and output supply functions imposed by neoclassical theory. Using dual functions to derive a system of factor demand and output supply relationships automatically enforces these theoretical restrictions via the functional form of the relationships; however, when supply or demand functions are estimated as ad hoc separate single equations, these restrictions sometimes are not enforced. We will say more about this in the empirical section to follow.

Table 1 summarizes the properties imposed by neoclassical theory on output supply and input demand functions. The nonnegative slope of output supply with respect to output price and the nonpositive slope of ordinary input demand functions with respect to their own prices are a result of the convexity of the profit function and Hotelling's Lemma. Hotelling's Lemma permits writing the slope of the supply function as the second derivative of the profit function with respect to output price. By convexity this second derivative and the slope must be nonnegative. Similarly, Hotelling's Lemma implies that the negative of the second derivative of the profit function with respect to the j'th input price equals the slope of the j'th ordinary input demand. Convexity of the profit function, implying nonnegativity of the second derivative, together with the minus sign applied to the second derivative require this slope to be nonpositive. In case the profit function is strictly convex, then the slope of the profit function is positive with respect to output price and the slope of the input demand functions are negative with respect to their own prices.

Table 1. Properties of Input Demand and Output Supply Functions as Derived from Duality Theory.

Function	Slope w/r/t Own Price[1]	Homogeneity of Degree....	Symmetry
Output Supply	Nonnegative	Zero in output price and input prices.	Derivative of supply function w/r/t i'th input price equals the negative derivative of i'th ordinary input demand w/r/t output price
Ordinary Input Demand	Nonpositive	Zero in output price and input prices	Derivative of i'th input demand w/r/t j'th input price equals the derivative of j'th input demand w/r/t i'th input price
Constant-Cost Input Demand	Positive, negative, or zero	Zero in input prices and cost	Derivative of i'th input demand w/r/t j'th input price does *not* necessarily equal the derivative of j'th input demand w/r/t i'th input price
Constant-Output Input Demand	Nonpositive	Zero in input prices	Derivative ot i'th input demand w/r/t j'th input price equals the derivative of j'th input demand w/r/t i'th input price

[1]w/r/t is the abbreviation for "with respect to."

Shephard's Lemma permits expressing the slope of the constant-output demand for X_i as the second derivative of the cost function with respect to r_i. The concavity of the cost function ensures that this derivative, and the slope of the demand function, will be nonpositive.

To demonstrate that the slope of a constant-cost input demand curve with respect to its own price can take any sign, first note the following identity

$$(69) \quad X^*_{j|q}(r_1,\ldots,r_n,q) \equiv X^*_{j|c}(r_1,\ldots,r_n,c^*(r_1,\ldots,r_n,q)), \quad j = 1,\ldots,n.$$

The expression (69) can actually be interpreted as a definition of the constant-output input demand function in terms of the constant-cost input demand function. To see this, first note that $X^*_{j|c}$ $(r_1,\ldots,r_n,c^*(r_1,\ldots,r_n,q))$ can be derived from the constrained output maximization problem

$$(70) \quad \text{Max } q(X_1,\ldots,X_n) + \lambda\,(c^*(r_1,\ldots,r_n,q_o) - \Sigma r_i X_i)$$

where q_o is a given level of output. The first-order conditions for (70) are given by

$$(71) \quad \partial q(X_1,\ldots,X_n)/\partial X_i - \lambda r_i = 0 \qquad i = 1,\ldots,n$$

$$(72) \quad c^*(r_1,\ldots,r_n,q_o) - \Sigma r_i X_i = 0.$$

Now recall that $X^*_{j|q}(r_1,\ldots,r_n,q_o)$ is derivable from the cost minimization problem (22) with q_o replacing q, i.e. from

$$(73) \quad \text{Min } \Sigma r_i X_i + \zeta(q_o - q(X_1,\ldots,X_n))$$

which has first-order conditions

$$(74) \quad r_i - \zeta \partial q/\partial X_i = 0 \qquad i = 1,\ldots,n$$

$$(75) \quad q_o - q(X_1,\ldots,X_n) = 0.$$

(Note we use ζ here to represent the Lagrangian multiplier.) By comparing (71) and (72) with (74) and (75), it is clear that $X^*_{j|q}(r_1,\ldots,r_n,q_o)$, $j = 1,\ldots,n$, which is the solution to the first order conditions (74) and (75), also solves the first order conditions (71) and (72) with $\zeta = \lambda^{-1}$ and with the minimum cost of producing q_o by definition equalling $c^*(r_1,\ldots,r_n,q_o) = \Sigma r_i X^*_{i|q}(r_1,\ldots,r_n,q_o)$. The choice of q_o is arbitrary, and thus (69) holds for all values of $r_1,\ldots,r_n,q$.

Differentiating (69) with respect to r_j obtains

$$(76) \quad \partial X^*_{j|q}(r_1,\ldots,r_n,q)/\partial r_j = \partial X^*_{j|c}(\cdot)/\partial r_j + (\partial X^*_{j|c}(\cdot)/\partial c^*(\cdot))(\partial c^*(\cdot)/\partial r_j)$$

where we use $X^*_{j|c}(\cdot)$ as an abbreviated notation for the right hand side of (69) and use $c^*(\cdot)$ to represent the cost function $c^*(r_1,\ldots,r_n,q)$.

By Shephard's Lemma, $\partial c^*(\cdot)/\partial r_j = X^*_{j|q}$ which we know by (69) is equal to $X^*_{j|c} = X^*_{j|c}$ $(r_1,\ldots,r_n,c^*(r_1,\ldots,r_n,q))$. Rearranging (76) obtains

$$(77) \quad \partial X^*_{j|c}(\cdot)/\partial r_j = (\partial X_{j|q}(r_1,\ldots,r_n,q)/\partial r_j). - (\partial X^*_{j|c}(\cdot)/\partial c^*(\cdot))X^*_{j|c}, \quad j = 1,\ldots,n.$$

(Note that this is analogous to Slutsky conditions that are derived in consumer theory.) Now we know from above that the slope of the constant-output input demand curve with respect to its own price is nonpositive. However, there is no general restriction on the sign of the derivative $\partial X^*_{j|c}(\cdot)/\partial c^*(\cdot)$ appearing in (77). For this reason, $\partial X^*_{j|c}(\cdot)/\partial r_j$ has the potential to assume any sign. (Again note the parallel with consumer theory, where $\partial X^*_{j|c}(\cdot)/\partial c^*(\cdot)$ is the analog to the "income effect", and in which case demand curves are positively sloped only when the commodity in question is a Giffen good.)

The homogeneity properties listed in Table 1 are all attributable to the linear homogeneity (homogeneity of degree one) of the parent dual functions and the mathematical result that first derivatives of homogeneous functions of degree k are homogeneous functions of degree (k-1).

Homogeneity of degree zero for the ordinary input demand and output supply functions means that optimal input use and output level for a profit maximizing firm will not be altered if both input and output prices change by the same proportional amount. This conclusion is consistent with the familiar proposition that optimal input use (and the corresponding output level) depends on the **ratio** of input and output prices. Of course, equiproportionate change of both prices does not alter this ratio. Similar interpretations hold for the other input demand function types.

The symmetry properties listed in Table 1 for output supply, ordinary input demand and constant-output input demand result from an assumption that the parent dual functions possess continuous second order derivatives. Young's Theorem from the calculus specifies that if Z is a function for which continuous second derivatives exist, then $\partial^2 Z/\partial X_i \partial X_j = \partial^2 Z/\partial X_j \partial X_i$. The cross-derivatives referred to in the last column of Table 1 for the output supply, ordinary input demand and constant-output input demand reflect the results of Young's Theorem applied to profit and cost functions possessing continuous second-order partial derivatives.

Note that equality of cross-price **derivatives** does not imply equality of cross-price **elasticities**; that is E_{ij} =

$(\partial X_i^*/\partial r_j)(r_j/X_i^*)$ will generally not equal $E_{ji} = (\partial X_j^*/\partial r_i)(r_i/X_j^*)$ because the second factor involving the fraction in each elasticity term will generally not be equal.

That constant-cost input demand functions do not possess the symmetry property can be verified by taking cross-partial derivatives of the set of identities in (67) to obtain relationships of the form

$$(78) \quad \partial X_{j|c}(\cdot)/\partial r_i = (\partial X_{j|q}(r_1,\ldots,r_n,q)/\partial r_i) - X_{i|c}^*(\partial X_{j|c}^*(\cdot)/\partial c^*(\cdot))$$

and

$$(79) \quad \partial X_{i|c}(\cdot)/\partial r_j = (\partial X_{i|q}(r_1,\ldots,r_n,q)/\partial r_j) - X_{j|c}^*(\partial X_{i|c}(\cdot)/\partial c^*(\cdot)).$$

Subtracting (79) from (78), and using the fact that conditional-output input demands **do** possess the symmetry property, obtains

$$(80) \quad (\partial X_{j|c}(\cdot)/\partial r_i)-(\partial X_{i|c}(\cdot)/\partial r_j) = X_{j|c}^*(\partial X_{i|c}(\cdot)/\partial c^*(\cdot)) - X_{i|c}^*(\partial X_{j|c}(\cdot)/\partial c^*(\cdot)).$$

Since there is no reason that the right hand side of (80) will always equal zero, the derivatives of the constant-cost input demand curves will not generally be equal, and the symmetry property does not obtain. (Again note the analogy to consumer theory, where cross price derivatives of consumer demand functions do not exhibit symmetry.)

Further Examples: Selected Factor Demand Functions for the Cobb-Douglas Technology

To further illustrate the use of the fundamental duality lemmas, we use them in this section to derive constant-output and ordinary input demand functions for the simple two-input Cobb-Douglas (C-D) technology. We retain the same notation and parameter restrictions for the C-D production function as in the earlier examples. We further verify that the required homogeneity and slope properties apply for the derived demand functions. The symmetry properties are a straightforward implication of the continuity of the second-order partial derivatives of the parent dual functions so symmetry is not formally verified here.

Initially, we use Shephard's Lemma to derive the constant-output demand for X_1 from the C-D cost function defined in (25):

$$(81) \quad \partial(Dr_1^{a/E}r_2^{b/E}q^{1/E})/\partial r_1 = X_{1|q}^* = (a/E)Dr_1^{(a/E)-1}r_2^{b/E}q^{1/E}$$

Multiplying r_1 and r_2 by λ and recalling $E = a+b$, we can verify the homogeneity of degree zero in input prices for X_1*:

$$(82) \quad X_{1|q}^*(\lambda r_1,\lambda r_2,q) = (a/E)D\lambda^{(a/E)-1}r_1^{(a/E)-1}\lambda^{b/E}r_2^{b/E}q^{1/E} = \lambda^{(a+b-E)/E}X_{1|q}^*(r_1,r_2,q) = \lambda^0 X_{1|q}^*(r_1,r_2,q)$$

Differentiating $X_{1|q}^*$ with respect to its own price, we can confirm the nonpositive slope of the C-D constant-output input demand:

$$(83) \quad \partial X_{1|q}^*/\partial r_1 = \underbrace{((a/E)-1)}_{-}\underbrace{(a/E)Dr_1^{(a/E)-2}r_2^{b/E}q^{1/E}}_{+} < 0$$

Next we use Hotelling's Lemma to derive the ordinary demand for X_1 from the C-D profit function defined in (32):

$$(84) \quad -\partial\pi^*/\partial r_1 = X_1^* = (a/g)A^{1/g}P^{1/g}a^{a/g}r_1^{-(a/g)-1}(b/r_2)^{b/g} - (a/g)\,VP^{1/g}r_1^{-(a/g)-1}r_2^{-b/g} - (a/g)WP^{1/g}r_1^{-(a/g)-1}r_2^{-b/g}$$

Multiplying P, r_1 and r_2 by λ and recalling $g = 1-a-b$ as previously defined, we can confirm that X_1^* is homogeneous of degree zero in output price and input prices:

$$(85) \quad X_1^*(\lambda P,\lambda r_1,\lambda r_2) = \lambda^{(1/g)-(a/g)-1-b/g}X_1^*(P,r_1,r_2) = \lambda^0 X_1^*(P,r_1,r_2)$$

Differentiating X_1^* with respect to its own price, we can verify the nonpositive slope of this C-D ordinary input demand:

$$(86) \quad \partial X_1^*/\partial r_1 = -\underbrace{((a/g)+1)(a/g)r_1^{-2}}_{+}\pi^* < 0$$

for positive π^*.

As noted for the earlier example, a positive profit will always exist for this simple decreasing returns to scale C-D technology. Consequently, X_1^* is negatively sloped with respect to its own price.

The Multiple-Product Case

Hotelling's Lemma can be extended directly to the multiple-product profit function introduced in (6). Direct estimation of $\pi^* = \pi^*(P_1, \ldots, P_m, r_1, \ldots, r_n)$ and using Hotelling's Lemma to extract output supply and input demand functions permits an elegant approach to analyzing an entire system of output supply and input demand relationships. As an example of a recent agricultural application of this methodology, Weaver simultaneously estimated supply relationships for three outputs (food grains, feed grains and livestock) and demand relationships for five inputs (labor, fertilizer, capital services, materials and petroleum products) in the U.S. midwestern spring wheat region. This approach permits calculating all own- and cross-price elasticities of supply, of demand and between supply and demand. For example, the impact of an increase in the price of livestock on the demand for fertilizer can be computed as $-\partial^2\pi^*/(\partial P_{fertilizer})(\partial P_{livestock})$ in accordance with Hotelling's Lemma. Furthermore, all homogeneity, symmetry and curvature restrictions can be imposed and (if desired) tested throughout the entire system.

DERIVING PRODUCTION FUNCTIONS FROM COST FUNCTIONS

One of the most important implications of duality is that the economically relevant aspects of a production technology can be described interchangeably by either a production function or a cost function. Each of these functions is simply a different way of expressing the same economically relevant information concerning input-output combinations, and it is possible to derive the production function from the cost function or vice versa.

Deriving a cost function from a production function, as described by our discussion of expressions (21)-(23), is generally well understood and treated in most microeconomic theory texts (e.g., Ferguson, pp. 163-166). However, deriving a production function from a cost function is less familiar to many economists. Shephard's Lemma, introduced in the previous section, makes this derivation possible. We will explain the general procedure for the two-input case and follow the explanation with an example using the Cobb-Douglas production technology. The case of more than two inputs is an extension of the procedure described here.

We begin by applying Shephard's Lemma to the cost function (21) to derive the constant-output input demand functions (23), which are repeated here for convenience in the two input case

$$(87)\qquad X^*_{i|q} = X^*_{i|q}(r_1,r_2,q) \qquad i = 1,2.$$

Since we know by our previous discussion of the properties of these constant-output input demand functions that they are homogeneous of degree zero in input prices, we can write (87) equivalently as

$$(88)\qquad X^*_{i|q} = X^*_{i|q}(1,r',q) \qquad i=1,2$$

where $r' = r_2/r_1$ (i.e. we have divided the input prices in (87) by $r_1>0$, and then used homogeneity of degree zero to write (88).

Note that (88) is a set of two equations in the two unknowns r' and q. Assuming for the moment that (88) is solvable for r' and q in terms of $X^*_{1|q}$ and $X^*_{2|q}$ (i.e., assuming that the inverse function system to (88) exists) we obtain

$$(89)\qquad q = f_q(X^*_{1|q}, X^*_{2|q}),$$

and

$$(90)\qquad r' = f_{r'}(X^*_{1|q}, X^*_{2|q}).$$

Note that (89) expresses output as a function of input levels, and is thus a production function of sorts. However, the fact that $f_q(\cdot)$ has as its arguments the (optimal) levels of inputs 1 and 2 that minimize the cost of producing the output level q at prices r_1 and r_2 and not just arbitrary input levels, can in some instances distinguish $f_q(\cdot)$ from the ordinary production function $q(X_1,X_2)$. The difference arises when there are some (X_1,X_2) input combinations in the input space that are not **cost-minimizing** input combinations for producing **some** output level q at some nonnegative input prices (r_1,r_2). In this instance, the non-optimal (X_1,X_2) points are simply not in the range of the vector function (88), and so these points are not in the domain of the inverse function system (89) and (90). However, note that both $f_q(\cdot)$ and $q(\cdot)$ coincide for all $(X^*_{1|q},X^*_{2|q})$, i.e. $f_q(X^*_{1|q},X^*_{2|q}) = q(X^*_{1|q},X^*_{2|q})$, since the input combinations that satisfy the cost minimizing problem (22) necessarily satisfy the production function constraint. In view of the distinction between the domains of $f_q(\cdot)$ and $q(\cdot)$ indicated above, $f_q(\cdot)$ might best be thought of as the production function for **economically relevant** input combinations (where economically irrelevant input combinations are those that are not minimum cost combinations for producing some output level q at some input prices (r_1,r_2)— which are irrelevant input combinations to the input-output decisions of the profit-maximizing firm).

As we alluded in the above discussion, there are instances where the domains of $f_q(\cdot)$ and $q(\cdot)$ coincide, so that a distinction between the two production functions is unnecessary. Such a case is identified when the constant-output input demand functions derived from the cost function have a range that is equal to the nonnegative quadrant of two-dimensional Euclidean space, which is equal to the domain of the production function $q(\cdot)$. In this case the functions $f_q(\cdot)$ and $q(\cdot)$

coincide for all (X_1,X_2) values in the input space, and in addition, all input combinations are economically relevant.

When will (88) be solvable for its inverse function system (89) and (90) so that the production function $f_q(\cdot)$ is recoverable from the cost function $c(r_1,\ldots,r_n,q)$? It is known by the Inverse Function Theorem (see Bartle, p. 381) that the two equations (88) can be solved for r′ and q in terms of points in a neighborhood of $(X^*_{1|q},X^*_{2|q})$ if the Jacobian matrix (the matrix of first-order partial derivatives)

$$(91)\qquad \begin{bmatrix} [\partial X^*_{1|q}(1,r',q)/\partial r'] & [\partial X_{1|q}(1,r',q)/\partial q] \\ [\partial X^*_{2|q}(1,r',q)/\partial r'] & [\partial X_{2|q}(1,r',q)/\partial q] \end{bmatrix}$$

is nonsingular. In general, it can be expected that (91) will be nonsingular. Suppose that both inputs are normal factors, so that the second column of (91) contains two positive entries. Note further that

$$(92)\qquad \partial X^*_{1|q}(r_1,r_2,q)/\partial r_1 = (\partial X^*_{1|q}(1,r',q)/\partial r')(\partial r'/\partial r_1) < 0$$

$$(93)\qquad \partial X^*_{1|q}(r_1,r_2,q)\partial/r_2 = (\partial X^*_{1|q}(1,r',q)/\partial r')(\partial r'/\partial r_2) < 0$$

since the slopes of the conditional factor demand functions with respect to own-input prices are negative. Then since $\partial r'/\partial r_1 = -r_2/r_1^2 < 0$ and $\partial r'/\partial r_2 = 1/r_1 > 0$, we know that the first column of (91) has a positive and a negative entry. The sign pattern of the entries in (91) is thus $[{}^{+}_{-}\,{}^{+}_{+}]$, so the determinant of the matrix is clearly positive implying the matrix is nonsingular. In any case, if inferior factors are admitted, it would be pure coincidence that the determinant of (91) was exactly zero, and thus (89) will generally imply (90) and (91).

We illustrate the procedure described above by the simple two-input Cobb-Douglas production technology used in previous examples. Retaining our previous notation from (25), we can write the Cobb-Douglas cost function as:

$$(94)\qquad c = Dr_1^{a/E}r_2^{b/E}q^{1/E}.$$

Applying Shephard's Lemma, we obtain the following constant-output input demands:

$$(95)\qquad \partial c/\partial r_1 = X^*_{1|q} = (a/E)D\ r_1^{(a/E)-1}r_2^{b/E}q^{1/E}$$

$$(96)\qquad \partial c/\partial r_2 = X^*_{2|q} = (b/E)D\ r_1^{a/E}r_2^{(b/E)-1}q^{1/E}.$$

Recalling that $E = a+b$, we rewrite (95) and (96) as follows:

$$(97)\qquad X^*_1 = (a/E)\ D\ (r_2/r_1)^{b/E}\ q^{1/E}$$

$$(98)\qquad X^*_2 = (b/E)\ D\ (r_2/r_1)^{-a/E}\ q^{1/E}.$$

Note here that the range of (97) and (98) is the positive quadrant of two-dimensional Euclidean space plus the origin (i.e. at the origin, or for points in the positive quadrant, there exist nonnegative levels of q, r_1 and r_2 that, when used in (97) and (98), will generate the coordinates of the point), and the points in the range define the economically relevant input combinations. Solving for $(r_2/r_1)^{-a/E}$ from (97) and substituting into (98) yields:

$$(99)\qquad X_2 = (b/E)\ D\ X_1^{-a/b}\,(a/E)^{a/b}\,D^{a/b}\,q^{1/b}$$

Asterisks have been dropped from the X_i's above to simplify notation. Now, by substituting $E(Aa^ab^b)^{-1/E}$ for D in (99), solving for q, and simplifying, we obtain:

$$(100)\qquad q = AX_1^aX_2^b$$

We could alternatively define the production function $f_q(\cdot)$ to clarify the domain of economically relevant input combinations as

$$(101)\qquad q = AX_1^aX_2^b \text{ for } (X_1,X_2) = (0,0) \text{ or both } X_1 > 0 \text{ and } X_2 > 0.$$

For readers interested in a graphical presentation of duality between production and cost functions, we provide Appendix A which presents an accessible exposition of Shephard's original work. That work illustrated that a set of isocost curves can be derived from the associated production function isoquants. Note that this is a derivation opposite in direction to that presented above.

AN EMPIRICAL RESEARCH ILLUSTRATION: AGGREGATE COST FUNCTION ANALYSIS

In this section we present an empirical application of duality theory involving an investigation by Rostamizadeh et al. of the production technology of U.S. agriculture. The application illustrates a number of facets of econometric research based on duality theory, including consistency of objectives with the dual approach, data availability and variable construction, choice of functional form, econometric estimation procedures and presenting empirical results together with hypothesis testing of regularity conditions required for the validity of the dual approach. In particular, the dual cost function approach was used in an attempt to obtain information on direct and cross-price elasticities of demand for inputs in U.S. agriculture.

We forewarn the reader that the illustration is sobering in the sense that despite the attempt to carefully conceptualize and execute the analysis, the empirical results do not support the contention that the function originally postulated represented a theoretically valid cost function. We present a post-mortem discussion of the empirical difficulties to provide the reader with a perspective on the types of problems that a researcher may encounter in empirical applications of duality theory.

Objectives

The overall objective of this application is to obtain information on the effect that changing factor prices has on factor employment in U.S. agriculture. This includes an assessment of own-price effects and cross- price effects. In addition, we examine changing factor shares of total output. The factors analyzed included land, labor, capital and fertilizer.

Is the dual function approach appropriate for meeting objectives of the type indicated above? The answer is yes. Three types of price effect assessments can be undertaken using a dual approach. The indirect production function approach can be used to assess own-price and cross-price elasticities of demand for production factors conditional on a given total expenditure for the factors. The cost function approach can be used to obtain information on the demand elasticities conditional on a given level of agricultural output. Finally, the profit function approach could be used to assess the ordinary elasticities of factor demand.

In this application, we examined constant-output price effects, and thus utilized a dual cost function approach.

Functional Form of the Cost Function

Finding an algebraic form for a cost or a production function which satisfies, or can be made to satisfy, the assumptions of duality theory and is also consistent with real world behavior is an important and difficult task. In regard to the task, credit is due to economists who have introduced functional forms for the production and cost functions. Functional forms that have been utilized in past empirical work include the Cobb-Douglas, the CES, the translog, the generalized Leontief, the Uzawa and the generalized linear function.

In this application, we chose the translog cost function as the functional form to employ in the econometric analysis. The choice was made for a number of reasons. First, the translog cost function can be viewed as a local, second-order approximation to an arbitrary cost function. Second, the translog function can be made to satisfy a subset of the regularity conditions for the validity of the dual approach by imposing linear constraints on the parameters of the function. This facilitates estimation of the parameters via a restricted least squares technique and allows statistical tests of regularity conditions by testing linear restrictions on model parameters. Third, factor share equations implied by the translog cost function are linear in the parameters of the model, so that linear least squares techniques can be used to estimate parameters.

The translog cost function can be written as

$$\ln C = \ln A + \ln Y + \sum_i a_i \ln W_i + (1/2) \sum_i \sum_j b_{ij} \ln W_i \ln W_j + cT \tag{102}$$

where $i,j = 1,\ldots,n$; $b_{ij} = b_{ji}$ $\forall i \neq j$; $\sum_i a_i = 1$, and

$\sum_j b_{ij} = 0$ for $i=1,\ldots,n$.

The symbol C stands for cost of production, A is a constant parameter, Y is the level of output, W_i is the price of the i'th factor of production, T is time and is used as a proxy for neutral technological progress over time and the a_i's, b_{ij}'s and c are constant parameters of the translog function.

For the function to be used in the dual approach, it must satisfy the regularity conditions presented earlier. First, the function must be continuous in input prices, and this obviously holds for the logarithmic transformation of prices as well as for the product of logarithmic-transformed prices used in the translog specification. Second, the linear restrictions placed on the parameters ensures that the cost function is linearly homogeneous in input prices, since if all prices are multiplied by $\lambda > 0$, then

$$\ln(\lambda C) = \ln A + \ln Y + \sum_i a_i \ln (\lambda W_i) + (1/2)\sum_i\sum_j b_{ij} \ln(\lambda W_i) \ln(\lambda W_j) + cT \tag{103}$$

and thus it is evident that cost has also changed by the factor λ when one recalls the restrictions imposed on the coefficients of (103).

The third requirement is that cost be nondecreasing in input prices. Note that

$$(\partial C/\partial W_i)(W_i/C) = a_i + \sum_j b_{ij} \ln W_j, \tag{104}$$

and by Shephard's Lemma $\partial C/\partial W_i = X_i$, so that

(105) $M_i = X_iW_i/C = a_i + \Sigma b_{ij} \ln W_j$

where M_i is the i'th factor share. Since factor shares must be nonnegative, and since $\partial C/\partial W_i = M_i(C/W_i)$, we have the theoretical restriction that $\partial C/\partial W_i \geqslant 0$. However, this derivation indicates that the required restriction on the translog function parameters is

(106) $M_i = a_i + \sum_j b_{ij} \ln W_j \geq 0 \ \forall\, i,$

which **cannot** be met globally, i.e., for all nonnegative values of $W_1,..,W_n$. For example, if any $b_{ij} \neq 0$, then M_i can be made negative if $b_{ij} > 0$ and $W_j \to 0$, or $b_{ij} < 0$ and $W_j \to \infty$. Thus, the translog function must be viewed as a local approximation to a theoretically valid cost function, and its use constrained to values of $W_1,\ldots,W_n$ for which $M_i \geqq 0 \ \forall$ i.

Finally, we require that the function be concave in input prices, which in turn requires that the Hessian matrix of second-order derivatives of C with respect to input prices be negative semidefinite. The i'th diagonal entry of the Hessian matrix has the form $C(b_{ii} + M_i^2 - M_i)/W_i^2$, while the $(i,j)^{th}$ off-diagonal entry of the matrix takes the form $C(b_{ij} + M_iM_j)/(W_iW_j)$. As in the case of the nonnegativity restriction above, the translog function cannot meet the concavity restriction globally. It thus must be viewed as a local approximation to be used for values of $W_1,\ldots,W_n$ for which nonnegativity of the Hessian matrix is attained.

Regarding elasticities of demand for the various factors of production, we have that

(107) $E_{ij} = (\partial X_i/\partial W_j)(W_j/X_i) = (\partial^2 C/\partial W_i \partial W_j)(W_j/X_i) \ \forall\, i,j$

by Shephard's Lemma, and thus the entries of the Hessian matrix described above are relevant here. By substitution, we have that

(108) $E_{ii} = (b_{ii}/M_i) + M_i - 1$

and

(109) $E_{ij} = (b_{ij}/M_i) + M_j.$

The constant-output factor demand functions to which the elasticities refer are representable as

(110) $X_i^* = M_i(C_i/W_i) = (a_i + \Sigma\, b_{ij} \ln W_j)C_i/W_i.$

Data Requirements and Econometric Estimation

The approach used to estimate the parameters of the translog cost function is the estimation of the set of factor share equations

(111) $M_i = a_i + \sum_j b_{ij} \ln W_j + e_i \ i = 1,\ldots,4$

subject to the linear restrictions

(112) $\sum_i a_i = 1,\ b_{ij} = b_{ji} \ \forall\, i \neq j$, and $\sum_j b_{ij} = 0$ for $i = 1,\ldots,4.$

Since by definition the factor shares must sum to 1, an additional restriction on the estimation of the system of factor share equations is

that $\sum_i M_i = 1.$

Two important questions that must be addressed before proceeding further with the analysis are: 1. what data are used in the estimation of the factor share equations, and 2. what econometric estimation technique should be used to estimate the factor share equations?

Data Requirements

The data required for estimating the factor share equations include time series on factor prices and total expenditure on all factors of production. The time series used for estimation were annual observations spanning the years 1960 through 1979. The sources and construction of the data used to represent prices and expenditures on land, labor, capital and fertilizer are briefly described below. A complete listing of the data can be found in Rostamizadeh.

The annual average wage for hired farm workers was used as the price of agricultural labor. The wage data was taken from USDA's *Farm Labor Monthly Reports*. Total labor expenditures were calculated by multiplying hired labor expenditures by the ratio of total farm labor to hired farm labor. The hired labor expenditure used in calculating the total expenditures was taken from various issues of USDA's *State Farm Income Statistics*.

Fertilizer data were readily accessible. The quantity and expenditures for commercial fertilizer were taken from various issues of *Agricultural Statistics*. The prices of fertilizer were calculated by dividing fertilizer expenditures by the quantity of fertilizer consumed.

Land value data were taken from various issues of USDA's Farm *Real Estate Market Developments*. Building values are included in the total land value. Land expenditures were calculated as 6% of the land value plus real estate taxes assessed on land (Binswanger, p. 285). The real estate taxes are taken from various issues of *Agricultural Statistics.*

A number of problems are associated with compiling capital data. Data for this variable are not directly available and there is not a unique method of measuring capital. However, a proxy variable for capital was proposed by Lianos and by Vathana which was used in this study as well. Capital expenditures consist of the following items: feed, livestock, seed, repairs and operation of capital items, depreciation and other consumption of farm capital. The time series data for capital has been taken from various issues of the *State Farm Income Statistics.* The Production Credit Association's Average Cost of Loans is used as a proxy for the price of capital (Lianos).

Total expenditure is the sum of expenditures for capital, land, labor and fertilizer. The factor shares were obtained by dividing the expenditures of each factor by the total expenditures and were used as the dependent variables in estimating share equations.

Econometric Estimation Technique

The econometric problem is one of estimating a system of four equations that are linear in parameters and, in addition, have linear restrictions on the parameters. The problem would be amenable to a straightforward application of joint restricted generalized least squares (JRGLS) (Theil, 1971, p. 312-317) were it not for the restriction $\Sigma M_i = 1$. The factor share restriction, together with the other restrictions imposed on the parameters of the system, imply that $\Sigma e_i = 0$, and thus the error terms of the system are linearly dependent. The contemporaneous covariance matrix of the disturbance terms is then singular, preventing the calculation of JRGLS estimates. The problem can be resolved by eliminating one of the factor share equations from the system. In this study, the land equation was eliminated, and the three share equations for labor, capital and fertilizer were estimated by the JRGLS method. The standard SAS program was used in calculating the estimates.

Although the land factor share equation was not estimated directly via JRGLS, estimates of the parameters of the equation were constructed by using a subset of the restrictions imposed on the parameters of the system of factor share equations. In particular, letting subscript 4 refer to the land equation, we have that

$$(113)\quad a_4 = 1 - \sum_{i=1}^{3} a_i$$

$$(114)\quad b_{4i} = b_{i4} \text{ for } i = 1,2,3$$

$$(115)\quad b_{44} = -\sum_{j=1}^{3} b_{4j}$$

and since all parameter values on the right-hand side of the equalities are estimated via the JRGLS method, the parameters in the land factor share equation can be estimated by forming the indicated linear combinations of the parameter estimates. The variances of the parameter estimates in the land factor share equation are calculated using the standard formula for the variance of a linear combination of random variables.

A final consideration concerns the exogeneity of the explanatory variables. If the explanatory variables were considered endogenous, then the JRGLS technique would require the introduction of instrumental variables for these endogenous variables, effectively resulting in a three stage least squares estimation procedure. Following the approach used by Binswanger and Lopez, it is assumed that the input prices are exogenous variables. When the capital used in the agricultural sector is a small portion of the total capital used in the whole economy, the price of capital will not be determined endogenously by the agricultural sector. Fertilizer is very energy intensive and the exogenous price of energy provides some justification for considering fertilizer price to be exogenous. The assumption of exogeneity of the labor price can be defended in the sense that the wage rate in the agricultural sector usually follows the same pattern as the industrial wage rate. Additionally the minimum wage rate can be looked at as an exogenously (politically) determined base farm wage rate. In regard to land price, where agriculture is the main user of land, its price may be endogenous. However, to avoid introducing additional complexity into the model, the land price is treated as an exogenous variable. Overall, the exogeneity assumptions imply that farmers face horizontal supply curves for inputs.

Empirical Results and Hypothesis Tests

The set of factor share equations were estimated according to the method described in the previous section, and the results are displayed in Table 2. Only one of the estimated parameter values was not significant at the 90% level of confidence. Among other things, the estimates indicate that the factor shares of labor, fertilizer and land increase as their respective factor prices increase, while the capital share decreases with increases in the price of capital, *ceteris paribus.*

The parameter estimates were used to predict factor shares for each of the years 1960 through 1979, as well as own-price elasticities for the same years. The results are displayed in Table 3. In particular, the results indicate that the factor shares of land and capital increased from 1960 through 1979, the factor share of labor declined over the same time period and the factor share of fertilizer fluctuated but settled at a slightly higher level in the late 1970's as opposed to

Table 2. JRGLS Estimation of Factor Share Coefficients.

Dependent Variables	Coefficients of Independent Variables				
M_i	lnP_N	lnP_K	lnP_F	lnP_L	Intercept
M_N	.07905 (3.99)[1]	.04215 (6.83)	.01855 (2.0761)	−.13974 (−9.51)	.79858 (9.01)
M_K		−.01526 (−2.17)	−.02991 (−9.04)	.00303 (.49)[2]	.56856 (18.73)
M_F	Symmetric		.03483 (6.42)	−.02347 (−3.83)	.07618 (1.8637)
M_L				.16018 (9.28)	−.44332 (−36.43)

[1]Asymptotic t ratios in parentheses.

[2]Not significant at 90 percent confidence level.

the early 1960's. In addition, the demand for each of the factors is inelastic, with the highest absolute elasticity attributed to the demand for capital, and the lowest elasticity alternating over the years between fertilizer and land demand.

To analyze the compatibility of the estimated model with the regularity conditions required for the validity of the dual approach, the restrictions placed on the model due to the regularity conditions were examined. First, an F-test of the six linear restrictions

$$(116) \quad \sum_{j=1}^{4} b_{ij} = 0; \; i=1,2,3$$

$$(117) \quad b_{ij} = b_{ji}; \; i < j, \; (i,j) = 1,2,3$$

was performed using the standard F-statistic having, in this case, 6 and 45 degrees of freedom in the numerator and denominator, respectively (Theil, 1971, p. 312-314). The calculated F-value was 4.11, indicating that the linear restrictions are rejected, even at the 0.005 level of significance (critical value = 3.7). Second, the Hessian matrix described earlier is required to be negative semidefinite, since the cost function is required to be concave in input prices for duality theory to apply. The Hessian matrix was constructed for each of the years 1960-1979, and the characteristic roots of the matrices were extracted. In all cases, at least one of the roots was positive, indicating that the matrices were not negative semidefinite and that cost was not concave in the input prices. Finally, cost is required to be nondecreasing in input prices. Since $\partial C/\partial W_i = M_i(C/W_i)$, the sign of the derivatives of cost with respect to input prices was judged by the sign of the predicted M_i values in Table 3. Since all $\hat{M}_i \geq 0$, $i=1,\ldots,4$, cost was judged to be nondecreasing in input prices.

An Evaluation of the Empirical Research Illustration

Overall, regularity conditions for the applicability of duality theory in our problem could not be met. The conclusion is that the particular translog function implied by the parameter estimates is not a theoretically valid cost function, even as an approximation, and thus applying Shephard's Lemma in this problem appears inappropriate. In particular, the theoretical basis for the interpretation of the estimated quantities presented in Tables 2 and 3 as representing factor share equations, predicted factor shares, and elasticities of factor demand is questionable.

While the above example has illustrated some of the steps required in an actual empirical analysis based on duality theory, it is admittedly not very good advertisement for empirical applications of the dual approach. What should we conclude at this point concerning the use of the dual approach to obtain information on elasticities of demand for inputs in U.S. agriculture? **Additional empirical and theoretical effort is warranted!**

Using data that more accurately reflect our input categories or which are less aggregated could possibly improve the results. It is also possible that another cost function specification such as the generalized Leontief model or some other functional form would be more compatible with the data in the sense of satisfying regularity conditions. In addition, no account of nonneutral technological progress has been incor-

Table 3. Predicted Factor Shares and Own Price Elasticities.

	Factor Share				Own-Price Elasticity			
Year	Labor $\hat{M}_N$	Capital $\hat{M}_K$	Fertilizer $\hat{M}_F$	Land $\hat{M}_L$	Labor $\hat{E}_N$	Capital $\hat{E}_K$	Fertilizer $\hat{E}_F$	Land $\hat{E}_L$
1960	.28462	.42713	.04449	.24376	−.43764	−.60860	−.17264	−.09908
1961	.27924	.42901	.04711	.24464	−.43767	−.60656	−.21356	−.10056
1962	.27013	.43433	.04324	.25230	−.43723	−.60080	−.15126	−.11278
1963	.26491	.43535	.04299	.25675	−.43669	−.59970	−.14682	−.11934
1964	.25961	.43380	.04306	.26353	−.43589	−.60138	−.14807	−.12861
1965	.25522	.43785	.03985	.26708	−.43505	−.59700	−.08612	−.13314
1966	.25713	.44144	.04034	.26109	−.43544	−.59313	−.09625	−.12537
1967	.25175	.43951	.03853	.27021	−.43425	−.59521	−.05750	−.13695
1968	.24973	.44321	.03831	.26875	−.43373	−.59122	−.05253	−.13519
1969	.25254	.44900	.03499	.26347	−.43444	−.58499	−.03042	−.12853
1970	.25923	.45053	.03123	.25901	−.43583	−.58334	−.14650	−.12252
1971	.26023	.45312	.04707	.23958	−.43600	−.58056	−.21297	−.09179
1972	.25240	.45741	.04645	.24374	−.43411	−.57595	−.20371	−.09904
1973	.24817	.45740	.04227	.25216	−.43330	−.57596	−.13374	−.11257
1974	.24143	.44140	.05526	.26191	−.43115	−.59317	−.31445	−.12647
1975	.23374	.44049	.06403	.26174	−.42806	−.59415	−.39201	−.12624
1976	.22139	.45182	.05890	.26789	−.42155	−.58195	−.34976	−.13414
1977	.20488	.45760	.05699	.28053	−.40928	−.57575	−.33185	−.14844
1978	.20526	.45533	.05703	.28238	−.40962	−.57818	−.33224	−.15033
1979	.20146	.45716	.04925	.29213	−.40615	−.57622	−.24354	−.15952

porated in the cost function specification, and a respecified translog function might be more compatible with the dual approach.

The point is that we have not rejected the applicability of duality theory to our problem, but we have rejected our particular implementation of the approach. In this respect, the dual approach to econometric modelling is no different than any other approach. Empirical applications of the theory necessarily force the researcher to commit to specific functional forms, estimation techniques and data types even though the field of candidates for these various facets of the analysis can be large despite the narrowing of the number of possibilities via the implications of the theory. We elaborate on some of the advantages and limitations of the use of the dual approach in applied production economics research in the final section of this report.

ADVANTAGES AND LIMITATIONS OF DUAL APPROACHES: AN APPRAISAL[5]

Having completed both an overview of duality theory in production and also a summary of an aggregate dual cost function application, we will attempt in this section to identify some general advantages and limitations of dual approaches to applied production economics research problems.

Advantages of Dual Approaches

Dual approaches may be attractive to applied researchers for what will be loosely classified as "theoretical" and "practical statistical" advantages. Among the former category, four are particularly important:

1. Applying dual functions properly can ensure that demand, supply and other economic relationships derived from them are consistent with the restrictions imposed by economic theory including slope, homogeneity and symmetry properties. Statistical tests of these restrictions also permit checking the validity of the implementation of the theory for the problem.

[5]This section draws extensively from an earlier discussion article (Young).

2. Dual approaches facilitate analyses of entire systems of input demand and output supply functions in which the impacts of all relevant input and output prices and other exogenous variables are included, and all theoretical restrictions across equations are observed.
3. The flexible functional forms typically used for dual cost or profit functions generally impose fewer restrictions on the nature of technology than do popular and mathematically tractable production function forms like the Cobb-Douglas or CES. The dual approaches "let the data speak" with regard to input substitution possibilities, homotheticity, constancy of input shares and other properties. Although more flexible production function forms such as the translog could be used, often they are avoided because of their algebraic complexity.
4. Dual profit functions facilitate empirical analysis of multiple-product firms and industries.

Potential computational and statistical advantages of dual approaches include:

1. Dual approaches promote computational ease in deriving input demand, output supply and other economic relationships as compared to deriving these from primal objective functions. Simplicity in calculations can reduce the opportunity for human error. Also, the risk of exacerbating estimation error by rounding error is greater with the more complex primal derivations. For example, rounding errors can be significant in inverting large matrices as is required for computing Allen partial elasticities of substitution from the production function.
2. Elasticities and other response measures can often be calculated as simple, frequently linear, functions of the parameters of dual functions. This facilitates deriving the statistical distributions of these expressions for purposes of hypothesis testing.
3. Estimating cost or profit functions with price data as independent variables may permit more precise econometric estimates of technology parameters in cases where multicollinearity among factor prices is less than that among factor quantities.
4. Prices, used as regressors in dual approaches, are more likely to be truly exogenous than are input quantities, thereby avoiding correlation of regressors with the error term. This precludes the need to use simultaneous equation regression techniques which probably should be used, but rarely are, to estimate firm or aggregate production functions.
5. A final practical advantage of dual approaches is that data on factor and output prices, total costs and annual profits will often be more readily available, and possibly more accurate, than data on output and input quantities.

Some Theoretical Limitations

Dual functions contain information on both the nature of technology and **assumed** rational behavioral response in accordance with specified objective functions and constraints. Consequently, valid results require identifying correctly both the objective function and constraints applying to the real world problem to be analyzed and selecting the appropriate dual approach. For example, if producers in a particular industry generally behave as rational profit maximizers to varying competitive market prices, fitting a dual profit function to a time series of annual profits, output prices and input prices might be justified. However, if profit maximizing behavior in the industry is precluded by the absence of effective markets, the existence of unknown constraints, or producer irrationality, any supposed estimated **maximum** profit function will be meaningless. The weakness of competitive markets and the potential existence of myriad cultural, subsistence, credit and other constraints in many developing countries' agricultural sectors, for example, could weaken the applicability of dual models to these settings.

Even in economies where functioning markets and rational behavior exists, it is critical to select the dual function that fits the institutional realities of the problem. Institutional restrictions imposed by farm programs or contracting practices could make cost minimization subject to an output constraint more realistic than unconstrained profit maximization.

It is also important to match the dual function which is to be estimated to the researchers' ultimate objectives. For example, consider the objective of providing information on the elasticity of demand for labor in a particular agricultural sector to evaluate the long run consequences of minimum wage policies or projected labor union supported wage increases. The customary dual approach to analyzing such agricultural input demand elasticities has been by estimating an aggregate dual cost function and applying Shephard's Lemma (e.g., Binswanger; Lopez and Tung; Kako; and Rostamizadeh). Although the constant-output nature of the calculated elasticities is sometimes acknowledged in developing the theory, there generally has been little emphasis in the interpretations/policy implications of these studies on the distinctly short run nature of these elasticities. Over the longer run when producers have time to adjust output level as well as input combinations in response to input price changes, elasticities will often be considerably higher in absolute value. For example, Hammonds, Yadav and Vathana reported that several studies of the U.S. hired farm labor market showed long run labor demand elasticities to be three or four times higher than the comparable short run elasticities.

Recently, Chambers computed both constant-cost and constant-output input demand elasticities from a cost function by transforming the cost function to the indirect (cost constrained) production function. For the U.S. meat products industry, Chambers' results showed capital, labor and energy demand elasticities to be slightly higher, and materials demand elasticity to be several times higher, for the constant-cost compared to the constant-output input demand. Constant-cost input demands represent an intermediate length of run, or adjustment potential, between constant-output and ordinary input demands. Of course, longer run ordinary demands could be estimated from the dual profit function by use of Hotelling's Lemma. The key is to select the dual function which fits the intended policy interpretation of the results.

As a final theoretical observation, note that dual approaches overcome no fundamental objections to neoclassical economics as a paradigm. For example, they do not resolve the venerable debate as to whether production functions (or equivalently cost functions) are valid constructs for *aggregate* economic analysis (Robinson; Harcourt; Pasinetti). This debate revolves around questions of whether a meaningful measure of capital distinct from relative prices is possible in the aggregate, issues of capital switching and reswitching and the nature of technical progress.

Empirical Problems

Econometric estimation of dual cost, profit, or indirect production functions requires reasonably accurate price data that exhibits some dispersion. For some applications, such data may not be available. For example, cross-sectional farm survey data from a single year may show all farms confronting basically the same vector of input and output prices. Time series of prices will generally exhibit more dispersion, but dispersion may still be limited for some industries.

Estimates of elasticities, technical change coefficients and other measures of policy interest from dual function parameters will often be very sensitive to data composition and variable construction procedures. Using an aggregate cost function, Lopez noted how the switch from time series to pooled cross-section and time series data led to a reversal of the conclusion that non-neutral technical progress had been a significant factor in Canadian agriculture. The demand elasticities for land in our cost function analysis reported above were considerably lower than those obtained by Binswanger or Lopez. This result may have been partially due to data inadequacies.

Of course, the sensitivity of results to practical data composition and variable construction problems applies equally to approaches using primal functions. The preceding discussion serves only to warn potential dualists that their "ammunition" may critically affect their success.

Imposing and testing required regularity conditions on empirically estimated dual functions can also present problems for the applied researcher. First, imposing parameter restrictions across equations will, as noted in the cost function example reported above, generally require use of jointly restricted generalized least squares or maximum likelihood estimation procedures. Some researchers may not have ready access to software or the expertise needed to use such procedures.

Secondly, statistical tests of the parameter restrictions may reveal that they are not consistent with the data as was reported in our cost function application above. Table 4 summarizes, among other information, the results of tests of certain necessary theoretical restrictions reported in several recent studies that have utilized a cost function approach to examine agricultural production. Although necessary symmetry and homogeneity parameter restrictions were generally imposed, the compatibility of these restrictions with the data were consistently tested statistically only in the recent work by Lopez and ourselves. Some researchers have also failed to test the local concavity of the cost function. This test involves checking whether the Hessian matrix of the cost function is negative semidefinite at each observation point.

Pope (p. 346) observes that "testing curvature [concavity] conditions is a cumbersome matter and is usually dispensed with." Our recent experience confirms that these tests are tedious, but given that these conditions are intrinsic to the theory and to the validity of the results, it is important that they be made.

As noted in Table 4, our experience with the Rostamizadeh study strongly indicates that concavity, symmetry, and homogeneity conditions certainly will not always hold, at least not with the translog specification applied to recent U.S. agriculture data. In an attempt to improve the specification, we tried alternative assumptions regarding technical change, homotheticity and structural breaks with little success. Lopez's success in accepting the concavity, symmetry and homogeneity tests using the generalized Leontief function on Canadian data provides encouraging support for the flexibility of the generalized Leontief specification.

What is the researcher to conclude if statistical tests show one or more of the regularity conditions is violated for his/her application? At face value, this would seem to imply rejection of the postulates of the underlying theory. For example, it could mean for a cost function that the data do not reflect rational cost minimizing responses to a well-behaved production technology. However, such conclusions are clouded by

Table 4. Selected Properties of Some Aggregate Cost Function Studies of Agricultural Production.

Study	Data	Function	Technical Change	Homotheticity of P.F.	Estimation Technique[1]	All Demands Negative?	Locally Concave	Test of Symmetry	Test of Homogeneity
Binswanger, 1974	U.S. Ag: cross-section + time series	Translog	Nonneutral	Homothetic	RGLS	Yes	No test	Accepted	None
Kako, 1978	Japan rice farms: cross-section + time series, 1953-70	Translog	Nonneutral	Homothetic	FIQML	Yes	Accepted	None	None
Lopez, 1980	Canada Ag: time series 1946-77	General Leontief	Nonneutral (Not Sig.)	Nonhomothetic	FIML	Yes	Accepted	Accepted	NA
Chambers, 1982	U.S. Meat processing: time series, 1954-76	General Leontief	Neutral	Nonhomothetic	FIML	Yes	Implicitly Rejected	None	NA
Babin, Willis and Allen, 1982	U.S. pooled industries: cross section	Translog	NA	Homothetic	Zellner	No	No test	None	None
Lopez and Tung, 1982	Canada Ag: cross section + time series, 1961-79	General Leontief	Nonneutral (Sig.)	Nonhomothetic	FIML	Yes	Accepted	Accepted	NA
Rostamizadeh et al., 1982	U.S. Ag: time series, 1960-79	Translog	Neutral	Homothetic	JRGLS	Yes	Rejected	Rejected	Rejected
Ray, 1982	U.S. Ag: time series for crop & livestock output, 1939-77	Translog (Multi-output)	Neutral (Sig.)	Nonhomothetic	JRGLS	Yes	No test	None	None
Ball and Chambers, 1982	U.S. meat processing: time series, 1954-76	Generalized translog	Nonneutral	Nonhomothetic	FIML	Yes (in 1965)	No test	None	None

Table 4 continued on next page.

Table 4 (continued).

Study	Data	Function	Technical Change	Homotheticity of P.F.	Estimation Technique[1]	All Demands Negative?	Locally Concave	Test of Symmetry	Test of Homogeneity
Kim, 1983	Gulf of Mexico shrimp fishery: cross-section, 1977	Translog	NA	Homothetic	Zellner	Yes	Yes	None	None
Antle and Aitah, 1983	Egypt rice farms: cross section, 1976–77	Translog	NA	Homothetic	Zellner	Yes	Inconclusive	Accepted	None
Chambers and Vasavada, 1983	U.S. Ag: time series, 1947–80	General Leontief	Nonneutral	Homothetic	MLE	Yes	Yes, in the majority	None	None

[1]RGLS = Restricted Generalized Least Squares; FIML = Full Information Maximum Likelihood; Zellner = Zellner's Seemingly Unrelated Regression; JRGLS = Joint Restricted Generalized Least Squares; MLE = Maximum Likelihood Estimation; and FIQML = Full Information Quasi-Maximum Likelihood.

specification issues. These tests are in fact **conditional** upon the particular functional form chosen by the researcher. For example, one or more regularity conditions might be violated by a translog cost function but a generalized Leontief cost function fitted to the same data could meet all regularity conditions.

SUMMARY RECOMMENDATIONS

Researchers should consider dual techniques when the real world research problem conforms to the theoretical requirements of the theory and available data are adequate. Are the optimization postulates of the dual function (e.g., profit maximization or cost minimization) reasonable? Do adequate markets or other price-setting institutions exist? Is there an absence of difficult-to-model confounding constraints which could hinder optimizing behavior by producers? Is a set of price data with reasonable dispersion available? If these questions can be answered in the affirmative, the researcher then should choose carefully the particular dual function which best suits his research objectives and behavioral/ institutional realities of the problem. Do cost, indirect production function, or profit function optimization postulates best describe behavior of the target group?

Careful attention should also be devoted to selection of an appropriate functional form. Review of previous literature can provide guidance on this choice. Research objectives can also influence this choice. For example, the estimating equations for a generalized Leontief cost function technology impose homogeneity of degree zero on factor demand functions in a manner that precludes statistically testing this property. If homogeneity tests are an objective of the research, another functional form should be chosen.

Researchers should choose the best data available and use logical well-documented variable construction procedures. Variables included should correspond with the objectives of the study.

Statistical estimation procedures should incorporate any necessary theoretical restrictions. As noted earlier, this will often require using jointly restricted generalized least squares to obtain efficient estimates.

Finally, researchers should ensure that the research resources at their disposal are adequate before using dual approaches. This includes access to the necessary econometric software and theoretical-econometric expertise.

In summary, we offer our judgement on types of problems where dual approaches are likely to be most and least fruitful. Among the types of problems agricultural economists typically consider, the following may possibly be productively examined by dual approaches:

- analysis of aggregate factor demand and substitution, output supply, technical change, and returns to scale at regional or national levels for particular sectors and industries,
- analysis of welfare and allocation impacts of policy or environmentally-induced price or constraint changes on typical firms or industries,
- analysis of multiple-product firm and multiple-product industry responses.

Production analyses where dual approaches are not relevant or appropriate include:

- analysis of pure biological or physical response relationships,
- "quick and dirty" studies where time and research resources are limited,
- problems with little dispersion in price data across observations,
- analysis of environmental or non-market goods for which no price history exists,
- problems in which assumed optimization behavior is in doubt or unknown constraints confound optimization.

For these problems, traditional primal or institutional approaches will most likely be more fruitful.

REFERENCES

Babin, Frederick G., Cleve E. Willis and P. Geoffrey Allen. "Estimation of Substitution Possibilities Between Water and Other Production Inputs." *American Journal of Agricultural Economics* 64(1982):149–152.

Ball, V. Eldon and Robert G. Chambers, "An Economic Analysis of Technology in the Meat Products Industry." *American Journal Agricultural Economics* 64(1982):699–709.

Bartle, Robert G. *The Elements of Real Analysis*, second edition. New York: John Wiley and Sons, 1976.

Berndt, E. R. and Christensen, L. R. "The Internal Structure of Functional Relationships: Separability, Substitution, and Aggregation." *Rev. Econ. Studies,* Vol. 40 (1973):403–410.

Binswanger, H. "A Cost Function Approach to the Measurement of Elasticities of Substitution." *American Journal Agricultural Economics* 56(1974):377–386.

Chambers, Robert G. "Duality, the Output Effect, and Applied Comparative Statics." *American Journal Agricultural Economics* 64(1982):152–56.

Christensen, L. R. and W. H. Green. "Economies of Scale in U.S. Electric Power Generation." *Journal of Political Economy*, Vol. 84 (1976):655–676.

Deaton, A. and J. Muellbauer. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press, 1980.

Diewert, W. E. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function." *Journal of Political Economy* 79(1971):481–507.

Ferguson, C. E. *The Neoclassical Theory of Production and Distribution*. Cambridge: Cambridge University Press, 1971.

Fuss, Melvyn A. "Factor Substitution in Electricity Generation: A Test of the Putty-Clay Hypothesis." *Production Economics: A Dual Approach to Theory and Applications*, ed. M. Fuss and D. McFadden. Amsterdam: North Holland Publishing Co., 1978.

Hammonds, T. M., R. Yadav and C. Vathana. "The Elasticity of Demand for Hired Farm Labor." *American Journal Agricultural Economics* 55(1973):242–245.

Harcourt, G. C. "Some Cambridge Controversies in the Theory of Capital." *Journal of Economic Literature,* Vol. VII, No. 2, June 1969; also by Cambridge University Press.

Heien, Dale. "The Structure of Food Demand: Interrelatedness and Duality." *American Journal Agricultural Economics* 64(1982):213–221.

Hotelling, H. "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions." *Journal of Political Economy*, Vol. 40 (1932):577–616.

Kako, T. "Decomposition Analysis of Derived Demand for Factor Inputs: The Case of Rice Production in Japan." *American Journal Agricultural Economics* 60(1978):628–635.

Kay, D. C. *College Geometry*. New York: Holt, Rinehart and Winston, 1969.

Lau, L. J. *A Characterization of the Normalized Restricted Profit Function*. Technical Report No. 134, Stanford University, June 1974.

Lau, L. J. and P. A. Yotopoulos. "Profit, Supply and Factor Demand Functions." *American Journal Agricultural Economics* 54(1972):11–18.

Lianos, T. P. "The Relative Share of Labor in United States Agriculture, 1949–1968." *American Journal Agricultural Economics* 53(1971):411–422.

Lopez, R. E. "The Structure of Production and the Derived Demand for Inputs in Canadian Agriculture." *American Journal Agricultural Economics* 62(1980):38–45.

Lopez, R. E. "Analysis of a Small Open Economy: The Case of Energy Prices in Canada." *American Journal Agricultural Economics* 64(1982):510–19.

Lopez, R. E. and F. L. Tung. "Energy and Non-Energy Input Substitution Possibilities and Output Scale Effects in Canadian Agriculture." *Canadian Journal Agricultural Economics* 30(1982):87–105.

McFadden, D. L. *Cost, Revenue, and Profit Functions*. Amsterdam: North-Holland Publishing Company, 1972.

Pasinetti, L. L. *Lectures on the Theory of Production*. New York: The MacMillan Press Ltd., 1977.

Pope, Rulon. "To Dual or not to Dual." *Western Journal Agricultural Economics* 7(1982):337–351.

Ray, S. "Translog Cost Analysis of U. S. Agriculture." *American Journal Agricultural Economics* 64(1982):490–498.

Robinson, Joan. *The Accumulation of Capital*. Homewood, Illinois: Richard D. Irwin, Inc., 1956.

Rostamizadeh, Ahmad. *The Duality Between Cost and Production Functions with Applications to the Agricultural Labor Markets in the United States, Washington State and California*. Unpublished Ph.D. dissertation, Washington State University, 1981.

Rostamizadeh, Ahmad, David Holland, Ron Mittelhammer, and Douglas Young. "The Duality Between Cost and Production Functions: Elements of Theory and an Application in Agricultural Economics Research." Unpublished typescript, Department of Agricultural Economics, Washington State University, 1982.

Shephard, R. W. *Cost and Production Functions.* Princeton: Princeton University Press, 1953.

Silberberg, E. *The Structure of Economics.* New York: McGraw Hill Book Co., 1978.

Stein, S. K. *Calculus and Analytic Geometry.* New York: McGraw Hill, Inc. 1977.

Theil, H. *Principles of Econometrics.* New York: John Wiley & Sons, Inc. 1971.

Theil, H. *The System-Wide Approach to Microeconomics.* Chicago: University of Chicago Press, 1980.

U.S. Department of Agriculture. *Agricultural Statistics.* Washington, D.C. Various issues, 1960–79.

________. *Farm Labor.* Washington, D.C. Various issues, 1960–79.

________. *Farm Real Estate Market Developments.* Washington, D.C. Various issues, 1960–79.

________. *State Farm Income Statistics.* Washington, D.C. Various issues, 1960–79.

Varian, H. *Microeconomic Analysis.* New York: W. W. Norton and Co., Inc. 1978.

Vathana, C. *A Study of Demand and Supply Relationships in the Agricultural Labor Market.* Unpublished M.S. thesis, Oregon State University, 1972.

Weaver, Robert D. "Multiple Input, Multiple Output Production Choices and Technology in the U.S. Wheat Region." *American Journal Agricultural Economics* 65(1983):45–56.

Young, D. L. "Relevance of Duality Theory to the Practicing Agricultural Economist: Discussion." *Western Journal of Agricultural Economics* 7(1982):367–372.

APPENDIX A: SHEPHARD'S GRAPHICAL ILLUSTRATION OF THE DUALITY BETWEEN PRODUCTION AND COST FUNCTIONS

Shephard (1953) was the first to present a rigorous graphical demonstration of the duality between production and cost functions. His demonstration utilized the geometric concept of polarity which may be unfamiliar to many economists. We attempt here to review the seminal ideas of Shephard, making the exposition as accessible as possible. The demonstration is in two dimensions, but the approach can be extended to higher dimensions. The point of the demonstration is that knowledge of the isoquants of a production technology imply knowledge of the isocost curves of a cost function in input price space and vice versa.[6]

A key concept in Shephard's graphical approach is that of the "unit minimum cost" (UMC) function, which represents the combinations of input prices, r_1 and r_2, for which the minimum cost of producing a given level of output, q, is unity. To clarify what is meant by the UMC function, first recall that the standard minimum cost function for the production of output level q can be defined as

$$\text{(A1)} \qquad c(q,r_1,r_2) = r_1X_1^*(r_1,r_2,q) + r_2X_2^*(r_1,r_2,q)$$

where $X_i^*(r_1,r_2,q)$ is the constant-output demand curve for input i defined by the solution to the problem

$$\text{(A2)} \qquad \min_{X_1,X_2} r_1X_1 + r_2X_2 \text{ s.t. } q = q(X_1,X_2),$$

$q(X_1,X_2)$ representing the production function. The unit minimum cost function is then implicitly defined by

$$\text{(A3)} \qquad c(q,r_1,r_2) = 1,$$

for a given q. The equation can be solved for (r_1,r_2) points that represent prices at which q can be produced at a minimum cost of 1 unit.

The demonstration of duality between isoquants and isocosts can be restricted to an analysis between isoquants and the UMC function. That minimum cost curves other than the UMC curve need not be referenced explicitly in the demonstration of duality can be motivated as follows. First, recall that minimum cost functions are homogeneous of degree one in input prices, assuming continuity and concavity of the production function, i.e.

$$\text{(A4)} \qquad c(q,\lambda r_1,\lambda r_2) = \lambda c(q,r_1,r_2)$$

for $\lambda > 0$. Now instead of defining a UMC function, implicitly define a minimum cost function representing combinations of input prices, r''_1 and r''_2, for which the minimum cost of producing output level q is some arbitrary positive value k, i.e.

$$\text{(A5)} \qquad c(q,r''_1,r''_2) = k.$$

By homogeneity of degree one in input prices, it is then possible to write

$$\text{(A6)} \qquad c(q,r''_1/k,r''_2/k) = 1 = c(q,r_1,r_2)$$

where $r_1 = r''_1/k$ and $r_2 = r''_2/k$ are prices on the UMC curve. Thus, once the points on the UMC curve for given output level q are known, the points on the minimum cost for any other cost level k, with output level q, can be found by scaling as $r''_1 = kr_1$ and $r''_2 = kr_2$. In other words, once you know the UMC curve, you know all of the minimum cost curves.

In Figure 2, the isoquant implicitly defined by $h(z_1,z_2) = \overline{q}$ and the unit cost function implicitly

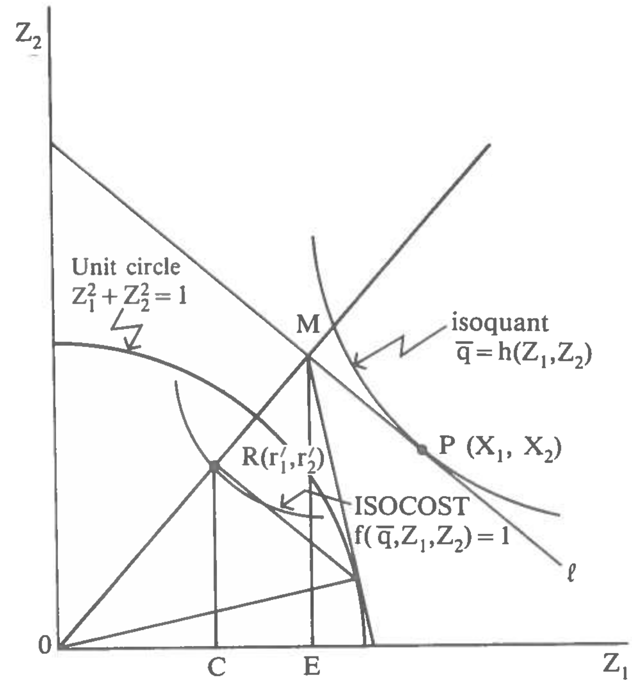

Figure 2. Duality Between the Cost Function and Production Function.

[6]The "isocost curve in input prices space" should not be confused with the more familiar isocost curve defined by a "capital constraint" in input quantity space whose tangency with an isoquant identifies the point of minimum cost production. To clearly distinguish between these entirely different concepts, we use the term "budget line" in this paper to refer to the isocost curve in input space and reserve the term "isocost curve" for the input-price space concept.

defined by $f(\overline{q},z_1,z_2) = 1$ are sketched where the horizontal and vertical axes are used interchangeably for the amount of factor inputs X_1 and X_2, and input price levels r_1 and r_2 (z_1 and z_2 act as dummy arguments in the production and cost functions). The units of measurement for (r_1,r_2,X_1,X_2) are chosen to relate to the z_1 and z_2 axes in such a way that the UMC isocost curve lies below the associated isoquant as well as to allow the unit circle to be drawn as in Figure 2. In the figure, let line ℓ be tangent to the isoquant at point P and the line OM be at a right angle to the line ℓ at point M. Let r''_1 and r''_2 be input prices such that the price ratio $T = r''_1/r''_2$ is equal to the absolute slope of the line . Then recalling the familiar graphical cost minimizing condition that the budget line should be tangent to the isoquant, the equation for the line ℓ can be written as

$$\text{(A7)} \qquad r''_1 z_1 + r''_2 z_2 = c(q,r''_1,r''_2)$$

where $c(r''_1,r''_2,q)$ defines the minimum cost of producing output level q given input prices r''_1 and r''_2.

Now note that the line OM intersects the UMC isocost curve at the point R having coordinates $(r_1',r_2') = (kr''_1,kr''_2)$ for a fixed constant k. In order to show this, recall that by the similar triangle theorem (see geometry review in Appendix B) OE/ME = r''_1/r''_2, which is the inverse slope of the line OM. The equation for the line OM can thus be represented by $z_2 = (r''_2/r''_1)z_1$, and it is easy to verify by substitution that all points $(z_1,z_2) = (kr''_1,kr''_2)$ for $k\varepsilon[0,1]$ are on this line. Since $f(q,r''_1,r''_2)$ is homogeneous of degree one in prices, there exists a value of k, namely $k = (f(q,r''_1,r''_2))^{-1}$, such that

$$\text{(A8)} \qquad k\,c(q,r''_1,r''_2) = c(q,kr''_1,kr''_2) = c(q,r_1',r_2') = 1,$$

and thus OM intersects the UMC cost curve at $(r_1',r_2') = (kr''_1,kr''_2)$.

Now define the normalized equation for the line ℓ as

$$\text{(A9)} \qquad \frac{r''_1X_2}{[(r''_1)^2+(r''_2)^2]^{1/2}} + \frac{r''_2X_2}{[(r''_1)^2+(r''_2)^2]^{1/2}} = \frac{c(q,r''_1,r''_2)}{[(r''_1)^2+(r''_2)^2]^{1/2}} = OM$$

where the last equality results from the fact that the value of $c(q,r''_1,r''_2)/[(r''_1)^2+(r''_2)^2]^{1/2}$ measures the distance between line ℓ and the origin, the distance clearly being OM.[7] Recalling that $r_1'=kr''_1$ and $r_2' = kr''_2$, substitution for (r''_1,r''_2) in the last equality above yields

$$\begin{aligned}\text{(A10)} \qquad OM &= f(q,r_1'/k,r_2'/k)/(k^{-1}[(r_1')^2+(r_2')^2]^{1/2}) \\ &= k^{-1}f(q,r_1',r_2')/(k^{-1}[(r_1')^2+(r_2')^2]^{1/2}) \\ &\quad \text{(by homogeneity)} \\ &= 1/OR\end{aligned}$$

where the last equality results from the fact that $(r_1')^2 + (r_2')^2 = (OR)^2$ by the Pythogorean theorem since $r_1' = OC$ and $r_2' = RC$ form the base and height of the right triangle OCR, and since $f(r_1',r_2',q) = 1$ because r_1' and r_2' are points on the UMC isocost curve. Thus OM = 1/OR.

The entire analysis can be repeated choosing a different slope for the line ℓ tangent to the isoquant, leading to another point identified on the UMC isocost curve, and so on. Thus, there is a one-to-one correspondence between points on the isoquant and points on the UMC isocost curve, and the procedure could be followed to construct the UMC isocost curve (and implicitly all minimum cost curves via the linear homogeneity property discussed earlier) from knowledge of the isoquant curve.

In terms of polarity, the above results indicate that the point R with coordinates (r_1',r_2') on the UMC isocost curve is the pole of the polar line ℓ with respect to the unit circle ($z_1^2 + z_2^2 = 1$) centered at the origin (see geometry review in appendix). Since the choice of the slope of line ℓ was arbitrary, it is clear that the UMC cost curve defined by $f(r_1,r_2,q) = 1$ is the locus of the poles of all polar lines, such as ℓ, tangent to the isoquant defined by $q = h(z_1,z_2)$ and defined by all input-price ratios $\lambda = r_1/r_2$. Thus, a dual relationship is demonstrated between isoquants and isocost curves, and knowledge of isoquants implies knowledge of isocost.

The entire approach described above can be repeated interchanging the roles of the isoquant and the UMC isocost curve to show that the isoquant curve can be constructed from knowledge of the UMC isocost curve. In terms of the polarity concept, it can be shown that the isoquant defined by $q = h(X_1,X_2)$ is the locus of the poles of all polar lines tangent to the UMC curve and defined by all choices of input ratios $T = X_1/X_2$. The isoquant defined by $q = h(X_1,X_2)$ and the UMC isocost curve defined by $f(q,r_1,r_2) = 1$ are thus polar reciprocal curves, and knowledge of isocost curves implies knowledge of the production technology.

---

[7]Assuming a linear function such as $ax+by = c$, the normalized equation is written as

$$\frac{ax}{(a^2+b^2)^{1/2}} + \frac{bx}{(a^2+b^2)^{1/2}} = \frac{c}{(a^2+b^2)^{1/2}},$$ and the absolute value of the RHS term measures the distance from the origin to the line (see Stein, p. 727).

APPENDIX B: A REVIEW OF SOME CONCEPTS IN GEOMETRY

Similar Triangles Theorem

Let the two triangles AOB and OME (Figure A1) be similar triangles. Then, OE/ME = OA/OB, and since the equation for line AB is

$P_1X_1 + P_2X_2 = \bar{C}$, implying OB = $\bar{C}/P_1$ and OA = $\bar{C}/P_2$, Then, OE/ME = P_1/P_2

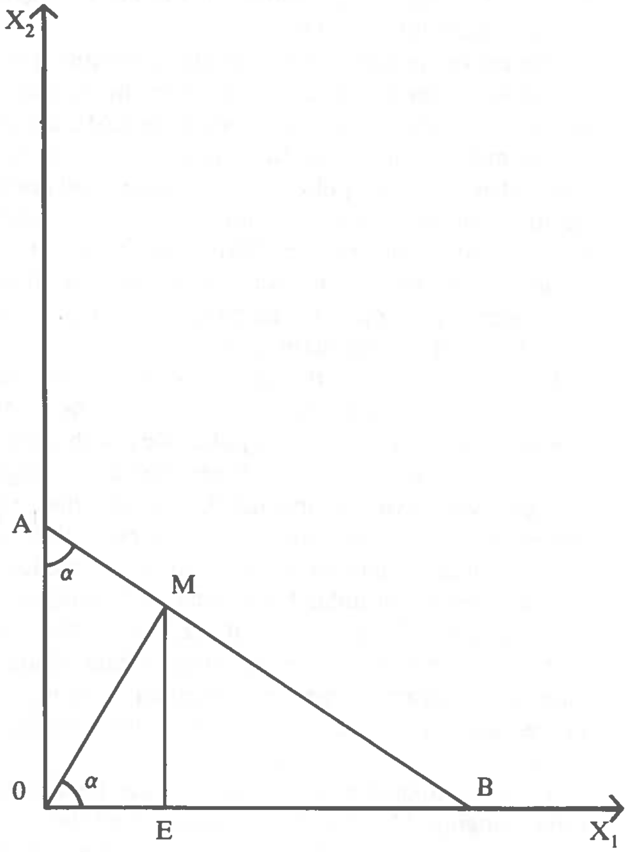

Figure A-1. Similar Triangles

Polarity

As a definition of the polarity concept, let O(r) be a circle of radius r and P be any point outside O(r) (Figure A2). Then P′ is defined to be the inverse of P and vice versa. The line MN is called the polar of P for circle O(r). If we draw a line perpendicular to OP′P at the point P, this line (a) is called the polar of P′ and P′ is the pole of the polar (a).

Based on the polarity concept the polarity theorem will be stated and proved. However, before proving the polarity theorem, two other relevant theorems from geometry, the Pythagorean Theorem and the Chord-Secant Theorem will be stated for use in the proof of the polarity theorem.

Pythagorean Theorem

In any triangle where one of the three angles is 90 degrees, the sum of the squared lengths of the two legs is equal to the hypotenuse squared, and conversely if the sum of the squared lengths of the two legs is equal to the hypotenuse squared then the triangle is a right triangle. This theorem can be illustrated by referring to Figure A3. In this figure, AB is the hypotenuse, CA and CB are the two legs. This theorem says that $(AB)^2 = (AC)^2 + (CB)^2$ (for proof see Kay, 1969, p. 173).

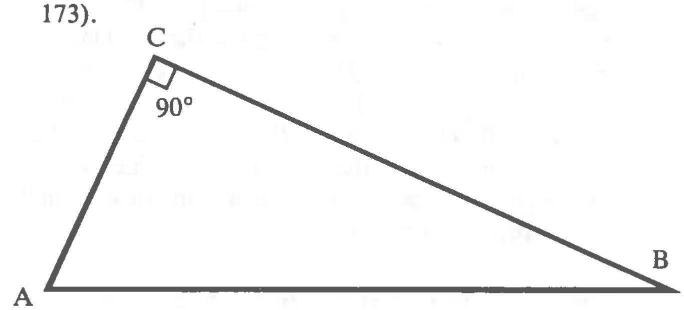

Figure A-3.Right Triangle

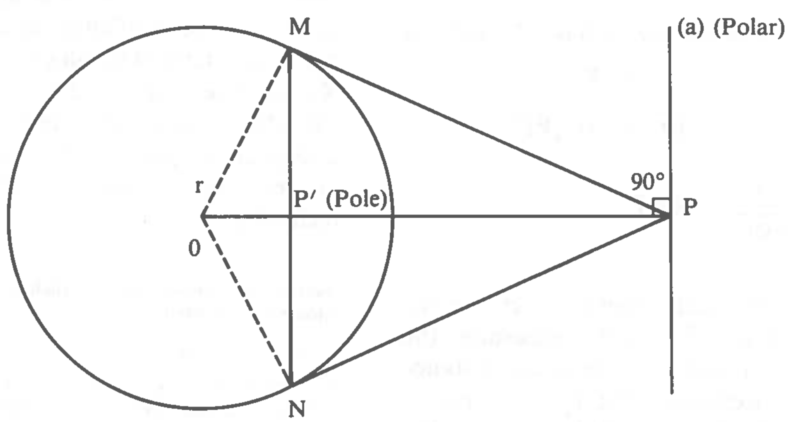

Figure A-2.Polarity

Chord-Secant Theorem

The Chord-Secant Theorem applies to a circle of radius r and a point, such as P, outside the circle (Figure A4). If a line goes through the point P and intersects the circle at two points A and B, then in magnitude and in sign $PA \cdot PB = (PO)^2 - r^2$ (for proof see Kay, 1969, pp. 231–232).

Also, if PT is tangent to the circle in Figure A4, then $(PT)^2 + r^2 = (PO)^2$ and the triangle PTO is a right triangle. Proof: Follows from Chord-Secant theorem upon recognizing that $PA \cdot PB = PT$.

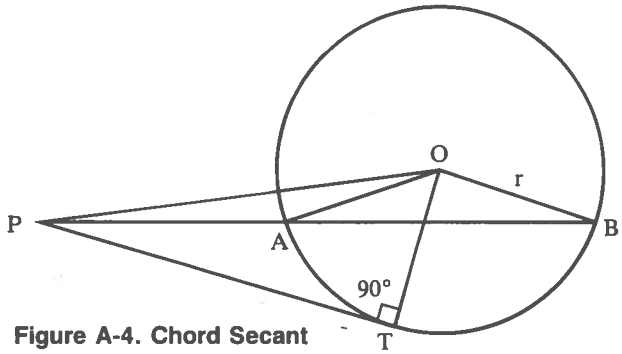

Figure A-4. Chord Secant

Polarity Theorem

A line PDO (Figure A5) intersecting a circle having a radius of one and going through the center of the circle, point 0, has distance from point D to the center of the circle equal to 1/OP.

Proof: In Figure A5, D is the pole of the line (a) and r (radius of the circle) is one. OTP is a right angle, because PT is tangent to the circle and OT is a radius of a circle which is drawn from 0 (center) to the point of tangency (see Chord-Secant Theorem and its corollary).

Now by the Pythagorean theorem, we know that

$$(OP)^2 = (OT)^2 + (TP)^2$$

and

$$(OD)^2 = (OT)^2 - (TD)^2,$$

so that
$$\begin{aligned}(OP)^2 \cdot (OD)^2 &= (OT)^4 + (TP)^2(OT)^2 \\ &\quad - (TD)^2(OT)^2 + (TP)^2 \\ &= (OT)^4 + (TP)^2(OT)^2 \\ &\quad - (TD)^2(OP)^2.\end{aligned}$$

Noting that the area of triangle OTP can be written equivalently as either $OT \cdot TP/2$ or $TD \cdot OP/2$, we have that $(OT)^2(TP)^2 = (TD)^2(OP)^2$, in which case substitution for $(TD)^2(OP)^2$ above results in

$$(OP)^2 \cdot (OD)^2 = (OT)^4$$

or

$$OP \cdot OD = (OT)^2.$$

Since $OT = 1$ because the radius of the circle is 1, we finally have that $OD = 1/OP$.

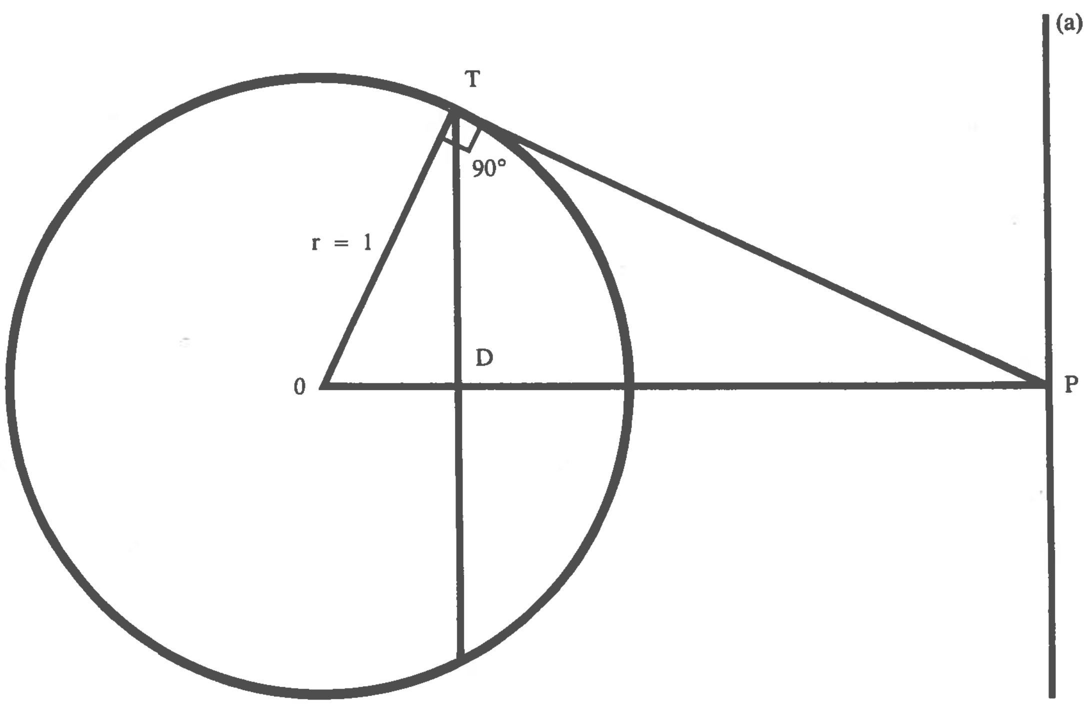

Figure A-5. Pole and polar with the circle having radius one.