ECONOMICS 581: LECTURE NOTES

CHAPTER 4: MICROECONOMIC THEORY: A DUAL APPROACH

W. Erwin Diewert March 2011.

1. Introduction

In this chapter, we will show how the theory of convex sets and concave and convex functions can be useful in deriving some theorems in microeconomics. Section 2 starts off by developing the properties of cost functions. It is shown that without assuming any regularity properties on an underlying production function, the corresponding function satisfies a large number of regularity properties. Section 3 shows how the cost function can be used to determine a production function that is consistent with a given cost function satisfying the appropriate regularity conditions. Section 4 establishes the derivative property of the cost function: it is shown that the first order partial derivatives of the cost function generate the firm's system of cost minimizing input demand functions. Section 5 shows how the material in the previous sections can be used to derive the comparative statics properties of the producer's system of cost minimizing input demand functions. Section 6 asks under what conditions can we assume that the technology exhibits constant returns to scale. Section 7 indicates that price elasticities of demand will tend to decrease in magnitude as a production model becomes more aggregated.

Section 8 notes that the duality between cost and production functions is isomorphic or identical to the duality between utility and expenditure functions. In this extension of the previous theory, the output level of the producer is replaced with the utility level of the consumer, the production function of the producer is replaced with the utility function of the consumer and the producer's cost minimization problem is replaced by the problem of the consumer minimizing the expenditure required to attain a target utility level. Thus the results in the first 5 sections have an immediate application to the consumer's system of Hicksian demand functions.

The final sections of the chapter return to producer theory but it is no longer assumed that only one output is produced; we extend the earlier analysis to the case of multiple output and multiple input technologies.

2. Properties of Cost Functions

The *production function* and the corresponding *cost* function play a central role in many economic applications. In this section, we will show that under certain conditions, the cost function is a sufficient statistic for the corresponding production function; i.e., if we know the cost function of a producer, then this cost function can be used to generate the underlying production function.

Let the producer's *production function* f(x) denote the maximum amount of output that can be produced in a given time period, given that the producer has access to the nonnegative vector of inputs, $x \equiv [x_1,\ldots,x_N] \geq 0_N$. If the production function satisfies certain regularity conditions,[1] then given any positive output level y that the technology can produce and any strictly positive vector of input prices $p \equiv [p_1,\ldots,p_N] >> 0_N$, we can calculate the producer's *cost function* C(y,p) as the solution value to the following constrained minimization problem:

(1) $C(y,p) \equiv \min_x \{p^T x : f(x) \geq y \; ; x \geq 0_N\}$.

It turns out that the cost function C will satisfy the following 7 properties, irrespective of the properties of the production function f.

Theorem 1; Diewert (1982; 537-543)[2]: Suppose f is continuous from above. Then C defined by (1) has the following properties:

Property 1: C(y,p) is a *nonnegative* function.
Property 2: C(y,p) is *positively linearly homogeneous in p* for each fixed y; i.e.,

(2) $C(y,\lambda p) = \lambda C(y,p)$ for all $\lambda > 0$, $p >> 0_N$ and $y \in$ Range f (i.e., y is an output level that is producible by the production function f).

Property 3: C(y,p) is *nondecreasing in p* for each fixed $y \in$ Range f; i.e.,

(3) $y \in$ Range f, $0_N << p^1 < p^2$ implies $C(y,p^1) \leq C(y,p^2)$.

Property 4: C(y,p) is a *concave function of p* for each fixed $y \in$ Range f; i.e.,

(4) $y \in$ Range f, $p^1 >> 0_N$; $p^2 >> 0_N$; $0 < \lambda < 1$ implies
$$C(y,\lambda p^1+(1-\lambda)p^2) \geq \lambda C(y,p^1) + (1-\lambda)C(y,p^2).$$

Property 5: C(y,p) is a *continuous function of p* for each fixed $y \in$ Range f.
Property 6: C(y,p) is *nondecreasing in y* for fixed p; i.e.,

(5) $p >> 0_N$, $y^1 \in$ Range f, $y^2 \in$ Range f, $y^1 < y^2$ implies $C(y^1,p) \leq C(y^2,p)$.

Property 7: For every $p >> 0_N$, C(y,p) is *continuous from below in y*; i.e.,

(6) $y^* \in$ Range f , $y^n \in$ Range f for n = 1,2,…, $y^n \leq y^{n+1}$, $\lim_{n\to\infty} y^n = y^*$ implies
$$\lim_{n\to\infty} C(y^n,p) = C(y^*,p).$$

---

[1] We require that f be *continuous from above* for the minimum to the cost minimization problem to exist; i.e., for every output level y that can be produced by the technology (so that $y \in$ Range f), we require that the set of x's that can produce at least output level y (this is the upper level set $L(y) \equiv \{x : f(x) \geq y\}$) is a closed set in R^N.
[2] For the history of closely related results, see Diewert (1974a; 116-120).

Proof of Property 1: Let $y \in$ Range f and $p >> 0_N$. Then

$$\begin{aligned} C(y,p) &\equiv \min_x \{p^T x : f(x) \geq y \, ; \, x \geq 0_N\} \\ &= p^T x^* && \text{where } x^* \geq 0_N \text{ and } f(x^*) \geq y \\ &\geq 0 && \text{since } p >> 0_N \text{ and } x^* \geq 0_N. \end{aligned}$$

Proof of Property 2: Let $y \in$ Range f, $p >> 0_N$ and $\lambda > 0$. Then

$$\begin{aligned} C(y,\lambda p) &\equiv \min_x \{\lambda p^T x : f(x) \geq y \, ; \, x \geq 0_N\} \\ &= \lambda \min_x \{p^T x : f(x) \geq y \, ; \, x \geq 0_N\} && \text{since } \lambda > 0 \\ &= \lambda C(y,p) && \text{using the definition of } C(y,p). \end{aligned}$$

Proof of Property 3: Let $y \in$ Range f, $0_N << p^1 < p^2$. Then

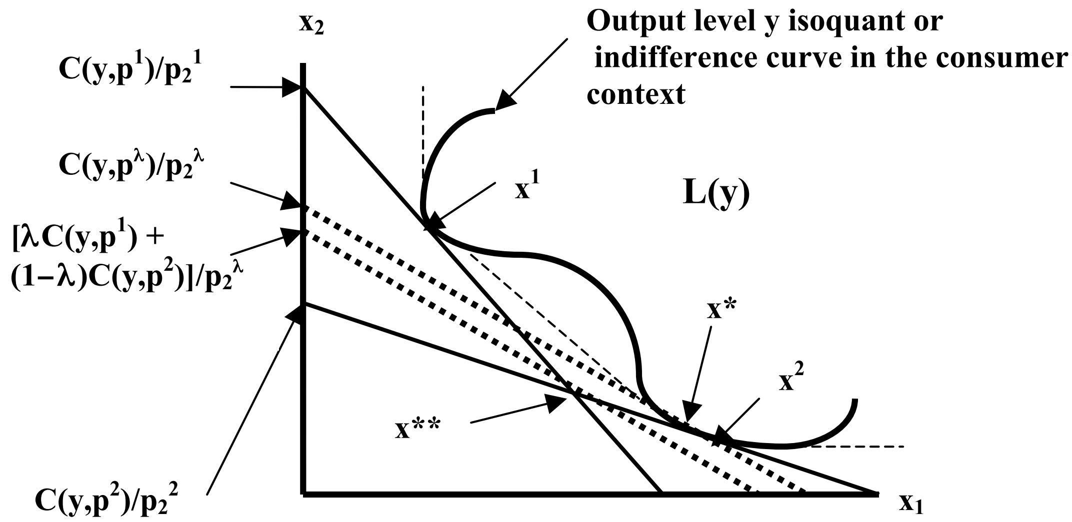
$$\begin{aligned} C(y,p^2) &\equiv \min_x \{p^{2T} x : f(x) \geq y \, ; \, x \geq 0_N\} \\ &= p^{2T} x^* && \text{where } f(x^*) \geq y \text{ and } x^* \geq 0_N \\ &\geq p^{1T} x^* && \text{since } x^* \geq 0_N \text{ and } p^2 > p^1 \\ &\geq \min_x \{p^{1T} x : f(x) \geq y \, ; \, x \geq 0_N\} && \text{since } x^* \text{ is feasible for this problem} \\ &\equiv C(y,p^1). \end{aligned}$$

Proof of Property 4: Let $y \in$ Range f, $p^1 >> 0_N$; $p^2 >> 0_N$; $0 < \lambda < 1$. Then

$$\begin{aligned} C(y,\lambda p^1+(1-\lambda)p^2) &\equiv \min_x \{[\lambda p^1+(1-\lambda)p^2]^T x : f(x) \geq y \, ; \, x \geq 0_N\} \\ &= [\lambda p^1 + (1-\lambda)p^2]^T x^* \quad \text{where } x^* \geq 0_N \text{ and } f(x^*) \geq y \\ &= \lambda p^{1T} x^* + (1-\lambda)p^{2T} x^* \\ &\geq \lambda \min_x \{p^{1T} x : f(x) \geq y \, ; \, x \geq 0_N\} + (1-\lambda)p^{2T} x^* \\ &\qquad \text{since } x^* \text{ is feasible for the cost minimization problem that uses} \\ &\qquad \text{the price vector } p^1 \text{ and using also } \lambda > 0 \\ &= \lambda C(y,p^1) + (1-\lambda)p^{2T} x^* \quad \text{using the definition of } C(y,p^1) \\ &\geq \lambda C(y,p^1) + (1-\lambda) \min_x \{p^{2T} x : f(x) \geq y \, ; \, x \geq 0_N\} \\ &\qquad \text{since } x^* \text{ is feasible for the cost minimization problem that uses} \\ &\qquad \text{the price vector } p^2 \text{ and using also } 1-\lambda > 0 \\ &= \lambda C(y,p^1) + (1-\lambda)C(y,p^2) \quad \text{using the definition of } C(y,p^2). \end{aligned}$$

Figure 1 below illustrates why this concavity property holds.

Figure 1: The Concavity in Prices Property of the Cost Function



In Figure 1, the isocost line $\{x: p^{1T}x = C(y,p^1)\}$ is tangent to the production possibilities set $L(y) \equiv \{x: f(x) \geq y, x \geq 0_N\}$ at the point x^1 and the isocost line $\{x: p^{2T}x = C(y,p^2)\}$ is tangent to the production possibilities set L(y) at the point x^2. Note that the point x** belongs to both of these isocost lines. Thus x** will belong to any weighted average of the two isocost lines. The λ and 1−λ weighted average isocost line is the set $\{x: [\lambda p^1+(1-\lambda)p^2]^T x = \lambda C(y,p^1) + (1-\lambda)C(y,p^2)\}$ and this set is the dotted line through x** in Figure 1. Note that this dotted line lies *below*[3] the parallel dotted line that is just tangent to L(y), which is the isocost line $\{x: [\lambda p^1+(1-\lambda)p^2]^T x = [\lambda p^1+(1-\lambda)p^2]^T x^* = C(y,\lambda p^1+(1-\lambda)p^2)\}$ and it is this fact that gives us the concavity inequality (4).

Proof of Property 5: Since C(y,p) is a concave function of p defined over the open set of p's, $\Omega \equiv \{p: p >> 0_N\}$, it follows that C(y,p) is also continuous in p over this domain of definition set for each fixed $y \in$ Range f.[4]

Proof of Property 6: Let $p >> 0_N$, $y^1 \in$ Range f, $y^2 \in$ Range f, $y^1 < y^2$. Then

$C(y^2,p) \equiv \min_x \{p^T x : f(x) \geq y^2 ; x \geq 0_N\}$
$\quad \geq \min_x \{p^T x : f(x) \geq y^1 ; x \geq 0_N\}$
since if $y^1 < y^2$, the set $\{x : f(x) \geq y^2\}$ is a subset of the set $\{x : f(x) \geq y^1\}$ and the minimum of a linear function over a bigger set cannot increase
$\quad \equiv C(y^1,p)$.

Proof of Property 7: The proof is rather technical and may be found in Diewert (1993; 113-114). Q.E.D.

3 It can happen that the two dotted lines coincide.
4 See Fenchel (1953; 75) or Rockafellar (1970; 82).

Problems

1. In industrial organization,[5] it used to be fairly common to assume that a firm's cost function had the following linear functional form: $C(y,p) \equiv \alpha + \beta^T p + \gamma y$ where α and γ are scalar parameters and β is a vector of parameters to be estimated econometrically. What are sufficient conditions on these N+2 parameters for this cost function to satisfy properties 1 to 7 above? Is the resulting cost function very realistic?

2. Suppose a producer's production function, $f(x)$, defined for $x \in S$ where $S \equiv \{x: x \geq 0_N\}$ satisfies the following conditions:
(i) f is continuous over S;
(ii) $f(x) > 0$ if $x >> 0_N$ and
(iii) f is positively linearly homogeneous over S; i.e., for every $x \geq 0_N$ and $\lambda > 0$, $f(\lambda x) = \lambda f(x)$.
Define the producer's unit cost function $c(p)$ for $p >> 0_N$ as follows:
(iv) $c(p) \equiv C(1,p) \equiv \min_x \{p^T x : f(x) \geq 1 ; x \geq 0_N\}$;
i.e., $c(p)$ is the minimum cost of producing one unit of output if the producer faces the positive input price vector p. For $y > 0$ and $p >> 0_N$, *show that*
(v) $C(y,p) = c(p)y$.
Note: A production function f that satisfies property (iii) is said to exhibit *constant returns to scale*. The interpretation of (v) is that if a production function exhibits constant returns to scale, then total cost is equal to unit cost times the output level.

3. Shephard (1953; 4) defined a production function F to be *homothetic* if it could be written as
(i) $F(x) = g[f(x)]$; $x \geq 0_N$
where f satisfies conditions (i)-(iii) in Problem 2 above and $g(z)$, defined for all $z \geq 0$, satisfies the following regularity conditions:
(ii) $g(z)$ is positive if $z > 0$;
(iii) g is a continuous function of one variable and
(iv) g is monotonically increasing; i.e., if $0 \leq z^1 < z^2$, then $g(z^1) < g(z^2)$.
Let $C(y,p)$ be the cost function that corresponds to $F(x)$. *Show that* under the above assumptions, for $y > 0$ and $p >> 0_N$, we have
(v) $C(y,p) = g^{-1}(y)c(p)$
where $c(p)$ is the unit cost function that corresponds to the linearly homogeneous f and g^{-1} is the inverse function for g; i.e., $g^{-1}[g(z)] = z$ for all $z \geq 0$. Note that $g^{-1}(y)$ is a monotonically increasing continuous function of one variable.

3. The Determination of the Production Function from the Cost Function

The material in the previous section shows how the cost function can be determined from a knowledge of the production function. We now ask whether a knowledge of the cost

[5] For example, see Walters (1961).

function is sufficient to determine the underlying production function. The answer to this question is *yes*, but with some qualifications.

To see how we might use a given cost function (satisfying the 7 regularity conditions listed in the previous section) to determine the production function that generated it, pick an arbitrary feasible output level $y > 0$ and an arbitrary vector of positive prices, $p^1 \gg 0_N$ and use the given cost function C to define the following isocost surface: $\{x: p^{1T}x = C(y,p^1)\}$. This isocost surface must be tangent to the set of feasible input combinations x that can produce at least output level y, which is the upper level set, $L(y) \equiv \{x: f(x) \geq y; x \geq 0_N\}$. It can be seen that this isocost surface and the set lying above it must contain the upper level set L(y); i.e., the following *halfspace* $M(y,p^1)$, contains L(y):

(7) $M(y,p^1) \equiv \{x: p^{1T}x \geq C(y,p^1)\}$.

Pick another positive vector of prices, $p^2 \gg 0_N$ and it can be seen, repeating the above argument, that the halfspace $M(y,p^2) \equiv \{x: p^{2T}x \geq C(y,p^2)\}$ must also contain the upper level set L(y). Thus L(y) must belong to the intersection of the two halfspaces $M(y,p^1)$ and $M(y,p^2)$. Continuing to argue along these lines, it can be seen that L(y) must be contained in the following set, which is the intersection of all of the supporting halfspaces to L(y):

(8) $M(y) \equiv \bigcap_{p \gg 0_N} M(y,p)$.

Note that M(y) is defined using just the given cost function, C(y,p). Note also that since each of the sets in the intersection, M(y,p), is a convex set, then M(y) is also a convex set. Since L(y) is a subset of each M(y,p), it must be the case that L(y) is also a subset of M(y); i.e., we have

(9) $L(y) \subset M(y)$.

Is it the case that L(y) is equal to M(y)? In general, the answer is *no*; M(y) forms an *outer approximation* to the true production possibilities set L(y). To see why this is, see Figure 1 above. The boundary of the set M(y) partly coincides with the boundary of L(y) but it encloses a bigger set: the backward bending parts of the isoquant $\{x: f(x) = y\}$ are replaced by the dashed lines that are parallel to the x_1 axis and the x_2 axis and the inward bending part of the true isoquant is replaced by the dashed line that is tangent to the two regions where the boundary of M(y) coincides with the boundary of L(y). However, if the producer is a price taker in input markets, then it can be seen that *we will never observe the producer's nonconvex portions or backwards bending parts of the isoquant*. Thus under the assumption of competitive behavior in input markets, there is no loss of generality in assuming that the producer's production function is *nondecreasing* (this will eliminate the backward bending isoquants) or in assuming that the upper level sets of the production function are convex sets (this will eliminate the nonconvex portions of the upper level sets). Recall that a function has convex upper level sets if and only if it is *quasiconcave*.

Putting the above material together, we see that conditions on the production function f(x) that are necessary for the sets M(y) and L(y) to coincide are:

(10) f(x) is defined for $x \geq 0_N$ and is continuous from above[6] over this domain of definition set;
(11) f is nondecreasing and
(12) f is quasiconcave.

Theorem 2: *Shephard Duality Theorem*:[7] If f satisfies (10)-(12), then the cost function C defined by (1) satisfies the properties listed in Theorem 1 above and the upper level sets M(y) defined by (8) using only the cost function coincide with the upper level sets L(y) defined using the production function; i.e., under these regularity conditions, the production function and the cost function determine each other.

We now consider how an explicit formula for the production function in terms of the cost function can be obtained. Suppose we have a given cost function, C(y,p), and we are given a strictly positive input vector, $x \gg 0_N$, and we ask what is the maximum output that this x can produce. It can be seen that

(13) $f(x) = \max_y \{y: x \in M(y)\}$
$= \max_y \{y: C(y,p) \leq p^T x \text{ for every } p \gg 0_N\}$ using definitions (7) and (8).
$= \max_y \{y: C(y,p) \leq 1 \text{ for every } p \gg 0_N \text{ such that } p^T x = 1\}$

where the last equality follows using the fact that C(y,p) is linearly homogeneous in p as is the function $p^T x$ and hence we can normalize the prices so that $p^T x = 1$.

We now have to make a bit of a digression and consider the continuity properties of C(y,p) with respect to p. We have defined C(y,p) for all strictly positive price vectors p and since this domain of definition set is open, we know that C(y,p) is also continuous in p over this set, using the concavity in prices property of C. We now would like to extend the domain of definition of C(y,p) from the strictly positive orthant of prices, $\Omega \equiv \{p: p \gg 0_N\}$, to the nonnegative orthant, Clo $\Omega \equiv \{p: p \geq 0_N\}$, which is the closure of Ω. It turns out that it is possible to do this if we make use of some theorems in convex analysis.

Theorem 3: *Continuity from above of a concave function using the Fenchel closure operation*: Fenchel (1953; 78): Let f(x) be a concave function of N variables defined over

[6] Since each of the sets M(y,p) in the intersection set M(y) defined by (8) are closed, it can be shown that M(y) is also a closed set. Hence if M(y) is to coincide with L(y), we need the upper level sets of f to be closed sets and this will hold if and only if f is continuous from above.

[7] Shephard (1953) (1967) (1970) was the pioneer in establishing various duality theorems between cost and production functions. See also Samuelson (1953-54), Uzawa (1964), McFadden (1966) (1978), Diewert (1971) (1974a; 116-118) (1982; 537-545) and Blackorby, Primont and Russell (1978) for various duality theorems under alternative regularity conditions. Our exposition follows that of Diewert (1993; 123-132). These duality theorems are global in nature; i.e., the production and cost functions satisfy their appropriate regularity conditions over their entire domains of definition. However, it is also possible to develop duality theorems that are local rather than global; see Blackorby and Diewert (1979).

the open convex subset S of R^N. Then there exists a unique extension of f to Clo S, the closure of S, which is concave and continuous from above.

Proof: By the second characterization of concavity, the hypograph of f, $H \equiv \{(y,x): y \le f(x); x \in S\}$, is a convex set in R^{N+1}. Hence the closure of H, Clo H, is also a convex set. Hence the following function f* defined over Clo S is also a concave function:

$$(14)\ f^*(x) \equiv \max_y \{y: (y,x) \in \text{Clo } H\}; \qquad x \in \text{Clo } S.$$
$$= f(x) \qquad \text{for } x \in S.$$

Since Clo H is a closed set, it turns out that f* is continuous from above. Q.E.D.

To see that the extension function f* need not be continuous, consider the following *example*, where the domain of definition set is $S \equiv \{(x_1,x_2); x_2 \in R^1, x_1 \ge x_2^2\}$ in R^2:

$$(15)\ f(x_1,x_2) \equiv - x_2^2/x_1 \text{ if } x_2 \ne 0, x_1 \ge x_2^2;$$
$$\equiv 0 \qquad \text{if } x_1 = 0 \text{ and } x_2 = 0.$$

It is possible to show that f is concave and hence continuous over the interior of S; see problem 5 below. However, we show that f is not continuous at (0,0). Let (x_1,x_2) approach (0,0) along the line $x_1 = x_2 > 0$. Then

$$(16)\ \lim_{x_1 \to 0} f(x_1,x_2) = \lim_{x_1 \to 0} [- x_1^2/x_1] = \lim_{x_1 \to 0} [- x_1] = 0.$$

Now let (x_1,x_2) approach (0,0) along the parabolic path $x_2 > 0$ and $x_1 = x_2^2$. Then

$$(17)\ \lim_{x_2 \to 0;\ x_1 = x_2^2} f(x_1,x_2) = \lim_{x_2 \to 0} - x_2^2/x_2^2 = -1.$$

Thus f is not continuous at (0,0). It can be verified that restricting f to Int S and then extending f to the closure of S (which is S) leads to the same f* as is defined by (15). Thus the Fenchel closure operation does not always result in a continuous concave function.

Theorem 4 below states sufficient conditions for the Fenchel closure of a concave function defined over an open domain of definition set to be continuous over the closure of the original domain of definition. Fortunately, the hypotheses of this Theorem are weak enough to cover most economic applications. Before stating the theorem, we need an additional definition.

Definition: A set S in R^N is a *polyhedral set* iff S is equal to the intersection of a *finite* number of halfspaces.

Theorem 4: *Continuity of a concave function using the Fenchel closure operation*; Gale, Klee and Rockafellar (1968), Rockafellar (1970; 85): Let f be a concave function of N variables defined over an open convex polyhedral set S. Suppose f is bounded from below over every bounded subset of S. Then the Fenchel closure extension of f to the closure of S results in a continuous concave function defined over Clo S.

The proof of this result is a bit too involved for us to reproduce here but we can now apply this result.

Applying Theorem 4, we can extend the domain of definition of C(y,p) from strictly positive price vectors p to nonnegative price vectors using the Fenchel closure operation and hence C(y,p) will be continuous and concave in p over the set $\{p: p \geq 0_N\}$ for each y in the interval of feasible outputs.[8]

Now we can return to the problem where we have a given cost function, C(y,p), we are given a strictly positive input vector, $x \gg 0_N$, and we ask what is the maximum output that this x can produce. Repeating the analysis in (13), we have

(18) $f(x) = \max_y \{y: x \in M(y)\}$
$= \max_y \{y: C(y,p) \leq p^T x \text{ for every } p \gg 0_N\}$ using definitions (7) and (8).
$= \max_y \{y: C(y,p) \leq 1 \text{ for every } p \gg 0_N \text{ such that } p^T x = 1\}$
where we have used the linear homogeneity in prices property of C
$= \max_y \{y: C(y,p) \leq 1 \text{ for every } p \geq 0_N \text{ such that } p^T x = 1\}$
where we have extended the domain of definition of C(y,p) to nonnegative prices from positive prices and used the continuity of the extension function over the set of nonnegative prices
$= \max_y \{y: G(y,x) \leq 1\}$

where the function G(y,x) is defined as follows:

(19) $G(y,x) \equiv \max_p \{C(y,p): p \geq 0_N \text{ and } p^T x = 1\}$.

Note that the maximum in (19) will exist since C(y,p) is continuous in p and the feasible region for the maximization problem, $\{p: p \geq 0_N \text{ and } p^T x = 1\}$, is a closed and bounded set.[9] Property 7 on the cost function C(y,p) will imply that the maximum in the last line of (18) will exist. Property 6 on the cost function will imply that for fixed x, G(y,x) is nondecreasing in y. Typically, G(y,x) will be continuous in y for a fixed x and so the maximum y that solves (18) will be the y* that satisfies the following equation:[10]

(20) $G(y^*,x) = 1$.

Thus (19) and (20) implicitly define the production function $y^* = f(x)$ in terms of the cost function C.

[8] If $f(0_N) = 0$ and f(x) tends to plus infinity as the components of x tend to plus infinity, then the feasible y set will be $y \geq 0$ and C(y,p) will be defined for all $y \geq 0$ and $p \geq 0_N$.
[9] Here is where we use the assumption that $x \gg 0_N$ in order to obtain the boundedness of this set.
[10] This method for constructing the production function from the cost function may be found in Diewert (1974a; 119).

Problems

4. Show that the $f(x_1,x_2)$ defined by (15) above is a concave function over the interior of the domain of definition set S. You do not have to show that S is a convex set.

5. In the case where the technology is subject to constant returns to scale, the cost function has the following form: $C(y,p) = yc(p)$ where $c(p)$ is a unit cost function. For $x >> 0_N$, define the function $g(x)$ as follows:
(i) $g(x) \equiv \max_p \{c(p): p^T x = 1; p \geq 0_N\}$.
Show that in this constant returns to scale case, the function $G(y,x)$ defined by (19) reduces to
(ii) $G(y,x) = yg(x)$.
Show that in this constant returns to scale case, the production function that is dual to the cost function has the following explicit formula for $x >> 0_N$:
(iii) $f(x) = 1/g(x)$.

6. Let $x \geq 0$ be input (a scalar number) and let $y = f(x) \geq 0$ be the maximum output that could be produced by input x, where f is the production function. Suppose that f is defined as the following *step function*:

(i) $f(x) \equiv 0$ for $0 \leq x < 1$;
$\equiv 1$ for $1 \leq x < 2$;
$\equiv 2$ for $2 \leq x < 3$;

and so on. Thus the technology cannot produce fractional units of output and it takes one full unit of input to produce each unit of output. It can be verified that this production function is continuous from above.
(a) Calculate the cost function $C(y,1)$ that corresponds to this production function; i.e., set the input price equal to one and try to determine the corresponding total cost function $C(y,1)$. (It will turn out that this cost function is continuous from below in y but it is not necessary to prove this).
(b) Graph both the production function $y = f(x)$ and the cost function $c = C(y,1)$.

7. Suppose that a producer's cost function is defined as follows for $y \geq 0$, $p_1 > 0$ and $p_2 > 0$:

(i) $C(y,p_1,p_2) \equiv [b_{11}p_1 + 2b_{12}(p_1p_2)^{1/2} + b_{22}p_2]y$

where the b_{ij} parameters are all positive.
(a) Show that this cost function is concave in the input prices p_1,p_2. *Note*: this is the two input case of the Generalized Leontief cost function defined by Diewert (1971).
(b) Calculate an explicit functional form for the corresponding production function $f(x_1,x_2)$ where we assume that $x_1 > 0$ and $x_2 > 0$.

4. The Derivative Property of the Cost Function

Up to this point, Theorem 2, the Shephard Duality Theorem, is of mainly academic interest: if the production function f satisfies properties (10)-(12), then the corresponding cost function C defined by (1) satisfies the properties listed in Theorem 1 above and moreover completely determines the production function. However, it is the next property of the cost function that makes duality theory so useful in applied economics.

Theorem 5: Shephard's (1953; 11) *Lemma*: If the cost function C(y,p) satisfies the properties listed in Theorem 1 above and in addition is once differentiable with respect to the components of input prices at the point (y*,p*) where y* is in the range of the production function f and $p^* \gg 0_N$, then

(21) $x^* = \nabla_p C(y^*,p^*)$

where $\nabla_p C(y^*,p^*)$ is the vector of first order partial derivatives of cost with respect to input prices, $[\partial C(y^*,p^*)/\partial p_1,..,\partial C(y^*,p^*)/\partial p_N]^T$, and x* is any solution to the cost minimization problem

(22) $\min_x \{p^{*T}x : f(x) \geq y^*\} \equiv C(y^*,p^*)$.

Under these differentiability hypotheses, it turns out that the x* solution to (22) is unique.

Proof: Let x* be any solution to the cost minimization problem (22). Since x* is feasible for the cost minimization problem when the input price vector is changed to an arbitrary $p \gg 0_N$, it follows that

(23) $p^T x^* \geq C(y^*,p)$ for every $p \gg 0_N$.

Since x* is a solution to the cost minimization problem (22) when p = p*, we must have

(24) $p^{*T}x^* = C(y^*,p^*)$.

But (23) and (24) imply that the function of N variables, $g(p) \equiv p^T x^* - C(y^*,p)$ is nonnegative for all $p \gg 0_N$ with $g(p^*) = 0$. Hence, g(p) attains a global minimum at p = p* and since g(p) is differentiable with respect to the input prices p at this point, the following first order necessary conditions for a minimum must hold at this point:

(25) $\nabla_p\, g(p^*) = x^* - \nabla_p C(y^*,p^*) = 0_N$.

Now note that (25) is equivalent to (21). If x** is any other solution to the cost minimization problem (22), then repeat the above argument to show that

(26) $x^{**} = \nabla_p C(y^*,p^*)$
$\quad\quad = x^*$

where the second equality follows using (25). Hence x** = x* and the solution to (22) is unique. Q.E.D.

The above result has the following implication: postulate a differentiable functional form for the cost function C(y,p) that satisfies the regularity conditions listed in Theorem 1 above. Then differentiating C(y,p) with respect to the components of the input price vector p generates the firm's system of cost minimizing input demand functions, $x(y,p) \equiv \nabla_p C(y,p)$.

Shephard (1953) was the first person to establish the above result starting with just a cost function satisfying the appropriate regularity conditions.[11] However, Hotelling (1932; 594) stated a version of the result in the context of profit functions and Hicks (1946; 331) and Samuelson (1953-54; 15-16) established the result starting with a differentiable utility or production function.

One application of the above result is its use as an aid in generating systems of cost minimizing input demand functions that are linear in the parameters that characterize the technology. For example, suppose that the cost function had the following *Generalized Leontief functional form*:[12]

(27) $C(y,p) \equiv \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2}\, y$; $\qquad b_{ij} = b_{ji}$ for $1 \leq i < j \leq N$

where the $N(N+1)/2$ independent b_{ij} parameters are all nonnegative. With these nonnegativity restrictions, it can be verified that the C(y,p) defined by (27) satisfies properties 1 to 7 listed in Theorem 1.[13] Applying Shephard's Lemma shows that the system of cost minimizing input demand functions that correspond to this functional form are given by:

(28) $x_i(y,p) = \partial C(y,p)/\partial p_i = \sum_{j=1}^{N} b_{ij}\, (p_j/p_i)^{1/2}\, y$; $\qquad i = 1,2,\ldots,N.$

Errors can be added to the system of equations (28) and the parameters b_{ij} can be estimated using linear regression techniques if we have time series or cross sectional data on output, inputs and input prices.[14] If all of the b_{ij} equal zero for $i \neq j$, then the demand functions become:

(29) $x_i(y,p) = \partial C(y,p)/\partial p_i = b_{ii}\, y$; $\qquad i = 1,2,\ldots,N.$

---

[11] See also Fenchel (1953; 104). We have used the technique of proof used by McKenzie (1956-57).

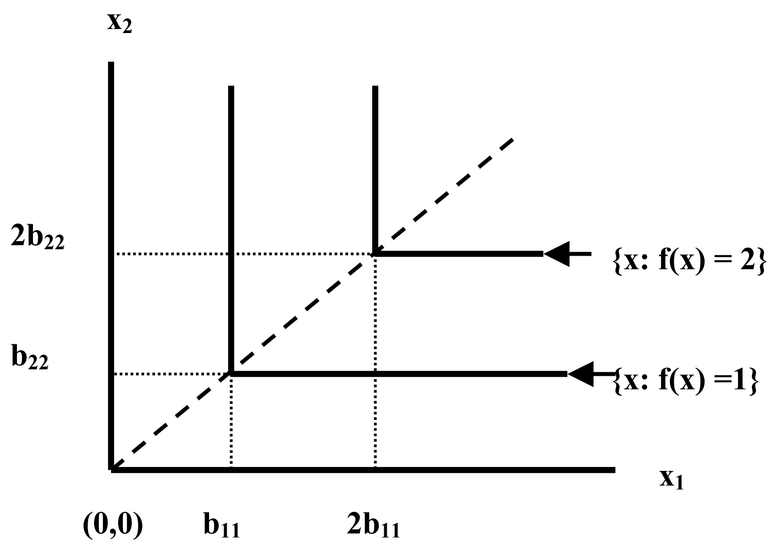
[12] See Diewert (1971).

[13] Using problem 7 above, it can be seen that if the b_{ij} are nonnegative and y is positive, then the functions $b_{ij}\, p_i^{1/2}\, p_j^{1/2}\, y$ are concave in the components of p. Hence, since a sum of concave functions is concave, it can be seen that the C(y,p) defined by (27) is concave in the components of p.

[14] Note that b_{12} will appear in the first input demand equation and in the second as well using the cross equation symmetry condition, $b_{21} = b_{12}$. There are $N(N-1)/2$ such cross equation symmetry conditions and we could test for their validity or impose them in order to save degrees of freedom. The nonnegativity restrictions that ensure global concavity of C(y,p) in p can be imposed if we replace each parameter b_{ij} by a squared parameter, $(a_{ij})^2$. However, the resulting system of estimating equations is no longer linear in the unknown parameters.

Note that input prices do not appear in the system of input demand functions defined by (29) so that input quantities do not respond to changes in the relative prices of inputs. The corresponding production function is known as the Leontief (1941) production function.[15] Hence, it can be seen that the production function that corresponds to (28) is a generalization of this production function. The unit output isoquant for the Leontief production function is graphed below in Figure 2.

Figure 2: The Two Input Leontief Production Function



5. The Comparative Statics Properties of Input Demand Functions

Before we develop the main result in this section, it will be useful to establish some results about the derivatives of a twice continuously differentiable linearly homogeneous function of N variables. We say that f(x), defined for $x >> 0_N$ is *positively homogeneous of degree* α iff f has the following property:

(30) $f(\lambda x) = \lambda^{\alpha} f(x)$ for all $x >> 0_N$ and $\lambda > 0$.

A special case of the above definition occurs when the number α in the above definition equals 1. In this case, we say that f is (positively) *linearly homogeneous*[16] iff

(31) $f(\lambda x) = \lambda f(x)$ for all $x >> 0_N$ and $\lambda > 0$.

---

[15] The Leontief production function can be defined as $f(x_1,\ldots,x_N) \equiv \min_i \{x_i/b_{ii} : i = 1,\ldots,N\}$. It is also known as the no substitution production function. Note that this production function is not differentiable even though its cost function is differentiable.

[16] Usually in economics, we omit the adjective "positively" but it is understood that the λ which appears in definitions (30) and (31) is restricted to be positive.

Theorem 6: *Euler's Theorems on Differentiable Homogeneous Functions*: Let f(x) be a (positively) linearly homogeneous function of N variables, defined for $x >> 0_N$. Part 1: If the first order partial derivatives of f exist, then the first order partial derivatives of f satisfy the following equation:

(32) $f(x) = \sum_{n=1}^{N} x_n\, \partial f(x_1,\ldots,x_N)/\partial x_n = x^T\nabla f(x)$ for all $x >> 0_N$.

Part 2: If the second order partial derivatives of f exist, then they satisfy the following equations:

(33) $\sum_{j=1}^{N} [\partial^2 f(x_1,\ldots,x_N)/\partial x_n\partial x_j]x_j = 0$ for all $x >> 0_N$ and $n = 1,\ldots,N$.

The N equations in (33) can be written using matrix notation in a much more compact form as follows:

(34) $\nabla^2 f(x)x = 0_N$ for all $x >> 0_N$.

Proof of Part 1: Let $x >> 0_N$ and $\lambda > 0$. Differentiating both sides of (31) with respect to λ leads to the following equation using the composite function chain rule:

(35) $f(x) = \sum_{n=1}^{N} [\partial f(\lambda x_1,\ldots,\lambda x_N)/\partial(\lambda x_n)][\partial(\lambda x_n)/\partial\lambda]$
$= \sum_{n=1}^{N} [\partial f(\lambda x_1,\ldots,\lambda x_N)/\partial(\lambda x_n)]x_n$.

Now evaluate (35) at $\lambda = 1$ and we obtain (32).

Proof of Part 2: Let $x >> 0_N$ and $\lambda > 0$. For $n = 1,\ldots,N$, differentiate both sides of (31) with respect to x_n and we obtain the following N equations:

(36) $f_n(\lambda x_1,\ldots,\lambda x_N)\partial(\lambda x_n)/\partial x_n = \lambda f_n(x_1,\ldots,x_N)$ for $n = 1,\ldots,N$ or
$f_n(\lambda x_1,\ldots,\lambda x_N)\lambda = \lambda f_n(x_1,\ldots,x_N)$ for $n = 1,\ldots,N$ or
$f_n(\lambda x_1,\ldots,\lambda x_N) = f_n(x_1,\ldots,x_N)$ for $n = 1,\ldots,N$

where the nth first order partial derivative function is defined as $f_n(x_1,\ldots,x_N) \equiv \partial f(x_1,\ldots,x_N)/\partial x_n$ for $n = 1,\ldots,N$.[17] Now differentiate both sides of the last set of equations in (36) with respect to λ and we obtain the following N equations:

(37) $0 = \sum_{j=1}^{N} [\partial f_n(\lambda x_1,\ldots,\lambda x_N)/\partial x_j][\partial(\lambda x_j)/\partial\lambda]$ for $n = 1,\ldots,N$
$= \sum_{j=1}^{N} [\partial f_n(\lambda x_1,\ldots,\lambda x_N)/\partial x_j]x_j$.

Now evaluate (37) at $\lambda = 1$ and we obtain the N equations (33). Q.E.D.

[17] Using definition (30) for the case where $\alpha = 0$, it can be seen that the last set of equations in (36) shows that the first order partial derivative functions of a linearly homogenous function are homogeneous of degree 0.

The above results can be applied to the cost function, C(y,p). From Theorem 1, C(y,p) is linearly homogeneous in p. Hence by part 2 of Euler's Theorem, if the second order partial derivatives of the cost function with respect to the components of the input price vector p exist, then these derivatives satisfy the following restrictions:

(38) $\nabla^2_{pp}C(y,p)p = 0_N$.

Theorem 7: Diewert (1982; 567): Suppose the cost function C(y,p) satisfies the properties listed in Theorem 1 and in addition is twice continuously differentiable with respect to the components of its input price vector at some point, (y,p). Then the system of cost minimizing input demand equations, $x(y,p) \equiv [x_1(y,p),\ldots,x_N(y,p)]^T$, exists at this point and these input demand functions are once continuously differentiable. Form the N by N matrix of input demand derivatives with respect to input prices, $B \equiv [\partial x_i(y,p)/\partial p_j]$, which has ij element equal to $\partial x_i(y,p)/\partial p_j$. Then the matrix B has the following properties:

(39) $B = B^T$ so that $\partial x_i(y,p)/\partial p_j = \partial x_j(y,p)/\partial p_i$ for all $i \neq j$;[18]
(40) B is negative semidefinite[19] and
(41) $Bp = 0_N$.[20]

Proof: Shephard's Lemma implies that the firm's system of cost minimizing input demand equations, $x(y,p) \equiv [x_1(y,p),\ldots,x_N(y,p)]^T$, exists and is equal to

(42) $x(y,p) = \nabla_p C(y,p)$.

Differentiating both sides of (42) with respect to the components of p gives us

(43) $B \equiv [\partial x_i(y,p)/\partial p_j] = \nabla^2_{pp}C(y,p)$.

Now property (39) follows from Young's Theorem in calculus. Property (40) follows from (43) and the fact that C(y,p) is concave in p and the fourth characterization of concavity. Finally, property (41) follows from the fact that the cost function is linearly homogeneous in p and hence (38) holds. Q.E.D.

Note that property (40) implies the following properties on the input demand functions:

(44) $\partial x_n(y,p)/\partial p_n \leq 0$ for $n = 1,\ldots,N$.

Property (44) means that input demand curves cannot be upward sloping.

---

[18] These are the Hicks (1946; 311) and Samuelson (1947; 69) symmetry restrictions. Hotelling (1932; 549) obtained analogues to these symmetry conditions in the profit function context.
[19] Hicks (1946; 311) and Samuelson (1947; 69) also obtained versions of this result by starting with the production (or utility) function F(x), assuming that the first order conditions for solving the cost minimization problem held and that the strong second order sufficient conditions for the primal cost minimization problem also held.
[20] Hicks (1946; 331) and Samuelson (1947; 69) also obtained this result using their primal technique.

If the cost function is also differentiable with respect to the output variable y, then we can deduce an additional property about the first order derivatives of the input demand functions. The linear homogeneity property of C(y,p) in p implies that the following equation holds for all $\lambda > 0$:

$$(45)\ C(y,\lambda p) = \lambda C(y,p) \qquad \text{for all } \lambda > 0 \text{ and } p >> 0_N.$$

Partially differentiating both sides of (45) with respect to y leads to the following equation:

$$(46)\ \partial C(y,\lambda p)/\partial y = \lambda \partial C(y,p)/\partial y \qquad \text{for all } \lambda > 0 \text{ and } p >> 0_N.$$

But (46) implies that the function $\partial C(y,p)/\partial y$ is linearly homogeneous in p and hence part 1 of Euler's Theorem applied to this function gives us the following equation:

$$(47)\ \partial C(y,p)/\partial y = \sum_{n=1}^{N} p_n \partial^2 C(y,p)/\partial y \partial p_n = p^T \nabla^2_{yp} C(y,p).$$

But using (42), it can be seen that (47) is equivalent to the following equation:[21]

$$(48)\ \partial C(y,p)/\partial y = \sum_{n=1}^{N} p_n \partial x_n(y,p)/\partial y.$$

Problems

8. For $i \neq j$, the inputs i and j are said to be *substitutes* if $\partial x_i(y,p)/\partial p_j = \partial x_j(y,p)/\partial p_i > 0$, *unrelated* if $\partial x_i(y,p)/\partial p_j = \partial x_j(y,p)/\partial p_i = 0$[22], and *complements* if $\partial x_i(y,p)/\partial p_j = \partial x_j(y,p)/\partial p_i < 0$. (a) If N = 2, show that the two inputs cannot be complements. (b) If N = 2 and $\partial x_1(y,p)/\partial p_1 = 0$, then show that all of the remaining input demand price derivatives are equal to 0; i.e., show that $\partial x_1(y,p)/\partial p_2 = \partial x_2(y,p)/\partial p_1 = \partial x_2(y,p)/\partial p_2 = 0$. (c) If N = 3, show that at most one pair of inputs can be complements.[23]

9. Let $N \geq 3$ and suppose that $\partial x_1(y,p)/\partial p_1 = 0$. Then show that $\partial x_1(y,p)/\partial p_n = 0$ as well for n = 2,3,…,N. *Hint*: You will need to use the definition of negative semidefiniteness in a strategic way. This problem shows that if the own input elasticity of demand for an input is 0, then that input is unrelated to all other inputs.

10. Recall the definition (27) of the Generalized Leontief cost function where the parameters b_{ij} were all assumed to be nonnegative. Show that under these nonnegativity restrictions, every input pair is either unrelated or substitutes. *Hint*: Simply calculate $\partial^2 C(y,p)/\partial p_i \partial p_j$ for $i \neq j$ and look at the resulting formula. *Comment*: This result shows

---

[21] This method of deriving these restrictions is due to Diewert (1982; 568) but these restrictions were originally derived by Samuelson (1949; 66) using his primal cost minimization method.
[22] Pollak (1969; 67) uses the term "unrelated" in a similar context.
[23] This result is due to Hicks (1946; 311-312): "It follows at once from Rule (5) that, while it is possible for all other goods consumed to be substitutes for x_r, it is not possible for them all to be complementary with it."

that if we impose the nonnegativity conditions $b_{ij} \geq 0$ for $i \neq j$ on this functional form in order to ensure that it is globally concave in prices, then we have a priori ruled out any form of complementarity between the inputs. This means if the number of inputs N is greater than 2, this nonnegativity restricted functional form cannot be a *flexible functional form*[24] for a cost function; i.e., it cannot attain an arbitrary pattern of demand derivatives that are consistent with microeconomic theory, since the nonnegativity restrictions rule out any form of complementarity.

11. Suppose that a producer's three input production function has the following Cobb Douglas (1928) functional form:

(a) $f(x_1,x_2,x_3) \equiv x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}$ where $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_3 > 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Let the positive input prices $p_1 > 0$, $p_2 > 0$, $p_3 > 0$ and the positive output level $y > 0$ be given. (i) Calculate the producer's cost function, $C(y,p_1,p_2,p_3)$ along with the three input demand functions, $x_1(y,p_1,p_2,p_3)$, $x_2(y,p_1,p_2,p_3)$ and $x_3(y,p_1,p_2,p_3)$. *Hint*: Use the usual Lagrangian technique for solving constrained minimization problems. You need not check the second order conditions for the problem. The positive constant $k \equiv \alpha_1^{-\alpha_1} \alpha_2^{-\alpha_2} \alpha_3^{-\alpha_3}$ will appear in the cost function.

(ii) Calculate the input one demand elasticity with respect to output $[\partial x_1(y,p_1,p_2,p_3)/\partial y][y/x_1(y,p_1,p_2,p_3)]$ and the three input one demand elasticities with respect to input prices $[\partial x_1(y,p_1,p_2,p_3)/\partial p_n][p_n/x_1(y,p_1,p_2,p_3)]$ for $n = 1,2,3$.

(iii) Show that $-1 < [\partial x_1(y,p_1,p_2,p_3)/\partial p_1][p_1/x_1(y,p_1,p_2,p_3)] < 0$.
(iv) Show that $0 < [\partial x_1(y,p_1,p_2,p_3)/\partial p_2][p_2/x_1(y,p_1,p_2,p_3)] < 1$.
(v) Show that $0 < [\partial x_1(y,p_1,p_2,p_3)/\partial p_3][p_3/x_1(y,p_1,p_2,p_3)] < 1$.

(vi) Can any pair of inputs be complementary if the technology is a three input Cobb Douglas?

Comment: The Cobb Douglas functional form is widely used in macroeconomics and in applied general equilibrium models. However, this problem shows that it is not satisfactory if $N \geq 3$. Even in the $N = 2$ case where analogues to (iii) and (iv) above hold,

[24] Diewert (1974; 115 and 133) introduced the term "flexible functional form" to describe a functional form for a cost function (or production function) that could approximate an arbitrary cost function (consistent with microeconomic theory) to the second order around any given point. The Generalized Leontief cost function defined by (27) above is flexible for the class of cost functions that are dual to linearly homogeneous production functions if we do not impose any restrictions on the parameters b_{ij}; see Diewert (1971) and section 9 below for a proof of this fact. However, if we do not impose the nonnegativity restrictions $b_{ij} \geq 0$ for $i \neq j$ on this functional form, it will frequently turn out that when these parameters are econometrically estimated, the resulting cost function fails the concavity restrictions, $\nabla^2_{pp}C(y^t,p^t)$ is negative semidefinite, at one or more points (y^t,p^t) in the observed data set that was used in the econometric estimation. Thus finding flexible functional forms where the restrictions implied by microeconomic theory can be *imposed* on the functional form without destroying its flexibility is a nontrivial task.

it can be seen that this functional form is not consistent with technologies where the degree of substitution between inputs is very high or very low.

12. Suppose that the second order partial derivatives with respect to input prices of the cost function C(y,p) exist so that the nth cost minimizing input demand function $x_n(y,p) = \partial C(y,p)/\partial p_n > 0$ exists for n = 1,…,N. Define the input n elasticity of demand with respect to input price k as follows:

(a) $e_{nk}(y,p) \equiv [\partial x_n(y,p)/\partial p_k][p_k/x_n(y,p)]$ for n = 1,..,N and k = 1,…,N.

Show that for each n, $\sum_{k=1}^{N} e_{nk}(y,p) = 0$.

13. Let the producer's cost function be C(y,p), which satisfies the regularity conditions in Theorem 1 and, in addition, is once differentiable with respect to the components of the input price vector p. Then the nth input demand function is $x_n(y,p) \equiv \partial C(y,p)/\partial p_n$ for n = 1,…,N. Input n is defined to be *normal* at the point (y,p) if $\partial x_n(y,p)/\partial y = \partial^2 C(y,p)/\partial p_n \partial y > 0$; i.e., if the cost minimizing demand for input n increases as the target output level y increases. On the other hand, input n is defined to be *inferior* at the point (y,p) if $\partial x_n(y,p)/\partial y = \partial^2 C(y,p)/\partial p_n \partial y < 0$. Prove that not all N inputs can be inferior at the point (y,p). *Hint*: Make use of (48).

14. If the production function f dual to the differentiable cost function C(y,p) exhibits *constant returns to scale* so that $f(\lambda x) = \lambda f(x)$ for all $x \geq 0_N$ and all $\lambda > 0$, then show that for each n, the input n elasticity of demand with respect to the output level y is 1; i.e., show that for n = 1,…,N, $[\partial x_n(y,p)/\partial y][y/x_n(y,p)] = 1$.

6. When is the Assumption of Constant Returns to Scale in Production Justified?

In many areas of applied economics, constant returns to scale in production is frequently assumed. In this section, we present a justification for making this assumption. The basic ideas are due to Samuelson (1967) but some of the technical details are due to Diewert (1981).

Assume that the technology of a single plant in an industry can be described by means of the production function y = F(x) where y is the maximum plant output producible by the input vector $x \equiv [x_1,…,x_N]$. We will eventually make three assumptions about the plant production function F; the first two are listed below.

Assumption 1 on F: F is *continuous from above* over the domain $x \geq 0_N$.

Assumption 2 on F: F is a *nonnegative function* with $F(0_N) = 0$ and $F(x^*) > 0$ for at least one $x^* > 0_N$.

As was shown in Theorem 1 above, Assumption 1 is sufficient to imply that the cost function C(y,p):

(49) $C(y,p) \equiv \min_x \{p^T x: F(x) \geq y\}$

is well defined for all strictly positive input price vectors $p >> 0_N$ and all output levels $y \in Y$ where Y is the smallest convex set containing the range of F.

Assumption 2 on F implies that total plant cost will be positive for positive output y and positive input price vectors p; that is:

(50) $C(y,p) > 0$ for $y \in Y$, $y > 0$ and $p >> 0_N$.

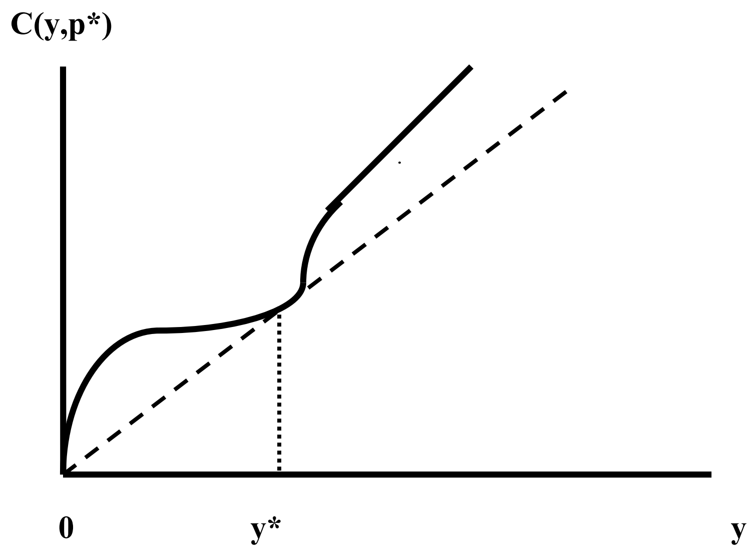
The first two assumptions on F are extremely weak. However, our next assumption, while reasonably weak, does not always hold.[25]

Assumption 3 on F: F is such that for every $p^* >> 0_N$, a solution y^* to the following *average cost minimization problem* exists:

(51) $\min_y \{C(y,p^*)/y: y > 0; y \in Y\} \equiv c(p^*)$.

Figure 3 illustrates the geometry of the average cost minimization problem (51).

Figure 3: The Geometry of Average Cost Minimization



The solid curve in Figure 3 is the graph of the cost $C(y,p^*)$ as a function of y. The slope of the dashed line in Figure 3 is equal to $C(y^*,p^*)/y^*$ and note that this line is the line through the origin that has the lowest slope and is also tangent to the graph of $C(y,p^*)$ as

[25] However, one could argue that it will always hold in real world situations. Eventually, as the target output level y becomes very large, total costs C(y,p) will increase at ever increasing rates (due to the finiteness of world resources) and Assumption 3 will be satisfied.

a function of y. Thus in this case, average plant cost is uniquely minimized at the output level y*.

However, one can construct examples of production functions F where Assumption 3 does not hold. For example, consider the following one output, one input production function F defined as follows:

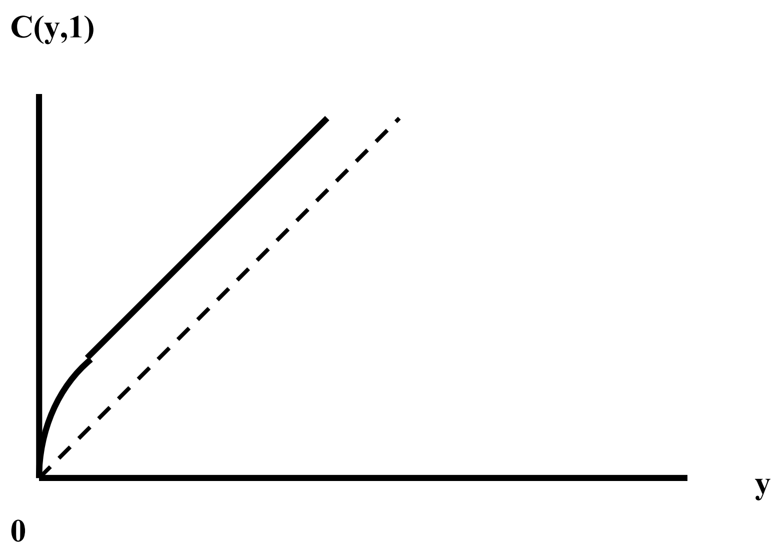
(52) $F(x) \equiv (1/2)x^2$ for $0 \le x \le 1$;
$\equiv x - (1/2)$ for $x > 1$.

If we set the single input price p_1 equal to 1, then the cost function C(y,1) that corresponds to this F is defined as follows:

(53) $C(y,1) \equiv [2y]^{1/2}$ for $0 \le y \le ½$;
$\equiv y + (1/2)$ for $y > ½$.

The cost function defined by (53) is graphed in Figure 4.

Figure 4: A Cost Function that does not Satisfy Assumption 3 on F



It can be seen that in this case, average cost is minimized only asymptotically at an infinite output level. Hence, there is no *finite* y* > 0 that minimizes average cost for this example and Assumption 3 is not satisfied for this particular production function.

Let y* solve the average cost minimization problem defined by (51) so that the minimized average cost is C(y*,p*)/y*. We can regard this minimized value as a function of the chosen input price vector p* and in (51), we have defined this function as c(p*). The following result shows that this function c has the properties of a unit cost function.

Theorem 8: Diewert (1981; 80): If F(x) satisfies Assumptions 1-3 listed above, then the minimum average cost function c(p) defined by (51) is a (i) positive, (ii) linearly homogeneous and (iii) concave function of p for $p >> 0_N$.

Proof of (i): Let $p^* >> 0_N$ and let $y^* > 0$ be a solution to (51). Then Assumptions 2 and 3 imply that $C(y^*,p^*) > 0$ and $y^* > 0$ so that $c(p^*) = C(y^*,p^*)/y^* > 0$.

Proof of (ii): Let $p^* >> 0_N$ and let $\lambda^* > 0$. Then by the definition of $c(\lambda^*p^*)$, we have:

(54) $c(\lambda^*p^*) \equiv \min_y \{C(y,\lambda^*p^*)/y: y > 0; y\in Y\}$
$= \min_y \{\lambda^*C(y, p^*)/y: y > 0; y\in Y\}$ using the linear homogeneity property of C(y,p) in p
$= \lambda^*\min_y \{C(y, p^*)/y: y > 0; y\in Y\}$ using $\lambda^* > 0$
$= \lambda^* c(p^*)$ using (51), the definition of $c(p^*)$.

Proof of (iii): Let $p^1 >> 0_N$, $p^2 >> 0_N$ and $0 < \lambda < 1$. Then using definition (51) for c, it can be seen that for every $y > 0$ such that $y\in Y$, we have:

(55) $C(y,p^1)/y \geq c(p^1)$; $C(y,p^2)/y \geq c(p^2)$.

Thus for every $y > 0$ such that $y\in Y$, we have, using the concavity in prices property of C(y,p):

(56) $C(y,\lambda p^1+(1-\lambda)p^2)/y \geq [\lambda C(y,p^1) + (1-\lambda)C(y,p^2)]/y$
$= \lambda[C(y,p^1)/y] + (1-\lambda)[C(y,p^2)]/y]$
$\geq \lambda c(p^1) + (1-\lambda)c(p^2)$ using (55) and $\lambda > 0$ and $(1-\lambda) > 0$.

Using the definition of $c(\lambda p^1+(1-\lambda)p^2)$, we have:

(57) $c(\lambda p^1+(1-\lambda)p^2) \equiv \min_y \{C(y,\lambda p^1+(1-\lambda)p^2)/y: y > 0; y\in Y\}$
$= C(y^*,\lambda p^1+(1-\lambda)p^2)/y^*$ for some $y^* > 0$; $y^*\in Y$
$\geq \lambda c(p^1) + (1-\lambda)c(p^2)$ using (56) for $y = y^*$. Q.E.D.

Since c(p) is concave over the open domain of definition set, $\{p: p >> 0_N\}$, we know that it is also continuous over this set.[26] We also know that the domain of definition of c can be extended to the nonnegative orthant, $\Omega \equiv \{p: p \geq 0_N\}$ using the Fenchel (1953; 78) closure operation and the resulting extension is continuous (and concave) over Ω.

Now we are in a position to apply the results of problem 5 above. We have just shown that the c(p) defined by (51) has all of the properties of a unit cost function and hence,

[26] We also know that the properties of concavity, linear homogeneity and positivity are sufficient to imply that c(p) is nondecreasing in the components of p.

there is a constant returns to scale production function f(x) that is dual to c(p). Using problem 5, for $x >> 0_N$, define the function g(x) as follows:

$$(58)\ g(x) \equiv \max_p \{c(p): p^Tx = 1; p \geq 0_N\}.$$

The function g may be used to define the dual to c production function f as follows:

$$(59)\ f(x) \equiv 1/g(x).$$

It can be shown[27] that if we define the unit cost function that corresponds to the production function defined by (59), c*(p) say, then this cost function coincides with the c(p) that was used in (58), which in turn was defined using the original cost function C(y,p) via definition (51); i.e., we have for each $p >> 0_N$:

$$(60)\ c^*(p) \equiv \min_x \{p^Tx: f(x) \geq 1\} = c(p) \equiv \min_y \{C(y,p)/y: y > 0; y \in Y\}.$$

As the above material is a bit abstract, we will indicate how the constant returns to scale production function f(x) can be constructed from the initially given plant production function F(x). For each positive plant output level y > 0, we can define the corresponding *upper level set* L(y) in the usual way:

$$(61)\ L(y) \equiv \{x : F(x) \geq y\}.$$

Now define the *family of scaled upper level sets* M(y) for each y > 0 with $y \in Y$ as follows:

$$(62)\ M(y) \equiv \{x/y : F(x) \geq y\}.$$

Thus to determine M(y) for a given y > 0, we find all of the input vectors x that can produce at least the output level y using the plant production function (this is the set L(y)) and then we divide all of those input vectors by the positive output level y. The continuity from above property of F implies that the level sets L(y) and the scaled level sets M(y) are all nonempty closed sets for each y such that y > 0 and $y \in Y$. Now define the input set U as the *union* of all of these sets M(y):

$$(63)\ U \equiv \cup_{y > 0;\, y \in Y} M(y).$$

Define the unit output upper level set of the constant returns to scale production function f as

$$(64)\ L \equiv \{x : f(x) \geq 1\}.$$

It turns out that the set U defined by (63) is a subset of the upper level set L defined by (64); i.e., we have

[27] See Diewert (1974; 110-112).

(65) $U \subset L$.

However, it turns out that U is in fact a sufficient statistic for L; i.e., we can perform some simple operations on U and transform it into L. We need to define a couple of set operations before we do this.

Let S be an arbitrary set in R^N. Then the *convex hull of S*, Con S, is defined as follows:

(66) $\text{Con } S \equiv \{x : x = \lambda x^1 + (1-\lambda)x^2 \,;\, x^1 \in S \,;\, x^2 \in S \text{ and } 0 < \lambda < 1\}$.

Thus the convex hull of S consists of all of the points belonging to S plus all of the line segments joining any two points belonging to S.

The *free disposal hull of S*, Fdh S, is defined as follows:

(67) $\text{Fdh } S \equiv \{x : x \geq x^* ;\, x^* \in S\}$.

Thus the free disposal hull of S consists of all of the points in S plus all points that lie above any point belonging to S.

It turns out that the *closure of the free disposal, convex hull of the set U* defined by (63) is in fact equal to the unit output upper level set L defined by (64). Thus define $U^* \equiv$ Con U, then define $U^{**} \equiv$ Fdh U^* and finally define $U^{***} \equiv$ Clo U^{**}. Then $U^{***} = L$. The boundary of the set U can have regions of nonconvexity and backward bending regions.[28] The operation of taking the convex hull of U eliminates these regions of nonconvexity and the operation of taking the free disposal convex hull eliminates any backward bending regions. Finally, since the union of an infinite number of closed sets is not necessarily closed, taking the closure of U^{**} ensures that this transformed U set is closed.

What is the significance of the unit cost function c defined by (51) or its dual constant returns to scale production function f defined by (58) and (59)? Samuelson (1967; 155-161) showed that this f represents the *asymptotic technology* which is available to the firm if firm output is large, plants can be *replicated* and the plant production function F satisfies Assumptions 1-3 listed above.

The plant replication idea of Samuelson can be explained as follows. Let $p^* \gg 0_N$ be given and let $y^* > 0$ be a solution to the average cost minimization problem (51). For each positive integer n, define the set of outputs Y(n) as follows:

(68) $Y(n) \equiv \{y: (n-1)y^* < y \leq ny^*\}$.

[28] Recall Figure 1 above where the boundary of L(y) had regions of nonconvexity and backward bending regions.

Now consider a situation where a firm has access to the plant technology production function F(x) that satisfies Assumptions 1-3 above and has the dual cost function C(y,p). Suppose that the firm wants to produce some positive output level y where y belongs to the set of outputs Y(n) defined by (68) for some positive integer n. Then in theory, the firm could build n plants and have each of them produce 1/n of the desired output level y. The firm's *average cost* of production using this plant replication strategy will be equal to:

(69) $c^n(y,p^*) \equiv C(y/n,p^*)/[y/n] \geq c(p^*)$

where the inequality in (69) follows from definition (51),which defined c(p*) as a minimum.[29]

The following result shows that as the target output level y becomes large, the average cost $c^n(y,p^*)$ using the plant replication strategy approaches the unit cost c(p*) where the unit cost function c is dual to the constant returns to scale production function f defined earlier by (58) and (59).

Theorem 9: Samuelson (1967; 159), Diewert (1981; 82): As firm output y becomes large, average cost using replicable plants $c^n(y,p^*)$ approaches the minimum average cost c(p*) defined by (51); i.e., for every $p^* \gg 0_N$,

(70) $c(p^*) = \lim_{n\to\infty} \{c^n(y,p^*): y\in Y(n)\}$.

Proof: If $y\in Y(n)$, then $(n-1)y^* < y \leq ny^*$ or

(71) $[(n-1)/n]y^* < y/n \leq y^*$ or
(72) $n/y^*(n-1) > n/y \geq 1/y^*$.

Thus for $y\in Y(n)$,

(73) $c^n(y,p^*) = C(y/n,p^*)/[y/n]$ using definition (69)
$\leq C(y^*,p^*)n/y$
using (71) and the nondecreasing in y property of cost functions
$< [n/(n-1)]C(y^*,p^*)/y^*$ using (72)
$= [n/(n-1)]\, c(p^*)$

since y* is a solution to the average cost minimization problem defined by (51). The inequalities (69) and (73) imply that

(74) $c(p^*) \leq c^n(y,p^*) < [n/(n-1)]\, c(p^*)$.

Taking limits of (74) as n tends to plus infinity gives us (70). Note that as n increases in the limit (70), y must also increase in order to remain in the set Y(n). Q.E.D.

[29] We also need $y/n\in Y$.

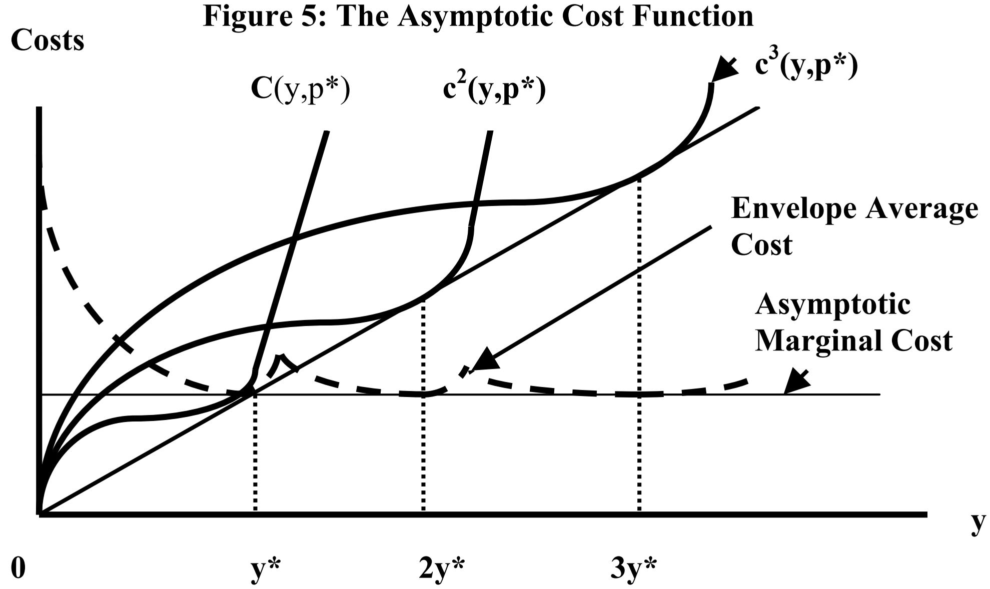
The significance of the above result can be explained as follows. If firm output is large relative to a minimum average cost output y* when input prices p* prevail, then the firm's total cost function will be approximately equal to yc(p*), where the unit cost function c(p) defined by (51) is dual to a well behaved constant returns to scale production function f(x). Thus if the firm is behaving competitively in input markets so that the firm regards input prices as fixed and beyond their control, then the firm's total demand for inputs can be generated (approximately) by solving the following cost minimization problem:

(75) $\min_x \{p^{*T}x: f(x) \geq y\} = yc(p^*)$

where the asymptotic (or large output) production function f is positive, linearly homogeneous and concave over the positive orthant,[30] even though the underlying plant production function F satisfies only Assumptions 1-3.

The geometry of Theorem 9 is illustrated in Figure 5.

Figure 5: The Asymptotic Cost Function



In Figure 5, the original plant cost function is C(y,p*) regarded as a function of y. The straight line through the origin is tangent to this curve at the minimum average cost output level y*. Note that this straight line represents the minimum average cost that is attainable for this particular plant technology. The function $c^2(y,p^*) \equiv 2C(y/2,p^*)$ can be

[30] Production functions that have these properties are often called *neoclassical* production functions. Of course, the function f can be extended in a continuous manner to the nonnegative orthant using the Fenchel closure operation.

graphed using the original cost curve C(y,p*) and so can the function $c^3(y,p^*) \equiv 3C(y/3,p^*)$. The dashed line labelled as the envelope average cost is the minimum average cost that corresponds to the three total cost curves that appear in Figure 5. Obviously, if the firm can replicate plants, this dashed line will represent the minimum average cost of producing any output level up to 3y*. Note that as the target output level y increases, this scalloped average cost line gets closer and closer to the straight line that is labelled as the asymptotic marginal cost line. Thus *if* (a) firm output y is large relative to the smallest minimum average cost output level y*, (b) the firm behaves competitively in input markets and (c) plants can be replicated without difficulties, then we can closely approximate the firm's behavior in input markets by assuming that it possesses a constant returns to scale production function f defined above by (58) and (59), rather than the nonconstant returns to scale plant production function F.

7. Aggregation and the Size of Input Price Elasticities of Demand

A great many problems in cost benefit analysis and applied economics in general hinge on the size of various *elasticities of demand or supply*. In this section, we will show that *increasing* the degree of aggregation in a production model will generally lead to elasticities of derived demand that are *smaller in magnitude* than the average of the micro elasticities of demand in the aggregate.

Consider a one output technology, y = F(z,x), that uses combinations of M + N inputs, $z \geq 0_M$ and $x \geq 0_N$, to produce output $y \geq 0$. Let the cost function that corresponds to this technology be the twice differentiable function, C(y,w,p), defined in the usual way as follows:

$$(76)\ C(y,w,p) \equiv \min_{z,x} \{w^Tz + p^Tx: F(z,x) \geq y\}$$

where $y > 0$ is the target output level and $w >> 0_M$ and $p >> 0_N$ are strictly positive input price vectors. This cost function will satisfy the regularity conditions in Theorem 1 above. The two sets of cost minimizing input demand functions, z(y,w,p) and x(y,w,p), can be obtained by using Shephard's Lemma:

$$(77)\ z(y,w,p) = \nabla_w C(y,w,p)\ ;$$
$$(78)\ x(y,w,p) = \nabla_p C(y,w,p).$$

We now introduce the assumption that the prices in the vector w *move proportionally* over time[31] (or space if we are in a cross sectional context); i.e., we assume that

$$(79)\ w = \alpha p_0\ ;\ \ \alpha \equiv [\alpha_1,\ldots,\alpha_M]^T >> 0_M.$$

[31] This is the framework used by Hicks (1946; 312-313) in his Composite Commodity Aggregation Theorem: "Thus we have demonstrated mathematically the very important principle, used extensively in the text, that if the prices of a group of goods change in the same proportion, that group of goods behaves just as if it were a single commodity."

We now construct an aggregate of the z inputs. Usually, we set one of the components of α equal to unity (e.g., set $\alpha_1 = 1$) so that the remaining α_m tell us how many units of input m are equivalent to one unit of input 1 in the z group of inputs. The "quantity" of the z aggregate, x_0, is defined in practice by deflating observed expenditure on the z inputs by the numeraire price p_0; i.e., we have

$$
\begin{aligned}
(80)\ x_0(y,w,p) &\equiv \textstyle\sum_{m=1}^M w_m z_m(y,w,p)/p_0 \\
&= \textstyle\sum_{m=1}^M p_0\alpha_m z_m(y,\alpha p_0,p)/p_0 && \text{using (79)} \\
&= \alpha^T z(y,\alpha p_0,p).
\end{aligned}
$$

To see how Hicks' Aggregation Theorem works in this context, we use the aggregation vector α, which appears in (79) in order to construct an *aggregate input requirements function*,[32] $G(y,x,\alpha)$, as follows:[33]

$$(81)\ x_0 = G(y,x,\alpha) \equiv \min_z \{\alpha^T z: F(z,x) \geq y\}.$$

The above aggregate input requirements function G can be used in order to define the following *aggregate cost function*, C*:

$$
\begin{aligned}
(82)\ C^*(y,p_0,p) &\equiv \min_{x_0,x} \{p_0 x_0 + p^T x: x_0 = G(y,x,\alpha)\} \\
&= \min_x \{p_0 G(y,x,\alpha) + p^T x\} && \text{using the constraint to eliminate } x_0 \\
&= \min_x \{p_0[\min_z \{\alpha^T z: F(z,x) \geq y\}] + p^T x\} && \text{using (81)} \\
&= \min_{x,z} \{p_0\alpha^T z + p^T x: F(z,x) \geq y\} && \text{using } p_0 > 0 \\
&= \min_{x,z} \{w^T z + p^T x: F(z,x) \geq y\} && \text{using (79)} \\
&\equiv C(y,w,p) && \text{using (76).}
\end{aligned}
$$

The string of equalities in (82) shows that if (z^*,x^*) solves the original micro cost minimization problem defined by (76), then $x_0^* \equiv \alpha^T z^*$ and x^* solve the macro cost minimization problem defined by the first line in (82). Thus if the z input prices vary in strict proportion over time, then these inputs can be aggregated using the construction in (80), and the resulting x_0 aggregate will obey the usual properties of an input demand function that is consistent with cost minimizing behavior. This is a version of Hicks' (1946; 312-313) Aggregation Theorem.

We now want to explore the relationship of the price elasticities of demand for the aggregate input compared to the underlying micro cross elasticities of demand.

The microeconomic matrices of *input price elasticities of demand* can be defined as follows:[34]

---

[32] An input requirements function, $x_0 = g(y,x)$, gives the minimum amount of an input x_0 that is required to produce the output level y given that the vector of other inputs x is available for the production function constraint $y = f(x_0,x)$. Thus g is a (conditional on x) inverse function for y regarded as a function of x_0, holding x constant. Input requirements functions were studied by Diewert (1974c).

[33] If there is no $z \geq 0_M$ such that $F(z,x) \geq y$, then $G(x,y,\alpha)$ is defined to equal plus infinity.

[34] All of these elasticities are evaluated at an initial point (y,w,p).

$$(83)\ E_{xp} \equiv [e^{nk}] \qquad n = 1,\ldots,N\ ;\ k = 1,\ldots,N$$
$$\equiv [(p_k/x_n)\partial x_n(y,w,p)/\partial p_k]$$
$$= [(p_k/x_n)\partial^2 C(y,w,p)/\partial p_n\partial p_k] \qquad \text{using (78)}$$
$$= \hat{x}^{-1}\nabla^2_{pp}C(y,w,p)\,\hat{p}$$

where $\hat{x}$ and $\hat{p}$ denote N by N diagonal matrices with the positive elements of the x and p vectors running down the main diagonal respectively;

$$(84)\ E_{zw} \equiv [e_{mk}] \qquad m = 1,\ldots,M\ ;\ k = 1,\ldots,M$$
$$\equiv [(w_k/z_m)\partial z_m(y,w,p)/\partial w_k]$$
$$= [(w_k/z_m)\partial^2 C(y,w,p)/\partial w_m\partial w_k] \qquad \text{using (77)}$$
$$= \hat{z}^{-1}\nabla^2_{ww}C(y,w,p)\,\hat{w}$$

where $\hat{z}$ and $\hat{w}$ denote M by M diagonal matrices with the positive elements of the z and w vectors running down the main diagonal respectively;

$$(85)\ E_{zp} \equiv [e_m{}^n] \qquad m = 1,\ldots,M\ ;\ n = 1,\ldots,N$$
$$\equiv [(p_n/z_m)\partial z_m(y,w,p)/\partial p_n]$$
$$= [(p_n/z_m)\partial^2 C(y,w,p)/\partial w_m\partial p_n] \qquad \text{using (77)}$$
$$= \hat{z}^{-1}\nabla^2_{wp}C(y,w,p)\,\hat{p}\,.$$

Using the linear homogeneity property of the cost function C(y,w,p) in the components of (w,p), it can be shown that the elements matrices E_{zw} and E_{zp} satisfy the following restrictions:[35]

$$(86)\ E_{zw}1_M + E_{zp}1_N = \hat{z}^{-1}\nabla^2_{ww}C(y,w,p)\,\hat{w}\,1_M + \hat{z}^{-1}\nabla^2_{wp}C(y,w,p)\,\hat{p}\,1_N$$
using (84) and (85)
$$= \hat{z}^{-1}\{\nabla^2_{ww}C(y,w,p)w + \nabla^2_{wp}C(y,w,p)p\}$$
$$= \hat{z}^{-1}\{0_M\}$$
using Part 2 of Euler's Theorem on homogeneous functions
$$= 0_M.$$

Thus the elements in each row of the M by M+N elasticity matrix $[E_{zw},E_{zp}]$ sum to zero.

Now we are ready to calculate the price elasticity of demand of the input aggregate x_0 with respect to its own price p_0. We first differentiate the last equation in (80) with respect to p_0:

$$(87)\ \partial x_0(y,\alpha p_0,p)/\partial p_0 = \textstyle\sum_{m=1}^{M} \alpha_m \sum_{k=1}^{M} [\partial z_m(y,\alpha p_0,p)/\partial w_k][\partial(\alpha_k p_0)/\partial p_0]$$
$$= \textstyle\sum_{m=1}^{M} \alpha_m \sum_{k=1}^{M} [\partial z_m(y,\alpha p_0,p)/\partial w_k]\alpha_k$$
$$= \textstyle\sum_{m=1}^{M} \sum_{k=1}^{M} \alpha_m[\partial^2 C(y,\alpha p_0,p)/\partial w_m\partial w_k]\alpha_k \qquad \text{using (77)}$$

[35] Notation: 1_M and 1_N are vectors of ones of dimension M and N respectively.

$$= \sum_{m=1}^{M} \sum_{k=1}^{M} \alpha_m[\partial^2 C(y,w,p)/\partial w_m \partial w_k]\alpha_k \qquad \text{using (79)}$$
$$= \alpha^T \nabla^2_{ww} C(y,w,p)\alpha$$
$$\le 0$$

where the last inequality follows from the negative semidefiniteness of $\nabla^2_{ww}C(y,w.p)$, which in turn follows from the fact that C(y,w,p) is concave in the components of w. Thus the own price elasticity of demand for the input aggregate will be negative or 0.

Now calculate the price elasticity of demand of the input aggregate x_0 with respect to the prices p_n for n = 1,…,N. We first differentiate the last equation in (80) with respect to p_n:

$$(88)\ \partial x_0(y,\alpha p_0,p)/\partial p_n = \sum_{m=1}^{M} \alpha_m\ \partial z_m(y,\alpha p_0,p)/\partial p_n \qquad \text{for } n = 1,\ldots,N$$
$$= \sum_{m=1}^{M} \alpha_m \partial^2 C(y,w,p)/\partial w_m \partial p_n \qquad \text{using (77)}$$
$$= \alpha^T \nabla^2_{wp} C(y,w,p)e_n$$

where e_n is an N dimensional unit vector; i.e., it has all elements equal to 0 except that the nth component is equal to 1.

We are ready to convert the derivatives defined by (87) and (88) into the *own and cross price elasticities of demand for the aggregate*, ε_{00} and ε_{0n} for n = 1,…,N:

$$(89)\ \varepsilon_{00} \equiv [p_0/x_0][\partial x_0(y,\alpha p_0,p)/\partial p_0]$$
$$= [p_0/x_0][\alpha^T \nabla^2_{ww} C(y,w,p)\alpha] \qquad \text{using (87)}$$
$$= p_0\alpha^T \nabla^2_{ww} C(y,w,p)p_0\alpha/p_0x_0$$
$$= w^T \nabla^2_{ww} C(y,w,p)w/p_0x_0 \qquad \text{using (79)}$$
$$= w^T \nabla^2_{ww} C(y,w,p)w/ \textstyle\sum_{m=1}^{M} w_m z_m(y,w,p) \qquad \text{using (80)}$$
$$= w^T \hat{z}\ \hat{z}^{-1} \nabla^2_{ww} C(y,w,p)w/w^T z \qquad \text{since } \hat{z}\ \hat{z}^{-1} = I_M$$
$$= w^T \hat{z}\ \hat{z}^{-1} \nabla^2_{ww} C(y,w,p)\ \hat{w}\ 1_M/w^T z \qquad \text{since } w = \hat{w}\ 1_M$$
$$= w^T \hat{z}\ E_{zw} 1_M/w^T z \qquad \text{using (84)}$$
$$= s^T\ E_{zw} 1_M$$
$$\le 0$$

where the vector of expenditure shares on the components of z is defined as $s^T \equiv [s_1,\ldots,s_M]$ where

$$(90)\ s_m \equiv w_m z_m(y,w,p)/w^T z(y,w,p) \qquad \text{for } m = 1,\ldots,M.$$

The inequality in (89) follows from (87), $\partial x_0(y,\alpha p_0,p)/\partial p_0 \le 0$, and the positivity of p_0 and x_0.

Converting the derivatives in (88) into elasticities leads to the following equations:

$$(91)\ \varepsilon_{0n} \equiv [p_n/x_0][\partial x_0(y,\alpha p_0,p)/\partial p_n] \qquad \text{for } n = 1,\ldots,N$$
$$= [p_n/x_0][\alpha^T \nabla^2_{wp} C(y,w,p)e_n] \qquad \text{using (88)}$$
$$= p_0\alpha^T \nabla^2_{wp} C(y,w,p)p_n e_n/p_0x_0$$

$= w^T\nabla^2_{wp}C(y,w,p)p_ne_n/p_0x_0$ using (79)
$= w^T\nabla^2_{wp}C(y,w,p)p_ne_n/ \sum_{m=1}^M w_mz_m(y,w,p)$ using (80)
$= w^T \hat{z}\ \hat{z}^{-1}\nabla^2_{wp}C(y,w,p)p_ne_n /w^Tz$ since $\hat{z}\ \hat{z}^{-1} = I_M$
$= w^T \hat{z}\ \hat{z}^{-1}\nabla^2_{wp}C(y,w,p)\ \hat{p}\ e_n /w^Tz$ since $p_ne_n = \hat{p}\ e_n$
$= w^T \hat{z}\ E_{zp}e_n/w^Tz$ using (85)
$= s^T E_{zp}e_n$.

Using the above formulae for the ε_{0n}, we can compute the sum of the ε_{0n} as follows:

(92) $\sum_{n=1}^N \varepsilon_{0n} = \sum_{n=1}^N s^T E_{zp}e_n$
$= s^TE_{zp}1_N$.

Now use (89) and (92) in order to compute the sum of all of the price elasticities of demand of the input aggregate with respect to its own price p_0 as well as the other input prices outside of the aggregate, $p_1,\ldots,p_N$:

(93) $\varepsilon_{00} + \sum_{n=1}^N \varepsilon_{0n} = s^TE_{zw}1_M + s^TE_{zp}1_N$
$= s^T\{E_{zw}1_M + E_{zp}1_N\}$
$= s^T\ \{0_M\}$ using (86)
$= 0.$

Using (89), we have $\varepsilon_{00} \leq 0$. Hence (93) implies that

(94) $\sum_{n=1}^N \varepsilon_{0n} = -\ \varepsilon_{00} \geq 0.$

Consider the cross elasticity of demand of the input aggregate with the input price p_n for some $n = 1,\ldots,N$. Using (91), we have

(95) $\varepsilon_{0n} = \sum_{m=1}^M s_me_m^n \leq \sum_{m=1}^M s_m\,|e_m^n|$ for $n = 1,\ldots,N$

where the inequality follows since the shares s_m are always positive and the cross elasticity of demand for the micro input z_m with respect to the price p_n, e_m^n, is always equal to or less than its absolute value, $|e_m^n|$. Hence if *any* of the inputs m in the group of inputs being aggregated (the z_m) are *complementary* to the input x_n, then the corresponding e_m^n will be negative and the inequality (95) will be *strict* for that n.

In general, it can be seen that the cross elasticities of aggregate input demand ε_{0n} are weighted averages of the micro cross elasticities of demand e_m^n. If all of these micro cross elasticities of demand are nonnegative (so that there are no complementary input pairs between the z and x groups of inputs), then it can be seen that the aggregate cross elasticities of demand ε_{0n} will be weighted averages of the e_m^n and will be roughly comparable in magnitude to the average magnitude of the e_m^n. However, *the greater the degree of complementarity between the z and x groups of inputs, the greater will be the reduction in the magnitudes of the* ε_{0n} *compared to the magnitudes of the* e_m^n, which are the absolute values $|e_m^n|$.[36] How likely is complementarity in empirical applications? Most empirical applications of production theory *impose* substitutability between every pair of inputs and so it is frequently thought that complementarity is a somewhat rare phenomenon. However, if flexible functional form techniques are used, then typically, if the number of inputs is greater than 3, complementarity is encountered.[37]

Finally, we compare the magnitude of the own price elasticity of demand of the input aggregate, ε_{00}, with the weighted average of the micro own price elasticities of demand in the aggregate, $\sum_{m=1}^M s_m e_{mm}$. Using (89), we have $\varepsilon_{00} \leq 0$ and the micro own price elasticities of demand, e_{mm}, are also nonpositive.[38] Hence

$$
\begin{aligned}
(96)\ 0 &\leq -\varepsilon_{00} \\
&= -\sum_{m=1}^M \sum_{k=1}^M s_m e_{mk} && \text{using (89)} \\
&= -\sum_{m=1}^M s_m e_{mm} - \sum_{m=1}^M \sum_{k=1\ m\neq k}^M s_m e_{mk} \\
&< -\sum_{m=1}^M s_m e_{mm}
\end{aligned}
$$

where the inequality follows provided that

$$(97)\ \sum_{m=1}^M \sum_{k=1\ m\neq k}^M s_m e_{mk} > 0.$$

The strict inequality in (96) says that the magnitude (or absolute value) of ε_{00} is *smaller* than the magnitude of the weighted average of the micro own price elasticities of demand in the input aggregate, $\sum_{m=1}^M s_m e_{mm}$. However the strict inequality in (96) will hold only if the strict inequality in (97) holds. We cannot guarantee that (97) will hold but it is very likely that it will hold. In particular, (97) will hold if all of the input pairs in the z group of inputs are *substitutes or are unrelated* so that in this case, $e_{mk} \geq 0$ for all $m \neq k$.[39]

We can summarize the above results as follows: if we estimate price elasticities of input demand for an aggregated model and compare the resulting elasticities with the elasticities obtained from the more disaggregated model, *there will be a strong tendency for the elasticities in the aggregated model to be smaller in magnitude than those in the disaggregated model*.[40]

---

[36] This point was made by Diewert (1974b; 16) many years ago in the elasticity of substitution context: "Taking a weighted average of both positive and negative micro elasticities of substitution σ_m^n will tend to give rise to aggregate elasticities of substitution which are considerably *smaller* in magnitude than an average of the absolute values of the micro elasticities of substitution."

[37] For example, see the 4 input evidence on the incidence of complementarity tabled in Diewert and Wales (1987; 63).

[38] This follows from the fourth characterization of concavity and the fact that C(y,w,p) is concave in w.

[39] Strictly speaking, we need at least one input pair to be substitutes so that $e_{mk} > 0$ for this pair of inputs and the other input pairs could be unrelated.

[40] For the 4 input models estimated in Diewert and Wales (1987; 63), the input price elasticities were all less than 1.2 in magnitude. For the 8 output and input model estimated by Diewert and Wales (1992; 716-717), all of the tabled price elasticities were less than 3.43 in magnitude. For the 12 output and input model estimated by Diewert and Lawrence (2002; 154), all of the tabled price elasticities (excluding inventory change which was very volatile) were less than 8.99 in magnitude.

In actual empirical examples, the strict proportionality assumptions made in (79) will not hold exactly and the aggregates will be constructed using an index number formula that will be consistent with the assumptions (79) if they happen to hold.[41] However, even if the assumption of price proportionality (79) holds only approximately, empirical evidence suggests that elasticities do decline in magnitude as the degree of aggregation increases. We conclude with a quotation that summarizes some early empirical evidence on this phenomenon:[42]

"Conversely, as we disaggregate, we can expect to encounter increasingly large elasticities of substitution. Two recent papers confirm this statement. Berndt and Christensen (1974) in their 'two types of labour, one type of capital" disaggregation of US manufacturing industries found that the mean partial elasticities of substitution were 7.88 (between blue and white collar workers), 3.72 (between blue collar workers and capital) and −3.77 (between white collar workers and capital). However, when they fitted a model which aggregated the two types of labour into a single labour factor, they found that the aggregate labour-capital elasticity of substitution was approximately 1.42, which is considerably smaller that an average of the three 'micro' elasticities of substitution. Similarly, Woodland (1972) found partial elasticities of substitution in Canadian manufacturing ranging from −11.16 to 2.18 in his 'four types of capital, one type of labour' disaggregated results, but he found that the aggregate capital-labour elasticity of substitution was only 0.39." W.E. Diewert (1974b; 16).

Problems

15. The N by M matrix of cross elasticities of demand of the x inputs with respect to the prices of the z inputs can be defined as follows:

$$
\begin{aligned}
\text{(i) } E_{xw} &\equiv [e^n{}_m] && n = 1,\ldots,N\ ;\ m = 1,\ldots,M \\
&\equiv [(w_m/x_n)\partial x_n(y,w,p)/\partial w_m] \\
&= [(w_m/x_n)\partial^2 C(y,w,p)/\partial p_n \partial w_m] && \text{using (78)} \\
&= \hat{x}^{-1}\nabla^2_{pw}C(y,w,p)\,\hat{w}\,.
\end{aligned}
$$

Show that the matrices of elasticities E_{xp} defined by (83) and E_{xw} defined by (i) above satisfy the following restriction:

(ii) $E_{xp}1_N + E_{xw}1_M = 0_N$.

16. Suppose that the N+M by N+M symmetric matrix C is negative semidefinite. Write the matrix C in partitioned form as follows:

(i) $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{bmatrix}$

[41] In fact, it is useful to aggregate commodities whose prices move almost proportionally over time since the resulting aggregates will be approximately consistent with Hicks' Aggregation Theorem.

[42] Part (b) of problem 17 below shows that elasticities of substitution are equal to price elasticities of demand divided by cost shares and hence will be considerably larger than price elasticities of demand so that the numbers in the quotation below are a bit misleading.

where C_{11} is N by N and symmetric and C_{22} is M by M and symmetric. Show that C_{11} and C_{22} are also negative semidefinite matrices.

17. Let C(y,p) be a twice continuously differentiable cost function that satisfies the regularity conditions listed in Theorem 1 in section 2 above. By Shephard's Lemma, the input demand functions are given by

(i) $x_n(y,p) = \partial C(y,p)/\partial p_n > 0$; $n = 1,\ldots,N$.

The Allen (1938; 504) Uzawa (1962) *elasticity of substitution* σ_{nk} between inputs n and k is defined as follows:

(ii) $\sigma_{nk}(y,p) \equiv \{C(y,p)\partial^2C(y,p)/\partial p_n\partial p_k\}/\{[\partial C(y,p)/\partial p_n][\partial C(y,p)/\partial p_k]\}$ $\quad 1 \le n,k \le N$
$= \{C(y,p)\partial^2C(y,p)/\partial p_n\partial p_k\}/x_n(y,p)\, x_k(y,p)$ $\quad$ using (i).

Define $\Sigma \equiv [\sigma_{nk}(y,p)]$ as the N by N matrix of elasticities of substitution.
(a) Show that Σ has the following properties:

(iii) $\Sigma = \Sigma^T$;
(iv) Σ is negative semidefinite and
(v) $\Sigma s = 0_N$

where $s \equiv [s_1,\ldots,s_N]^T$ is the vector of cost shares; i.e., $s_n \equiv p_n\, x_n(y,p)/C(y,p)$ for $n = 1,\ldots,N$. Now define the N by N matrix of cross price elasticities of demand E in a manner analogous to definition (83) above:

(vi) $E \equiv [e^{nk}]$ $\quad n = 1,\ldots,N$; $k = 1,\ldots,N$
$\equiv [(p_k/x_n)\partial x_n(y,p)/\partial p_k]$
$= [(p_k/x_n)\partial^2C(y,p)/\partial p_n\partial p_k]$ $\quad$ using (i)
$= \hat{x}^{-1}\nabla^2_{pp}C(y,p)\,\hat{p}$.

(b) Show that $E = \Sigma\hat{s}$ where $\hat{s}$ is an N by N diagonal matrix with the elements of the share vector s running down the main diagonal.

18. Suppose a firm's cost function has the following Constant Elasticity of Substitution (CES) functional form:[43]

(i) $C(y,p_1,\ldots,p_N) \equiv ky[\sum_{n=1}^N \alpha_n p_n^r]^{1/r}$; $\quad k > 0$; $r \le 1$, $r \ne 0$; $\alpha_n > 0$ and $\sum_{n=1}^N \alpha_n = 1$.

Thus the cost function is equal to a positive constant k times the output level y times a mean of order r. From the chapter on inequalities, we know that C(y,p) is a concave function of p provided that r is equal to or less than one. Show that

[43] This functional form was introduced into the production literature by Arrow, Chenery, Minhas and Solow ((1961).

(ii) $\sigma_{nk}(y,p) = -(r-1)$ for all n,k such that $n \neq k$

where $\sigma_{nk}(y,p)$ is the elasticity of substitution between inputs n and k defined above in problem 17, part (ii). *Comment*: This problem shows why the CES functional form is unsatisfactory if the number of inputs N exceeds two, since it is a priori unlikely that all elasticities of substitution between every pair of inputs would equal the same number.

8. The Application of Cost Functions to Consumer Theory

The cost function and production function framework described in the previous sections can be readily adapted to the consumer context: simply replace output y by utility u, reinterpret the production function F as a utility function, reinterpret the input vector x as a vector of commodity demands and reinterpret the vector of input prices p as a vector of commodity prices. With these changes, the producer's cost minimization problem (1) becomes the following problem of *minimizing the cost or expenditure of attaining a given level of utility u*:

(98) $C(u,p) \equiv \min_x \{p^T x : F(x) \geq u\}$.

If the cost function is differentiable with respect to the components of the commodity price vector p, then Shephard's (1953; 11) Lemma applies and the consumer's system of commodity demand functions as functions of the chosen utility level u and the commodity price vector p, x(u,p), is equal to the vector of first order partial derivatives of the cost or expenditure function C(u,p) with respect to the components of p:

(99) $x(u,p) = \nabla_p C(u,p)$.

The demand functions $x_n(u,p)$ defined in (99) are known as *Hicksian*[44] *demand functions*.

Thus it seems that we can adapt the theory of cost and production functions used in section 2 above in a very straightforward way, replacing output y by utility u and reinterpreting all of our previous results. However, there is a problem: the output level y is an *observable* variable but the corresponding utility level u is *not observable*!

However, this problem can be solved (but as we will see, some of the details are rather complex). We need only equate the cost function C(u,p) to the consumer's *observable expenditure* in the period under consideration, Y say, and solve the resulting equation for u as a function of Y and p, say $u = g(Y,p)$. Thus $u = g(Y,p)$ is the u solution to the following equation:

(100) $C(u,p) = Y$

[44] See Hicks (1946; 311-331).

and the resulting solution function u = g(Y,p) is the *consumer's indirect utility function*. Now replace the u in the system of Hicksian demand functions (99) by g(Y,p) and we obtain the consumer's system of (observable) *market demand functions*:

(101) $d(Y,p) = \nabla_p C(g(Y,p),p)$.

We illustrate the above procedure for the generalized Leontief cost function defined by (27) above. For this functional form, equation (100) becomes:

(102) $u \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2} = Y$; $\qquad$ ($b_{ij} = b_{ji}$ for all i and j)

and the u solution to this equation is:

(103) $u = g(Y,p) = Y/[\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2}]$.

The Hicksian demand functions for the C(u,p) defined by the left hand side of (102) are:

(104) $x_n(u,p) \equiv \partial C(u,p)/\partial p_n = [\sum_{j=1}^{N} b_{nj}\, (p_j/p_n)^{1/2}]u$; $\qquad$ n = 1,…,N.

Substituting (103) into (104) leads to the following system of market demand functions:

(105) $d_n(Y,p) = [\sum_{j=1}^{N} b_{nj}\, (p_j/p_i)^{1/2}]\; Y/[\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2}]$; $\qquad$ n = 1,...,N.

Equations (105) can be used as the basis for the econometric estimation of preferences. Suppose that we have collected data on the quantities x_n^t purchased over T time periods for a household as well as the corresponding commodity prices p_n^t. Then we can define period t "income"[45] or expenditure on the n commodities as Y^t:

(106) $Y^t \equiv \sum_{n=1}^{N} p_n^t x_n^t$; $\qquad$ t = 1,…,T.

Evaluating (105) at the period t data and adding a stochastic error term e_n^t to equation n in (105) for n = 1,...,N leads to the following system of estimating equations:

(107) $x_n^t = [\sum_{j=1}^{N} b_{nj}\, (p_j^t/p_n^t)^{1/2}]\; Y^t/[\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, (p_i^t)^{1/2}\, (p_j^t)^{1/2}] + e_n^t$; t = 1,...,T; n = 1,...,N.

Not all N equations in (107) can have independent error terms since if we multiply both sides of equation n in (107) by p_n^t and sum over n, we obtain the following equation:

(108) $\sum_{n=1}^{N} p_n^t x_n^t = Y^t + \sum_{n=1}^{N} p_n^t e_n^t$.

Using (106), we find that the period t errors e_n^t satisfy the following linear restriction exactly:

---

[45] Strictly speaking, a household's income will also include savings in addition to expenditures on current goods and services.

(109) $\sum_{n=1}^{N} p_n^t e_n^t = 0$; $t = 1,\ldots,T$.

There is one other factor that must be taken into account in doing an econometric estimation of preferences using the system of estimation equations (107). Note that if we multiply all of the b_{ij} parameters by the positive number λ, the right hand sides of each equation in (107) will remain unchanged; i.e., the demand functions are homogeneous of degree 0 in the b_{ij} parameters. Thus these parameters *will not be identified* as matters stand. Hence, it will be necessary to impose a normalization on these parameters. One normalization that is frequently used in applied economics is to set unit cost[46] equal to 1 for some set of reference prices p^0 say.[47] Thus we impose the following normalization on the b_{ij}:

(110) $1 = c(p^0) = \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij} (p_i^0)^{1/2} (p_j^0)^{1/2}$.

Equation (110) can be used to solve for say b_{11} in terms of the other b_{ij} and then this equation can be used to eliminate b_{11} from the $N-1$ independent estimating equations in (107) and the remaining parameters can be estimated using nonlinear regression techniques.

The technique suggested above for the econometric estimation of preferences is a special case of the following general strategy: (i) Assume that the consumer's preferences can be represented by the cost function $C(u,p)$ that has the following form:

(111) $C(u,p) = uc(p)$

where $c(p)$ is a suitable differentiable unit cost function. (ii) Differentiate (111) with respect to the components of the commodity price vector p to form the following system of Hicksian demand functions:

(112) $x(u,p) = u\, \nabla_p c(p)$.

(iii) Equate cost $uc(p)$ to expenditure or income Y and solve for u as a function of Y and p to get the consumer's indirect utility function $u = g(y,p)$:

(113) $u = Y/c(p)$.

(iv) Substitute (113) into the right hand side of (112) in order to obtain the following system of consumer demand functions:

(114) $d(Y,p) = \nabla_p c(p)Y/c(p)$.

(v) Finally, impose the normalization (110), $c(p^0) = 1$, in order to identify all of the unknown parameters in (114).

---

[46] The Generalized Leontief cost function defined by (27) has the form $C(u,p) = uc(p)$ where $c(p) \equiv C(1,p)$.

[47] Usually, p^0 is taken to be p^1, the vector of prices that prevailed in the base period.

Unfortunately, there is a problem with the above strategy for estimating a consumer's preferences. The problem is the same one that occurred in problem 14 above; with the nth consumer demand function, $d_n(Y,p)$ defined by the nth equation in (114), we find that *all income elasticities of demand are equal to one*; i.e., we have:

(115) $[Y/d_n(Y,p)]\partial d_n(Y,p)/\partial Y = 1$; $n = 1,\ldots,N$.

But (115) contradicts *Engel's Law*, which says that the income elasticity of demand for food is less than one.

In the following two sections, we show how this problem of unitary income elasticities can be solved.

9. Flexible Functional Forms and Nonunitary Income Elasticities of Demand

We first define what it means for a unit cost function, c(p), to be a *flexible functional form*. Let c*(p) be an arbitrary unit cost function that satisfies the appropriate regularity conditions on unit cost functions and in addition, is twice continuously differentiable around a point $p^* \gg 0_N$. Then we say that a given unit cost function c(p) that is also twice continuously differentiable around the point p* is *flexible* if it has enough free parameters so that the following $1 + N + N^2$ equations can be satisfied:

(116) $c(p^*) = c^*(p^*)$;
(117) $\nabla c(p^*) = \nabla c^*(p^*)$;
(118) $\nabla^2 c(p^*) = \nabla^2 c^*(p^*)$.

Thus c(p) is a flexible functional form if it has enough free parameters to provide a second order Taylor series approximation to an arbitrary unit cost function.

At first glance, it looks like c(p) will have to have at least $1 + N + N^2$ independent parameters in order to be able to satisfy all of the equations (116)-(118). However, since both c and c* are assumed to be twice continuously differentiable, Young's Theorem in calculus implies that $\partial^2 c(p^*)/\partial p_i \partial p_j = \partial^2 c(p^*)/\partial p_j \partial p_i$ for all $i \neq j$ (and of course, the same equations hold for the second order partial derivatives of c*(p) when evaluated at p = p*). Thus the N^2 equations in (118) can be replaced with the following N(N+1)/2 equations:

(119) $\partial^2 c(p^*)/\partial p_i \partial p_j = \partial^2 c^*(p^*)/\partial p_j \partial p_i$ for $1 \leq i \leq j \leq N$.

Another property that both unit cost functions must have is homogeneity of degree one in the components of p. By part 1 of Euler's Theorem on homogeneous functions, c and c* satisfy the following equations:

(120) $c(p^*) = p^{*T}\nabla c(p^*)$ and $c^*(p^*) = p^{*T}\nabla c^*(p^*)$.

Thus if c and c* satisfy equations (117), then using (120), we see that c and c* automatically satisfy equation (116). By part 2 of Euler's Theorem on homogeneous functions, c and c* satisfy the following equations:

(121) $\nabla^2 c(p^*)p^* = 0_N$ and $\nabla^2 c^*(p^*)p^* = 0_N$.

This means that if we have $\partial^2 c(p^*)/\partial p_i\partial p_j = \partial^2 c^*(p^*)/\partial p_i\partial p_j$ for all $i \neq j$, then equations (121) will imply that $\partial^2 c(p^*)/\partial p_j\partial p_j = \partial^2 c^*(p^*)/\partial p_j\partial p_j$ as well, for $j = 1,\ldots,N$.

Summarizing the above material, if c(p) is linearly homogeneous, then in order for it to be flexible, c(p) needs to have only enough parameters so that the N equations in (117) can be satisfied and so that the following N(N−1)/2 equations can be satisfied:

(122) $\partial^2 c(p^*)/\partial p_j\partial p_j = \partial^2 c^*(p^*)/\partial p_i\partial p_j \equiv c_{ij}^*$ for $1 \leq i < j \leq N$.

Thus in order to be flexible, c(p) must have at least $N + N(N-1)/2 = N(N+1)/2$ independent parameters.

Now consider the Generalized Leontief unit cost function defined as follows:[48]

(123) $c(p) \equiv \sum_{i=1}^N \sum_{j=1}^N b_{ij}\, p_i^{1/2}\, p_j^{1/2}$; $\qquad b_{ij} = b_{ji}$ for all i and j.

Note that there are exactly N(N+1)/2 independent b_{ij} parameters in the c(p) defined by (123). For this functional form, the N equations in (117) become:

(124) $\partial c(p^*)/\partial p_n = \sum_{j=1}^N b_{nj}\, (p_j^*/p_n^*)^{1/2} = \partial c^*(p^*)/\partial p_n \equiv c_n^*$; $\qquad n = 1,\ldots,N$.

The N(N−1)/2 equations in (122) become:

(125) $(1/2)b_{ij}/(p_i^*p_j^*)^{1/2} = c_{ij}^*$; $\qquad 1 \leq i < j \leq N$.

However, it is easy to solve equations (125) for the b_{ij}:

(126) $b_{ij} = 2c_{ij}^*(p_i^*p_j^*)^{1/2}$; $\qquad 1 \leq i < j \leq N$.

Once the b_{ij} for $i < j$ have been determined using (126), we set $b_{ji} = b_{ij}$ for $i < j$ and finally the b_{ii} are determined using the N equations in (124).

The above material shows how we can find a flexible functional form for a unit cost function[49]. We now turn our attention to finding a flexible functional form for a general cost function C(u,p). Let C*(u,p) be an arbitrary cost function that satisfies the appropriate regularity conditions on cost functions listed in Theorem 1 above and in

[48] We no longer restrict the b_{ij} to be nonnegative.
[49] This material can be adapted to the case where we want a flexible functional form for a linearly homogeneous utility or production function f(x): just replace p by x and c(p) by f(x).

addition, is twice continuously differentiable around a point (u*,p*) where u* > 0 and p* >> 0_N. Then we say that a given cost function C(u,p) that is also twice continuously differentiable around the point (u*,p*) is *flexible* if it has enough free parameters so that the following $1 + (N+1) + (N+1)^2$ equations can be satisfied:

(127) $C(u^*,p^*) = C^*(u^*,p^*)$; (1 equation)
(128) $\nabla_p C(u^*,p^*) = \nabla_p C^*(u^*,p^*)$; (N equations)
(129) $\nabla^2_{pp} C(u^*,p^*) = \nabla^2_{pp} C^*(u^*,p^*)$; (N^2 equations)
(130) $\nabla_u C(u^*,p^*) = \nabla_u C^*(u^*,p^*)$; (1 equation)
(131) $\nabla^2_{pu} C(u^*,p^*) = \nabla^2_{pu} C^*(u^*,p^*)$; (N equations)
(132) $\nabla^2_{up} C(u^*,p^*) = \nabla^2_{up} C^*(u^*,p^*)$; (N equations)
(133) $\nabla^2_{uu} C(u^*,p^*) = \nabla^2_{uu} C^*(u^*,p^*)$ (1 equation).

Equations (127)-(129) are the counterparts to our earlier unit cost equations (116)-(118). As was the case with unit cost functions, equation (127) is implied by the linear homogeneity in prices of the cost functions and Part 1 of Euler's Theorem on homogeneous functions. Young's Theorem on the symmetry of cross partial derivatives means that the lower triangle of equations in (129) is implied by the equalities in the upper triangle of both matrices of partial derivatives. Part 2 of Euler's Theorem on homogeneous functions implies that if all the off diagonal elements in both matrices in (129) are equal, then so are the diagonal elements. Hence, in order to satisfy all of the equations in (127)-(129), we need only satisfy the N equations in (128) and the N(N−1)/2 in the upper triangle of equations (129). Young's Theorem implies that if equations (131) are satisfied, then so are equations (132). However, Euler's Theorem on homogeneous functions implies that

(134) $\partial C(u^*,p^*)/\partial u = p^{*T}\nabla^2_{pu} C(u^*,p^*) = p^{*T}\nabla^2_{pu} C^*(u^*,p^*) = \partial C^*(u^*,p^*)/\partial u$.

Hence, if equations (131) are satisfied, then so is the single equation (130). Putting this all together, we see that in order for C to be flexible, we need enough free parameters in C so that the following equations can be satisfied:

- Equations (128); N equations;
- The upper triangle in equations (129); N(N−1)/2 equations;
- Equations (131); N equations; and
- Equation (133); 1 equation.

Hence, in order for C to be a flexible functional form, it will require a minimum of 2N + N(N−1)/2 +1 = N(N+1)/2 + N +1 parameters. Thus a fully flexible cost function, C(u,p), will require N + 1 additional parameters compared to a flexible unit cost function, c(p).

Suppose the unit cost function is the Generalized Leontief unit cost function c(p) defined by (123) above. We now show how terms can be added to it in order to make it a fully flexible cost function. Thus define C(u,p) as follows:

(135) $C(u,p) \equiv uc(p) + b^T p + (1/2)a_0\alpha^T p u^2$

where $b^T \equiv [b_1,\ldots,b_N]$ is an N dimensional vector of new parameters, a_0 is a new parameter and $\alpha^T \equiv [\alpha_1,\ldots,\alpha_N] > 0_N$ is a vector of predetermined parameters.[50] Using (135) as our C, equations (128), (129), (131) and (133) become:

$$(136)\quad u^*\nabla_p c(p^*) + b + (1/2)a_0\alpha u^{*2} = \nabla_p C^*(u^*,p^*) ;$$
$$(137)\quad u^*\nabla^2_{pp} c(p^*) = \nabla^2_{pp} C^*(u^*,p^*) ;$$
$$(138)\quad \nabla_p c(p^*) + a_0\alpha u^* = \nabla^2_{pu} C^*(u^*,p^*) ;$$
$$(139)\quad a_0\alpha^T p^* = \nabla^2_{uu} C^*(u^*,p^*) .$$

Use equations (137) in order to determine the b_{ij} for $i \neq j$. Use (139) in order to determine the single parameter a_0. Use equations (138) in order to determine the b_{ii}. Finally, use equations (136) in order to determine the parameters b_n in the b vector. Thus the cost function C(u,p) defined by (135), which uses the Generalized Leontief unit cost function c(p) defined by (123) as a building block, is a parsimonious flexible functional form for a general cost function.

In fact, it is not necessary to use the Generalized Leontief unit cost function in (135) in order to convert a flexible functional form for a unit cost function into a flexible functional form for a general cost function. Let c(p) be any flexible functional form for a unit cost function and define C(u,p) by (135). Use equation (139) to determine the parameter a_0. Once a_0 has been determined, equations (137) and (138) can be used to determine the parameters in the unit cost function c(p).[51] Finally, equations (136) can be used to determine the parameters in the vector b.

Obviously, the material in this section can be applied to the problems involved in estimating a flexible cost function in the production context: simply replace utility u by output y and reinterpret the commodity price vector p as an input price vector. Differentiating (135) leads to the following system of estimating equations, where x(y,p) $= \nabla_p C(y,p)$ is the producer's system of cost minimizing input demand functions:

$$(140)\quad x(y,p) = y\nabla c(p) + b + (1/2)a_0\alpha y^2.$$

In order to obtain estimating equations for the general cost function defined by (135), there are some normalization issues that need to be discussed. We do this in the following section.

10. Money Metric Utility Scaling and Other Methods of Cardinalizing Utility

[50] The parameter a_0 could be set equal to 1 and the vector of parameters α could be estimated econometrically. We have defined the cost function C in this manner so that it has the *minimal* number of parameters required in order to be a flexible functional form. Thus it is a *parsimonious* flexible functional form.

[51] It can be seen that equations (137) and (138) have the same structure as equations (117) and (118). Hence if c(p) has enough free parameters to satisfy (117) and (118), then it has enough free parameters to satisfy (137) and (138) once a_0 has been determined.

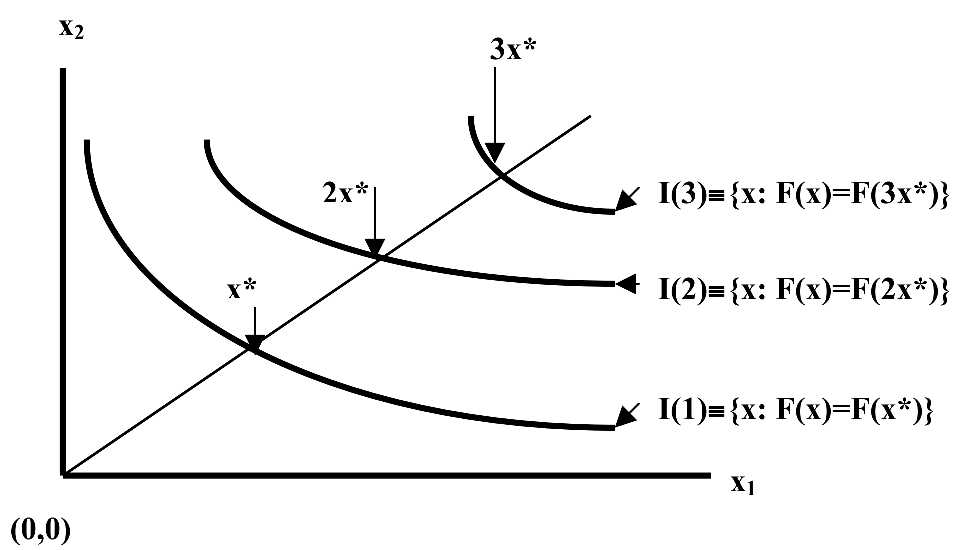
Since utility is unobservable, in order to estimate econometrically a consumer's utility function, it will be necessary to pick a utility scale for that consumer; i.e., it will be necessary to *cardinalize* the consumer's utility function.[52]

There are two commonly used methods that have been used to pick a cardinal utility scale for a consumer. The first method is used when we are working with the consumer's direct utility function, F(x) say. We simply pick a strictly positive *reference consumption vector*, $x^* >> 0_N$ say, set the utility of this vector equal to some positive number $F(x^*)$ and scale the level of utility along the ray through the point x as follows:[53]

(141) $F(\lambda x^*) = \lambda F(x^*)$; $\lambda \geq 0$.

Thus all consumption vectors $x \geq 0_N$ such that they yield the same utility as x^* are assigned the utility level $F(x^*)$; this is the indifference curve or surface $\{x: F(x) = F(x^*)\}$. Then all consumption vectors x that are on the same indifference surface as $2x^*$ are given the utility level $2F(x^*)$; this is the indifference surface $\{x: F(x) = F(2x^*) = 2F(x^*)\}$, and so on. Thus the ray through the origin and the reference consumption vector x^* is used to scale utility levels. Figure 6 illustrates how this cardinalization method works.

Figure 6: Scaling Utility by a Reference Ray through the Origin



[52] If preferences can be represented by the utility function u = F(x), then they can be equally well represented by the utility function g{F(x)} where g(u) is a monotonically increasing function of one variable.
[53] This is the type of utility scaling recommended by Blackorby (1975) and other welfare economists because this form of scaling does not depend on prices.

Of course, different choices of the reference consumption vector x* will lead to different cardinalizations of the consumer's utility function. Usually, x* will be chosen to be the observed consumption vector of a consumer in some base period or situation.

We turn now to a second method of utility scaling that was referred to as *money metric utility scaling* by Samuelson (1974; 1262). For this method of utility scaling, we choose a reference set of prices, say $p^* >> 0_N$, and if these reference prices face the consumer, we normalize the consumer's cost function, C(u,p), so that the following restriction holds:

(142) $C(u,p^*) = u$ for all $u > 0$.

Typically, we choose p* to be the prices facing the consumer in some base period situation when the consumer spends the "income" Y* on the N commodities and has utility level u* so that equating expenditure to income in this base period, we have

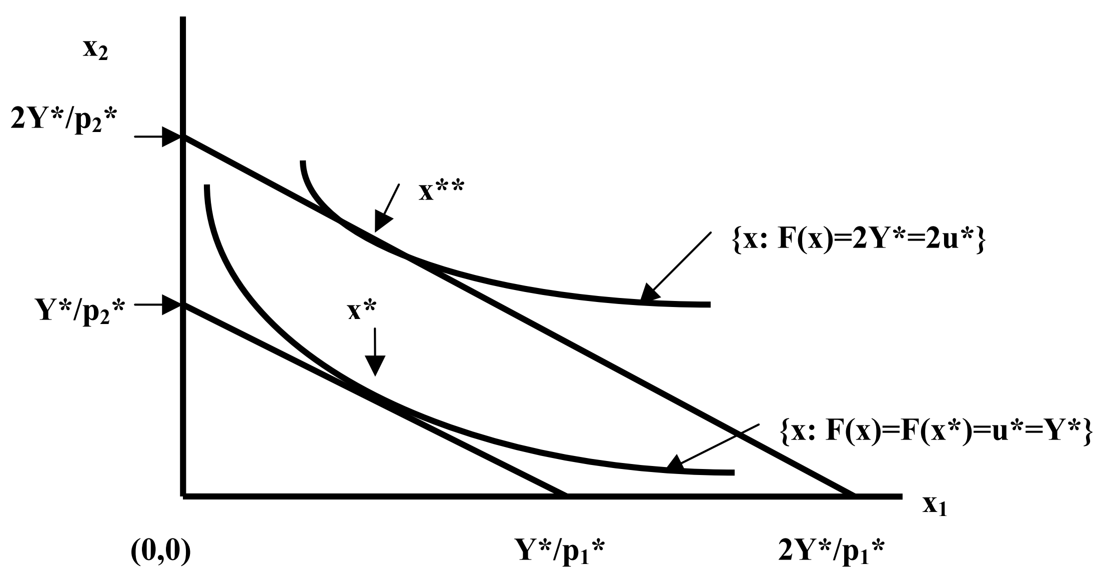
(143) $C(u^*,p^*) = Y^*$.

Combining (142) and (143), we see that for this base period situation, we have

(144) $u^* = Y^*$;

i.e., utility equals expenditure in this base period. Thus the money metric utility scaling convention (142) has the effect of making nominal "income" Y equal to utility u provided that the consumer is facing the reference prices p*. The geometry of this scaling method is illustrated in Figure 7.

Figure 7: Money Metric Utility Scaling



In Figure 7, x* solves the cost minimization problem, $\min_x \{p^{*T}x: F(x) = u^*\} \equiv C(u^*,p^*)$, which in turn is equal to observed expenditure, Y*. We scale utility so that u* is set equal to Y* in this base period situation. Thus all x combinations that yield the utility level u* = Y* are assigned this utility level Y*. This is the set $\{x: F(x) = F(x^*)\}$ that is labelled in Figure 7. Now double the initial utility level to 2u* = 2Y* and solve the cost minimization problem $\min_x \{p^{*T}x: F(x) = 2u^*\} \equiv C(2u^*,p^*)$, which in turn is equal to twice the initial expenditure, 2Y*. The solution to this cost minimization problem is x** in Figure 7. All points x on the corresponding indifference curve, $\{x: F(x) = F(x^{**}) = 2u^*\}$, are assigned the utility level 2u*, which in turn is equal to 2Y*.

In general, money metric utility scaling works as follows. For each positive "income" level Y > 0, define the budget set B(Y) as follows:

(145) $B(Y) \equiv \{x: p^{*T}x = Y; x \geq 0_N\}$.

For each Y greater than zero, an indifference surface will be tangent to the budget set B(Y).[54] All points on this indifference surface are assigned the utility level Y.

Money metric utility scaling suffers from the same disadvantage that ray scaling had; i.e., different choices of the reference vector of consumer prices p* will give rise to different utility scales. However, both money metric and ray scaling are acceptable methods of scaling utility; neither method of scaling can be contradicted by observable data on a consumer.

The money metric utility scaling assumption (142) implies additional restrictions on the derivatives of the cost function. Differentiating both sides of (142) with respect to u gives us the following equation:

(146) $\partial C(u,p^*)/\partial u = 1$ for all u > 0.

Differentiating (146) with respect to u again leads to the following equation:

(147) $\partial^2 C(u,p^*)/\partial u^2 = 0$ for all u > 0.

Euler's Theorem on homogeneous functions and (146) imply the following additional restriction on the second order partial derivatives of the cost function:

(148) $p^{*T}\nabla^2_{pu}C(u,p^*) = \partial C(u,p^*)/\partial u = 1$ for all u > 0.

We shall use money metric utility scaling quite frequently in this course.

The restrictions (143) and (146)-(148) imposed by money metric utility scaling have an impact on our earlier discussion of flexible functional forms for the cost function, C(u,p). Since empirically, it is harmless to impose money metric utility scaling, we can impose

[54] If the cost function is differentiable, the tangent indifference surface is $\{x: F(x) = F[\nabla_p C(Y,p^*)]\}$.

money metric scaling on both C(u,p) and C*(u,p) at the point (u*,p*) when we are attempting to find a flexible cost function C(u,p). This means that equations (130) and (133) become:

(149) $\nabla_u C(u^*,p^*) = \nabla_u C^*(u^*,p^*) = 1;$ (1 equation)
(150) $\nabla^2_{uu} C(u^*,p^*) = \nabla^2_{uu} C^*(u^*,p^*) = 0;$ (1 equation).

Thus if money metric utility scaling using the reference prices p* is imposed on both C and C*, equations (149) and (150) will be satisfied automatically. This reduces the number of free parameters that are required for C(u,p) to be a flexible functional form by 2 compared to our earlier discussion. The restriction (143) reduces the required number of parameters by 1 as well. Hence, in order for C to be a flexible functional form under these money metric utility scaling assumption, C will require a minimum of $N(N+1)/2 + N - 2$ parameters. Recall the Generalized Leontief cost function C(u,p) defined earlier by (135). Under our new money metric utility scaling assumptions, it is evident that we can set the parameter a_0 equal to 0. Thus define C(u,p) as follows:

(151) $C(u,p) \equiv uc(p) + b^T p$

where c(p) is defined by (123). The system of equations that we now have to satisfy in order for C defined by (151) to be flexible is the following one:

(152) $u^*\nabla_p c(p^*) + b = \nabla_p C^*(u^*,p^*)$;
(153) $u^*\nabla^2_{pp} c(p^*) = \nabla^2_{pp} C^*(u^*,p^*)$;
(154) $\nabla_p c(p^*) = \nabla^2_{pu} C^*(u^*,p^*)$;

Use equations (153) in order to determine the b_{ij} for $i \neq j$. Use equations (154) in order to determine the b_{ii}. Finally, use equations (152) in order to determine the parameters b_n in the b vector. However, using equation (148) applied to both C and C* means that c(p*) satisfies the following restriction:

(155) $c(p^*) = p^{*T}\nabla_p c(p^*) = p^{*T}\nabla^2_{pu} C^*(u^*,p^*) = 1.$

Thus c(p*) = 1, which means that we can impose a restriction on the b_{ij} such as:

(156) $b_{11} = \{1-[\sum_{n=2}^N b_{nn}p_n^* + \sum_{i=1}^N\sum_{j=1}^N {}_{i\neq j}\, b_{ij}\,(p_i^*p_j^*)^{1/2}]\}/p_1^*$.

Recall the normalization (110) in section 8 above, which was similar to (156) except that the reference prices p^0 were used in place of the reference prices p*. Now premultiply both sides of (152) by p^{*T} in order to obtain the following equation:

(157) $u^*c(p^*) + p^{*T}b = C^*(u^*,p^*)$ using Euler's Theorem
$= u^*$ using (142).

Using (155), namely that c(p*) = 1, we find that equation (157) becomes $u^* + p^{*T}b = u^*$ or

(158) $p^{*T}b = 0$.

Hence we can impose a restriction on the components of b such as

(159) $b_1 = -\sum_{n=2}^{N} p_n^* b_n / p_1^*$

without destroying the flexibility of the functional form defined by (151). Thus there are N(N+1)/2 − 1 independent b_{ij} parameters in the Generalized Leontief unit cost function c(p) defined by (123) and N−1 independent b_n parameters in the b vector, which is just the right number for (151) to be a parsimonious flexible functional form for a cost function in the context of money metric utility scaling.

In fact, it is not necessary to use the Generalized Leontief unit cost function in (151).[55] Let c(p) be any flexible functional form for a unit cost function and define C(u,p) by (151). Use equations (153) and (154) to determine the parameters in the unit cost function c(p). Then use equations (152) to determine b. Finally, repeat the arguments around equations (155), (157) and (158) to show that c(p) and b satisfy the additional restrictions $c(p^*) = 1$ and $p^{*T}b = 0$.

The reader will note that our suggested flexible functional form for C(u,p) defined by (151) reduces to uc(p) if the parameters b_n in the b vector all turn out to be zero. If c(p) is a flexible functional form for a unit cost function, then when $b = 0_N$, our general flexible functional form C(u,p) can model homothetic (or linearly homogeneous) preferences in a flexible manner. Put another way, we found a flexible functional form for a general cost function, C(u,p), by simply adding an extra parameter vector b to a cost function that was flexible for homothetic preferences, namely uc(p). The indirect utility function that corresponds to the cost function C(u,p) defined by (151) can be obtained by setting the right hand side of (151) to Y and then solving the resulting equation for u = g(Y,p), which results in the following formula for g:

(160) $u = g(Y,p) = [Y - b^Tp]/c(p)$.

In order for utility to be positive (and meaningful in this model), we require that the consumer's income Y be greater than or equal to *committed expenditures*, b^Tp.[56] The system of Hicksian demand functions that corresponds to (151) is:

(161) $x(u,p) = b + \nabla_p c(p)u$.

Substituting (160) into (161) leads to the following system of market demand functions:

(162) $d(Y,p) \equiv x[g(Y,p),p] = b + \nabla_p c(p)[Y - b^Tp]/c(p)$.

[55] This general argument is due to Diewert (1980; 597).

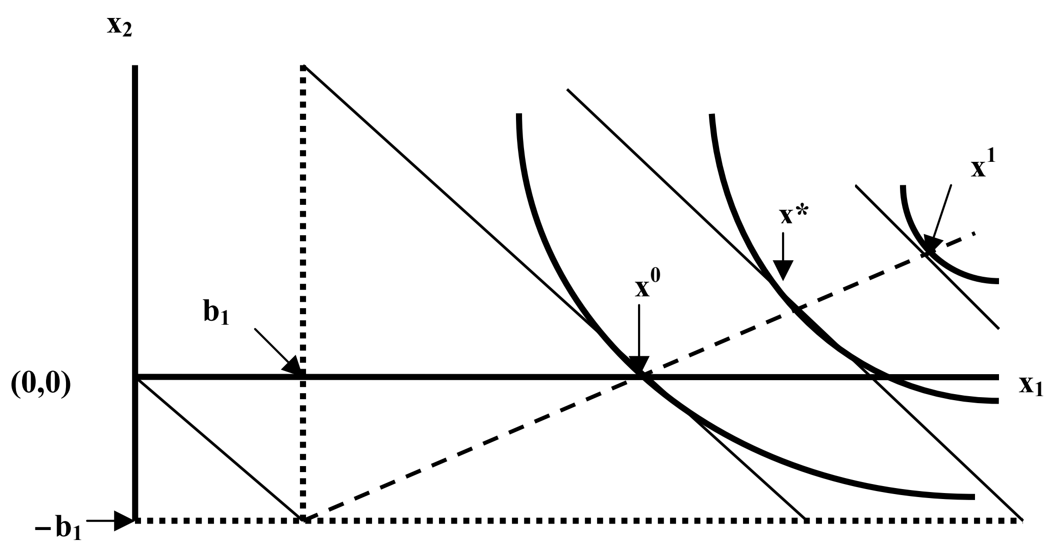
[56] In the case where one or more components of b are negative, we will require income to be large enough so that the demands d_n defined by (162) are nonnegative. Thus we require Y to be large enough so that $b + \nabla_p c(p)[Y - b^Tp]/c(p) \geq 0_N$.

To see how the geometry of this method of adding the vector of committed expenditures b to a homothetic preferences cost function works, consider the case N = 2 and let the vector of reference prices $[p_1^*,p_2^*]$ be [1,1]. In this case, the constraint (158) implies that

(163) $b_2 = -b_1$.

In Figure 8 below, we assumed that b_1 is positive so that $b_2 = -b_1$ is negative. We drew a quadrant in Figure 8 with an origin at the point $b = [b_1,b_2] = [b_1,-b_1]$. Now fill in this quadrant with the family of indifference curves that are dual to the unit cost function c(p). Three of these indifference curves are graphed in Figure 8 that pass through the x points, x^0, x^* (our point of approximation that has utility level u^*) and x^1.

Figure 8: The Addition of b to a Flexible Unit Cost Function



There are also 4 parallel budget lines that correspond to income Y = 0 (this is the line that passes through (0,0) and (b_1,b_2)), $Y = Y^0$ (the corresponding consumer demand vector $x^0 = [x_1^0,x_2^0]$ has its x_2 component equal to 0, i.e., $x_2^0 = 0$),[57] $Y = Y^* \equiv p^{*T}x^*$ where $x^* = \nabla_p C(u^*,p^*)$ is the consumer demand vector at the point of approximation for the flexible functional form, and $Y = Y^1 = p^{*T}x^1$ where $Y^1 > Y^*$. Note that the points x^0, x^* and x^1 lie on the dashed line that starts at the point b. Draw a straight line starting at the origin (0,0) and passing through the point x^*. It can be seen that this straight line is below the dashed line to the right of x^*; this means that the income elasticity of demand for commodity 1 is less than one while the income elasticity of demand for commodity 2 is greater than one. Thus it can be seen how this model can be consistent with arbitrary income elasticities of demand around a point of approximation, x^*.

[57] In order to obtain nonnegative demands, we require that income Y be equal to or greater than Y^0.

Problems

19. Instead of estimating preferences using dual methods, it is possible to estimate primal utility functions directly. Suppose that $x^t >> 0_N$ solves the consumer's period t utility maximization problem:

(i) $\max_x \{F(x): p^{tT}x = Y^t; x \geq 0_N\}$

where $p^t >> 0_N$, $Y^t > 0$ and F(x) is the consumer's differentiable utility function. (a) Show that the consumer's period t normalized price vector, p^t/Y^t, satisfies the following system of equations (Hotelling (1935; 71), Wold (1944; 69-71) (1953; 145)):

(ii) $p^t/Y^t = \nabla F(x^t)/x^{tT}\nabla F(x^t)$.

Hint: Set up the Lagrangian for the constrained maximization problem and look at the resulting first order necessary conditions for an interior solution. Eliminate the Lagrange multiplier from these N+1 equations. The remaining N equations can be rewritten in the form (ii).

(b) Now assume that the utility function is g{F(x)}, where g(u) is a monotonic once differentiable function of one variable with $g'(u) > 0$ for all $u > 0$. Find the counterparts to equations (ii) above.

(c) Find a flexible functional form for F(x) in the class of functions with no restrictions on F. You do not have to formally prove its flexibility; just exhibit what you think might be a flexible functional form. What system of equations does F have to satisfy in order to be flexible at a point $x^* >> 0_N$? Thus what is the minimal number of parameters that F must have in order to be flexible?

(d) Substitute this candidate flexible functional form into the system of econometric estimating equations defined above by (ii). Can all of the parameters of F be identified?

(e) The material in this section treats commodity demand vectors x^t as the dependent variables in a system of econometric estimating equations while income Y^t and the commodity price vector p^t are regarded as independent variables. The parameters of a cost function are estimated using this framework. However, the material in this problem treats the commodity price vectors deflated by income or expenditure, p^t/Y^t, as the dependent variables and the commodity demand vectors x^t as the independent variables. The parameters of a utility function are estimated in this framework. Which framework is preferable in applied work?

20. Assume that the twice continuously differentiable utility function F(x) satisfies the ray scaling assumption, (141). (a) Show that the utility function and its derivatives must satisfy the following 2 restrictions:

(i) $x^{*T}\nabla F(x^*) = F(x^*)$;

(ii) $x^{*T}\nabla^2F(x^*)x^* = 0$.

Hint: define the functions $g(\lambda) \equiv F(\lambda x^*)$ and $h(\lambda) \equiv \lambda F(x^*)$. Since (141) holds, $g(\lambda) = h(\lambda)$ for $\lambda \geq 0$. Now differentiate g and h with respect to λ once and then again and set $\lambda = 1$.

(b) In view of part (a) of this problem, how many independent parameters must F(x) have in order to be a parsimonious flexible functional form at the point x* where F and F* both satisfy the ray scaling assumption (141)?

(c) Use the flexible functional form for F that you suggested in problem 19 above and impose ray scaling on it. What additional restrictions on the parameters does ray scaling imply on your suggested functional form?

11. Variable Profit Functions

Up to now, we have only considered technologies that produce one output. In reality, firms (and industries) usually produce many outputs. Hence, in this section, we consider technologies that produce many outputs while using many inputs.

Let S denote the technology set of a firm. We decompose the inputs and outputs of the firm into two sets of commodities: variable and fixed. Let $y \equiv [y_1,\ldots,y_M]$ denote a vector of *variable net outputs* (if $y_m > 0$, then commodity m is an output while if $y_m < 0$, then commodity m is an input) and let $x \equiv [x_1,\ldots,x_N]$ denote a nonnegative vector of *"fixed" inputs*[58]. Thus the technology set S is a set of feasible variable net output and fixed input vectors, (x,y).

Let $p >> 0_M$ be a strictly positive vector of variable net output prices that the firm faces during a production period. Then conditional on a given vector of fixed inputs x, we assume that the firm attempts to solve the following *variable profit maximization problem*:

(164) $\max_y \{p^Ty: (y,x)\in S\} \equiv \pi(p,x)$.

Some regularity conditions on the technology set S are required in order to ensure that the maximum in (164) exists. A simple set of sufficient conditions are:[59]

(165) S is a closed set in R^{M+N};

[58] These "fixed' inputs may only be fixed in the short run.

[59] Let $x \geq 0_N$. Then by (166), there exists y_x such that $(y_x,x)\in S$. Define the closed and bounded set $B(x,p) \equiv \{y: y \leq b(x)1_M;\ p^Ty \geq p^Ty_x\}$. It can be seen that the constraint $(y,x)\in S$ in (164) can be replaced by the constraint $(y,x)\in S\cap B(x,p)$. Using (165), $S\cap B(x,p)$ is a closed and bounded set so that the maximum in (164) will exist.

(166) for each $x \geq 0_N$, the set of y such that $(y,x)\in S$ is not empty and is bounded from above; i.e., for each $x \geq 0_N$ and y such that $(y,x)\in S$, there exists a number b(x) such that $y \leq b(x)1_M$.

Condition (166) means that for each vector of fixed inputs, $x \geq 0_N$, the amount of each variable net output that can be produced by the technology is bounded from above, which is not a restrictive condition.

Note that (164) serves to define the firm's *variable profit function*,[60] $\pi(p,x)$; i.e., $\pi(p,x)$ is equal to the optimized objective function in (164) and is regarded as a function of the net output prices for variable commodities that the firm faces, p, as well as a function of the vector of fixed inputs, x, that the firm has at its disposal. Just as in section 2 above where we showed that the cost function C(y,p) satisfied a number of regularity conditions without assuming much about the production function, we can now show that the profit function $\pi(p,x)$ satisfies some regularity conditions without assuming much about the technology set S.

Theorem 10: McFadden (1966) (1978), Gorman (1968), Diewert (1973): Suppose the technology set S satisfies (165) and (166). Then the variable profit function $\pi(p,x)$ defined by (164) has the following properties with respect to p for each $x \geq 0_N$:

Property 1: $\pi(p,x)$ is *positively linearly homogeneous in p* for each fixed $x \geq 0_N$; i.e.,

(167) $\pi(\lambda p,x) = \lambda\pi(p,x)$ for all $\lambda > 0$, $p >> 0_N$ and $x \geq 0_N$.

Property 2: $\pi(p,x)$ is a *convex function of p* for each $x \geq 0_N$; i.e.,

(168) $x \geq 0_M$, $p^1 >> 0_M$; $p^2 >> 0_N$; $0 < \lambda < 1$ implies
$$\pi(\lambda p^1+(1-\lambda)p^2,x) \leq \lambda\pi(p^1,x) + (1-\lambda)\pi(p^2,x).$$

Problem

21. Prove Theorem 10. *Hint*: Properties 1 and 2 above for $\pi(p,x)$ are analogues to Properties 2 and 4 for the cost function C(y,p) in Theorem 1 above and can be proven in the same manner.

We now ask whether a knowledge of the profit function $\pi(p,x)$ is sufficient to determine the underlying technology set S. As was the case in section 3 above, the answer to this question is *yes*, but with some qualifications.

---

[60] This concept is due to Hicks (1946; 319) and Samuelson (1954-54), who determined many of its properties using primal optimization techniques. For more general approaches to this function using duality theory, see Gorman (1968), McFadden (1966) (1978) and Diewert (1973). McFadden used the term "conditional profit function" while Diewert used the term "variable profit function".

To see how to use a given profit function satisfying the 2 regularity conditions listed in Theorem 10 to determine the technology set that generated it, pick an arbitrary vector of fixed inputs $x \geq 0_N$ and an arbitrary vector of positive prices, $p^1 >> 0_M$. Now use the given profit function π to define the following isoprofit surface: $\{y: p^{1T}y = \pi(p^1,x)\}$. This isoprofit surface must be tangent to the set of net output combinations y that are feasible, given that the vector of fixed inputs x is available to the firm, which is the conditional on x production possibilities set, $S(x) \equiv \{x: (y,x)\in S\}$. It can be seen that this isoprofit surface and the set lying below it must contain the set S(x); i.e., the following *halfspace* $M(x,p^1)$, contains S(x):

$$(169)\ M(x,p^1) \equiv \{y: p^{1T}y \leq \pi(p^1,x)\}.$$

Pick another positive vector of prices, $p^2 >> 0_M$ and it can be seen, repeating the above argument, that the halfspace $M(x,p^2) \equiv \{y: p^{2T}y \leq \pi(p^2,x)\}$ must also contain the conditional on x production possibilities set S(x). Thus S(x) must belong to the intersection of the two halfspaces $M(x,p^1)$ and $M(x,p^2)$. Continuing to argue along these lines, it can be seen that S(x) must be contained in the following set, which is the intersection over all $p >> 0_M$ of all of the supporting halfspaces to S(x):

$$(170)\ M(x) \equiv \bigcap_{p>>0_M} M(x,p).$$

Note that M(x) is defined using just the given profit function, $\pi(p,x)$. Note also that since each of the sets in the intersection, M(x,p), is a convex set, then M(x) is also a convex set. Since S(x) is a subset of each M(x,p), it must be the case that S(x) is also a subset of M(x); i.e., we have

$$(171)\ S(x) \subset M(x).$$

Is it the case that S(x) is equal to M(x)? In general, the answer is *no*; M(x) forms an *outer approximation* to the true conditional production possibilities set S(x). To see why this is, see Figure 9 below. The boundary of the set M(x) partly coincides with the boundary of S(x) but it encloses a bigger set: the backward bending parts of the true production frontier are replaced by the dashed lines that are parallel to the y_1 axis and the y_2 axis and the inward bending part of the true production frontier is replaced by the dashed line that is tangent to the two regions where the boundary of M(x) coincides with the boundary of S(x). However, if the producer is a price taker in the two output markets, then it can be seen that *we will never observe the producer's nonconvex or backward bending parts of the production frontier*.

Figure 9 illustrated the case where the two variable commodities were both outputs. Figure 10 illustrates the one variable output, one variable input geometry that corresponds to (171). In Figure 10, y_1 is the variable input and y_2 is the variable output. Again, the boundary of the set M(x) partly coincides with the boundary of S(x) but it encloses a bigger set: the downward bending part of the true production frontier is replaced by the dashed line that is parallel to the y_1 axis and the nonconvex part of the true production frontier is replaced by the dashed line that is tangent to the two regions where the boundary of M(x) coincides with the boundary of S(x). Again, if the producer is a price taker in the two variable markets, then it can be seen that *we will never observe the producer's nonconvex or downward bending parts of the production frontier*.

Figure 9: The Geometry of the Two Output Maximization Problem

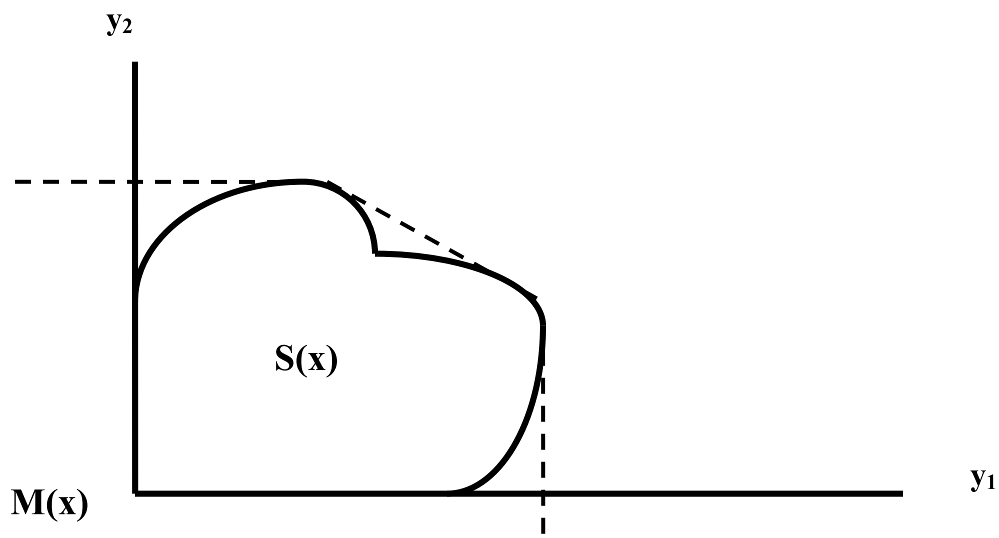
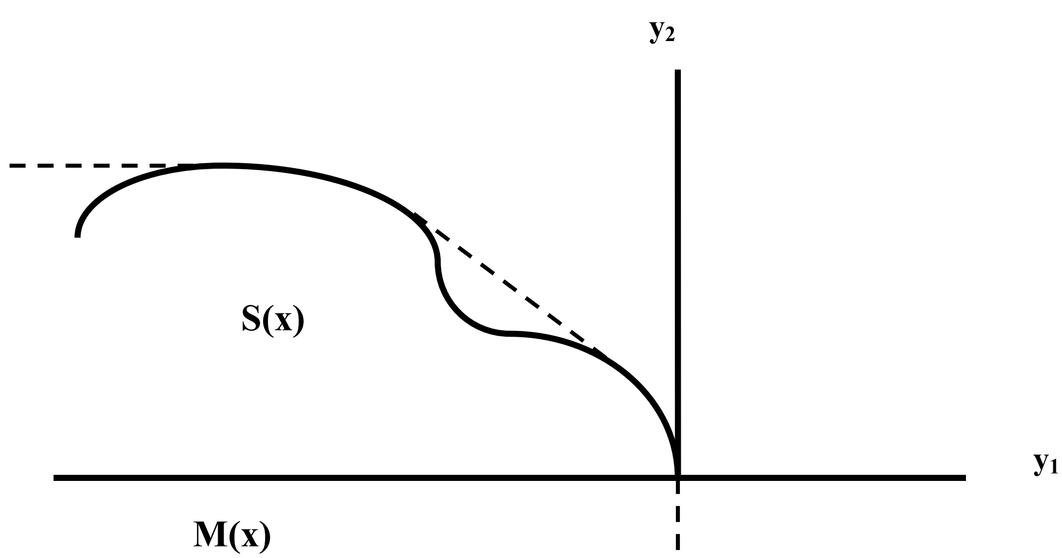


Figure 10: The Geometry of the One Output One Input Maximization Problem



What are conditions on the technology set S (and hence on the conditional technology sets S(x)) that will ensure that the outer approximation sets M(x), constructed using the variable profit function $\pi(p,x)$, will equal the true technology sets S(x)? It can be seen that the following two conditions on S (in addition to conditions (165) and (166)) are the required conditions:

(172) For every $x \geq 0_N$, the set $S(x) \equiv \{x: (y,x)\in S\}$ has the following *free disposal property*: $y^1\in S(x)$, $y^2 \leq y^1$ implies $y^2\in S(x)$;

(173) For every $x \geq 0_N$, the set $S(x) \equiv \{y: (y,x)\in S\}$ is convex.[61]

Conditions (172) and (173) are the conditions on the technology set S that are counterparts to the two regularity conditions of nondecreasingness and quasiconcavity[62] that were made on the production function, F(x), in section 3 above in order to obtain a duality between cost and production functions. If the firm is behaving as a price taker in variable commodity markets, it can be seen that it is not restrictive from an empirical point of view to assume that S satisfies conditions (172) and (173), just as it was not restrictive to assume that the production function was nondecreasing and quasiconcave in the context of the producer's (competitive) cost minimization problem studied earlier.

The next result provides a counterpart to Shephard's Lemma, Theorem 5 in section 4 above.

Theorem 11: Hotelling's (1932; 594) *Lemma*:[63] If the profit function $\pi(p,x)$ satisfies the properties listed in Theorem 10 above and in addition is once differentiable with respect to the components of the variable commodity prices at the point (p^*,x^*) where $x^* \geq 0_N$ and $p^* \gg 0_M$, then

(174) $y^* = \nabla_p\pi(p^*,x^*)$

where $\nabla_p\pi(p^*,x^*)$ is the vector of first order partial derivatives of variable profit with respect to variable commodity prices and y^* is any solution to the profit maximization problem

(175) $\max_y \{p^{*T}y: (y,x^*)\in S\} \equiv \pi(p^*,x^*)$.

Under these differentiability hypotheses, it turns out that the y^* solution to (175) is unique.

[61] If N = 1 so that there is only one fixed input, then given a producible net output vector $y\in R^M$, we can define the (fixed) *input requirements function* that corresponds to the technology set S as $g(y) \equiv \min_x \{x: (y,x)\in S\}$. In this case, condition (172) becomes the following condition: the input requirements function g(y) is *quasiconvex* in y. For additional material on this one fixed input model, see Diewert (1974c).

[62] Recall conditions (11) and (12) in section 3.

[63] See also Gorman (1968) and Diewert (1974a, 137).

Proof: Let y* be any solution to the profit maximization problem (175). Since y* is feasible for the profit maximization problem when the variable commodity price vector is changed to an arbitrary $p >> 0_N$, it follows that

$$(176)\ p^T y^* \leq \pi(p,x^*) \qquad \text{for every } p >> 0_M.$$

Since y* is a solution to the profit maximization problem (175) when p = p*, we must have

$$(177)\ p^{*T} y^* = \pi(p^*,x^*).$$

But (176) and (177) imply that the function of M variables, $g(p) \equiv p^T y^* - \pi(p,x^*)$ is nonpositive for all $p >> 0_M$ with $g(p^*) = 0$. Hence, g(p) attains a global maximum at p = p* and since g(p) is differentiable with respect to the variable commodity prices p at this point, the following first order necessary conditions for a maximum must hold at this point:

$$(178)\ \nabla_p\, g(p^*) = y^* - \nabla_p \pi(p^*,x^*) = 0_M.$$

Now note that (178) is equivalent to (174). If y** is any other solution to the profit maximization problem (175), then repeat the above argument to show that

$$(179)\ y^{**} = \nabla_p \pi(p^*,x^*)$$
$$= y^*$$

where the second equality follows using (178). Hence y** = y* and the solution to (175) is unique. Q.E.D.

Hotelling's Lemma may be used in order to derive systems of variable commodity output supply and input demand functions just as we used Shephard's Lemma to generate systems of cost minimizing input demand functions; for examples of this use of Hotelling's Lemma, see Diewert (1974a; 137-139).

If we are willing to make additional assumptions about the underlying firm production possibilities set S, then we can deduce that π(p,x) satisfies some additional properties. One such additional property is the following one: S is subject to the *free disposal of fixed inputs* if it has the following property:

$$(180)\ x^2 > x^1 \geq 0_N \text{ and } (y,x^1)\in S \text{ implies } (y,x^2)\in S.$$

The above property means if the vector of fixed inputs x^1 is sufficient to produce the vector of variable inputs and outputs y and if we have at our disposal a bigger vector of fixed inputs x^2, then y is still producible by the technology that is represented by the set S.

Theorem 12:[64] Suppose the technology set S satisfies assumptions (165) and (166) above. (a) If in addition, S has the following property:[65]

(181) For every $x \geq 0_N$, $(0_M,x)\in S$;

then for every $p >> 0_M$ and $x \geq 0_N$, $\pi(p,x) \geq 0$; i.e., the variable profit function is *nonnegative* if (181) holds.
(b) If S is a convex set, then for each $p >> 0_M$, then $\pi(p,x)$ is a *concave function* of x over the set $\Omega \equiv \{x: x \geq 0_N\}$.
(c) If S is a cone so that the technology is subject to constant returns to scale, then $\pi(p,x)$ is (positively) *homogeneous of degree one* in the components of x.
(d) If S is subject to the free disposal of fixed inputs, then

(182) $p >> 0$, $x^2 > x^1 \geq 0_N$ implies $\pi(p,x^2) \geq \pi(p,x^1)$;

i.e., $\pi(p,x)$ is *nondecreasing* in the components of x.

Proof of (a): Let $p >> 0_M$ and $x \geq 0_N$. Then

(183) $\pi(p,x) \equiv \max_y \{p^Ty: (y,x)\in S\}$
$\geq p^T 0_M$ since by (181), $(0_M,x)\in S$ and hence is feasible for the problem
$= 0.$

Proof of (b): Let $p >> 0_M$, $x^1 \geq 0_N$, $x^2 \geq 0_N$ and $0 < \lambda < 1$. Then

(184) $\pi(p,x^1) \equiv \max_y \{p^Ty: (y,x^1)\in S\}$
$= p^Ty^1$ where $(y^1,x^1)\in S$;

(185) $\pi(p,x^2) \equiv \max_y \{p^Ty: (y,x^2)\in S\}$
$= p^Ty^2$ where $(y^2,x^2)\in S$.

Since S is assumed to be a convex set, we have

(186) $\lambda(y^1,x^1) + (1-\lambda)(y^2,x^2) = [\lambda y^1 + (1-\lambda)y^2, \lambda x^1 + (1-\lambda)x^2]\in S$.

Using the definition of π, we have:

(187) $\pi(p,\lambda x^1 + (1-\lambda)x^2) \equiv \max_y \{p^Ty: (y,\lambda x^1 + (1-\lambda)x^2)\in S\}$
$\geq p^T[\lambda y^1 + (1-\lambda)y^2]$ since by (186), $\lambda y^1 + (1-\lambda)y^2$ is feasible for the problem
$= \lambda p^Ty^1 + (1-\lambda)p^Ty^2$
$= \lambda\pi(p,x^1) + (1-\lambda)\pi(p,x^2)$ using (184) and (185).

[64] The results in this Theorem are essentially due to Samuelson (1953-54; 20), Gorman (1968) and Diewert (1973) (1974a; 136) but they are packaged in a somewhat different form in this chapter.
[65] This property says that the technology can always produce no variable outputs and utilize no variable inputs given any vector of fixed inputs x.

Proof of (c): Let $p >> 0_M$, $x^* \geq 0_N$ and $\lambda > 0$. Then

$$(188)\ \pi(p,x^*) \equiv \max_y \{p^Ty: (y,x^*)\in S\} = p^Ty^* \qquad \text{where } (y^*,x^*)\in S.$$

Since S is a cone and since $(y^*,x^*)\in S$, then we have $(\lambda y^*,\lambda x^*)\in S$ as well. Hence, using a feasibility argument:

$$(189)\ \pi(p,\lambda x^*) \equiv \max_y \{p^Ty: (y,\lambda x^*)\in S\} \geq p^T\lambda y^* \quad \text{since } (\lambda y^*,\lambda x^*)\in S \text{ and hence is feasible for the problem} \\ = \lambda p^Ty^*.$$

Now *suppose* that the strict inequality in (189) holds so that

$$(190)\ \pi(p,\lambda x^*) \equiv \max_y \{p^Ty: (y,\lambda x^*)\in S\} = p^Ty^{**} \quad \text{where } (y^{**},\lambda x^*)\in S \\ > \lambda p^Ty^*.$$

Since S is a cone and since $(y^{**},\lambda x^*)\in S$, then we have $(\lambda^{-1}y^{**},x^*)\in S$ as well. Thus $\lambda^{-1}y^{**}$ is feasible for the maximization problem (188) that defined $\pi(p,x^*)$ and so

$$(191)\ p^Ty^* = \max_y \{p^Ty: (y,x^*)\in S\} \quad \text{using (188)} \\ \geq p^T\lambda^{-1}y^{**} \quad \text{since } \lambda^{-1}y^{**} \text{ is feasible for the problem} \\ = \lambda^{-1}p^Ty^{**}$$

or since $\lambda > 0$, (191) is equivalent to

$$(192)\ \lambda p^Ty^* \geq p^Ty^{**} > \lambda p^Ty^* \qquad \text{using (190).}$$

But (192) implies that $\lambda p^Ty^* > \lambda p^Ty^*$, which is impossible and hence our *supposition* is false and the desired result follows.

Proof of (d): Let $p >> 0$, $x^2 > x^1 \geq 0_N$. Using the definition of $\pi(p,x^1)$, we have

$$(193)\ \pi(p,x^1) \equiv \max_y \{p^Ty: (y,x^1)\in S\} = p^Ty^1 \qquad \text{where } (y^1,x^1)\in S.$$

Using the free disposal property (180) for S, since $(y^1,x^1)\in S$ and $x^2 > x^1$, we have

$$(194)\ (y^1,x^2)\in S.$$

Using the definition of $\pi(p,x^2)$, we have

$$(195)\ \pi(p,x^2) \equiv \max_y \{p^Ty: (y,x^2)\in S\}$$

$\geq p^T y^1$ since by (194), (y^1,x^2) is feasible
$= \pi(p,x^1)$ using (193).

Q.E.D.

Note that if the technology set S satisfies the minimal regularity conditions (165) and (166) plus all of the additional conditions that are listed in Theorem 12 above (we shall call such a technology set a *regular technology set*), then the associated variable profit function $\pi(p,x)$ will have *all* of the regularity conditions with respect to its fixed input vector x that a nonnegative, nondecreasing, concave and linearly homogeneous production function f(x) possesses with respect to its input vector x.

Hotelling's Lemma enabled us to interpret the vector of first order partial derivatives of the variable profit function with respect to the components of the variable commodity price vector p, $\nabla_p\pi(p,x)$, as the producer's vector of variable profit maximizing output supply (and the negative of variable input demand) functions, y(p,x), provided that the derivatives existed. If the first order partial derivatives of the variable profit function $\pi(p,x)$ with respect to the components of the fixed input vector x exist, then this vector of derivatives, $\nabla_x\pi(p,x)$, can also be given an economic interpretation as a vector of shadow prices or imputed contributions to profit of adding marginal units of fixed inputs. The following result also shows that these derivatives can be interpreted as competitive input prices for the "fixed" factors if they are allowed to become variable.

Theorem 13:[66] Suppose the technology set S satisfies assumptions (165) and (166) above and in addition is a convex set. Suppose in addition that $p^* \gg 0_M$, $x^* \geq 0_N$ and that the vector of derivatives, $\nabla_x\pi(p^*,x^*) \equiv w^*$, exists. Then x* is a solution to the following *long run profit maximization problem* that allows the "fixed" inputs x to be variable:

(196) $\max_x \{\pi(p^*,x) - w^{*T}x: x \geq 0_N\}$.

Proof: Part (b) of Theorem 12 above implies that $\pi(p^*,x)$ is a concave function of x over the set $\Omega \equiv \{x: x \geq 0_N\}$. The function $- w^{*T}x$ is linear in x and hence is also a concave function of x over Ω. Hence f(x) defined for $x \geq 0_N$ as

(197) $f(x) \equiv \pi(p^*,x) - w^{*T}x$

is also a concave function in x over the set Ω. Since $x^* \geq 0_N$, $x^*\in\Omega$. Hence using the third characterization of concavity and the differentiability of f(x) with respect to x at x*, we have:

(198) $f(x) \leq f(x^*) + \nabla_x f(x^*)^T(x - x^*)$ for all $x \geq 0_N$
$= \pi(p^*,x^*) - w^{*T}x^* + 0_N^T(x - x^*)$ since $\nabla_x f(x^*) = \nabla_x\pi(p^*,x^*) - w^* = 0_N$
$= \pi(p^*,x^*) - w^{*T}x^*$.

[66] Related results can be found in Samuelson (1953-54; 10) and Diewert (1974a; 140).

But (197) and (198) show that x* solves the profit maximization problem (196). Q.E.D.

Corollary: If in addition to the above assumptions, $\pi(p,x)$ is differentiable with respect to the components of p at the point (p^*,x^*), so that $y^* \equiv \nabla_p\pi(p^*,x^*)$ exists, then (y^*,x^*) solves the following long run profit maximization problem:

$$(199)\ \Pi(p^*,w^*) \equiv \max_{y,x} \{p^{*T}y - w^{*T}x: (y,x)\in S\}.$$

Proof: Using Hotelling's Lemma, we know that y* solves the following variable profit maximization problem:

$$(200)\ \pi(p^*,x^*) \equiv \max_{y} \{p^{*T}y: (y,x^*)\in S\} = p^{*T}y^*.$$

Now look at the long run profit maximization problem defined by (199):

$$\begin{aligned}(201)\ \Pi(p^*,w^*) &\equiv \max_{y,x} \{p^{*T}y - w^{*T}x: (y,x)\in S\} \\ &= \max_x [\max_y \{p^{*T}y: (y,x)\in S\} - w^{*T}x] && \text{where we have rewritten the maximization problem as a two stage maximization problem} \\ &= \max_x [\pi(p^*,x) - w^{*T}x] && \text{using the definition of } \pi(p^*,x) \\ &= \pi(p^*,x^*) - w^{*T}x^* && \text{using Theorem 13.}\end{aligned}$$

Hence with x = x* being an x solution to (201), we must have

$$\begin{aligned}(202)\ \Pi(p^*,w^*) &\equiv \max_{y,x} \{p^{*T}y - w^{*T}x: (y,x)\in S\} \\ &= [\max_y \{p^{*T}y: (y,x^*)\in S\} - w^{*T}x^*] && \text{letting } x = x^* \\ &= p^{*T}y^* - w^{*T}x^* && \text{using (200).}\end{aligned}$$

Q.E.D.

Hotelling's Lemma and Theorem 13 can be used as a convenient method for obtaining econometric estimating equations for determining the parameters that characterize a producer's technology set S. Assuming that S satisfies (165) and (166), we need only postulate a differentiable functional form for the producer's variable profit function, $\pi(p,x)$, that is linearly homogeneous and convex in p. Suppose that we have collected data on the fixed input vectors used by the firm in period t, x^t, and the net supply vectors for variable commodities produced in period t, y^t, for t = 1,…,T time periods as well as the corresponding variable commodity price vectors p^t. Then the following M equations can be used in order to estimate the unknown parameters in $\pi(p,x)$:

$$(203)\ y^t = \nabla_p\pi(p^t,x^t) + u^t\ ; \qquad t = 1,\ldots,T$$

where u^t is a vector of errors. If in addition, it can be assumed that the firm is optimizing with respect to its vector of fixed inputs in each period, where it faces the fixed input price vector w^t in period t, then the following N equations can be added to (203) as additional estimating equations:

(204) $w^t = \nabla_x \pi(p^t,x^t) + v^t$; $t = 1,\ldots,T$

where v^t is a vector of errors.[67]

12. The Comparative Statics Properties of Net Supply and Fixed Input Demand Functions

From Theorem 10 above, we know that the firm's variable profit function $\pi(p,x)$ is convex and linearly homogeneous in the components of the vector of variable commodity prices p for each fixed input vector x. Thus if $\pi(p,x)$ is twice continuously differentiable with respect to the components of p at some point (p,x), then using Hotelling's Lemma, we can prove the following counterpart to Theorem 7 for the cost function.

Theorem 14: Hotelling (1932; 597), Hicks (1946; 321), Diewert (1974a; 142-146): Suppose the variable profit function $\pi(p,x)$ is linearly homogeneous and convex in p and in addition is twice continuously differentiable with respect to the components of p at some point, (p,x). Then the *system of variable profit maximizing net supply functions*, $y(p,x) \equiv [y_1(p,x),\ldots,y_M(p,x)]^T$, exists at this point and these net supply functions are once continuously differentiable. Form the M by M matrix of net supply derivatives with respect to variable commodity prices, $B \equiv [\partial y_i(p,x)/\partial p_j]$, which has ij element equal to $\partial y_i(p,x)/\partial p_j$. Then the matrix B has the following properties:

(205) $B = B^T$ so that $\partial y_i(p,x)/\partial p_j = \partial y_j(p,x)/\partial p_i$ for all $i \neq j$;[68]
(206) B is positive semidefinite and
(207) $Bp = 0_M$.

Proof: Hotelling's Lemma implies that the firm's system of variable profit maximizing net supply functions, $y(p,x) \equiv [y_1(p,x),\ldots,y_M(p,x)]^T$, exists and is equal to

(208) $y(p,x) = \nabla_p \pi(p,x)$.

Differentiating both sides of (208) with respect to the components of p gives us

(209) $B \equiv [\partial y_i(p,x)/\partial p_j] = \nabla^2_{pp}\pi(p,x)$.

Now property (205) follows from Young's Theorem in calculus. Property (206) follows from (209) and the fact that $\pi(p,x)$ is convex in p and the fourth characterization of convexity. Finally, property (207) follows from the fact that the profit function is linearly homogeneous in p and hence, using Part 2 of Euler's Theorem on homogeneous functions, (207) holds. Q.E.D.

---

[67] If the technology set S is subject to constant returns to scale and the data reflect this fact by "adding up" (so that $p^{tT}y^t = w^{tT}x^t$ for $t = 1,\ldots,T$), then the error vectors u^t and v^t in (203) and (204) cannot be statistically independent since they will satisfy the constraint $p^{tT}u^t = w^{tT}v^t$ for $t = 1,\ldots,T$. Hence, under these circumstances, one of the M+N equations in (203) and (204) must be dropped in the system of estimating equations.

[68] These are the Hotelling (1932; 549) and Hicks (1946; 321) symmetry restrictions on supply functions.

Note that property (206) implies the following properties on the net supply functions:

(210) $\partial y_m(p,x)/\partial p_m \geq 0$ for m = 1,…,M.

Property (210) means that output supply curves cannot be downward sloping. However, if variable commodity m is an input, then $y_m(p,x)$ is negative. If we define the positive input demand function as

(211) $d_m(p,x) \equiv - y_m(p,x) \geq 0$,

then the restriction (210) translates into $\partial d_m(p,x)/\partial p_m \leq 0$, which means that variable input demand curves cannot be upward sloping.

Obviously, if the technology set is a convex cone, then the firm's competitive fixed input price functions, $w(p,x) \equiv \nabla_x\pi(p,x)$, will satisfy properties analogous to the properties of cost minimizing input demand functions in Theorem 7.

Theorem 15: Samuelson (1953-54; 10), Diewert (1974a; 144-146): Suppose that the firm's technology set S is regular. Define the firm's variable profit function $\pi(p,x)$ by (164). Suppose that $\pi(p,x)$ is twice continuously differentiable with respect to the components of x at some point (p,x) where $p >> 0_M$ and $x \geq 0_N$. Then the *system of fixed input price functions*[69], $w(p,x) \equiv [w_1(p,x),\ldots,w_N(p,x)]^T$, exists at this point[70] and these input price functions are once continuously differentiable. Form the N by N matrix of fixed input price derivatives with respect to the fixed inputs, $C \equiv [\partial w_i(p,x)/\partial x_j]$, which has ij element equal to $\partial w_i(p,x)/\partial x_j$. Then the matrix C has the following properties:

(212) $C = C^T$ so that $\partial w_i(p,x)/\partial x_j = \partial w_j(p,x)/\partial x_i$ for all $i \neq j$;
(213) C is negative semidefinite and
(214) $Cx = 0_N$.

Proof: Using the results of Theorem 13, the firm's system of fixed input price functions, $w(p,x) \equiv [w_1(p,x),\ldots,w_N(p,x)]^T$, exists and is equal to

(215) $w(p,x) = \nabla_x\pi(p,x)$ (*Samuelson's Lemma*).

Differentiating both sides of (215) with respect to the components of x gives us

(216) $C \equiv [\partial w_i(p,x)/\partial x_j] = \nabla^2_{xx}\pi(p,x)$.

[69] The functions w(p,x) can also be interpreted as the producer's system of *inverse demand functions for fixed inputs*.
[70] The assumption that S is regular implies that S has the free disposal property in fixed inputs property (180), which implies by part (d) of Theorem 12 that $\pi(p,x)$ is nondecreasing in x and this in turn implies that $w(p,x) \equiv \nabla_x\pi(p,x)$ is nonnegative.

Now property (212) follows from Young's Theorem in calculus. Property (213) follows from (216) and the fact that $\pi(p,x)$ is concave in x[71] and the fourth characterization of concavity. Finally, property (214) follows from the fact that the profit function is linearly homogeneous in x[72] and hence, using Part 2 of Euler's Theorem on homogeneous functions, (214) holds. Q.E.D.

Note that property (213) implies the following properties on the fixed input price functions:

(217) $\partial w_n(p,x)/\partial x_n \leq 0$ for $n = 1,\ldots,N$.

Property (217) means that the inverse fixed input demand curves cannot be upward sloping.

If the firm's production possibilities set S is regular and if the corresponding variable profit function $\pi(p,x)$ is twice continuously differentiable with respect to all of its variables, then there will be additional restrictions on the derivatives of the variable net output supply functions $y(p,x) = \nabla_p\pi(p,x)$ and on the derivatives of the fixed input price functions $w(p,x) = \nabla_x\pi(p,x)$. Define the M by N *matrix of derivatives of the net output supply functions y(p,x) with respect to the components of the vector of fixed inputs x* as follows:

(218) $D \equiv [\partial y_i(p,x)/\partial x_j] = \nabla^2_{px}\pi(p,x)$; $i = 1,\ldots,M; j = 1,\ldots,N$,

where the equalities in (218) follow by differentiating both sides of the Hotelling's Lemma relations, $y(p,x) = \nabla_p\pi(p,x)$, with respect to the components of x. Similarly, define the N by M *matrix of derivatives of the fixed input price functions w(p,x) with respect to the components of the vector of variable commodity prices p* as follows:

(219) $E \equiv [\partial w_i(p,x)/\partial p_j] = \nabla^2_{xp}\pi(p,x)$; $i = 1,\ldots,N; j = 1,\ldots,M$,

where the equalities in (219) follows by differentiating both sides of the Samuelson's Lemma relations, $w(p,x) = \nabla_x\pi(p,x)$, with respect to the components of p.

Theorem 16: Samuelson (1953-54; 10), Diewert (1974a; 144-146): Suppose that the firm's technology set S is regular. Define the firm's variable profit function $\pi(p,x)$ by (164). Suppose that $\pi(p,x)$ is twice continuously differentiable with respect to the components of x at some point (p,x) where $p \gg 0_M$ and $x \geq 0_N$ and define the matrices of derivatives D and E by (218) and (219) respectively. Then these matrices have the following properties:

[71] The assumption that S is regular implies that S is a convex set and this in turn implies that $\pi(p,x)$ is concave in x.

[72] The assumption that S is regular implies that S is a cone and this in turn implies that $\pi(p,x)$ is linearly homogeneous in x.

(220) $D = E^T$ so that $\partial y_m(p,x)/\partial x_n = \partial w_n(p,x)/\partial x_m$ for m = 1,…,M and n = 1,…,N;
(221) $w(p,x) = Ep \geq 0_N$;
(222) $y(p,x) = Dx$.

Proof: The symmetry restrictions (220) follow from definitions (218) and (219) and Young's Theorem in calculus.

Since $\pi(p,x)$ is linearly homogeneous in the components of p, we have

(223) $\pi(\lambda p,x) = \lambda\pi(p,x)$ for all $\lambda > 0$.

Partially differentiate both sides of (223) with respect to x_n and we obtain:

(224) $\partial\pi(\lambda p,x)/\partial x_n = \lambda\partial\pi(p,x)/\partial x_n$ for all $\lambda > 0$ and n = 1,…,N.

But (224) implies that the functions $w_n(p,x) \equiv \partial\pi(p,x)/\partial x_n$ are homogeneous of degree one in p. Hence, we can apply Part 1 of Euler's Theorem on homogeneous functions to these functions $w_n(p,x)$ and conclude that

(225) $w_n(p,x) = \sum_{m=1}^M [\partial w_n(p,x)/\partial p_m]p_m$; n = 1,…,N.

But equations (225) are equivalent to the equations in (221). The inequality in (221) follows from $w(p,x) = \nabla_x\pi(p,x) \geq 0_N$, which in turn follows from the fact that regularity of S implies that $\pi(p,x)$ is nondecreasing in the components of x.

Since S is regular, part (c) of Theorem 12 implies that $\pi(p,x)$ is linearly homogeneous in x, so that

(226) $\pi(p,\lambda x) = \lambda\pi(p,x)$ for all $\lambda > 0$.

Partially differentiate both sides of (226) with respect to p_m and we obtain:

(227) $\partial\pi(p,\lambda x)/\partial p_m = \lambda\partial\pi(p,x)/\partial p_m$ for all $\lambda > 0$ and m = 1,…,M.

But (227) implies that the functions $y_m(p,x) \equiv \partial\pi(p,x)/\partial p_m$ are homogeneous of degree one in x. Hence, we can apply Part 1 of Euler's Theorem on homogeneous functions to these functions $y_m(p,x)$ and conclude that

(228) $y_m(p,x) = \sum_{n=1}^N [\partial y_m(p,x)/\partial x_n]x_n$; m = 1,…,M.

But equations (228) are equivalent to equations (222). Q.E.D.

Problems

22. Under the hypotheses of Theorem 16, show that $y(p,x)$ and $w(p,x)$ satisfy the following equation:

(i) $p^T y(p,x) = x^T w(p,x)$.

23. Let S be a technology set that satisfies assumptions (165) and (166) and let $\pi(p,x)$ be the corresponding differentiable variable profit function defined by (164). Variable commodities m and k (where $m \neq k$) are said to be *substitutes* if (i) below holds, *unrelated* if (ii) below holds and *complements* if (iii) below holds:

(i) $\partial y_m(p,x)/\partial p_k < 0$;
(ii) $\partial y_m(p,x)/\partial p_k = 0$;
(iii) $\partial y_m(p,x)/\partial p_k > 0$.

(a) If the number of variable commodities M = 2, then show that the two variable commodities cannot be complements.
(b) If M = 2 and the two variable commodities are unrelated, then show that:

(iv) $\partial y_1(p,x)/\partial p_1 = \partial y_2(p,x)/\partial p_2 = 0$.

(c) If M = 3, then show that at most one pair of variable commodities can be complements.[73]

24. Let S be a regular technology and let $\pi(p,x)$ be the corresponding differentiable variable profit function. Define the producer's system of inverse fixed input demand functions as $w(p,x) \equiv \nabla_x \pi(p,x)$. Fixed inputs n and k (where $n \neq k$) are said to be *substitutes* if (i) below holds, *unrelated* if (ii) below holds and *complements* if (iii) below holds:

(i) $\partial w_n(p,x)/\partial x_k > 0$;
(ii) $\partial w_n(p,x)/\partial x_k = 0$;
(iii) $\partial w_n(p,x)/\partial x_k < 0$.

(a) If the number of fixed inputs N = 2, then, assuming that $x_1 > 0$ and $x_2 > 0$, show that the two fixed inputs cannot be complements.
(b) If N = 2 and the two fixed inputs are unrelated, then show that (assume $x_1 > 0$ and $x_2 > 0$):

(iv) $\partial w_1(p,x)/\partial x_1 = \partial w_2(p,x)/\partial x_2 = 0$.

(c) If N = 3, then show that at most one pair of fixed inputs can be complements.

[73] This type of argument (that substitutability tends to be more predominant than complementarity) is again due to Hicks (1946; 322-323) but we have not followed his terminology exactly.

25. *Application to International Trade Theory*. Suppose that the technology set of a small open economy can be represented by a *regular production possibilities set*, S ≡ {(y,x)} where y is a vector of internationally traded goods (the components of C + I + G + X − M where imported commodities have negative signs) and $x \geq 0_N$ is a nonnegative vector of input factors that are available for use by the aggregate production sector. Let $p >> 0_M$ be a vector of international prices for traded goods that the economy faces. Thus in this case,

(i) $\pi(p,x) \equiv \max_y \{p^T y: (y,x) \in S\}$

is the *economy's GDP function*,[74] regarded as a function of the vector of world prices p that the economy faces and of the factor endowment vector or vector of primary resources x that the economy has available to produce goods and services. Assume that π(p,x) is twice continuously differentiable with respect to its variables at an initial equilibrium for the economy.

(a) Show that if the amount of the first primary input, x_1, increases a small amount, then GDP does not decrease; i.e., show that

(i) $\partial\pi(p,x)/\partial x_1 \geq 0$.

(b) Show that as the amount of the first primary input increases a small amount, then the corresponding factor price does not increase and an input quantity weighted sum of the other factor prices does not decrease; i.e., show that

(ii) $\partial w_1(p,x)/\partial x_1 \leq 0$ and
(iii) $\sum_{n=2}^{N} x_n\, \partial w_n(p,x)/\partial x_1 \geq 0$.

If the inequalities (ii) and (iii) hold strictly, then they show that input 1 experiences a decrease in its price as the amount of input 1 increases but at least one other input must gain as a result of this increase in input 1.

(c) Show that if the price of the first internationally traded good, p_1, increases a small amount and the first traded good is not imported,[75] then GDP increases; i.e., show that

(iv) $\partial\pi(p,x)/\partial p_1 > 0$.

(d) Continuation of (c). Show that as the first traded commodity price increases a small amount, then the production of commodity 1 does not decrease and a traded commodity price weighted sum of the other components of GDP does not increase; i.e., show that

[74] For applications of duality theory to the theory of international trade, see Samuelson (1953-54), Chipman (1972), Diewert (1974a; 142-146), Diewert and Woodland (1977), Kohli (1978) (1991) and Woodland (1982).
[75] In fact, we assume that in the initial equilibrium, a positive amount of this first traded commodity is produced by the aggregate production sector.

(v) $\partial y_1(p,x)/\partial p_1 \geq 0$ and
(vi) $\sum_{m=2}^{M} p_m\, \partial y_m(p,x)/\partial p_1 \leq 0$.

If the inequalities (v) and (vi) hold strictly, then they show that output 1 experiences an increase in production as the price of output 1 increases but at least one other output must decrease (or at least one other imported commodity must increase in magnitude) as a result of this increase in the price of output 1.

(e) Show that if the price of the first internationally traded good, p_1, increases a small amount and the first traded good is imported, then GDP decreases; i.e., show that

(vii) $\partial \pi(p,x)/\partial p_1 < 0$.

(f) Continuation of (e). Show that as the first traded commodity price increases a small amount, then the importation of commodity 1 does not increase and a traded commodity price weighted sum of the other components of GDP does not increase; i.e., show that[76]

(viii) $-\partial y_1(p,x)/\partial p_1 \leq 0$ and
(ix) $\sum_{m=2}^{M} p_m\, \partial y_m(p,x)/\partial p_1 \leq 0$.

If the inequalities (viii) and (ix) hold strictly, then they show that imports of traded commodity 1 decline as the price of output 1 increases and in addition, at least one output must decrease (or at least one other imported commodity must increase in magnitude) as a result of this increase in the price of output 1.

26. Let S be a regular technology set and let $\pi(p,x)$ be the corresponding twice continuously differentiable variable profit function defined by (164). Variable commodities m and fixed input n are said to be *normal* if (i) below holds, *unrelated* if (ii) below holds and *inferior* if (iii) below holds (we assume $p >> 0_M$ and $x >> 0_N$):

(i) $\partial y_m(p,x)/\partial x_n = \partial w_n(p,x)/\partial p_m > 0$;
(ii) $\partial y_m(p,x)/\partial x_n = \partial w_n(p,x)/\partial p_m = 0$;
(iii) $\partial y_m(p,x)/\partial x_n = \partial w_n(p,x)/\partial p_m < 0$.

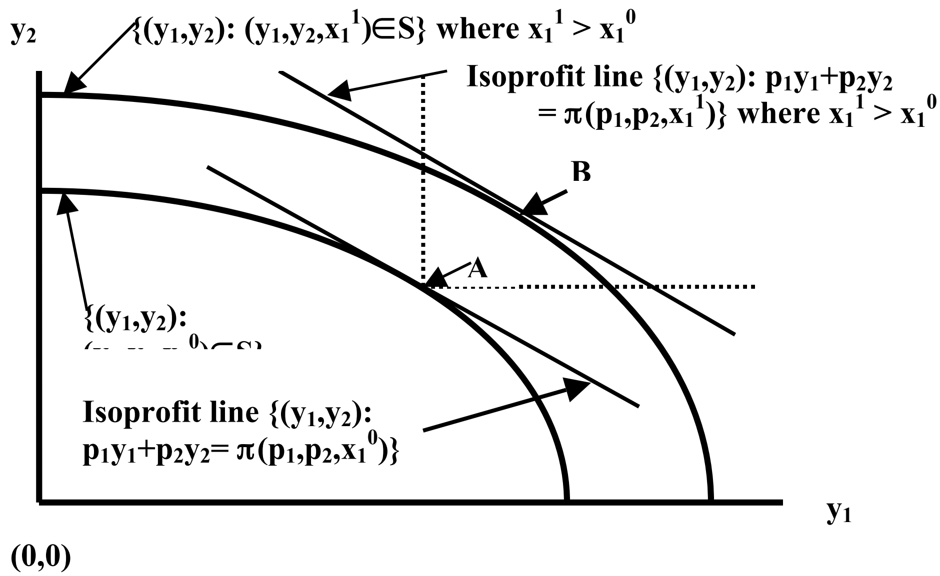
(a) If $w_n(p,x) > 0$, then there exists at least one variable commodity m such that commodity m and fixed input n are normal.
(b) If $w_n(p,x) \geq 0$, then there exists at least one variable commodity m such that commodity m and fixed input n are either normal or unrelated.
(c) If $y_m(p,x) > 0$, then there exists at least one fixed input n such that commodity m and fixed input n are normal.
(d) If $y_m(p,x) < 0$, then there exists at least one fixed input n such that commodity m and fixed input n are inferior.

---

[76] Since y_1 is imported in the initial equilibrium, $y_1(p,x) < 0$. Thus $-\,y_1(p,x) > 0$ is the magnitude of imports in the initial equilibrium.

We illustrate the concepts of normality and inferiority for the case of two variable outputs, y_1 and y_2, and a varying amount of the first fixed input x_1.[77] In Figure 11 below, the case where variable commodities 1 and 2, y_1 and y_2, are both normal with respect to the first fixed input x_1 is illustrated. In this figure, it can be seen that as x_1 increases from its initial level of x_1^0 to the greater level x_1^1, the production possibilities set $\{(y_1,y_2): (y_1,y_2,x_1^0)\in S\}$ shifts outwards to the production possibilities set $\{(y_1,y_2): (y_1,y_2,x_1^1)\in S\}$. The initial variable profit maximizing (y_1,y_2) point is at point A. After x_1 increases from x_1^0 to x_1^1, the new variable profit maximizing (y_1,y_2) point is at point B.

Figure 11: Variable Commodities 1 and 2 Normal with Fixed Input 1



It can be seen as long as the point B is to the northeast of the point A, both y_1 and y_2 will be normal with the fixed input x_1.[78]

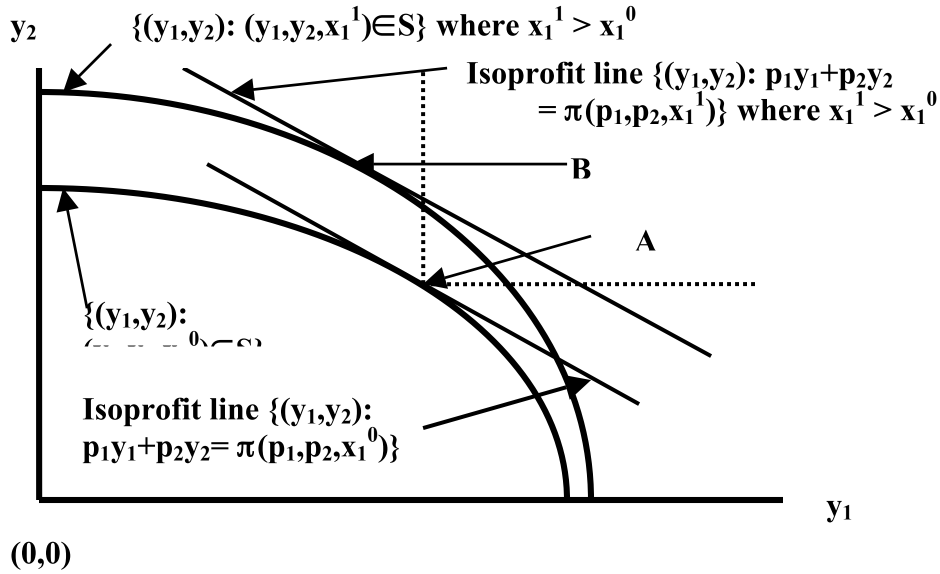
In Figure 12, as x_1 increases from the initial level x_1^0 to the higher level x_1^1, the revenue maximizing (y_1,y_2) point moves from A to B. It can be seen that y_1 decreases as x_1 increases and thus the first variable output and the first fixed input are an *inferior* pair of commodities. On the other hand, y_2 increases as x_1 increases and thus the second variable output and the first fixed input are a *normal* pair of commodities.

---

[77] Since the amounts of the other fixed inputs, $x_2,\ldots,x_N$ remain fixed in the figures to follow, they will be suppressed from the notation.

[78] The similarity of this normality concept in production theory with the corresponding normality concept in consumer theory can be seen by looking at Figures 11 and 12; the fixed input x_1 now plays the role of utility u and the frontiers of the production possibility sets $\{(y_1,y_2): (y_1,y_2,x_1)\in S\}$ indexed by x_1 now replace the indifference curves $\{(x_1,x_2): F(x_1,x_2) = u\}$ indexed by u.

Figure 12: Variable Commodity 1 Inferior with Respect to Fixed Input 1



Problem

27. Draw counterparts to Figures 11 and 12 to illustrate the concepts of normality and inferiority for the case where $y_1 < 0$ is a variable input and $y_2 > 0$ is a variable output.

13. The Gains from Trade and the Costs of Trade Restrictions

In this section, we show how the material on variable profit functions developed in the previous sections can be used to model the costs and benefits of imposing a tariff on imported goods and services.

We consider the simplest possible model of a small open economy where there is one domestic good that is produced by the private production sector, $C \geq 0$ with price $p_C > 0$, one export commodity, $X \geq 0$ with price $p_X > 0$, one imported commodity, $M \geq 0$ with price $p_M > 0$, and one primary factor of production, $L \geq 0$ with price $w > 0$.[79]

The technology available to the economy is the *regular* production possibilities set S, which is a feasible set of vectors of the form $(X,-M,C,L)$. We assume that the price of exports and imports, p_X and p_M, are fixed on international markets. We treat exports and

[79] We follow the example of Kohli (1978) (1991), who assumed that all trade flowed through the private production sector of the economy rather than flowing *directly* to consumers and other final demanders of goods and services. This is a fairly reasonable assumption because even if a household directly consumes an imported good, typically there are domestic transportation, wholesaling and retailing inputs that are required to deliver the good to the consumer.

imports as variable commodities and C and L as fixed inputs and define the economy's variable profit function as follows:[80]

(229) $\pi(p_X,p_M,C,L) \equiv \max_{X,M} \{p_X X - p_M M: (X,-M,C,L)\in S\}$.

We assume that the government imposes an ad valorem tariff, $t > 0$, on imported goods and services. Hotelling's Lemma gives us the economy's *export supply* and *import demand* functions as functions of the tariff t as follows:

(230) $X(t) \equiv \partial\pi(p_X,(1+t)p_M,C(t),L)/\partial p_X$;
(231) $M(t) \equiv -\partial\pi(p_X,(1+t)p_M,C(t),L)/\partial p_M$.[81]

Samuelson's Lemma gives us the economy's *price of domestic output* and *wage rate* functions as functions of the tariff t as follows:[82]

(232) $p_C(t) \equiv -\partial\pi(p_X,(1+t)p_M,C(t),L)/\partial C$;
(233) $w(t) \equiv \partial\pi(p_X,(1+t)p_M,C(t),L)/\partial L$.

In the above equations, domestic production, C(t), is regarded as a function of the tariff t. This function is found by solving the following *balance of trade equation* for C as a function of t:

(234) $p_X\partial\pi(p_X,(1+t)p_M,C(t),L)/\partial p_X + p_M\partial\pi(p_X,(1+t)p_M,C(t),L)/\partial p_M = 0$.

Using (230) and (231), we see that (234) simply sets the value of exports minus the value of imports equal to 0 and for each given tariff rate t, we solve for the C which makes the balance of trade equal to 0.

Equation (234) illustrates the advantages in using duality theory when building an economic model. The present model boils down to solving a single equation in a single unknown C, which is a considerable simplification over using traditional primal optimization techniques to set up the model.[83]

---

[80] C is obviously not an input but the analysis in the previous sections can be modified to deal with this complication. Only two modifications to the previous analysis are required: (i) In definition (229), if C is large relative to L, then there may not exist any X and M such that $(X,-M,C,L)\in S$. In this case, define $\pi(p_X,p_M,C,L) \equiv -\infty$. For additional details on how to model this situation, see Diewert (1973). (ii) $\pi(p_X,p_M,C,L)$ will now be nonincreasing in C so that in the differentiable case, $\partial\pi(p_X,p_M,C,L)/\partial C \equiv -p_C \le 0$.
[81] There is an abuse of notation here and in subsequent equations up to equation (255): the partial derivative of π with respect to the price of imports, p_M, $\partial\pi(p_X,(1+t)p_M,C(t),L)/\partial p_M$, should be interpreted as the partial derivative of π with respect to the entire import price, ,$(1+t)p_M$; i.e., the price of imports including the tariff.
[82] Because C is a "fixed" output instead of being a fixed input, the minus sign was inserted in the right hand side of (232) to make the price of the consumption commodity, p_C, positive.
[83] Our model implies that the tariff revenue is either transferred back to the household sector or spent on government goods and services, which are included in domestic final demand, C(t). To see this, note that the linear homogeneity of $\pi(p_X,p_M,C,L)$ in p_X and p_M implies that $\pi(p_X,(1+t)p_M,C,L) = p_X\partial\pi(p_X,(1+t)p_M,C,L)/\partial p_X + (1+t)p_M\partial\pi(p_X,(1+t)p_M,C,L)/\partial p_M = p_X X(p_X,(1+t)p_M,C,L) - (1+t)p_M M(p_X,(1+t)p_M,C,L) \equiv p_X X(t) - (1+t)p_M M(t)$. The linear homogeneity of $\pi(p_X,p_M,C,L)$ in C and L

We are interested primarily in what happens to domestic consumption C(t) as the tariff t increases from an initial level of 0. Hence, we are interested in calculating the first and second derivatives of C(t), evaluated at t = 0, so that we can form first and second order Taylor series approximations to C(t). Differentiating both sides of (234) with respect to t leads to the following equation:

$$(235)\ [p_X\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_X\partial C + p_M\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_M\partial C]C'(t)$$
$$= -\{p_X\,p_M\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_X\partial p_M + p_M^2\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_M^2\}.$$

Equations (207) in the present context are the following two equations:

$$(236)\ p_X\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_X^2 + p_M(1+t)\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_X\partial p_M = 0;$$
$$(237)\ p_X\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_M\partial p_X + p_M(1+t)\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_M^2 = 0.$$

Equations (221) in the present context are the following two equations:

$$(238)\ -p_C(t) = p_X\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial C\partial p_X + p_M(1+t)\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial C\partial p_M\ ;$$
$$(239)\quad w(t) = p_X\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial L\partial p_X + p_M(1+t)\,\partial^2\pi(p_X,(1+t)p_M,C,L)/\partial L\partial p_M\ .$$

The following symmetry conditions are also valid:

$$(240)\ \partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_X\partial p_M = \partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_M\partial p_X\ ;$$
$$(241)\ \partial^2\pi(p_X,(1+t)p_M,C,L)/\partial C\partial p_X = \partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_X\partial C\ ;$$
$$(242)\ \partial^2\pi(p_X,(1+t)p_M,C,L)/\partial C\partial p_M = \partial^2\pi(p_X,(1+t)p_M,C,L)/\partial p_M\partial C\ .$$

Now substitute (237) and (238) into (235). Using also (240)-(242), we find that (235) becomes the following equation:

$$(243)\ C'(t) = -tp_M^2\,[\partial^2\pi(t)/\partial p_M^2]/[p_C(t) + tp_M\,\partial^2\pi(t)/\partial C\partial p_M]$$

where we have simplified the notation by defining

$$(244)\ \pi(t) \equiv \pi(p_X,(1+t)p_M,C(t),L).$$

Now evaluate (243) at t = 0 and we find that

---

implies that $\pi(p_X,(1+t)p_M,C,L) = C\partial\pi(p_X,(1+t)p_M,C,L)/\partial C + L\partial\pi(p_X,(1+t)p_M,C,L)/\partial L = -C(t)p_C(p_X,(1+t)p_M,C,L) + Lw(p_X,(1+t)p_M,C,L) \equiv -C(t)p_C(t) + Lw(t)$. Equating these two expressions for $\pi(p_X,(1+t)p_M,C,L)$ gives us the equation $p_XX(t) - (1+t)p_MM(t) = -C(t)p_C(t) + Lw(t)$. This equation can be rearranged to give us the following budget constraint for domestic final demanders and suppliers of primary inputs: $p_C(t)C(t) = w(t)L + tp_MM(t) - [p_XX(t) - p_MM(t)] = w(t)L + tp_MM(t)$ where we have used the balance of trade restriction, (234). Thus the value of final domestic demand, $p_C(t)C(t)$, is equal to the value of domestic primary input payments, $w(t)L$, plus the value of tariff revenue, $tp_MM(t)$. If the government does not have any expenditure programs, then the tariff revenue is simply rebated back to households in this very simple model.

(245) $C'(0) = 0.$

Differentiate both sides of (243) with respect to t and evaluate the resulting derivatives at t = 0. Using (245), we find that

$$(246)\ C''(0) = -\, p_M{}^2\, [\partial^2\pi(0)/\partial p_M{}^2]/p_C(0)$$
$$\leq 0$$

where the inequality follows from $\partial^2\pi(0)/\partial p_M{}^2 \geq 0$, which in turn follows from the fact that $\nabla_{pp}{}^2\, \pi(p_X,p_M,C(0),L)$ is a positive semidefinite matrix.

Using (245) and (246), we obtain the following second order Taylor series approximation to C(t):

$$(247)\ C(t) \cong C(0) + C'(0)(t-0) + (1/2)\, C''(0)(t-0)^2$$
$$= C(0) - t^2 p_M{}^2\, [\partial^2\pi(0)/\partial p_M{}^2]/2p_C(0).$$

Hence the *loss of domestic output due to a tariff* of size t is approximately equal to:

$$(248)\ C(0) - C(t) \cong t^2 p_M{}^2\, [\partial^2\pi(0)/\partial p_M{}^2]/2p_C(0) \geq 0.$$

It is more convenient to express this loss in terms of shares and elasticities. Define (the negative of) the elasticity of import demand at the free trade equilibrium as follows:

$$(249)\ \eta_M \equiv -\, [\partial M(p_X,p_M,C(0),L)/\partial p_M][p_M/M(p_X,p_M,C(0),L)]$$
$$= [\partial^2\pi(p_X,p_M,C(0),L)/\partial p_M{}^2][p_M/M(p_X,p_M,C(0),L)] \qquad \text{using (231)}$$
$$= p_M\, [\partial^2\pi(0)/\partial p_M{}^2]/M(0)$$
$$\geq 0.$$

In what follows, we will assume that the import elasticity of demand is not equal to 0 and hence (249) holds with a strict inequality. Now divide both sides of (248) through by C(0) and we obtain the following expression for the approximate percentage loss of domestic production due to a tariff of size t:

$$(250)\ DWL(t) \equiv [C(0) - C(t)]/C(0) \cong t^2 p_M{}^2\, [\partial^2\pi(0)/\partial p_M{}^2]/2p_C(0)C(0)$$
$$= t^2 p_M\, M(0)\, \eta_M\, /2p_C(0)C(0)$$
$$= (1/2)\, t^2\, s_M\, \eta_M$$
$$> 0$$

where $s_M \equiv p_M M(0)/p_C(0)C(0)$ is the share of imports in GDP at the free trade equilibrium.

We can now use formula (250) to make some rough estimates of what the cost of a tariff will be as the tariff increases. Suppose the elasticity of import demand is 1, the share of imports in GDP is 2/5 and the tariff rate is one percent or .01.[84] Then the approximate loss DWL(.01) is a miniscule .002 percentage points. If the tariff increases to 10 per cent or .1, then the loss increases to .2 percentage points. If the tariff rate increases to 50 % or .5, then the loss increases to .05 or 5 percentage points.[85]

Recall equations (230)-(233), which defined the export supply, import demand, price of domestic output and wage rate functions, X(t), M(t), $p_C(t)$ and w(t) respectively, as functions of the tariff rate t. Now that we know $C'(0) = 0$, we can differentiate equations (230)-(233) with respect to t and evaluate the resulting derivatives at t = 0 in order to see what happens to these variables as t increases from its initial level of 0. We obtain the following results:

$$(251)\; X'(0) = p_M\, \partial^2\pi(p_X,p_M,C(0),L)/\partial p_X\partial p_M + C'(0)\, \partial^2\pi(p_X,p_M,C(0),L)/\partial p_X\partial C$$
$$= p_M\, \partial^2\pi(0)/\partial p_X\partial p_M \qquad \text{using } C'(0) = 0$$
$$< 0$$

where the inequality follows from (237) evaluated at t = 0 and $\partial^2\pi(p_X,p_M,C,L)/\partial p_M^2 > 0$, which in turn follows from our assumption that the elasticity of demand for imports, η_M, is strictly positive.

$$(252)\; M'(0) = -\, p_M\, \partial^2\pi(p_X,p_M,C(0),L)/\partial p_M^2 - C'(0)\, \partial^2\pi(p_X,p_M,C(0),L)/\partial p_M\partial C$$
$$= -\, p_M\, \partial^2\pi(0)/\partial p_M^2 \qquad \text{using } C'(0) = 0$$
$$< 0$$

where the inequality follows from $\partial^2\pi(p_X,p_M,C,L)/\partial p_M^2 > 0$, which in turn follows from our assumption that the elasticity of demand for imports, η_M, is strictly positive.

The inequalities (251) and (252) are intuitively plausible: the effect of increasing the tariff rate is to restrict both exports and imports. If the tariff is pushed high enough, eventually international trade will cease.[86]

We turn now to the effects on domestic prices of increasing the tariff rate.

$$(253)\; p_C'(0) = -\, p_M\, \partial^2\pi(p_X,p_M,C(0),L)/\partial C\partial p_M - C'(0)\, \partial^2\pi(p_X,p_M,C(0),L)/\partial C^2$$
$$= -\, p_M\, \partial^2\pi(0)/\partial C\partial p_M \qquad \text{using } C'(0) = 0.$$

$$(254)\; w'(0) = p_M\, \partial^2\pi(p_X,p_M,C(0),L)/\partial L\partial p_M + C'(0)\, \partial^2\pi(p_X,p_M,C(0),L)/\partial L\partial C$$
$$= p_M\, \partial^2\pi(0)/\partial L\partial p_M \qquad \text{using } C'(0) = 0.$$

---

[84] Most OECD countries have tariff rates around one per cent.

[85] Most countries in the early stages of development have relatively high tariff levels of this order of magnitude because they cannot easily raise revenue by taxing components of final demand or by taxing primary inputs.

[86] This result can be turned around to show the benefits of a country opening up its borders to international trade so that as t decreases from an initial prohibitive tariff (where no trade takes place), there are *gains from opening up the economy to international trade*.

Equations (222) imply that the cross partial derivatives in (253) and (254) satisfy the following equation:

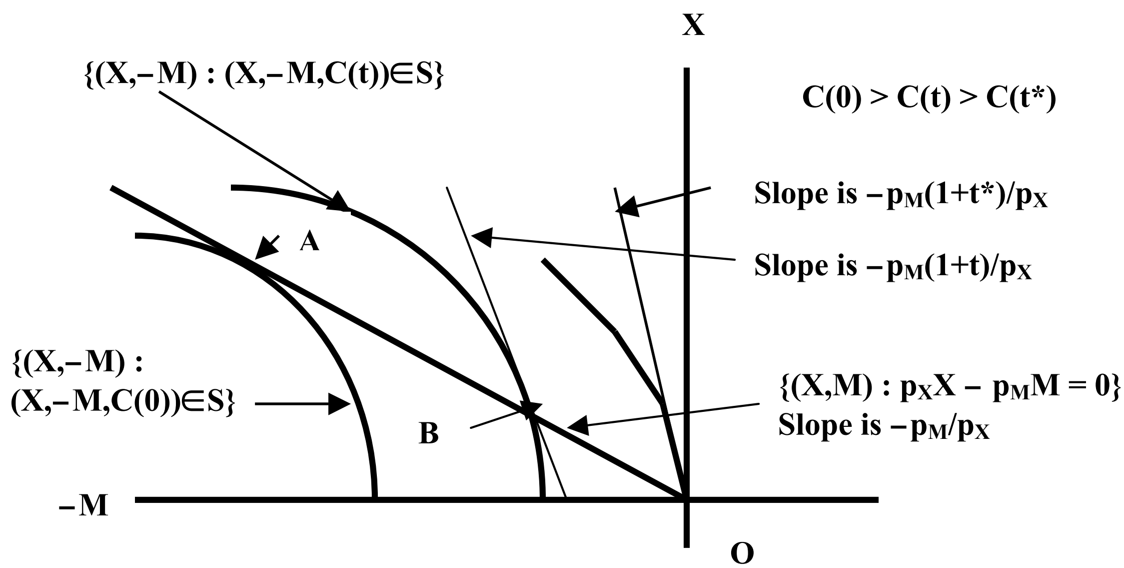
(255) $C(0)\, \partial^2\pi(0)/\partial C\partial p_M + L\, \partial^2\pi(0)/\partial L\partial p_M = \partial\pi(p_X,p_M,C,L)/\partial p_M = -\, M(0) < 0.$

Hence at least one of these two cross partial derivatives must be negative. If they are both negative (which is quite likely in practice), then $p_C'(0) > 0$ and $w'(0) < 0$; i.e., increasing the tariff increases the price of domestic consumption and decreases the return to primary inputs.

It seems to be extraordinarily difficult for the general public and politicians to understand the case for free trade. Thus every applied economist at least should be aware of the above argument for free trade.

The geometry of the above model is illustrated in Figure 13.

Figure 13: The Benefits of Free Trade and the Costs of a Tariff



The free trade equilibrium is at the point A where exports are equal to X(0), imports are equal to M(0) and domestic consumption is equal to C(0). When the tariff is increased from its initial level of 0 to some positive level t, exports and imports contract along the international trading line AO to the point B. Domestic production contracts as well to C(t) < C(0). As the tariff increases even more, trade continues to contract until the prohibitive tariff level t* is reached.

There are a number of limitations of the above model:

- We have assumed only a single finally demanded commodity. If there are many finally demanded commodities, then changing the tariff will in general change the relative prices of finally demanded goods and services and it becomes necessary to bring household preferences into the model.

- We have assumed only a single aggregate primary input for the economy. If there is more than one primary input, then it is quite possible that some factors will gain as the tariff increases and some will lose. Overall, losses in factor incomes will be bigger than the gains but if the losses are small and spread over many inputs while the gains due to increased tariffs are concentrated among only one or a few primary inputs, then these gainers will often successfully lobby governments for increased protection while the losers remain unorganized and powerless.
- We have not modelled the government sector adequately. In particular, it will always be necessary for governments to use some sort of distortionary taxes in order to raise revenue and it may be that taxing imports is the most effective way of raising revenue.[87] Also, the tariff distortions may offset other tax distortions in the economy and thus may not be as harmful as our model paints them.
- We have assumed competitive or price taking behavior in our model. Monopolistic behavior also introduces distortions into the economy and in some cases, increasing tariffs may offset these distortions to some extent.

Problems

28. Obtain counterparts to (245), (246), (251) and (252) if $t < 0$ so that imports are subsidized instead of taxed. As the subsidy level is increased in magnitude from 0 to $-t > 0$, what happens to the level of exports and imports? Illustrate your results in a counterpart to Figure 13.

29. Suppose that exports are taxed at the rate $t > 0$ instead of imports. The new equilibrium equation that is the counterpart to (234) is:

(i) $p_X\partial\pi((1-t)p_X,p_M,C(t),L)/\partial p_X + p_M\partial\pi((1-t)p_X,p_M,C(t),L)/\partial p_M = 0.$

(a) Calculate $C'(0)$ and $C''(0)$ for this new model.
(b) Calculate a counterpart to the deadweight loss formula (250), $DWL(t) \equiv [C(0) - C(t)]/C(0)$, for this new model.
(c) Calculate counterparts to (251) and (252); i.e., calculate the derivatives $X'(0)$ and $M'(0)$ and see if you can sign these derivatives.

14. Samuelson's Le Chatelier Principle and Profit Functions

As can be seen from the results in the previous section, the magnitude of various elasticities plays a key role in many economic models. In this section, we will show that

---

[87] This is particularly true for relatively poor countries where it may be extremely difficult to impose effective income taxes or commodity taxes.

elasticities tend to become larger in magnitude in long run models where "fixed" factors can adjust compared to short run models where many inputs are temporarily fixed.[88]

Suppose that the firm's technology set S satisfies assumptions (165) and (166) above. Then for variable net output price vector $p >> 0_M$ and fixed input vector $x \geq 0_N$, the firm's *short run profit function* $\pi(p,x)$ is defined as follows:

(256) $\pi(p,x) \equiv \max_y \{p^Ty: (y,x)\in S\}$.

Given a strictly positive vector of input prices $w >> 0_N$ for the "fixed" inputs, the firm's *long run profit function* $\Pi(p,w)$ is defined as follows:[89]

(257) $\Pi(p,w) \equiv \max_{y,x} \{p^Ty - w^Tx: (y,x)\in S\}$.

Thus the short run profit maximization problem holds the vector of inputs x fixed whereas the long run profit maximization problem allows x to be variable but holds the corresponding "fixed" input prices constant.

Let $p^* >> 0_M$ and $w^* >> 0_N$ and suppose that (y^*,x^*) solves the following long run profit maximization problem:

$$
\begin{aligned}
(258)\ \Pi(p^*,w^*) &\equiv \max_{y,x} \{p^{*T}y - w^{*T}x: (y,x)\in S\} \\
&= p^{*T}y^* - w^{*T}x^* \qquad \text{where } (y^*,x^*)\in S.
\end{aligned}
$$

Using (258), it is easy to show that y* solves the following maximization problem:

$$
\begin{aligned}
(259)\ \Pi(p^*,w^*) &= \max_y \{p^{*T}y - w^{*T}x^*: (y,x^*)\in S\} && \text{where we have fixed } x = x^* \text{ in (258)} \\
&= \max_y \{p^{*T}y: (y,x^*)\in S\} - w^{*T}x^* && \text{rearranging terms} \\
&= \pi(p^*,x^*) - w^{*T}x^* && \text{using definition (256).}
\end{aligned}
$$

Assume that $\Pi(p,w^*)$ and $\pi(p,x^*)$ are twice continuously differentiable with respect to the components of p at $p = p^*$. Now let $p >> 0_M$ and consider the following long run profit maximization problem:

$$
\begin{aligned}
(260)\ \Pi(p,w^*) &\equiv \max_{y,x} \{p^Ty - w^{*T}x: (y,x)\in S\} \\
&\geq \max_y \{p^Ty - w^{*T}x^*: (y,x^*)\in S\} \\
&= \pi(p,x^*) - w^{*T}x^* && \text{using definition (256)}
\end{aligned}
$$

---

[88] The original results along these lines were established by Samuelson (1947; 3-38) using primal optimization techniques. Pollak (1969; 75-77) and Diewert (1974a; 146-150) adapted Samuelson's results to the consumer and producer contexts using duality theory.

[89] In order to ensure that this maximum exists, some additional assumptions will generally be required; i.e., we need to assume that there are some fixed factors lurking in the background that prevent long run profits from becoming infinite.

where the inequality in (260) follows from the fact that the feasible set is larger in the first maximization problem than in the second problem and both problems have the same objective function.

Define the function of p for $p >> 0_M$ as follows:

(261) $f(p) \equiv \Pi(p,w^*) - [\pi(p,x^*) - w^{*T}x^*] \geq 0$

where the inequality follows using (260). Using (259), we have

(262) $f(p^*) = 0$.

But (261) and (262) show that f(p) attains a global minimum at p = p*. Since Π(p,w*) and π(p,x*) are assumed to be twice continuously differentiable with respect to the components of p at p = p*, the following *first and second order necessary conditions for minimizing a differentiable function* f(p) at p = p* hold:

(263) $\nabla_p f(p^*) = \nabla_p \Pi(p^*,w^*) - \nabla_p \pi(p^*,x^*) = 0_M$;
(264) $\nabla_{pp}^2 f(p^*) = \nabla_{pp}^2 \Pi(p^*,w^*) - \nabla_{pp}^2 \pi(p^*,x^*)$ is a positive semidefinite matrix.

The producer's system of *short run variable commodity net supply functions*, y(p,x), can be defined using Hotelling's Lemma:

(265) $y(p,x) \equiv \nabla_p \pi(p,x)$.

Similarly, the producer's system of *long run variable commodity net supply functions*, Y(p,x), can be defined using Hotelling's Lemma:

(266) $Y(p,w) \equiv \nabla_p \Pi(p,w)$.

Equations (263) show that at the initial equilibrium, which is both a short and long run equilibrium of the firm, long and short run net supplies are equal; i.e., we have

(267) $y^* \equiv y(p^*,x^*) = Y(p^*,w^*)$.

Differentiating (265) with respect to the components of p gives us the M by M matrix of *short run net supply derivatives with respect to variable prices*:

(268) $\nabla_p y(p,x) \equiv \nabla_{pp}^2 \pi(p,x)$.

Differentiating (266) with respect to the components of p gives us the M by M matrix of *long run net supply derivatives with respect to variable prices*:

(269) $\nabla_p Y(p,w) \equiv \nabla_{pp}^2 \Pi(p,w)$.

(264), (268) and (269) imply that the long run supply derivatives $\nabla_p Y(p^*,w^*)$ are related to the corresponding short run supply derivatives $\nabla_p y(p,x)$ as follows:

(270) $z^T\nabla_p Y(p,w)z \geq z^T\nabla_p y(p,x)z$ for all $z \neq 0_M$.

In (270), letting $z = e_m$, the mth unit vector, implies the following inequalities:

(271) $\partial Y_m(p^*,w^*)/\partial p_m \geq \partial y_m(p^*,x^*)/\partial p_m \geq 0$[90] ; $m = 1,\ldots,M$.

If variable commodity m is an output, then y_m^*, the mth component of y^* defined by (267), is *positive* and we can convert the mth inequality in (271) into the following *elasticity inequality*:

(272) $[p_m^*/y_m^*][\partial Y(p^*,w^*)/\partial p_m] \geq [p_m^*/y_m^*][\partial y(p^*,x^*)/\partial p_m] \geq 0$.

Thus the long run own price output supply elasticity for commodity m is equal to or greater than the corresponding short run own price output supply elasticity.

If variable commodity m is an input, then y_m^*, the mth component of y^* defined by (267), is *negative* and dividing through inequality m in (271) by the negative y_m^* will reverse the inequality. In this case, we obtain the following *elasticity inequality*:

(273) $[p_m^*/y_m^*][\partial Y(p^*,w^*)/\partial p_m] \leq [p_m^*/y_m^*][\partial y(p^*,x^*)/\partial p_m] \leq 0$.

Since the two input demand elasticities in (273) are negative (or zero), it can be seen that the *long run own price input demand elasticity for commodity m is equal to or greater in magnitude than the corresponding short run own price input demand elasticity*.

Thus no matter whether variable commodity m is an output or an input, own price elasticities of supply or demand will tend to increase in magnitude as additional inputs become variable in the long run. This is an important observation that should be kept in mind by the applied economist. The long run effects of policy changes can often be much larger than the short run effects.[91]

Samuelson explains the significance of his Le Chatelier principle as follows:

---

[90] This second set of inequalities follows from the fact that $\nabla_p y(p,x) \equiv \nabla_{pp}^2\pi(p,x)$ is a positive semidefinite matrix and hence the diagonal elements of this matrix are nonnegative.

[91] Hicks (1946; 206) hinted at this observation: "The additional output which can be produced in the current week, or planned for weeks in the near future, will usually be quite small. The initial equipment, which the entrepreneur possesses at the planning date, will generally contain, in a nearly finished form, most of the output which can be produced in the present and near future; since there can only exist a limited amount of these nearly finished goods, the flexibility of output in response to any change in price will necessarily be small. But there is no such check on the expansion of distant future outputs; or rather the check gets less and less strong as the output recedes into the future." Diewert (1985) adapts the Le Chatelier principle to the study of deadweight loss and concludes that dynamic deadweight losses can be considerably larger than discounted static losses since the distortions induce inappropriate investments.

"This explains why economically long run demands are more elastic than those in the short run. A lengthening of the time period so as to permit new factors to be varied will result in *greater* changes in the factor whose price has changed, *regardless* of whether the factors permitted to vary are complementary or competitive with the one whose price has changed." Paul A. Samuelson (1947; 38-39).

References

Allen, R.G.C. (1938), Mathematical Analysis for Economists, London: Macmillan.

Arrow, K.J., H.B. Chenery, B.S. Minhas and R.M. Solow (1961), "Capital-Labor Substitution and Economic Efficiency", Review of Economic Statistics 63, 225-250.

Berndt, E.R. and L.R, Christensen (1974), "Testing for the Existence of a Consistent Aggregate Index of Labor Inputs", *American Economic Review* 64, 391-404.

Blackorby, C. (1975), "Degrees of Cardinality and Aggregate Partial Orderings", *Econometrica* 43, 845-852.

Blackorby, C. and W.E. Diewert (1979), "Expenditure Functions, Local Duality and Second Order Approximations", *Econometrica* 47, 579-601.

Blackorby, C., D. Primont and R.R. Russell (1978), *Duality, Separability and Functional Structure: Theory and Economic Applications*, New York: North-Holland.

Chipman, J.S. (1966), "A Survey of the Theory of International Trade: Part 3: The Modern Theory", *Econometrica* 34, 18-76.

Cobb, C. and P.H. Douglas (1928), "A Theory of Production", *American Economic Review*, Supplement, 18, 139-165.

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

Diewert, W.E. (1973), "Functional Forms for Profit and Transformation Functions", *Journal of Economic Theory* 6, 284-316.

Diewert, W.E. (1974a), "Applications of Duality Theory", pp. 106-171 in *Frontiers of Quantitative Economics*, Volume 2, M.D. Intriligator and D.A. Kendrick (eds.), Amsterdam: North-Holland.

Diewert, W.E. (1974b), "A Note on Aggregation and Elasticities of Substitution", *Canadian Journal of Economics* 7, 12-20.

Diewert, W.E. (1974c), "Functional Forms for Revenue and Factor Requirements Functions", *International Economic Review* 15, 119-130.

Diewert, W.E. (1978), “Hicks' Aggregation Theorem and the Existence of a Real Value Added Function”, pp. 17-51, Vol. 2, *in Production Economics: A Dual Approach to Theory and Applications*, M. Fuss and D. McFadden, editors, North-Holland, Amsterdam.

Diewert, W.E. (1980), “Symmetry Conditions for Market Demand Functions”, *Review of Economic Studies* 47, 595-601.

Diewert, W.E. (1981), “The Comparative Statics of Industry Long Run Equilibrium”, *Canadian Journal of Economics* 14, 78-92.

Diewert, W.E. (1982), “Duality Approaches to Microeconomic Theory”, pp. 535-599 in *Handbook of Mathematical Economics*, Volume 2, K.J. Arrow and M.D. Intriligator (eds.), Amsterdam: North-Holland.

Diewert, W.E. (1985), “A Dynamic Approach to the Measurement of Waste in an Open Economy”, *Journal of International Economics* 19, 213-240.

Diewert, W.E. (1993), “Duality Approaches to Microeconomic Theory”, pp. 105-175 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland. This paper is a rewrite of Diewert (1982) but it also includes proofs.

Diewert, W.E. and D. Lawrence (2002), “The Deadweight Costs of Capital Taxation in Australia”, pp. 103-167 in *Efficiency in the Public Sector*, K.J. Fox (ed.), Boston: Kluwer Academic Publishers.

Diewert, W.E. and T.J. Wales (1987), “Flexible Functional Forms and Global Curvature Conditions”, *Econometrica* 55, 43-68.

Diewert, W.E. and T.J. Wales (1992), “Quadratic Spline Models for Producer’s Supply and Demand Functions”, *International Economic Review* 33, 705-722.

Diewert, W.E. and A.D. Woodland (1977), “Frank Knight’s Theorem in Linear Programming Revisited”, *Econometrica* 45, 375-398.

Fenchel, W. (1953), “Convex Cones, Sets and Functions”, Lecture Notes at Princeton University, Department of Mathematics, Princeton, N.J.

Gale, D, V.L. Klee and R.T. Rockafellar (1968), “Convex Functions on Convex Polytopes”, *Proceedings of the American Mathematical Society* 19, 867-873.

Gorman, W.M. (1968), “Measuring the Quantities of Fixed Factors”, pp. 141-172 in *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, J.N. Wolfe (ed.), Chicago: Aldine.

Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: Clarendon Press.

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

Hotelling, H. (1935), "Demand Functions with Limited Budgets", *Econometrica* 3, 66-78.

Kohli, U.R.J. (1978), "A Gross National Product Function and the Derived Demand for Imports and Supply of Exports", *Canadian Journal of Economics* 11, 167-182.

Kohli, U. (1991), *Technology, Duality and Foreign Trade: The GNP Function Approach to Modelling Imports and Exports*, Ann Arbor, MI: University of Michigan Press.

Leontief, W.W. (1941), *The Structure of the American Economy 1919-1929*, Cambridge, MA: Harvard University Press.

McFadden, D. (1966), "Cost, Revenue and Profit Functions: A Cursory Review", IBER Working Paper No. 86, University of California, Berkeley.

McFadden, D. (1978), "Cost, Revenue and Profit Functions", pp. 3-109 in *Production Economics: A Dual Approach*, Volume 1, M. Fuss and D. McFadden (eds.), Amsterdam: North-Holland.

Pollak, R.A. (1969), "Conditional Demand Functions and Consumption Theory", *Quarterly Journal of Economics* 83, 60-78.

Rockafellar, R.T. (1970), *Convex Analysis*, Princeton, N.J.: Princeton University Press.

Samuelson, P.A. (1947), *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Samuelson, P.A. (1953-54), "Prices of Factors and Goods in General Equilibrium", *Review of Economic Studies* 21, 1-20.

Samuelson, P.A. (1967), "The Monopolistic Competition Revolution", pp. 105-138 in *Monopolistic Competition Theory: Studies in Impact*, R.E. Kuenne (ed.), New York: John Wiley.

Samuelson, P.A. (1974), "Complementarity—An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory", *The Journal of Economic Literature* 12, 1255-1289.

Shephard, R.W. (1953), *Cost and Production Functions*, Princeton N.J.: Princeton University Press.

Shephard, R.W. (1967), "The Notion of a Production Function", *Unternehmenforschung* 11, 209-232.

Shephard, R.W. (1970), *Theory of Cost and Production Functions*, Princeton N.J.: Princeton University Press.

Uzawa, H. (1962), "Production Functions with Constant Elasticities of Substitution", *Review of Economic Studies* 29, 291-299.

Uzawa, H. (1964), "Duality Principles in the Theory of Cost and Production", *International Economic Review* 5, 291-299.

Walters, A.A. (1961), "Production and Cost Functions: An Econometric Survey", *Econometrica* 31, 1-66.

Wold, H. (1944), "A Synthesis of Pure Demand Analysis; Part 3", *Skandinanvisk Aktuarietidskrift* 27, 69-120.

Woodland, A.D. (1972), "Factor Demand Functions for Canadian Industries, 1946-1969", Department of Manpower and Immigration, Strategic Planning and Research, Ottawa.

Woodland, A.D. (1982), *International Trade and Resource Allocation*. Amsterdam: North Holland.