INTERNAL PRODUCTION AND COST FUNCTIONS IN CONNECTION WITH ECONOMIES OF SCALE

Kairat T. Mynbaev
Professor da Universidade Federal de Roraima

Resumo: *O princípio bem conhecido da dualidade entre funções de custo e produção afirma que algum conceito definido em termos de uma função de produção tem uma definição "dual", em termos da função de custo associada e vice-versa. Nós estudamos a relação da dualidade de duas formas: definimos várias noções de economias internas de escala (em termos da função de custo), para os níveis iniciais e finais "($\to \infty$)" de insumos, e encontramos a "dual" deles em termos da respectiva função de produção. Por exemplo, provamos que o custo médio cresce indefinidamente quando o produto tende ao infinito, se e somente se a função de produção "f" cresce ao infinito mais lentamente do que qualquer função linear inclinada positivamente. De outra forma, para uma tecnologia exibir economias iniciais de escala é necessário e suficiente que "f" seja "maiorizada" por qualquer função linear inclinada positivamente para todos os níveis iniciais de insumos. Esses resultados levam-nos a descrever exatamente todas as tecnologias que têm curvas de custo médio em forma de "U". Monotonicidade estrita e continuidade da função de produção são as mais fortes suposições que nós fazemos. Problemas desse tipo não podem ser resolvidos usando cálculo ou o Lema de Shepard, no que diz respeito à dualidade. A abordagem direta desenvolvida neste trabalho não requer diferenciabilidade ou concavidade (convexidade).*

Palavras-chave: *Função de Produção; Função de Custo; Economias de Escola; Teoria da Produção.*

1 INTRODUCTION

Various production processes in an economy can be described using a notion of a production function. A production function is a mathematical relationship showing the maximum output that can be produced using given inputs. We study only the case of a single-output technology; the number of inputs in this Introduction will also be assumed to be one for simplicity. Thus, a production function can be written as

$$y = f(x)$$

where:
y = output;
x = input (all variables are non-negative).

It only describes a purely physical ability of the production unit to transform inputs into outputs.

One of the notions which make it more economically relevant is that of a cost function. Given the input price "w" it shows the minimum cost of producing, at least the given level of output "y" and is formally written as

$$c(y, w) = \min\{wx: f(x) \geq y\}.$$

Mention that if "f" is continuous and monotonic; then in the one-input case the cost function can be easily found

$$c(y, w) = wf^{1}(x).$$

It is less obvious that under some restrictions on "f", it can be restored from its cost function. The way it can be restored depends on the assumptions which include either differentiability or concavity of "f". This important result (due to Shephard) is called duality between production functions and cost functions. More generally, the duality principle is understood as follows: any concept defined in terms of the production function has a "dual" definition in terms of the cost function and vice versa (Varian, p.81).

In the Main Results section we study duality relationships of two kinds. We define several notions of internal economies of scale for large and small inputs (in terms of the cost function) and find their duals in terms of the respective production function.

For example, in THEOREM 1 we prove that the average cost increases indefinitely as output goes to infinity, if and only if the production function "f" grows at infinity slower than any positively sloped linear function. On the other hand, one of the statements of THEOREM 2 says that *for a technology to exhibit increasing initial economies of scale, it is necessary* and sufficient that "f" be majorized by any positively sloped linear function for all small inputs. These results allow us to exactly describe in Theorem 3 all technologies that have U-shaped average cost curves (for the intuition behind this notion see, e.g., Varian, section 5.2.).

The basic intuition behind these results is simple: the faster grows the productive capacity of the production unit, the slower is the growth of the associated cost function. Under some very weak conditions, the properties of the cost function for large (small) outputs are only affected by a behavior of the production function for large (small, resp.) inputs. Exact quantitative statements, however, are less easy to explain.

The questions we ask have trivial answers for the Cobb-Douglas and some other production functions often used by economists. We try to obtain general answers under restriction as weak as possible. Problems of this kind cannot be solved using calculus or the Shephard's duality result. The direct approach developed in this paper does not require differentiability or concavity (convexity). Strict monotonicity and continuity of the production function are the strongest assumptions we make.

It turns out that functional inequalities are an appropriate way to answer the questions under consideration. The linear function

$$G(x) = x_1 + \ldots + x_n \qquad (1.1)$$

and the Leontief function

$$L_\alpha(x) = \min\{\alpha_1 x_1, \ldots, \alpha_n x_n\},\ \alpha_i > 0, \qquad (1.2)$$

arise on several occasions as characterizing "boundary" or limiting behavior of production functions.

NOTATION: Throughout the paper, "R^n" denotes an Euclidean space provided with the dot product

$$(x, y) = \sum_{i=1}^{n} x_i y_i$$

and the corresponding norm

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

"$R_+ = [0, \infty)$", "Q_r" is a quadrant

$\{x \in R^n: x_i \geq r, i = 1, ..., n\}, r \geq 0.$

For vectors "w", "v", the inequality "$w \geq v$" means that "$w_i \geq v_i$" for all "i". The notation "$w > v$" is used when "$w_i > v_i$" for all "i".

Theorems 1-3 have been proved by the author in his MS thesis (Mynbaev). THEOREM 4 and most of the remarks have been added for this publication.

2 MAIN RESULTS

First, we define various concepts of internal economies of scale in terms of cost functions and show that they can be equivalently expressed in terms of functional bounds on production functions. Then these results are applied to U-shaped average cost curves and production functions exhibiting increasing, constant, or decreasing returns to scale.

DEFINITION 1: by a production function we mean any function "f" which is defined and non-negative in "Q_0" and vanishes at the origin. We call it also a "physical production function" (PPF) when we want to emphasize that it does not necessarily possess the following "monotonicity" property:

$$f(x) \geq f(y), \forall x \geq y \geq 0. \quad (2.1)$$

If "f" is bounded on each set of the form "$\{x: \|x\| \leq N\}$", we say that "f" is "locally" bounded. Remind that for a "strictly monotonic" production function "x < y implies f(x) < f(y)".

REMARK 1: a production technology can be equivalently described in terms of input requirement sets. A production function "f" generates technology (a set of sets) "$\tau = \{T(y): y \geq 0\}$", where the "input requirement set"

$$T(y) = \{x: f(x) \geq y\}$$

shows all input vectors that are capable of producing "y". Conversely, if the set "τ" is given, then the corresponding production function can be found by

$$f(x) = \sup\{y: x \in T(y)\}.$$

It is useful to state properties of input requirement sets implied by our definition of a production function:

T.1 - for some "y > 0 T(y)" may be empty (*it happens if and only if* "f" is bounded);

T.2 - since "f" vanishes at the origin, we have "$0 \in T(0)$";

T.3 - for the same reason "$0 \notin T(y)$ if y > 0". It is possible, however, that "$0 \in \overline{T(y)}$" for some "y > 0", because input requirement sets are not required to be closed;

T.4 - *if "f" is monotonic, then from "$x \in T(y)$" and "$x' \geq x$" it always follows* that "$x' \in T(y)$". For a physical production function this property may not hold;

T.5 - input requirement sets are obviously non-increasing in "y": if "$y' \geq y$", then "$T(y') \subset T(y)$";

T.6 - the correspondence "τ" is upper semicontinuous for a strictly monotonic "f", but may not be such for a physical or monotonic production function;

T.7 - since "f" is finite everywhere in "Q_0", we have $\bigcap_{y>0} T(y) = \varnothing$.

Let "f" be a production function and

$$c(y, w) = \inf\{(w, x): f(x) \geq y\}, y \geq 0, w \geq 0$$

the associated cost function where

$$(w, x) = \sum_{i=1}^{n} w_i x_i$$

is a dot product. If the input requirement set "T(y)" is empty (for example, if "f" is bounded,

$$f(x) \leq M, \forall x \in Q_0.$$

and "y > M"), then we assume by definition

$$c(y, w) = \infty.$$

The average cost function "AC" is defined by

$$AC(y) = c(y, w)/y.$$

REMARK 2: here we want to comment on properties of the cost function implied by our definition of the production function:

C.1 - denote by "$\overline{R}_+$" the extended half-axis [0, ∞]. The cost function is a function of variables "$y \in R_+$" and "$w \in R_+^n$" with values in "$\overline{R}_+$". Defining "$c(y, w)$" to be equal to "∞" for some "y" and "w" has a prohibitive meaning: such "y" at such prices "w" cannot be produced at a finite cost and therefore will not be produced;

C.2 - monotonicity in factor prices: if "$w' \geq w$", then "$c(y, w') \geq c(y, w)$";

C.3 - homogeneity of degree 1 in prices:

$$c(y, tw) = t\, c(y, w), \forall\ t > 0;$$

C.4 - concavity in prices:

$$c(y, tw + (1-t)w') \geq tc(y,\ w) + (1-t)c(y,\ w'),\ \forall\ t \in [0,1];$$

C.5 - let "$y > 0$" be fixed. If "$c(y, w)$" is finite for one "$w > 0$", then it is finite for all "$w > 0$". We shall need a quantitative version of this statement. Let "$w^1 > 0$, $w^2 > 0$". By definition of "$c(y, w^1)$", for any "$\varepsilon > 0$" there exists "$x = x(y, w^1, \varepsilon)$" such that

$$(w^1, x) \leq c(y, w^1) + \varepsilon,\ f(x) \geq y.$$

Obviously,

$$(w^2 - w^1,\ x) = \sum \frac{w_i^2 - w_i^1}{w_i^1} w_i^1 x_i \leq \max_i \left|w_i^2 / w_i^1 - 1\right| (w^1,\ x).$$

Therefore,

$$c(y, w^2) \leq (w^2, x) = (w^1, x) + (w^2 - w^1, x) \leq$$

$$\leq c(y, w^1) + \varepsilon + \max_i \left|w_i^2/w_i^1 - 1\right| (c(y, w^1) + \varepsilon).$$

Since "ε" can be chosen arbitrarily small, this tells us that

$$c(y, w^2) \leq \lambda_{1,2} c(y, w^1)$$

where

$$\lambda_{1,2} \equiv \lambda(w^1, w^2) = 1 + \max_i \left|w_i^2/w_i^1 - 1\right|.$$

Thus, by symmetry:

$$\lambda_{2,1} c(y, w^1) \leq c(y, w^2) \leq \lambda_{1,2} c(y, w^1). \qquad (2.2)$$

It follows, in particular, that "$c(y, w)$" is continuous in "$w > 0$" because "$\lambda_{1,2} \to 1$" when "$w^2 \to w^1$";

C.6 - "$c(y, w)$" is monotonic in "y": if "$y' \geq y$", then "$c(y', w) \geq c(y, w)$". The cost function is not necessarily continuous in output.

In all equations in C.2 - C.6, if the smaller of the quantities compared is "∞", then the larger is too. We would like to stress that under such general conditions the duality result between cost and production functions may not hold. Remind also that the correspondence between cost and production functions is not one-to-one (see examples in Varian).

Denote

$$h(w) = \inf\{(w, x) : \sum x_i \geq 1\}, \qquad H(w) = \max_i w_i .$$

To prove THEOREM 1 below, we need a series of simple propositions:

1) The implication

$$\sum x_i \geq 1 \Rightarrow (w, x) \geq \min_i w_i \sum x_i \geq \min_i w_i$$

shows that "h(w)" cannot be less than "$\min_i w_i$". Combining this fact with the implication

$$x_j = 1,\ x_i = 0,\ \forall\, i \neq j, \Rightarrow (w, x) = w_j,\ \sum x_j = 1$$

we get:

$$h(w) = \min_i w_i \tag{2.3}$$

2) For any "$c_1 > 0, c_2 < y$" we have:

$$\inf\{(w, x): c_1 \sum x_i + c_2 \geq y\} = \tag{2.4}$$

$$= \inf\left\{\left\{w\frac{y-c_2}{c_1}, \frac{c_1}{y-c_2}x\right\}: \sum \frac{c_1}{y-c_2}x_i \geq 1\right\} = \frac{y-c_2}{c_1}h(w);$$

3) If "f "is strictly monotonic, then

$$\{x: f(x) \geq f(y)\} \subset \{x\ Q_0 : x_i \geq y_i \text{ for some "i"}\}$$

so that

$$\inf\{(w,x): f(x) \geq f(y)\} \geq$$

$$\geq \inf\{(w,x): x_i \geq y_i \text{ for some } i\} = \min_i m_i$$

where

$$m_i = \inf\{(w,x): x_i \geq y_i, x_j \geq 0 \ \forall j \neq i\} =$$

$$= \inf\left\{\sum_{j \neq i} w_j x_j + w_i (x_i - y_i) + w_i y_i : x_i - y_i \geq 0, x_j \geq 0 \ \forall j \neq i\right\} =$$

$$= w_i y_i + \inf\{(w,x): x \geq 0\} = w_i y_i.$$

Thus:

$$\inf\{(w,x): f(x) \geq f(y)\} \geq \min_i w_i y_i \geq h(w) \min_i y_i; \qquad (2.5)$$

4) If "$y \geq cr$", where "c" and "r" are some positive constants, then:

$$\inf\{(w,x): c \min_i x_i \geq y, \min_i x_i \geq r\} = \qquad (2.6)$$

$$= \inf\left\{\sum_i w_i (x_i - y/c) + y/c \sum_i w_i : \min_i (x_i - y/c) \geq 0\right\} =$$

$$= y/c \sum_i w_i + \inf\{(w,x): x \geq 0\} = y/c \sum_i w_i.$$

DEFINITION 2: we say that a production function "f" (or the technology "τ") exhibits "decreasing economies of scale" if

$$\text{(DES)} \qquad \lim_{y \to \infty} AC(y,w) = \infty, \forall w > 0,$$

"constant economies of scale" if both

(CES′) $$\liminf_{y \to \infty} AC(y, w) > 0, \forall w > 0,$$

and

(CES″) $$\limsup_{y \to \infty} AC(y, w) < \infty, \forall w > 0,$$

hold, and "increasing economies of scale" if

(IES) $$\lim_{y \to \infty} AC(y, w) = 0, \forall w > 0.$$

REMARK 3: to see if a technology exhibits some kind of economies of scale defined above it is in fact sufficient to check a corresponding condition just for one "w > 0". Indeed, because of (2.2) conditions (DES) through (IES) are equivalent to

(DES) $$\exists\, w > 0 \text{ such that } \lim_{y \to \infty} AC(y, w) = \infty,$$

(CES′) $$\exists\, w > 0 \text{ such that } \liminf_{y \to \infty} AC(y, w) > 0,$$

(CES″) $$\exists\, w > 0 \text{ such that } \limsup_{y \to \infty} AC(y, w) < \infty,$$

(IES) $$\exists\, w > 0 \text{ such that } \lim_{y \to \infty} AC(y, w) = 0,$$

respectively. A similar remark is true with respect to initial economies of scale introduced later in DEFINITION 3. We call this phenomenon a "crowd principle".

THEOREM 1: let f be a "strictly monotonic production function". Then (DES) is true if and only if

(DES.1)

$$\begin{cases} \text{for any } \varepsilon \in (0, 1) \text{ there exists } C(\varepsilon) > 0 \text{ such that} \\ \qquad f(x) \le \varepsilon \sum x_i + C(\varepsilon), \quad \forall x \ge 0. \end{cases}$$

(CES′) and (CES″) are equivalent to

(CES′.1)

$$\begin{cases} \text{there exists } c > 0 \text{ such that} \\ f(x) \leq c(\sum x_i + 1), \quad \forall x \geq 0, \end{cases}$$

and

(CES″.1)

$$\begin{cases} \text{there exists } c > 0 \text{ such that for some } r > 0 \\ f(x) \geq c \min_i x_i, \forall x \in Q_r, \end{cases}$$

respectively. Finally, the condition

(IES.1)

$$\begin{cases} \text{for any } \varepsilon \in (0, 1) \text{ there exists } r = r(\varepsilon) > 0 \text{ such that} \\ f(x) \geq \frac{1}{\varepsilon} \min_i x_i, \forall x \in Q_r, \end{cases}$$

is necessary and sufficient for (IES).

PROOF. if (DES.1) holds, then for any "$y \geq 2C(\varepsilon)$" we have "$y/2 - C(\varepsilon) \geq 0$" and

$$\{x: f(x) \geq y\} \subset \{x: \varepsilon \sum x_i + C(\varepsilon) \geq y\}$$

so that (2.4) gives

$$c(y, w) \geq \inf\{(w, x) : \varepsilon \sum x_i + C(\varepsilon) \geq y\} = \frac{y - C(\varepsilon)}{\varepsilon} h(w) =$$

$$= \frac{1}{\varepsilon}\left(\frac{y}{2} + \frac{y}{2} - C(\varepsilon)\right)h(w) \geq \frac{h(w)}{2\varepsilon} y, \quad \forall y \geq 2C(\varepsilon).$$

Hence

$$\liminf_{y \to \infty} AC(y, w) \geq h(w)/2\varepsilon$$

and (DES) follows because "$\varepsilon > 0$" can be chosen arbitrarily small.

Suppose that (DES.1) does not hold. Then there exists "$\varepsilon_0 > 0$" such that for any natural "$N > 0$" a vector "$x^N > 0$" can be found such that

$$f(x^N) > \varepsilon_0 \sum x_i^N + N.$$

Put

$$y_N = \varepsilon_0 \sum x_i^N + N.$$

Then

$$y_N \to \infty, f(x^N) > y_N$$

and by definition of infimum

$$c(y_N, w) \leq (w, x^N) \leq H(w) \sum x_i^N = \frac{H(w)}{\varepsilon_0} (y_N - N) \leq \frac{H(w)}{\varepsilon_0} y_N.$$

Thus,

$$\limsup_{N \to \infty} AC(y_N, w) \leq \frac{H(w)}{\varepsilon_0}$$

which contradicts (DES). It means that (DES) $\Rightarrow$ (DES.1).

To deduce (CES') from (CES'.1), mention that (CES'.1) and (2.4) give

$$c(y, w) \geq \inf\{ (w, x) : c(\sum x_i + 1) \geq y\} =$$

$$= \frac{y - c}{c} h(w) \geq \frac{h(w)}{2c} y, \ \forall y \geq 2c,$$

which implies (CES').

Conversely, suppose (CES'.1) is violated. Then for every natural "N" there exists "X^n" such that

$$f(x^N) > N(\sum x_i^N + 1) .$$

Put

$$y_N = N(\sum x_i^N + 1).$$

Then "$y_N \to \infty$" and

$$c(y_N , w) \leq (w, x^N) \leq H(w) \sum x_i^N = H(w) (\frac{y_N}{N} - 1) \leq \frac{H(w)}{N} y_N .$$

Thus, (CES') cannot be true. We have proved that (CES') is equivalent to (CES'.1).

If (CES".1) holds, then use (2.6) to get

$$c(y, w) \leq \inf\{ (w, x) : f(x) \geq y, \ x \in Q_r\} \leq$$
$$\leq \inf\{ (w, x) : c \min_i x_i \geq y, \ x \in Q_r\} = \frac{y}{c} \sum w_i$$

from which (CES") follows immediately.

Conversely, suppose (CES".1) is violated. Then for any "$\varepsilon \in (0,1)$" there exists a sequence "$\{x^N\}$" such that

$$f(x^N) < \varepsilon \min_i x_i^N, \quad \min_i x_i^N \to \infty \tag{2.7}$$

Consider two cases: a) suppose the sequence "$\{f(x^N)\}$" is bounded. Then from (2.1) and (2.7) it follows that f is bounded and (CES″) is rejected trivially:

$$AC(y,w) \equiv \infty, \forall y > \sup_x f(x);$$

b) now let "$\{f(x^N)\}$" be unbounded. Then using monotonicity we see that

$$y_N \equiv f(x^N) \to \infty.$$

Further, (2.5) leads to a lower bound

$$c(y_N, w) = \inf\{(w,x) : f(x) \geq f(x_N)\} \geq h(w) \min_i x_i^N > \frac{h(w)}{\varepsilon} y_N$$

which again allows us to reject (CES″).

From (IES.1) and (2.6) we have

$$c(y, w) \leq \inf\{(w,x) : f(x) \geq y, \quad x \in Q_r\} \leq$$

$$\leq \inf\{(w,x) : \frac{1}{\varepsilon} \min_i x_i \geq y, \quad \min_i x_i \geq r\} = \varepsilon y \sum w_i, \ \forall y \geq \frac{r}{\varepsilon},$$

so (IES.1) is sufficient for (IES).

To prove the necessity, suppose that (IES.1) is violated. Then there exist "$c > 0$" and a sequence "$\{x^N\}$" which satisfy

$$f(x^N) < c \min_i x_i^N, \quad \min_i x_i^N \to \infty.$$

Consider two cases as above: a) If "$\{f(x^N)\}$" is a bounded sequence, then "f" is bounded and we have DES instead of IES; b) If "$\{f(x^N)\}$" is unbounded, then.

$$y^N \equiv f(x^N) \to \infty$$

and using (2.5) we arrive at

$$c(y_N, w) \geq h(w) \min_i x_i^N > \frac{h(w)}{c} y_N$$

which excludes (IES). The proof is complete.

REMARK 4:

a) the proof of the equivalences

(DES) ⇔ (DES.1), (CES') ⇔ (CES'.1)

does not use the monotonicity assumption so that they hold for physical production functions;

b) (DES.1) means that "f" grows slower than any (positively sloped) linear function. If "f" is locally bounded, then because of the inequality

$$\frac{1}{\sqrt{n}} \|x\| \leq \sum x_i \leq \sqrt{n}\|x\|$$

(DES.1) is equivalent to

$$\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = 0 ;$$

c) (CES'.1) states that "f" is majorized by a linear function (1.1) for all inputs. If "f" is locally bounded, then (CES'.1) is equivalent to

$$\limsup_{\|x\| \to \infty} \frac{f(x)}{\|x\|} < \infty ;$$

d) (CES".1) allows to support "f" in some quadrant "Q" by a Leontief function (1.2). Define a function

$$h(x) = \left(\frac{1}{x_1}, \ldots, \frac{1}{x_n}\right).$$

Using the equations

$$\frac{1}{\min\limits_i x_i} = \max_i \frac{1}{x_i}, \qquad \max_i x_i \leq \|x\| \leq \sqrt{n} \max_i x_i,$$

it is easy to show that (CES″.1) is equivalent to

$$\liminf_{\|h(x)\| \to 0} f(x)\|h(x)\| > 0;$$

e) (IES.1) means that "f" grows in the north-east direction faster than any Leontief function. (IES.1) is equivalent to

$$\lim_{\|h(x)\| \to 0} f(x)\|h(x)\| = \infty;$$

f) In (DES.1) and (IES.1) we restrict values of "ε to (0, 1)" just to stress that zero is the point of interest.

We need a few more auxiliary propositions for the next theorem which gives a similar description of the average cost behavior for infinitesimal quantities of output.

5) The inequality

$$h(w) \inf\{\sum x_i : \ f(x) \geq y\} \leq c(y, w) \leq H(w) \inf\{\sum x_i : \ f(x) \geq y\}$$

leads to an equivalence

$$\lim_{y \to 0} \inf\{\sum x_i : \ f(x) \geq y\} = 0 \Leftrightarrow \lim_{y \to 0} c(y, w) = 0. \qquad (2.8)$$

6) Let us prove that

$$\inf\{ (w, x) : \ f(x) > 0 \} = \lim_{y \to +0} c(y, w). \qquad (2.9)$$

The quantity at the left-hand side does not exceed the limit at the right because

$$\{x: f(x) > 0\} \supset \{x: f(x) \geq y\}, \forall y > 0.$$

The opposite inequality is also true because if "x^n" is chosen in such a way that

$$(w, x^N) \leq \inf\{(w, x): f(x) > 0\} + \tfrac{1}{N}, \qquad f(x^N) > 0,$$

where "N" is natural, then with "$y_N = f(x^N)$"

$$c(y_N, w) \leq (w, x^N) \leq \inf\{(w, x): f(x) > 0\} + \tfrac{1}{N}$$

and the desired result follows by monotonicity of "c(y, w)" in "y".

(2.9) directly gives the equivalence

$$\lim c(y,w) > 0 \Leftrightarrow f = 0 \tag{2.10}$$

in some neighborhood of "x = 0".

7) If "y < cr/n", where "c" and "r" are some positive constants, then

$$\inf\left\{\Sigma x_i : c \min_i x_i \geq y, \quad \sum x_i \leq r\right\} = \tag{2.11}$$

$$= \inf\left\{\sum\left(x_i - \frac{y}{c}\right) + \frac{ny}{c}: \quad \min_i\left(x_i - \frac{y}{c}\right) \geq 0\right\} = \frac{ny}{c}.$$

8) We want to mention one more useful fact although we do not need it. Suppose "f" is monotonic and "$c(y_0, w) = 0$" for some "$y_0 > 0$". Then there exists a sequence "$\{x^N\}$" such that

$$(w, x^N) \to 0, \quad f(x^N) \geq y_0.$$

By monotonicity, "$f(x) \geq y_0, \forall x > 0$".

Conversely, if with some "$y_0 > 0$" we have "$f(x) \geq y_0, \forall x > 0$",then "$c(y,w) = 0, \forall y \leq y_0$". Thus, for a monotonic "f"

$$\exists\, y_0 : \ c(y, w) = 0, \ \forall y \in (0, y_0) \Leftrightarrow \inf_{x>0} f(x) > 0.$$

DEFINITION 3: a production function "f" exhibits increasing "initial economies of scale" if

(IIES) $$\lim_{y\to 0} AC(y, w) = \infty, \qquad \forall\, w > 0,$$

"constant initial economies of scale" if simultaneously

(CIES′) $$\liminf_{y\to 0} AC(y, w) > 0, \qquad \forall\, w > 0,$$

and

(CIES″) $$\limsup_{y\to 0} AC(y, w) < \infty, \qquad \forall\, w > 0,$$

hold, and "decreasing initial economies of scale" if

(DIES) $$\lim_{y\to 0} AC(y, w) = 0, \qquad \forall\, w > 0.$$

Denote by "S_r" the simplex "$\{x > 0: \sum x_i \leq r\}$".

THEOREM 2: let "f" be a strictly "monotonic production function". Then (IIES) is true if and only if

(IIES.1)

$$\begin{cases} \text{for any } \varepsilon \in (0, 1) \text{ there exists } r = r(\varepsilon) > 0 \text{ such that} \\ \qquad f(x) \leq \varepsilon \sum x_i, \quad \forall\, x \in S_r . \end{cases}$$

(CIES′) and (CIES″) are equivalent to

(CIES′.1)

$$\begin{cases} \text{there exist } c, r > 0 \text{ such that} \\ f(x) \le c \sum x_i \,, \quad \forall \ x \in S_r \,, \end{cases}$$

and

(CIES″.1)

$$\begin{cases} \text{there exist } c, r > 0 \text{ such that} \\ f(x) \ge c \min_i x_i, \quad \forall \ x_i \in S_r, \end{cases}$$

respectively. Finally, the condition

(FIES.1)

$$\begin{cases} \text{for any } \varepsilon \in (0,1) \text{ there exists } r = r(\varepsilon) > 0 \text{ such that} \\ f(x) \ge \dfrac{1}{\varepsilon} \min_i x_i, \quad \forall \ x \in S_r, \end{cases}$$

is necessary and sufficient for (DIES).

PROOF: when proving that (IIES.1) implies (IIES) we can assume that

$$\lim_{y \to +0} c(y, w) = 0$$

(2.12)

because otherwise this limit is positive and (IIES) is trivially fulfilled. Then from the equivalence (2.8) it follows that for all small "y's"

$$\{x: \ f(x) \ge y\} \cap S_r \ne \emptyset. \qquad (2.13)$$

For such "y" we can use (IIES.1) to get

$$c(y, w) \ge h(w) \inf\{\textstyle\sum x_i : \ f(x) \ge y\} =$$

$$= h(w)\,\mathrm{inf}\{\sum x_i : \ f(x) \geq y, \ \ x \in S_r\} \geq$$

$$\geq \ h(w)\,\mathrm{inf}\{\sum x_i : \ \varepsilon \sum x_i \geq y, \ \ \sum x_i \leq r\} = \frac{h(w)}{\varepsilon} y.$$

(IIES) follows since "$\varepsilon > 0$" can be arbitrarily small.

Conversely, if (IIES.1) does not hold, then there exist "$\varepsilon_0 > 0$" and a sequence "$\{x^N\}$" such that

$$f(x^N) > \varepsilon_0 \sum x_i^N , \quad x^N \to 0. \text{ Put}$$

$$y_N = \varepsilon_0 \sum x_i^N .$$

Then "$y_N \to 0$" and

$$c(y_N , w) \leq (w, x^N) \leq \frac{H(w)}{\varepsilon_0} y_N$$

so that (IIES) is not true.

Suppose, (CIES'.1) holds. As above, we can consider only the case (2.12) and hence use (2.13). For small enough "y"

$$c(y, w) \geq h(w)\,\mathrm{inf}\{\sum x_i : \ f(x) \geq y, \ \ x \in S_r\} \geq$$

$$\geq h(w)\,\mathrm{inf}\{\sum x_i : \ c \sum x_i \geq y, \ \ x \in S_r\} = \frac{h(w)}{c} y$$

which gives (CIES').

Conversely, if (CIES'.1) does not hold, then for any "$\varepsilon \in (0, 1)$" there exists a sequence "$\{x^N\}$" such that

$$f(x^N) > \frac{1}{\varepsilon} \sum x_i^N , \quad x^N \to 0.$$

Put

$$y_N = \frac{1}{\varepsilon} \sum x_i^N .$$

Then

$$c(y_N , w) \le (w, x^N) \le H(w) \sum x_i^N = \varepsilon H(w)\, y_N$$

and (CIES′) does not hold.

Suppose (CIES".1) holds. Then "f" cannot vanish in the neighborhood of "x = 0" and (2.10) shows that we can use (2.8)*. Hence, if "y" is small enough, we can apply (CIES".1) and (2.11) to estimate

$$c(y, w) \le H(w) \inf\{\sum x_i : \ f(x) \ge y, \ \ x \in Q_r\} \le \qquad (2.14)$$

$$\le H(w) \inf\left\{\sum x_i : \ c \min_i x_i \ge y, \quad \sum x_i \le r\right\} = \frac{nH(w)}{c}\, y.$$

Thus, (CIES") is proved.

Assume that the opposite of (CIES".1) holds, that is for any "$\varepsilon \in (0, 1)$" there exists a sequence "$\{x^N\}$" such that

$$f(x^N) < \varepsilon \min_i x_i^N , \quad x^N \to 0.$$

Denote "$y_N = f(x^N)$". Then "$y_N \to 0$" and (2.5) implies (remind that "f" is strictly monotonic)

$$c(y_N, w) = \inf\{ (w, x) : \ f(x) \ge f(x^N) \} \ge \ H(w) \min_i x_i^N > \frac{H(w)}{\varepsilon}\, y_N \qquad (2.15)$$

which in turn gives the opposite of (CIES").

Finally, if (DIES.1) holds, then an argument similar to that which led to (2.14) results in a bound

$$c(y, w) \le \varepsilon n H(w) y$$

---

* Strict monotonicity could also be used at this point but we would like to refer to this property as seldom as possible.

for small "y's". Hence, (DIES) is true.

Conversely, if (DIES.1) is violated, then there exist "c > 0" and a sequence "$\{x^N\}$" which satisfy

$$f(x^N) < c \min_i x_i^N, \quad x^N \to 0.$$

It allows us to prove the following analogue of (2.15):

$$c(y_N, w) > \frac{h(w)}{c} y_N, \quad y_N \to 0.$$

Hence, (DIES) cannot be true.

Thus, we have proved THEOREM 2.

REMARK 5:

a) equivalences

(IIES) ⇔ (IIES.1),(CIES') ⇔ (CIES'.1)

are valid for physical production functions;

b) (IIES.1) is equivalent to

$$\lim_{\|x\| \to 0} \frac{f(x)}{\|x\|} = 0;$$

c) (CIES´.1) is equivalent to

$$\limsup_{\|x\| \to 0} \frac{f(x)}{\|x\|} < \infty;$$

d) (CIES´´.1) is equivalent to

$$\liminf_{\|x\|\to 0} f(x)\|h(x)\| > 0\,^*;$$

e) (DIES.1) is equivalent to

$$\lim_{\|x\|\to 0} f(x)\|h(x)\| = \infty .$$

DEFINITION 4: an average cost curve "$AC(y,w) = c(y,w)/y$" is called U-shaped if:

$$\lim_{y\to 0} AC(y, w) = \lim_{y\to\infty} AC(y, w) = \infty.$$

Here "c(y, w)" denotes the total cost.

THEOREM 3. Let "f" be a "PPF". So:

a) If it is a long-run production function (fixed costs "F = 0"), then the "AC" curve is U-shaped if and only if the following two conditions hold:

$$\forall\ \varepsilon > 0 \quad \exists\ C(\varepsilon) > 0 : \quad f(x) \le \varepsilon \sum x_i + C(\varepsilon)\,, \quad \forall\ x \ge 0,$$

$$\forall\ \varepsilon > 0 \quad \exists\ r = r(\varepsilon) > 0 : \quad f(x) \le \varepsilon \sum x_i\,, \quad \forall\ x \in S_r\,.$$

b) If "f" is a short-run production function, then the AC curve "$(F + c(y, w))\ /\ y$" is U-shaped if and only if

$$\forall\ \varepsilon > 0\ \exists\ C(\varepsilon) > 0 : \quad f(x) \le \varepsilon \sum x_i + C(\varepsilon), \quad \forall\ x \ge 0.$$

This theorem is an immediate consequence of remarks 4 and 5. Note that it is reasonable to assume that a short-run production function is bounded in which case the "AC" curve is always U-shaped.

DEFINITION 5: "f" is said to exhibit increasing returns to scale, if:

---

* See REMARK 4 for the definition of the function "h".

$$f(tx) \geq tf(x), \quad \forall\ t > 1, \quad \forall\ x \geq 0,$$

decreasing returns to scale, if:

$$f(tx) \leq tf(x), \quad \forall\ t > 1, \quad \forall\ x \geq 0,$$

and constant returns to scale, if:

$$f(tx) = tf(x), \quad \forall\ t > 0, \quad \forall\ x \geq 0*$$

We would like to relate these notions to the notions of economies of scale. If "f" exhibits increasing returns to scale, then for "y' > y" we have "t = y'/y > 1" and:

$$\frac{c(y', w)}{y'} = \frac{1}{y'} \inf\{(w, x) : f(x) \geq y'\} = \frac{1}{y'} \inf\{(w, x) : \frac{1}{t} f(t \frac{x}{t}) \geq y\} \leq$$
$$\leq \frac{1}{y'} \inf\{(tw, \frac{x}{t}) : f(\frac{x}{t}) \geq y\} = t \frac{1}{y'} c(y, w) = \frac{c(y, w)}{y}.$$

This means that the average cost is non-increasing. Hence only two kinds of behavior at zero are possible:

$$\lim_{y \to 0} AC(y, w) = \infty \quad \text{or} \quad 0 < \lim_{y \to 0} AC(y, w) < \infty. \qquad (2.16)$$

Likewise, only two kinds of behavior at infinity are possible:

$$0 < \lim_{y \to \infty} AC(y, w) < \infty \quad \text{or} \quad \lim_{y \to \infty} AC(y, w) = 0. \qquad (2.17)$$

Let us restrict the argument to the behavior at zero. Then if "f" is in addition strictly monotonic, by THEOREM 2 we have only two mutually excluding possibilities: "f" satisfies either (IIES.1) or (CIES'.1) + (CIES".1). This is not at all obvious: in the definition of increasing returns to scale the values of "f" are compared to the values of the same "f", while in (IIES.1), (CIES'.1), and (CIES".1) "f" is compared to other functions.

* (see Varian, section 1.10).

Combining each possibility at zero with each possibility at infinity, we obtain four possible kinds of behavior each of which can be completely characterized using theorems 1 and 2. The cases of constant and decreasing returns to scale are considered in a similar fashion. Thus we obtain the following result (in which we do not distinguish between obvious and less obvious).

THEOREM 4: suppose "f" is a strictly monotonic production function. Then statements a) and b) are true:

a) if the technology exhibits increasing returns to scale, then only two kinds (2.16) of behavior of the "AC" curve at zero are possible and the production function "f" satisfies only one of conditions

(IIES.1) or (CIES′.1)+(CIES″.1)

for small inputs. Only two kinds (2.17) of behavior of the "AC" curve for large outputs are possible and, respectively, the function "f" can satisfy only one of the conditions

(CES′.1)+(CES″.1) *or* (IES.1);

b) in the case of decreasing returns to scale, the "AC" curve can satisfy only one of the following two conditions for small outputs

$$0 < \lim_{y \to 0} AC(y, w) < \infty \quad or \quad \lim_{y \to 0} AC(y, w) = 0$$

and only one of the two corresponding conditions in terms of "f"

(CIES′.1)+(CIES″.1) or (DIES.1).

For large outputs, there are only two possibilities:

$$\lim_{y \to \infty} AC(y, w) = \infty \quad or \quad 0 < \lim_{y \to \infty} AC(y, w) < \infty$$

which are equivalent to

(DES.1) and (CES′.1)+(CES″.1)

respectively;

c) If "f" exhibits constant returns to scale, then the average cost function is constant as a function of output and the production function satisfies the condition:

$$\begin{cases} \exists\ c_1, c_2 \text{ such that } 0 < c_1 < c_2 < \infty \text{ and} \\ c_1 \min_i x_i \leq f(x) \leq c_2 \sum x_i, \qquad \forall\, x > 0. \end{cases}$$

Abstract: A well-known duality principle between production and cost functions states that any concept defined in terms of a production function has a "dual" definition in terms of the associated cost function and vice versa. We study duality relationships of two kinds. We define several notions of internal economies of scale (in terms of the cost function) for large and small inputs and find their duals in terms of the respective production function. For example, we prove that the average cost increases indefinitely as output goes to infinity, if and only if the production function "f" grows at infinity slower than any positively sloped linear function. On the other hand, for a technology to exhibit increasing initial economies of scale it is necessary and sufficient that "f" be majorized by any positively sloped linear function for all small inputs. These results allow us to exactly describe all technologies that have U-shaped average cost curves. Strict monotonicity and continuity of the production function are the strongest assumptions we make. Problems of this kind cannot be solved using calculus or the Shephard's duality result. The direct approach developed in this paper does nor require differentiability or concavity (convexity).

Key Words: Production Function; Cost Function; Economies of Scale; Production Theory.

REFERENCES

1. ARROW, K.J., INTRILIGATOR, M.D., ed. *Handbook of mathematical economics*. Amsterdam: North-Holland, 1982.

2. JOHANSEN, L., *Production functions.* Amsterdam: North-Holland,1972.

3. MORRONI, M., *Production process and technical change.* Cambridge: Cambridge University Press, 1992.

4. MYNBAEV, K.T., Two Essays in microeconomic theory and econometrics. Thesis (Thesis in Economics), - Department of Economics, Oregon State University, Corvallis, Oregon, USA, 1995.

5. NADIRI, M.I., Producers theory. In: *Handbook of mathematical economics.* Amsterdam: North-Holland, 1982. v, 2.

6. SHEPHARD, R.W. *Theory of cost and production functions.* Princeton, New Jersey: Princeton University Press,1970.

7. VARIAN, H.L., *Microeconomic analysis.* New York: W.W. Norton & Company, 1992.

Recebido para publicação em 06.08.96.