Returns to Scale and the Ray Average Cost

Seung Hoon Lee*
Seoul National University

It is shown that a standard definition of the locally increasing returns to scale at a cost efficient input-output vector constitutes neither sufficient nor necessary condition for the ray average cost to decrease in a neighborhood of the current scale of outputs at the given input prices. And then the most general necessary and sufficient conditions are derived for the ray average cost globally to increase, remain constant or decrease for any given input prices as the level of outputs produced increases proportionlly.

I. Introduction

The concept of scale economies and diseconomies plays an important role in the theory of firm and market organization. It describes whether a proportionate increase in the uses of inputs can or cannot bring forth the corresponding increase in the outputs at a greater proportion. Naturally these technological properties will properly influence the structures of the associated cost functions. In fact the same terminologies are widely being used to describe whether the (ray) average cost curve slopes up- or down-ward. But these uses have not been fully justified yet. Is it really true in general that the ray average cost is increasing, constant, or decreasing as the outputs increase proportionately, if and only if the production technology exhibits the decreasing, constant, or increasing returns to scale respectively?

This equivalence relation was established long ago only for the cases where the production technologies are represented by homogeneous production functions with convex isoquants. Recently Baumol (1977) has shown that if a proportionate increase in the

* The author is professor at the department of economics, Seoul National University. He appreciates the comments of Professors W. J. Baumol and Dosung Lee. The financial aid from Asan Foundation is gratefully acknowledged.
[Seoul Journal of Economics 1990, Vol. 3, No. 1]

uses of inputs at a cost efficient point can bring forth the corresponding increase in the outputs at a greater proportion, then the ray average cost will always decrease as the outputs increase proportionately from that point. Panzar and Willig (1977) established an equivalence relation successfully between the technological properties and the structures of the associated ray average costs. But this equivalence is not quite the same as the one raised in the above paragraph[1] and furthermore is valid only when the production functions and the cost functions are differentiable with respect to the input and output variables.

In this paper, I will investigate the most general case in order to clarify some conceptual problems and to generalize the pre-existing results. It will be shown that a standard definition of the locally increasing returns to scale at a cost efficient input-output vector constitutes *neither sufficient nor necessary* condition for the ray average cost to decrease in a neighborhood of the current scale of outputs at the given input prices. And then *the* most general necessary and sufficient conditions will be derived for the ray average cost globally to increase, remain constant or decrease for any input prices as the level of outputs produced increases proportionately. In Section II, the basic notations and concepts will be introduced and discussed. The Panzar-Willig equivalence is reviewed in Section III, and our main results are reported in Section IV.

II. Notations, Definitions and Concepts

Consider the case where m outputs are produced from n inputs. Let $y \in R^m_+$ and $x \in R^n_+$ denote the output and the input vector respectively. Standard definition of the input requirement set $V(y)$ for producing the output vector y gives

$$V(y) \equiv \{x \in R^n_+ \mid y \text{ can be produced from } x\}.$$

We assume that the set $V(y)$ is nonempty and closed for every y in R^m_+. It is *the only* assumption regarding the production technology that I make in this analysis. It is also standard to define.

Definition 1
Returns to scale are said to be increasing, if and only if $x \in V(y)$ implies $\alpha x \in \text{Int } V(\alpha y)$ for every $\alpha > 1$ and for every $y \in R^m_+$.

[1]See Sections III and IV below

Returns to scale are said to be constant, if and only if $x \in V(y)$ implies $\alpha x \in V(\alpha y)$ for every $\alpha > 0$ and for every $y \in R^m_+$. And returns to scale are said to be decreasing, if and only if $x \in V(y)$ implies $\alpha x \in \text{Int } V(\alpha y)$ for every $\alpha \in (0, 1)$ and for every $y \in R^m_+$.[2]

Of course this definition is a global one. The following Lemma provides an alternative form of Definition 1.

Lemma

Returns to scale are increasing, if and only if $x \notin \text{Int } V(y)$ implies $\beta x \notin V(\beta y)$ for every $\beta \in (0, 1)$ and for every $y \in R^m_+$. Returns to scale are constant, if and only if $x \notin V(y)$ implies $\beta x \notin V(\beta y)$ for every $\beta > 0$ and for every $y \in R^m_+$. And returns to scale are decreasing, if and only if $x \notin \text{Int } V(y)$ implies $\beta x \notin V(\beta y)$ for every $\beta > 1$ and for every $y \in R^m_+$.

Proof: Only the proof for increasing returns to scale will be provided. Suppose the returns are increasing. If the Lemma is false, then there will be a number $\beta \in (0, 1)$ and $x \notin \text{Int } V(y)$ such that $\beta x \in V(\beta y)$ holds. Now by Definition 1, one has $(1/\beta)(\beta x) \in \text{Int } V((1/\beta)(\beta y))$, which is contradictory. Consider the converse. If there exists an $x \in V(y)$ and a number $\alpha > 1$ such that $\alpha x \notin \text{Int } V(\alpha y)$, then we have $(1/\alpha)(\alpha x) \notin V((1/\alpha)(\alpha y))$, which is again contradictory.

Q. E. D.

The expositions in Definition 1 and Lemma are mutually symmetric. For instance, the increasing returns to scale of Definition 1 describe the scale characteristic when one expands the scale while those of Lemma describe when one contracts. In view of this aspect, the local definitions of increasing, constant and decreasing returns to scale can also be given in two ways respectively. It is because, given a production vector (x, y), the scale characteristics when one expands the scale are not necessarily identical to those when one contracts. The following Definition 2 provides the detail.

Definition 2

Returns to scale are locally increasing *from* the input-output vector (x, y) with $x \in V(y)$, if and only if there exists a real number $u > 1$ such that $\alpha x \in \text{Int } V(\alpha y)$ for every $\alpha \in (1, u]$. Returns to scale are locally increasing *up to* the input-output vector (x, y) with

[2]Here the notation Int A denotes the interior of the set A.

$x \notin \text{Int } V(y)$, if and only if there exists a real number $d \in (0, 1)$ such that $\beta x \notin V(\beta y)$ for every $\beta \in [d, 1)$. Returns to scale are locally increasing at (x, y), if and only if they are locally increasing both *from* and *up to* (x, y). Notice that the returns to scale cannot increase locally at any (x, y), if either $x \in \text{Int } V(y)$ or $x \notin V(y)$ holds. Baumol's definition is that of the increasing returns to scale *from* only. Similar definitions can be provided for the cases of the locally constant and decreasing returns to scale.

Let $C(y)$ denote the cost necessary to produce the output vector y according to the technology $V(\cdot)$. It is assumed that each input is purchased at a constant price vector $w = (w_1, w_2, \cdots, w_n)$, $w_i > 0$ $(i = 1, \cdots, n)$, in the competitive markets. The ray average cost, according to Baumol, can be defined as $C(\lambda y)/\lambda$ for each output vector y, where the positive real number λ plays the role of a scale parameter.

The technology $V(\cdot)$ is not the only technology which will generate the cost function $C(\cdot)$. Let

$$V^*(y) \equiv \{x \in R^n_+ \mid \text{There exists } x' \text{ in co}V(y) \text{ such that } x' \leq x \text{ holds}\}$$

where co$V(y)$ denotes the closed convex hull of $V(y)$. It is easy to show that $V^*(y)$ is non-empty, closed, convex and monotone for every y in R^m_+. I will name this as the *convexified technology* of $V(\cdot)$. When $C^*(\cdot)$ denotes the cost function generated by the convexified technology $V^*(\cdot)$, it can be shown by a standard duality argument that $C(y) = C^*(y)$ holds for every y in R^m_+. Hence any technological characteristic of $V(\cdot)$ that will influence the structure of the cost function will be preserved in the convexified technology $V^*(\cdot)$ exactly and exhaustively. A natural step to infer the technological characteristics from a given cost function will narrow the scope to the investigation of the convexified technology $V^*(\cdot)$ rather than the original $V(\cdot)$.

III. Panzar-Willig Equivalence

It will be useful to clarify exactly what is the equivalence relation that Panzar and Willig (1977) established, before getting into our main analysis. They assumed a continuous production function $\phi(\cdot)$ such that an output vector y can be produced from the input vector x, if and only if $\phi(x, y) > 0$ holds. Their main interests were to

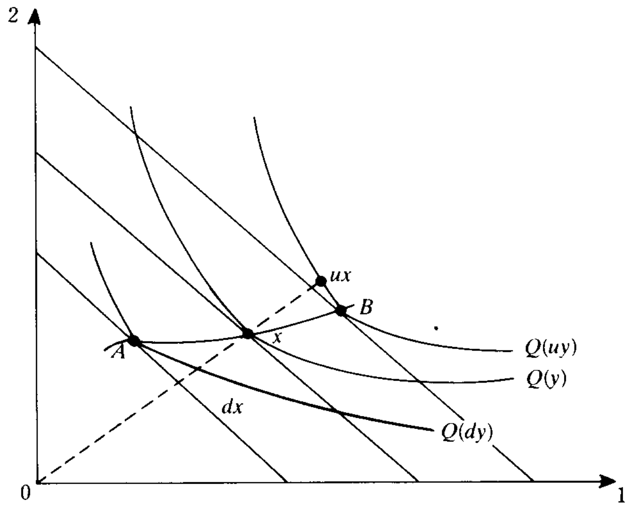

FIGURE 1

find out the conditions when the marginal cost pricing can cover at least the total production cost. In pursuing their goal, they derived an equivalence relation of our interest, when the production function and the cost function are differentiable with respect to the input and output variables. This relation can virtually be stated as

$$\frac{\partial \phi(\lambda x, \lambda y)}{\partial \lambda}\bigg|_{\lambda=1} \gtrless 0 \text{ iff } \frac{\partial}{\partial \lambda}\left(\frac{C(\lambda y)}{\lambda}\right)\bigg|_{\lambda=1} \lessgtr 0$$

if the input vector x is cost efficient for the output y. It is of course a very remarkable and powerful result. But it is not sufficient enough to conclude that the returns to scale are increasing, constant or decreasing, if and only if the ray average cost is decreasing, constant or increasing globally or locally. The following two examples will clarify this aspect.

Example 1.
Let $Q(y)$ be the isoquant for the output y. Suppose the returns to scale are locally constant at (x, y) in Figure 1, i.e., $\alpha x \in V(\alpha y)$ for every $\alpha \in [d, u]$ for some $d \in (0, 1)$ and $u > 1$. Thus $\frac{\partial \phi(\lambda x, \lambda y)}{\partial \lambda}\big|_{\lambda=1} = \frac{\partial}{\partial \lambda}(\frac{C(\lambda y)}{\lambda})\big|_{\lambda=1} = 0$ holds in this case. But if the expansion path passes through $A - x - B$ as depicted in Figure 1, then one must have $C(\lambda y)/\lambda < C(y)$ for every $\lambda \in [d, u]$ with $\lambda \neq 1$, and therefore the ray average cost is not constant in any neighborhood of the current scale of the output y.

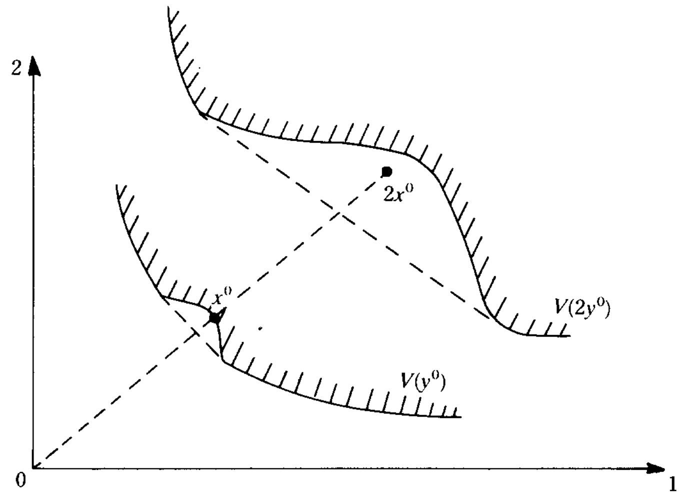

FIGURE 2

Example 2
Consider the non-convex technology of Figure 2. Its convexified technology $V^*(\cdot)$ exhibits globally increasing returns to scale. Then the ray average cost will decrease as the scale of production is augmented for any output vector $y > 0$,[3] i.e.,

$$\frac{\partial}{\partial \lambda}\left(\frac{C(\lambda y)}{\lambda}\right)\Bigg|_{\lambda=1} \leq 0$$

holds everywhere for every $w \gg 0$. Consequently

$$\frac{\partial \phi(\lambda x, \lambda y)}{\partial \lambda}\Bigg|_{\lambda=1} \geq 0$$

will hold for each cost efficient input-output vector (x, y) at any $w \gg 0$. But the input vector x^0 in Figure 2 can never be cost efficient for the output vector y^0 at any price vector $w \gg 0$. So it is not precluded that the returns to scale are locally decreasing at x^0 even when the ray average cost is decreasing in a neighborhood of the current scale of the output y^0.

IV. Theorems and Proofs

First I will investigate the possibility of the local equivalence

[3]See Theorem 2 in Section IV

and then derive the global equivalence relation. The following Theorem 1 summarizes the results for the local returns to scale and the slope of the ray average cost curve.

Theorem 1

Let $C(y) > 0$, $x \in V(y)$ and $C(y) = wx$.

1) If $\alpha x \in \operatorname{Int} V^*(\alpha y)$ holds for every $\alpha \in (1, u)$ for some number $u > 1$, then one has $C(y) > C(\alpha y)/\alpha$.

2) If $C(\beta y)/\beta > C(y)$ holds for every $\beta \in (l, 1)$ for some number $l \in (0, 1)$, then one has $\beta x \notin \operatorname{Int} V^*(\beta y)$.

Proof: 1) This was proved by Baumol originally. Since one has $\alpha x \in \operatorname{Int} V^*(\alpha y)$, it immediately follows $C(\alpha y) < w \alpha x = \alpha C(y)$.

2) Assume $\beta x \in \operatorname{Int} V^*(\beta y)$. Then one obtains $\beta C(y) = \beta wx > C(\beta y)$, which is a contradiction.

Q. E. D.

The first part implies that when x is cost efficient for y and the returns to scale are locally increasing from (x, y), then the ray average cost will decrease at least for a while as the level of outputs increases proportionately from y. The second part implies that when the ray average cost increases at least for a while as the level of outputs decreases proportionately from y, then the returns to scale are locally increasing up to (x, y). A theorem parallel to Theorem 1 can be easily established for the locally decreasing returns to scale and increasing ray average cost in a symmetric way. The converse of each part in Theorem 1 is not necessarily true. The following Examples 3 and 4 illustrate this.

Example 3

Consider the two isoquants $Q(y^0)$ and $Q(\alpha y^0)$ in Figure 3. The input vectors x^0 and x' are cost efficient for the output vectors y^0 and αy^0 respectively. The value of the real number α is given so that the point αx^0 lies between A and mx^0 in Figure 3. For each number $\gamma \in (0, 1)$ let the point $\gamma x^0 + (1 - \gamma)x'$ be the cost efficient input vector for the output $\gamma y^0 + (1 - \gamma)\alpha y^0$. Then one obtains

$$C([\gamma + (1 - \gamma)\alpha]y^0) = w[\gamma x^0 + (1 - \gamma)x'] < \\ [\gamma + (1 - \gamma)\alpha]wx^0 = [\gamma + (1 - \gamma)\alpha]C(y^0)$$

and so the ray average cost decreases at least for a while as the outputs increase proportionately from y^0. Now let the point $[\gamma +$

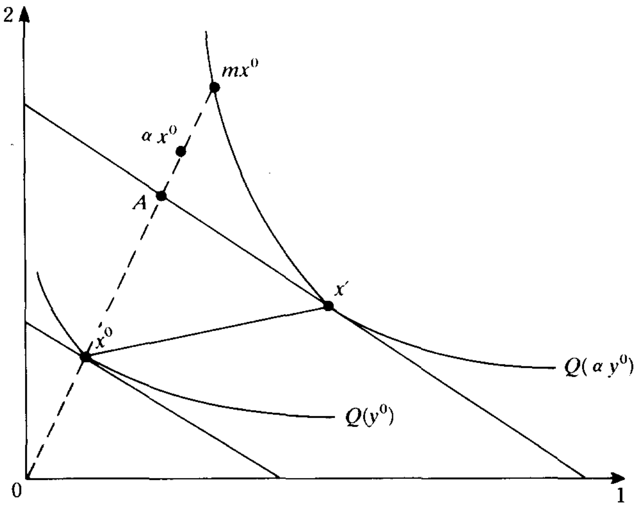

FIGURE 3

$(1 - \gamma)m]x^0$ lie on the isoquant $Q([\gamma + (1 - \gamma)\alpha]y^0)$ for each $\gamma \in (0, 1)$. Then one obtains a case where the returns to scale are locally decreasing from (x^0, y^0) in spite of the decreasing ray average cost.

Example 4

Consider the two isoquants $Q(y^0)$ and $Q(\beta y^0)$ in Figure 4. Again the input vectors x^0 and x' are cost efficient for the output vectors y^0 and βy^0 respectively. The value of the number β is chosen such that the point βx^0 lies between B and mx^0 in Figure 3. For each number $\gamma \in (0, 1)$ let the input vector $\gamma mx^0 + (1 - \gamma)x^0 = [\gamma m + (1 - \gamma)]x^0$ lie on the isoquant for the output $[\gamma \beta + (1 - \gamma)]y^0$. Then it is clear from the figure that the input $[\gamma \beta + (1 - \gamma)]x^0$ cannot produce the output $[\gamma \beta + (1 - \gamma)]y^0$, since we have $\gamma \beta + (1 - \gamma) < \gamma m + (1 - \gamma)$ for each $\gamma \in (0, 1)$. Therefore the technology exhibits locally increasing returns to scale up to (x^0, y^0). Now let the point $\gamma x' + (1 - \gamma)x^0$ be the cost efficient input vector for producing the output $[\gamma \beta + (1 - \gamma)]y^0$ respectively for each $\gamma \in (0, 1)$. Then one has

$$\begin{aligned} C([\gamma \beta + (1 - \gamma)]y^0) &= w[\gamma x' + (1 - \gamma)x^0] \\ &= \gamma wx' + (1 - \gamma)wx^0 < \gamma w \beta x^0 + (1 - \gamma)wx^0 \\ &= [\gamma \beta + (1 - \gamma)]C(y^0). \end{aligned}$$

Thus the ray average cost decreases here at least for a while as the

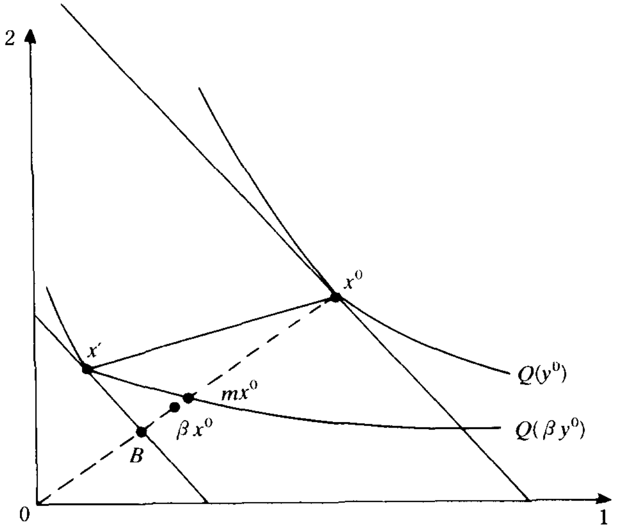

FIGURE 4

outputs decrease proportionately from y^0 even though the returns to scale are locally increasing up to (x^0, y^0).

Let the input vector x be the cost-minimizer for the output vector y at the current input prices. It is important to notice that *neither* "the locally increasing returns to scale at (x, y)" *nor* "the falling ray average cost in a neighborhood of the current production scale" implies each other. The above Examples 3 and 4 can be made use of to construct the relevant counter-examples. First, let the returns to scale increase locally at (x, y) but follow the manner of Example 4 up to (x, y). Then the ray average cost will not be monotone decreasing in any neighborhood of the current production scale. Second, let the ray average cost be monotone decreasing in a neighborhood of the current production scale but follow the manner of Example 3 when one expands the production scale. Then we have the locally decreasing returns to scale from (x, y) here.

So far we have not been able to establish the local equivalence relation between the returns to scale and the shape of ray average cost curves. But the situation changes when we investigate the global equivalence. The following Theorem 2 summarizes the most general necessary and sufficient conditions for the ray average cost globally to increase, to remain constant, or to decrease as the outputs increase proportionately at all outputs $y > 0$ and at all input prices $w > 0$.

Theorem 2

The ray average cost $C(\lambda y)/\lambda$ is globally decreasing, constant, or increasing for every output vector $y > 0$ and for every input price vector $w > 0$ as the scale parameter λ increases, if and only if the returns to scale are globally increasing, constant, or decreasing respectively in the convexified technology $V^*(\cdot)$.

Proof: The sufficient part for the ray average cost to be globally constant is widely known. Now suppose the returns to scale are globally increasing in $V^*(\cdot)$. Let x be cost efficient for some $y > 0$ at some input price $w > 0$. For any $\lambda > 1$ one has $w\lambda x > C(\lambda y)$, since it holds $\lambda x \in \mathrm{Int}\ V^*(\lambda y)$. Thus $\lambda C(y) > C(\lambda y)$ holds for every $y > 0$ and for every $\lambda > 1$, which implies a strictly decreasing ray average cost for every $y > 0$ at any $w > 0$. The proof for the ray average cost to increase can be obtained in a similar manner when one utilizes the definition given in the Lemma.

Now let us prove the necessary part for the ray average cost to decrease. Let $C(y) > C(\lambda y)/\lambda$ hold for every $y > 0$, for every $w > 0$, and for any $\lambda > 1$. Let $x \notin \mathrm{Int}\ V^*(y)$. It suffices to show that $\beta x \notin V^*(\beta y)$ for every $\beta \in (0, 1)$. By Minkowski's "Supporting Hyperplane Theorem" there exists a vector $w \neq 0$ such that $wx \leq wz$ for every $z \in V^*(y)$. From the monotonicity of $V^*(\cdot)$, one can conclude $w > 0$. Suppose that the input price vector is given as this w. Then clearly one has $wx \leq C(y)$. Now assume that there exists a number $\beta' \in (0, 1)$ such that $\beta'x \in V^*(\beta'y)$ holds. It follows that $C(\beta'y) \leq \beta'wx \leq \beta'C(y)$ and in turn $C(\beta'y)/\beta' \leq C(y)$, which is a contradiction to the falling ray average cost. So we must have $\beta'x \notin V^*(\beta'y)$ for any number $\beta' \in (0, 1)$. Similar proofs can be provided to the cases where the ray average cost is constant or increasing.

Q. E. D.

References

Baumol, W. J. "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry." *American Economic Review* 67 (1977) : 809-22.

Debreu, G. *Theory of Value*. New York: John Wiley, 1959.

Panzar, J. C. and Willig, R. D. "Economies of Scale in Multi-output Production." *Quarterly Journal of Economics* 91 (1977) : 481-93.

Uzawa, H. "Duality Principles in the Theory of Cost and Production Functions." *International Economic Review* 5 (1964) : 216-20.

Varian, H. R. *Microeconomic Analysis.* New York: Norton, 1978.