CALIFORNIA INSTITUTE OF TECHNOLOGY
Division of the Humanities and Social Sciences

The old fashioned approach to comparative statics of cost minimization

KC Border
Fall 1998, Revised Fall 2004
v. 2008.11.25.10.22

This is the the more or less traditional approach to cost minimization, as found in say Samuelson [1, pp. 57–69].

1 Production functions

The economists' somewhat lame standard approach to production is to assume that there is **production function** that relates output to inputs. That is, the maximal quantity of output y that can be produced with a list $x = (x_1, \ldots, x_n)$ of inputs is given by

$$y = f(x_1, \ldots, x_n).$$

This formulation has built into it the assumption of **no joint production**. There is only one output per producer. It is not that we cannot deal with joint production, I'll cover that in a separate note, it is merely a convenient benchmark, and it is a gentle introduction for students. We will assume the following usually unstated assumptions.

P.1 The production function $f \colon \boldsymbol{R}^{\mathrm{n}}_{+} \to \boldsymbol{R}$ is continuous, and twice continuously differentiable on $\boldsymbol{R}^{\mathrm{n}}_{++}$.

P.2 At each point $x \gg 0$, we have $f'(x) \gg 0$, which is a strong monotonicity condition.

P.3 The production function satisfies the following strong quasiconcavity condition. At each $x \gg 0$, the Hessian is negative definite on the subspace orthogonal to the gradient. That is, for all $v \in \boldsymbol{R}^{\mathrm{n}}$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(x) v_i v_j < 0 \qquad v \neq 0 \text{ and } f'(x) \cdot v = 0,$$

where $f_{ij}(x) = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$. This is equivalent to

$$(-1)^p \begin{vmatrix} f_{11} & \ldots & f_{1p} & f_1 \\ \vdots & & \vdots & \vdots \\ f_{p1} & \ldots & f_{pp} & f_p \\ f_1 & \ldots & f_p & 0 \end{vmatrix} > 0 \qquad p = 2, \ldots, n.$$

In particular,

$$\begin{vmatrix} f_{11} & \dots & f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ f_{n1} & \dots & f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{vmatrix} \neq 0.$$

P.4 If $f(x) > 0$, then $x \gg 0$,.

Under these conditions, for $w \gg 0$ and $y > 0$,there is always a unique cost minimizing input vector $\hat{x}$ and it satisfies $\hat{x} \gg 0$.

2 Cost minimization

Mathematically the cost minimization problem can be formulated as follows.

$$\underset{x}{\text{minimize}} \ w \cdot x \quad \text{subject to } f(x) \geqslant y, \ x \geqq 0,$$

where $w \gg 0$ and $y > 0$.

It is clear that we can replace the condition $f(x) \geqslant y$ by $f(x) - y = 0$ without changing the solution. Let $\hat{x}(w, y)$ solve this problem, and assume that $\hat{x} \gg 0$. The Lagrangean for this minimization problem is

$$w \cdot x - \lambda\big(f(x) - y\big).$$

The gradient of the constraint function (with respect to x) is just $f'(\hat{x})$, which is not zero. Therefore by the Lagrange Multiplier Theorem, there is a Lagrange multiplier $\hat{\lambda}$ (depending on w, y) so that locally the first order conditions

$$w_i - \hat{\lambda}(w, y) f_i\big(\hat{x}(w, y)\big) = 0, \quad i = 1, \dots, n, \tag{1}$$

where $f_i(x) = \frac{\partial f(x)}{\partial x_i}$, and the constraint

$$y - f\big(\hat{x}(w, y)\big) = 0 \tag{2}$$

hold for all w, y. Note that (1) implies that $\hat{\lambda} > 0$.

The second order condition is that

$$\hat{\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij}(\hat{x}) v_i v_j \leqslant 0, \tag{3}$$

for all $v \in \boldsymbol{R}^{\mathrm{n}}$ satisfying

$$f'(\hat{x}) \cdot v = \sum_{i=1}^{n} f_i(\hat{x}) v_i = 0.$$

Using the **method of implicit differentiation** with respect to each w_j on (1) yields:

$$\delta_{ij} - \frac{\partial \hat{\lambda}}{\partial w_j} f_i(\hat{x}) - \hat{\lambda} \sum_{k=1}^{n} f_{ik}(\hat{x}) \frac{\partial \hat{x}_k}{\partial w_j} = 0, \quad \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, n \end{matrix}, \tag{4}$$

where δ_{ij} is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Differentiating (1) with respect to y yields

$$-\frac{\partial \hat{\lambda}}{\partial y} f_i(\hat{x}) - \hat{\lambda} \sum_{k=1}^{n} f_{ik}(\hat{x}) \frac{\partial \hat{x}_k}{\partial y} = 0, \quad i = 1, \dots, n, \tag{5}$$

Now differentiate (2) with respect to each w_j to get

$$-\sum_{k=1}^{n} f_k(\hat{x}) \frac{\partial \hat{x}_k}{\partial w_j} = 0, \quad j = 1, \dots, n, \tag{6}$$

and with respect to y to get

$$-\sum_{k=1}^{n} f_k(\hat{x}) \frac{\partial \hat{x}_k}{\partial y} + 1 = 0. \tag{7}$$

We can rearrange equations (4) through (7) into one gigantic matrix equation:

$$\begin{bmatrix} \hat{\lambda} f_{11} & \dots & \hat{\lambda} f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \hat{\lambda} f_{n1} & \dots & \hat{\lambda} f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{x}_1}{\partial w_1} & \dots & \frac{\partial \hat{x}_1}{\partial w_n} & \frac{\partial \hat{x}_1}{\partial y} \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \frac{\partial \hat{x}_n}{\partial w_1} & \dots & \frac{\partial \hat{x}_n}{\partial w_n} & \frac{\partial \hat{x}_n}{\partial y} \\ \frac{\partial \hat{\lambda}}{\partial w_1} & \dots & \frac{\partial \hat{\lambda}}{\partial w_n} & \frac{\partial \hat{\lambda}}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 & 0 \\ 0 & \ddots & & & \vdots & \vdots \\ \vdots & & & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & \dots & & \dots & 0 & 1 \end{bmatrix}.$$

$$\left[\begin{array}{ccc|c} \ddots & & & \vdots \\ & (4) & & (5) \\ & & \ddots & \vdots \\ \hline \dots & (6) & \dots & (7) \end{array}\right]$$

Figure 1. The blocks in the matrix version of equations (4) through (7).

To see where this comes from, break up the $(n+1) \times (n+1)$ matrix equation into four blocks. The upper left $n \times n$ block comes from (4). The upper right $n \times 1$ block comes from (5). The lower left $1 \times n$ block comes from (6), and finally the lower right 1×1 block is just (7). What this tells us is that

$$\begin{bmatrix} \frac{\partial \hat{x}_1}{\partial w_1} & \dots & \frac{\partial \hat{x}_1}{\partial w_n} & \frac{\partial \hat{x}_1}{\partial y} \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \frac{\partial \hat{x}_n}{\partial w_1} & \dots & \frac{\partial \hat{x}_n}{\partial w_n} & \frac{\partial \hat{x}_n}{\partial y} \\ \frac{\partial \hat{\lambda}}{\partial w_1} & \dots & \frac{\partial \hat{\lambda}}{\partial w_n} & \frac{\partial \hat{\lambda}}{\partial y} \end{bmatrix} = \begin{bmatrix} \hat{\lambda} f_{11} & \dots & \hat{\lambda} f_{1n} & f_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \hat{\lambda} f_{n1} & \dots & \hat{\lambda} f_{nn} & f_n \\ f_1 & \dots & f_n & 0 \end{bmatrix}^{-1} . \tag{8}$$

Condition **P.3** implies that the $n \times n$ matrix

$$\begin{bmatrix} \frac{\partial \hat{x}_1}{\partial w_1} & \dots & \frac{\partial \hat{x}_1}{\partial w_n} \\ \vdots & & \vdots \\ \frac{\partial \hat{x}_n}{\partial w_1} & \dots & \frac{\partial \hat{x}_n}{\partial w_n} \end{bmatrix}$$

is negative semidefinite of rank $n-1$, being the upper left block of the inverse of a bordered matrix that is negative definite under constraint. (See my notes on quadratic forms.) It follows therefore that

$$\frac{\partial \hat{x}_i}{\partial w_i} \leqslant 0 \quad i = 1, \dots, n.$$

3 What the support function approach left out

Note that this approach provides us with conditions under which the cost function is twice continuously differentiable. It follows from (8) that if the bordered Hessian is invertible, the Implicit Function Theorem tells us that $\hat{x}$ and $\hat{\lambda}$ are C^1 functions of w and y (since f is C^2). On the other hand, if $\hat{x}$ and $\hat{\lambda}$ are C^1 functions of w and y, then (8) implies that the bordered Hessian is invertible. In either case, the marginal cost $\frac{\partial c}{\partial y} = \hat{\lambda}$, is a C^1 function of w and y, so the cost function is C^2, which is hard to establish by other means.

4 Reciprocity results

Returning now to (8), note that since the Hessian is a symmetric matrix, we have a number of **reciprocity** results. Namely:

$$\frac{\partial \hat{x}_i}{\partial w_j} = \frac{\partial \hat{x}_j}{\partial w_i} \qquad \begin{array}{l} i = 1, \dots, n, \\ j = 1, \dots, n, \end{array}$$

and

$$\frac{\partial \hat{x}_i}{\partial y} = \frac{\partial \hat{\lambda}}{\partial w_i} = \frac{\partial^2 c}{\partial w_i \partial y}.$$

5 The marginal cost function

Define the cost function c by

$$c(w, y) = \sum_{k=1}^{n} w_k \hat{x}_k(w, y).$$

Then

$$\frac{\partial c(w, y)}{\partial y} = \sum_{i=k}^{n} w_k \frac{\partial \hat{x}_k(w, y)}{\partial y},$$

and

$$\frac{\partial^2 c(w, y)}{\partial y^2} = \sum_{i=k}^{n} w_k \frac{\partial^2 \hat{x}_k(w, y)}{\partial y^2}. \tag{9}$$

From (1), we have $w_k = \hat{\lambda} f_k(\hat{x})$, so

$$\frac{\partial c(w, y)}{\partial y} = \hat{\lambda} \sum_{k=1}^{n} f_i(\hat{x}) \frac{\partial \hat{x}_k(w, y)}{\partial y} = \hat{\lambda}, \tag{10}$$

where the second equality is just (7). That is, *the Lagrange multiplier $\hat{\lambda}$ is the marginal cost.*

Now let's see whether the marginal cost is increasing or decreasing as a function of y. Differentiating (7) with respect to y yields

$$\sum_{j=1}^{n}\left(\frac{\partial \hat{x}_j}{\partial y}\sum_{i=1}^{n} f_{ij}(\hat{x})\frac{\partial \hat{x}_i}{\partial y} + f_j(\hat{x})\frac{\partial^2 \hat{x}_j}{\partial y^2}\right) = 0,$$

or rearranging,

$$\sum_{j=1}^{n} f_j(\hat{x})\frac{\partial^2 \hat{x}_j}{\partial y^2} = -\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(\hat{x})\frac{\partial \hat{x}_i}{\partial y}\frac{\partial \hat{x}_j}{\partial y}. \tag{11}$$

From (9) and (1) we have that the left-hand side of (11) is $\frac{1}{\hat{\lambda}}\frac{\partial^2 c}{\partial y^2}$. What is the right-hand side?

Fix w and consider the curve $y \mapsto \hat{x}(y)$. This is called an **expansion path**. It traces out the optimal input combination as a function of the level of output. The tangent line to this curve at $\hat{x}$ is just $\{\hat{x} + \alpha v : \alpha \in \boldsymbol{R}\}$, where

$$v_i = \frac{\partial \hat{x}_i}{\partial y}.$$

Write the output along this tangent line, $f(\hat{x} + \alpha v)$, as a function $\hat{f}$ of α. That is, $\hat{f}(\alpha) = f(\hat{x} + \alpha v)$. By the chain rule,

$$\hat{f}'(\alpha) = \sum_{j=1}^{n} f_j(\hat{x} + \alpha v)v_j,$$

and

$$\hat{f}''(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(\hat{x} + \alpha v)v_i v_j,$$

so

$$\hat{f}''(0) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(\hat{x})\frac{\partial \hat{x}_i}{\partial y}\frac{\partial \hat{x}_j}{\partial y}.$$

Thus (11) can be written as

$$\frac{\partial^2 c}{\partial y^2} = -\hat{\lambda}\hat{f}''(0).$$

In other words (11) asserts that *the slope of the marginal cost curve is increasing (that is, the cost function is a locally convex function of y) when the production function is locally concave on the line tangent to the expansion path*, and vice-versa.

6 Average cost and elasticity of scale

Recall that a production function f exhibits **constant returns to scale** if $f(\alpha x) = \alpha f(x)$ for all $\alpha > 0$. It exhibits **increasing returns to scale** if $f(\alpha x) > \alpha f(x)$ for $\alpha > 1$, and **decreasing returns to scale** if $f(\alpha x) < \alpha f(x)$ for $\alpha > 1$. If f is **homogeneous of degree *k***, that is, if

$$f(\alpha x) = \alpha^k f(x),$$

then the returns to scale are decreasing, constant, or increasing, as $k < 1$, $k = 1$, or $k > 1$. Define

$$h(\alpha, x) = f(\alpha x).$$

The **elasticity of scale** $e(x)$ of the production function at x is defined to be

$$D_1 h(1, x) \frac{1}{f(x)} = f'(x) \cdot x / f(x),$$

where D_1 denotes the partial derivative with respect to the first argument α, and which Varian [2] writes as

$$\left.\frac{df(\alpha x)}{d\alpha} \frac{\alpha}{f(x)}\right|_{\alpha=1}.$$

If f is homogeneous of degree k, then $e(x) = k$, as

$$D_1 h(\alpha, x) = k\alpha^{k-1} f(x).$$

Even if f is not homogeneous, following Varian, we can express the elasticity of scale in terms of the marginal and average cost functions, at least for points x that minimize cost uniquely for some (y, w):

$$\begin{aligned}
e\big(\hat{x}(y,w)\big) &= f'(\hat{x}) \cdot \hat{x} / f(\hat{x}) \\
&= f'(\hat{x}) \cdot \hat{x} / y && \text{as } y = f\big(\hat{x}(y,w)\big) \\
&= \frac{w}{\hat{\lambda}} \cdot \hat{x} / y && \text{by the first order condition } w = \hat{\lambda} f'(\hat{x}) \\
&= \frac{c(y,w)/y}{D_y c(y,w)} && \text{as } c(y,w) = w \cdot \hat{x}(y,w), \text{ and by (10) } \hat{\lambda} = D_y c(y,w) \\
&= \mathrm{AC}(y) / \mathrm{MC}(y).
\end{aligned}$$

Holding w fixed, and writing the cost simply as a function of y,

$$\frac{d}{dy} \mathrm{AC}(y) = \frac{d}{dy} \frac{c(y)}{y} = \frac{c'(y)y - c(y)}{y^2} = \frac{1}{y} \left(c'(y) - \frac{c(y)}{y} \right) = \frac{1}{y} \left(\mathrm{MC}(y) - \mathrm{AC}(y) \right).$$

Thus

$$\mathrm{AC}'(y) > 0 \iff \mathrm{MC}(y) > \mathrm{AC}(y) \iff e(\hat{x}) < 1.$$

References

[1] P. A. Samuelson. 1965. *Foundations of economic analysis*. New York: Athenaeum. Reprint of edition published by Harvard University Press, 1947.

[2] H. R. Varian. 1992. *Microeconomic analysis*, 3d. ed. New York: W. W. Norton & Co.