PRODUCTION THEORY[1]

The neoclassical theory of the firm is not a theory about industrial organization but rather a theory about the relation between input and output prices. There is no concern about organization or transactions costs. These issues are treated under the aegis of what I like to call the contracting cost theory of the firm. In the neoclassical theory we assume that there is an economic agent called the firm. We treat it like it is a real entity. It is assumed to maximize profits that are the difference between revenues and costs; profits are maximized subject to a technical relationship between inputs and output, which is written as $q = f(x_1, ..., x_n)$.

In this model, the level of profits has no effect on the relation between inputs and output. That is, profits are independent of $f(.)$. Moreover, we assume that resources are bought by the firm at fixed prices that are also independent of profits. Profit is not the payment to or for any resource. It is in this sense that we take about "economic" profits.

In a meat-&-potatoes kind of way, we can reduce the number of inputs to two, labor and capital. However, even in this setting the price of capital is interpreted as a rental rate on machine or plant units. There is no discounting in this problem. We assume that the firm goes to the capital rental store each morning and picks up its equipment before going to the meeting street to get its workers.

The model of the behavior of the firm is broken into two parts based on the following logical tautology: For the firm to be a profit maximizer, it must minimize cost at any given level of output. In this way we break the behavior of the firm into 1) Cost minimization given an output constraint, and then 2) profit maximization given cost minimizing behavior. Cost minimization analysis tells us how the firm chooses its input mix and, by inference, how output is related to cost. Our interest is in mapping the predictions that we can derive concerning the behavior of the firm in input space into functional relations between input price and input quantity, and between output price and output level. This latter point is most important. From cost minimization analysis we derive a *cost function* that represents the relation between output and the minimized cost at each level of output. This cost function can be differentiated w.r.t. output to yield *marginal cost* (MC) and can be divided by the level of output to yield *average cost* (AC). These are the functions that we then use to analyze profit maximization. Sometimes we even jump past profit maximization and go directly to output market equilibrium.

1. THE COST MINIMIZATION MODEL

The behavior of the profit maximizing firm starts with the analysis of economic efficiency in production. This behavior is described by the model of cost minimization subject to an output objective. Assume that the firm attempts to optimize according to the model:

$$\min_{\{x_i\}} C = \sum_{i=1}^{n} w_i x_i \text{ subj. to } q = f(x_1, \ldots, x_n)$$

Rewriting in Lagrangian terms gives:

[1]Silberberg (1990) Ch 8, esp. 8.11; (2001) Ch 8. esp. 8.11. Layard and Walters (1978) Ch, Appendix 7. Ch 9. Varian (1992) Ch 4 & 5. Nicholson (1998) Ch 11; (2000) Ch 7.

$$\min_{\{x_i\}} C = \sum_{i=1}^{n} w_i x_i \ + \mu(q - f(x_1, \ldots, x_n))$$

The FOC look like:

$$w_i - \mu f_i = 0,\ i = 1, n$$
$$q - f(.) = 0$$

The SSOC is described by the hessian determinate:

$$H = \begin{vmatrix} -\mu f_{ij} & \vdots & -f_i \\ \cdots & & \cdots \\ -f_j & \vdots & 0 \end{vmatrix},\ i \& j = 1,\ n$$

The FOC and SSOC imply input demand curves of the sort:

$$x_i^* = x_i^*(w_1, \ldots, w_{n,} q) \qquad 1$$

Recognize that the envelope theorem tells us that μ^* is marginal cost. That is, the comparative static result of a change in output on optimized cost is the value of the Lagrangian multiplier, i.e., $\partial C^*/\partial q = \mu^*(q) = \text{MC}(q)$.

Results

There are four fundamental theorems derivable from cost function analysis:

1. Input demand is downward sloping.
2. Input demand curves are symmetric.
   a) Cross price effects are symmetric.
   b) The shift in marginal cost w.r.t. an input price is equal to the shift in the input's demand w.r.t. output.
3. Input demand curves are homogeneous of degree zero with respect to input prices.
   a) The sum of own and cross price elasticities is equal to zero.
   b) A proportional increase in all input prices must shift cost by the same amount holding output constant. (This means that average and marginal cost shift vertically by the factor of proportion.)
4. The magnitude of the effect of input price on average cost is equal to the ratio of input usage to output (envelope result).

With these results we can map the behavior of the firm in the dimension of input usage into price and quantity of inputs and of output.

The comparative statics of this problem start with the following system of equations:

$$\begin{bmatrix} -\mu f_{11} & \cdots & -\mu f_{1i} & \cdots & -\mu f_{1n} & -f_1 \\ \vdots & \ddots & \vdots & & \vdots & \vdots \\ -\mu f_{i1} & \cdots & -\mu f_{ii} & \cdots & -\mu f_{in} & -f_i \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ -\mu f_{n1} & \cdots & -\mu f_{ni} & \cdots & -\mu f_{nn} & -f_n \\ -f_1 & \cdots & -f_i & \cdots & -f_n & 0 \end{bmatrix} \bullet$$

$$\begin{bmatrix} \partial x_1^* / \partial w_1 & \cdots & \partial x_1^* / \partial w_i & \cdots & \partial x_1^* / \partial w_n & \partial x_1^* / \partial q \\ \vdots & \ddots & \vdots & & \vdots & \vdots \\ \partial x_i^* / \partial w_1 & \cdots & \partial x_i^* / \partial w_i & \cdots & \partial x_i^* / \partial w_n & \partial x_i^* / \partial q \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ \partial x_n^* / \partial w_1 & \cdots & \partial x_n^* / \partial w_i & \cdots & \partial x_n^* / \partial w_n & \partial x_n^* / \partial q \\ \partial \mu^* / \partial w_1 & \cdots & \partial \mu^* / \partial w_i & \cdots & \partial \mu^* / \partial w_n & \partial \mu^* / \partial q \end{bmatrix} =$$

$$\begin{bmatrix} -1 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & \cdots & -1 & \cdots & 0 & 0 \\ \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & -1 & 0 \\ 0 & \cdots & 0_i & \cdots & 0 & -1 \end{bmatrix}$$

First, we are interested in the nature of input demand. The derived result is: $\frac{\partial x_i^*}{\partial w_i} = -\frac{H_{ii}}{H} < 0$. The sign of this derivative is negative because based on the SSOC for a minimization problem, the principal minors of H are all of the same sign.

The second proposition follows from the fact that the matrix of coefficients and the matrix of constants defining the comparative statics of this problem are symmetric. As a result, the matrix of comparative static results is symmetric. This means that $\frac{\partial x_i^*}{\partial w_j} = \frac{\partial x_j^*}{\partial w_i}$ and $\frac{\partial \mu^*}{\partial w_i} = \frac{\partial x_i^*}{\partial q}$.

The third proposition is that the input demand curves are homogeneous of degree zero in input prices. If all input prices increase by the same amount there is no effect on the demand for any input holding output constant. This is true because no relative price changes; the identical proportional increase in all input prices cancels out of all pairs of the first n FOC.

Based on this we can apply Euler's Theorem, which gives:

$$\sum_{j=1}^{n} \frac{\partial x_i^*}{\partial w_j} w_j = 0$$

Dividing through by x_i^*, we convert to elasticities:

$$\sum_{j=1}^{n}\frac{\partial x_i^*}{\partial w_j}\frac{w_j}{x_i^*}=\sum_j \varepsilon_{ij}=0$$

This homogeneity result also affects the cost functions. Since a proportionate increase in all input prices does not change behavior, optimized cost is still defined by the same input usage levels, only now at a higher input price level. The proportion of input price increase factors out, which shows that optimized cost increases by this same amount. Marginal and average cost shift up vertically by this amount.[2]

The last proposition is an envelope theorem result. It comes from differentiating the Lagrangian objective function evaluated at the optimal values of the inputs w.r.t. an input price. Since all terms involving comparative static results cancel out based on the FOC we are left with:

$$\frac{\partial C^*}{\partial w_i}=x_i^*$$

In addition to these four fundamental results there are a couple of other things that are interestingly observed about the nature of cost based on this analysis. From these comparative static results we can deduce how marginal and average cost change w.r.t. a change in an input price. Because we know the relation between marginal and average cost in general (when average is falling, marginal lies below, and when average is rising, marginal lies above), we can infer the change in the shape of average cost that results from a change in an input price. Proposition four says that average cost must increase w.r.t. an increase in an input price. But since marginal cost may shift up or down, we can combine these effects in order to paint a picture of the change in average cost as an input price changes. We pursue this in a moment.

Some Definitions

Elasticity of Substitution—Silberberg (p. 287) defines this as the percentage change in input ratio per percentage change in input price ratio. For the Cobb-Douglas production function it has a value of 1.

The elasticity of substitution is a term that is kicked around. Paul Samuelson made a big deal out of it and his treatment has had a sustaining influence. I don't think it is necessarily a big deal. However, you need to be able to deal with it. Here's the way I think about it.

From the Model of Cost Minimization we derive input demand functions. That is, we get $K^*(w_K,w_L,q)$ and $L^*(w_K,w_L,q)$. The ratio of these is K^*/L^*, and in the context of the elasticity of substitution, this can be thought of as $(K/L)^*$. The optimal input ratio, $(K/L)^*$, can be found by solving the first two of the FOC.[3] The elasticity of substitution can be found by differentiating with respect to the input price ratio as follows:

---

[2] Note that even though marginal cost may shift in either direction (up or down) w.r.t. a change in a single price, a proportional change in all prices must cause marginal cost to shift up.

[3] For the two input case. In general it is found by solving any pair of the first n of the FOC where there are n inputs.

$$Elasticity\ of\ \ Substitution = \frac{\partial \left({}^{K}\!/_{L} \right)^*}{\partial\, {}^{w_L}\!/_{w_K}} \frac{{}^{w_L}\!/_{w_K}}{{}^{K^*}\!/_{L^*}}$$

You should work through this using the Cobb-Douglas production function.

Output Elasticity of an Input versus ***Input Elasticity of Output***

These are two expressions that are confusing. They may mean the same thing to some commentators. They may mean different things. They may mean different things to different people.

There are two ideas that map into these two expressions. Let's start with the ideas and then try to label them. From the production function, we have the concept of marginal product. Marginal product is a widely used concept. It means the partial derivative of output with respect to an input. A partial derivative means that we measure the change in output with respect to a change in input holding all other inputs constant. Thus, from

$$q = f(x_1, x_2)$$

we have marginal product,

$$\frac{\partial q}{\partial x_1} = f_1(x_1, x_2)$$

which can be converted into elasticity form:

$$\frac{\partial q}{\partial x_1}\frac{x_1}{q} = f_1(x_1, x_2)\cdot\frac{x_1}{q} \qquad 2$$

The other idea that we must consider comes from the Model of Cost Minimization. From this model we get input demand functions, which express the optimal use of each input given the target level of output and the prices of all of the inputs. The input demand function for input 1,

$$x_1^* = x_1^*(w_1, w_2, q)$$

can be differentiated with respect to output, $\frac{\partial x_1^*}{\partial q}$. This derivative tells us how the input demand shifts as the target level of output changes, or in other words, it tells us how the optimal input usage changes as target output changes holding the input prices constant. It definitely assumes that there is a simultaneous change in the usage of the other input so that its level is optimized. The derivative of the input demand function with respect to output can be converted to elasticity form and is often signified as follows:

$$\varepsilon_{1q} = \frac{\partial x_1^*}{\partial q}\frac{q}{x_1^*} \qquad 3$$

Equations (2) and (3) represent the concepts. I am going to label them as follows:

(2): Input elasticity of output
(3): Output elasticity of input

I don't really care what you call them. At all events you can never be really sure what someone else means when they use these terms. However, it is important to remember the concepts whatever they are called.

2. THE STRUCTURE OF COSTS

We typically assume that AC is U-shaped. We do this because it is convenient in describing the output market equilibrium. We don't always do it as we will discuss in the next lecture, but we do it often enough that it is important to discuss the nature of U-shaped average cost. To say that AC is U-shaped because it gives a unique firm size for competitive firms is not an explanation of why AC is U-shaped, but rather a plea for there to be a good explanation of why it is.

U-shaped average cost is often described in terms of economies and diseconomies of scale. Declining average cost is labeled economies of scale. Economies of scale are described in terms of returns to scale in production. Returns to scale in production are elucidated by reference to homogeneous production functions. However, homogeneous production functions cannot give U-shaped average cost. This is not to say that the discussion is wrong, but rather that you must be careful when thinking about economies and diseconomies of scale along a U-shaped average cost curve.

The best explanation of why AC is U-shaped comes from Alchian in his paper on "Costs and Outputs." Alchian describes the causes of declining average cost. He characterizes this as "volume" effects. When the firm plans to produce a lot, it uses more durable production processes, and the cost of durability does not increase linearly. On the other hand, when the firm must produce a lot in a hurry, average cost rises. Alchian attributes this to "rate" effects. Motors burn up, people run into each other, inventories get snafu'ed. Holding time constant, increasing output increases *both* volume and rate. Hence, both effects operate in the normal production setting. Volume effects dominate at low levels of output and rate effects dominate at high levels.

Read the paper that Lindsay and I wrote on economies of scale. It is an easy paper. It contains some good references to history of thought. Economies of scale refer to the percentage change in total, optimized resource expenditures as the firm adjusts its target level of output. Economies of scale are sometimes associated with the characteristics of the underlying production function. In this sense we can say that economies of scale, as defined above, flow from the nature of the production process. It is this that Lindsay and I try to develop in the paper.

The Elasticity of Average Cost[4]

Consider the average cost curve as we examine the alternatives available to the firm across its scale of output. It is useful to draw the long and short run curves paying careful attention to the association of long-run and short-run marginal costs. Notice the particularly striking case where long-run average cost is falling. Short-run average cost is tangent to the long-run function to the left of the minimum of the short and long run curves. Average cost is more inelastic in the short run than in the long run. That is, the *constrained* average cost function, the function with some amount of one input held fixed, is more negatively sloped to the left and more positively sloped to the right of the tangency.[5]

The marginal costs associated with these two average cost functions are quite different. Constrained marginal cost is most likely positively sloping and unconstrained marginal cost may be negatively sloped. Constrained and unconstrained marginal costs are *equal* at the quantity where constrained and unconstrained average costs are equal. However the marginal cost functions are not shaped the same, and, while the slopes of the average cost functions are equal at their tangency, the marginal functions have different slopes. Indeed, they may be of opposite sign. It is the fact that constrained marginal cost is steeper than unconstrained marginal cost that proves the point that constrained average cost is everywhere higher than unconstrained average cost, except where the constraint is not binding.

Let's prove that constrained marginal cost is steeper than unconstrained marginal cost. In the unconstrained problem, marginal cost is

$$\mu^* = \mu^*(w_1,\ldots,w_n,q)$$

based on input demand functions:

$$x_i^* = x_i^*(w_1,\ldots,w_n,q)$$

Constrained cost holds one (or more) of the inputs constant at the level that is optimal for the output associated with the tangency of the short-run and long-run average cost. That is, let $\hat{q}$ be the output level at the tangency of a short run average cost with the long run average cost. Then we can write:

$$\hat{x}_i = x_i^*(w_1,\ldots,w_n,\hat{q})$$

More generally, we can say that total cost in the unconstrained case is equal to the sum of the optimal resource expenditures for a given level of output. By defining a level of input *i* as fixed at what would be an optimal choice in some setting, i.e., $\hat{x}_i = x_i^*(w_1,\ldots,w_n,\hat{q})$, we have definitionally equated short-run total cost and long-run total cost.

---

4 SR &LR input demand from Silberberg, ie. Le Chatlier Principle, Silberberg (1990) Section 8.8. Portes, Richard D. "LR Scale Adjustment of a Perfectly Competitive Firm and Industry: An Alternative Approach: in American Economic Review 1971. p 430-34. Layard & Walters (1978) Ch 7, Section 6.2; Varlain (1992) Section 13.12.

5 To say that constrained average cost is more inelastic is a bit of a misnomer. The statement is used because of the symmetry with demand, i.e., the percentage change in output with respect to a percentage change in price. However, if we think of cost as a function of quantity, then the short-run average cost is more cost responsive to changes in output than is unconstrained or long-run average cost.

$$\hat{C} = \sum_{j \neq i} x_j^* w_j + \hat{x}_i w_i = C^* = \sum_{i=1,n} x_i^* w_i$$

Thus, at this output level constrained marginal cost equals unconstrained marginal cost and the same is true for the average costs. Indeed, because the marginal costs are equal, constrained average cost has the same slope as unconstrained average cost.

Marginal cost in the constrained case is given by minimizing cost by the optimal choice of the other inputs for this target level of output holding the i^{th} input constant. Call this:

$$\hat{\mu} = \hat{\mu}(\{w_{j \neq i}\}, \hat{x}_i, q)$$

As stated above, constrained and unconstrained marginal costs are equal at $\hat{q}$:

$$\mu^*(\{w_{i;i=1,n}\}, \hat{q}) = \hat{\mu}(\{w_{j \neq i}\}, \hat{x}_i, \hat{q})$$

Differentiating with respect to q gives:

$$\frac{\partial \mu^*}{\partial q} = \frac{\partial \hat{\mu}}{\partial q} + \frac{\partial \hat{\mu}}{\partial \hat{x}_i} \frac{\partial x_i^*}{\partial q} \qquad 4$$

This equation says that the slope of unconstrained marginal cost is equal to the slope of constrained marginal cost plus a term that measures the effect of the constraint on marginal cost. This term is negative (as we will show). That is, if the firm were able to vary x_i marginal cost would fall.

To develop this effect, we differentiate the two marginal costs with respect to w_i:

$$\frac{\partial \mu^*}{\partial w_i} = \frac{\partial \hat{\mu}}{\partial x_i} \frac{\partial x_i^*}{\partial w_i}$$

Rewriting:

$$\frac{\partial \hat{\mu}}{\partial x_i} = \frac{\partial \mu^*}{\partial w_i} \bigg/ \frac{\partial x_i^*}{\partial w_i}$$

Substituting using the symmetry conditions from the cost minimization model concerning the shift in marginal cost w.r.t. an input price:

$$\frac{\partial \hat{\mu}}{\partial x_i} = \frac{\partial x_i^*}{\partial q} \bigg/ \frac{\partial x_i^*}{\partial w_i}$$

Finally, substitute back into (1). This gives:

$$\frac{\partial \mu^*}{\partial q} = \frac{\partial \hat{\mu}}{\partial q} + \left(\frac{\partial x_i^*}{\partial q}\right)^2 \Bigg/ \frac{\partial x_i^*}{\partial w_i}$$

Thus the difference between unconstrained and constrained marginal cost is given by the second term on the right-hand side. It is negative because the denominator is the slope of the demand curve and the numerator is squared. Thus, constrained marginal cost is steeper than unconstrained marginal cost.

The difference depends on the output elasticity of the input. The stronger is the reaction of the input to changes in output, the bigger the difference in the slopes of the two marginal costs. This makes sense because the more the firm would wish to change but cannot, the larger should be the effect on costs. Also the more price responsive is the input, the bigger the difference.

Remember, the difference in the slopes of marginal costs reflects the difference in the elasticities of average cost. The fact that constrained marginal cost lies below unconstrained to the left of the tangency of constrained and unconstrained average cost means that as output falls, cost does not go down as fast in the constrained case. Hence, constrained average cost lies above unconstrained AC. On the other side, constrained MC lies above unconstrained. As output rises, constrained cost goes up faster than unconstrained. Thus, unconstrained AC lies below constrained.