3. Cost and Technology

Varian, Chapters 1-5

I. Describing the Firm

A. The neoclassical description of the firm is really just a description of the firm's production possibilities.

1. Which outputs can be obtained from given inputs
2. How much has to be spent to get those inputs
3. How these production possibilities generate cost curves for the firm.
4. These cost curves themselves completely describe *everything we need to know about the firm*, if we are neoclassical.

B. Input/Output

1. The firm produces a vector $\mathbf{y}$ of product quantities.
   a. We'll usually focus on a firm with a single product with quantity y.
2. The firm has a set of inputs it can use to create these products.
   a. We describe these inputs as a vector (or a bundle) $\mathbf{x} = (x_1, x_2, ..., x_n)$.
      - The vector components are the quantities of each input being utilized in production.

C. Technology

1. What output y can the firm produce with a bundle $\mathbf{x}$?
2. This is described by the firm's technology.
3. The **input requirement set** V(y) consists of all of the bundles $\mathbf{x}$ that can produce output quantity y.

   Ex: Activity analysis and production plans.
   Basically, recipes. For spaghetti sauce, for 16Gb memory chips, for 100 rides to SFO, ...

4. The **production function** $y = f(\mathbf{x})$ describes the maximum output that can be produced with any input bundle .

   Ex: Cobb-Douglas technology, $y = a_o x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$.

   Ex: Leontief technology , $y = \min\{a_1 x_1, a_2 x_2, ..., a_n x_n\}$.

5. The **isoquant** for given output level y^* is the set of input bundles that can produce y^* .

   It is the lower boundary of $V(y^*)$.

6. Analogous to an indifference curve, but here the label (y^*) is meaningful.

D. Common assumptions about technology

   1. Monotone

      a. More input enables at least as much output.

      b. Say this using V's: if you can produce y' with $\mathbf{x}$ you can still produce y' with a bigger bundle $\mathbf{x}' > \mathbf{x}$.

      c. This is innocuous if extra inputs can be thrown away, "free disposal."

   2. Convex

      a. If plans $\mathbf{x}$ and $\mathbf{x}'$ are in $V(y)$ (i.e. can produce y), then so is the mixture $\alpha\mathbf{x} + (1-\alpha)\mathbf{x}'$, for any mixing proportion $0 < \alpha < 1$.

      b. If a production plan can be replicated, then it is reasonable to say that the technology is convex.

         Ex: Replicating two production plans to create any convex sets.

   3. Non-empty

      a. With enough of and the right kinds of inputs, you can create any level of output y.

   4. Closed: a boring technical condition.

E. Trade Offs in Production Plans

1. Assume we have a "smooth" production technology $y = f(x_1, ..., x_n)$.
2. At what rate can we substitute one of our inputs for another in producing a particular output, y?
   a. Called the **technical rate of substitution**.
   b. With a smooth technology with two inputs, it is just the slope of the isoquant.
   c. TRS is a direct analogue of -MRS (the sign, by convention, is negative).

   Ex: Using the production function and the **implicit function theorem** to find the technical rate of substitution in terms of marginal products (mp_i):

$$TRS_{ij} = \frac{dx_j}{dx_i}|_{[f(\cdot)=y^*]} = -\frac{\partial f}{\partial x_i}/\frac{\partial f}{\partial x_j} = -\frac{mp_i}{mp_j} \tag{1}$$

   Ex: A Cobb-Douglas example.

3. Elasticity of substitution σ is elasticity of $[x_j/x_i]$ wrt $|TRS|$. It is a measure of isoquant curvature. See Varian for ugly details (optional).

F. Returns to Scale

1. The returns to scale tells us what happens when we try to scale up a production plan.
2. If we multiply $\mathbf{x}$ by $t > 0$, what happens to y?
3. Three cases:
   a. Constant returns to scale.
      - If $y = f(x_1, x_2)$, then for $f(tx_1, tx_2) = ty$
      - Output is proportional to the inputs.
   b. Increasing returns to scale.
      - If $y = f(x_1, x_2)$, then $f(tx_1, tx_2) > ty$ for $t > 1$.
      - We get more bang for our buck (at fixed prices of course) at higher scales of production.

c. Decreasing returns to scale.

- If $y = f(x_1, x_2)$, then $f(tx_1, tx_2) < ty$ for $t > 1$.
- We get diminishing returns from scaling our plans up.
- A major reason for DRS: there is some fixed input (not in the list), such as CEO attention, or planetary resources, or ...

Ex: Cobb-Douglas and returns to scale.

4. Homogeneity degree 1 as CRS. Homogeneous functions of degree $d = 0, 1, ...$.

II. Cost Minimization

A. Behavior of the firm

1. We assume that firms economize in production:
2. they choose input bundles that minimize the cost of producing their chosen level of output.
3. When is this reasonable to assume? For competitive firms, and even for unrestrained monopolists. Only two exceptions come to mind:

   a. rogue managers pursue self-interest at the expense of firm owners, e.g., buy unnecessary corporate jets;

   b. old-fashioned regulators set price of a monopoly firm based on actual costs. Then it might be in the firm's interest to inflate costs.

B. The firm's problem.

1. To derive cost function, take as given the desired output quantity y, and the input price vector $\mathbf{w} = (w_1, w_2, ..., w_n)$. Sometimes also called factor prices.
2. Firms choose an input bundle $\mathbf{x}$.

   a. For convenience we will often write $\mathbf{x} = (x_1, x_2)$ and $\mathbf{w} = (w_1, w_2)$, but the reasoning extends to any finite vector of inputs.

3. The firm's main constraint (aside from factor prices) is technological.

a. Can be summarized with the production function: $y = f(x_1, x_2)$.

4. So the firm's problem is simply:

$$c(\mathbf{w}, y) = \min_{x_1, x_2 \geq 0} w_1 x_1 + w_2 x_2 \text{ s.t. } f(x_1, x_2) = y \tag{2}$$

5. This cost function is an exact analogue of the expenditure function for consumers.

6. The four conditions noted earlier, including a strict version of convexity, allow us to say that if the problem above has an interior solution x^*, then it is unique and is characterized by the first order conditions, $w_i = \lambda m p_i, \quad i = 1, ..., n$.

7. Taking ratios of these first order conditions show that the technical rate of substitution is equal to the ratio of the factor prices: $|TRS_{ij}(x^*)| = \frac{w_i}{w_j}$.

8. The intuition is appealing: at the optimum input vector x^*, the isocost curve has the same slope (i.e., market tradeoff rate given by the price ratio) as the isoquant (the production tradeoff rate, TRS).

9. So far, this is entirely analogous to consumer choice. But here we get a bonus: the Lagrange multiplier λ is equal to marginal cost! This can be seen from the general interpretation as shadow price, here of output in terms of expenditure on inputs. It can also be seen by solving any of the FOCs for λ, since w_i/mp_i is the cost of increasing output by a (micro) unit via increasing the input i. The insight (to be elaborated later) is that that cost must be the same for all inputs used in positive quantities.

C. Conditional factor demand

1. The firm's cost minimizing problem yields the firm's demand for each input as a function of prices and the scale of output.

2. Conditional factor demand for input i is $x_i^*(w_1, w_2, y)$.

D. The cost function

1. Cost functions represent the lowest cost of production available to a firm at a given set of factor prices. So we can rewrite equation (2) as

$$c(\mathbf{w}, y) = \mathbf{w} \cdot \mathbf{x}^*(\mathbf{w}, y) \tag{3}$$

2. With only two factors: $c(\mathbf{w}, y) = w_1 x_1^*(w_1, w_2, y) + w_2 x_2^*(w_1, w_2, y)$

Ex: Constant Elasticity technology.

Special cases: Cobb-Douglas technology, Leontief technology, Linear technology

E. Relationship between cost and conditional factor demand

1. If the cost function is differentiable, then you can use it to recover the input (or factor) demand functions.
2. This is known as **Shephard's lemma**: $x_i^*(\mathbf{w}, y) = \frac{\partial c(\mathbf{w}, y))}{\partial w_i}$.
3. To verify, differentiate equation (3) wrt w_i. The main effect is as in Shepard's lemma, but there is also an indirect effect $\mathbf{w} \cdot \frac{\partial x_i^*(\mathbf{w}, y))}{\partial w_i}$. By the FOC, this indirect effect is proportional to $(mp_1, ...mp_n) \cdot \frac{\partial x_i^*(\mathbf{w}, y))}{\partial w_i} = 0$, as can be seen by differentiating the isoquant identity $f(\mathbf{x}^*(\mathbf{w}, y)) = y$.
   Geometrically, the idea is that the production function gradient (i.e., marginal product vector) is normal to (aka orthogonal or perpendicular to) the isoquant surface, and therefore also normal to the isocost surface, so the indirect effect is zero. This is an example of the envelope theorem. See Varian p. 74 for further discussion.

F. Duality and properties of cost functions

1. Suppose that you have a function $c(\mathbf{w}, y)$ with the following properties
   a. monotone increasing in each argument,
   b. homogeneous degree 1 in $\mathbf{w}$,
   c. concave in $\mathbf{w}$, and
   d. continuous and (at least piecewise) differentiable

Then there is some nice [monotone, ..., closed] production function for which c is the cost function!

Conversely, if c is the cost function for some nice production function, then it satisfies the four properties just mentioned.

2. Implications for applied work:

   (a) usually you can skip estimating a production function, especially if not all data on input quantities are available, and just estimate the cost function directly, using data on input prices and output levels, which usually is easier to collect.

   (b) when estimating a cost function, consider imposing homogeneity and monotonicity as coefficient restrictions, and testing for concavity.

Ex: CES cost function — see handwritten notes, attached below.

Ex: Translog cost function

$$\ln c = a_0 + a_1 \ln w_1 + a_2 \ln w_2 + \frac{1}{2} b_{11} [\ln w_1]^2 + b_{12} \ln w_1 \ln w_2 + \frac{1}{2} b_{22} [\ln w_2]^2 \ , \quad (4)$$

where homogeneity of degree 1 implies (a) $a_1 + a_2 = 1$ and (b) $b_{11} + b_{22} + 2b_{12} = 0$. See Varian pp. 210 for factor share calculations.

III. Cost Curves

A. Cost curves

1. It turns out we can summarize all of the (neoclassical) facts about the firm using the cost functions we derived in the last section.
2. The main ideas come through most clearly when we assume just two inputs to production which we will rename f (for fixed in SR) and v (for freely variable).

   a. $\mathbf{x} = (x_f, x_v)$

   b. $\mathbf{w} = (w_f, w_v)$

Notes on CES Cost Functions.

Consider the cost function $c(y, w_1, w_2) = y\left((w_1/a_1)^r + (w_2/a_2)^r\right)^{1/r}$.

(output quantity) (input prices)

when:

$r = 1 \Rightarrow$ linear isocost curves (perfect subs)

$r = 0 \Rightarrow$ Cobb-Douglas isocost curves $\rightarrow$ 

(use L'Hospital's Rule)

$0 < r < 1$: Isocost curves intersect axes $\Rightarrow$ neither input essential, but imperfect substitutes.

$r \leq 0$: isocost curves don't intersect axes $\Rightarrow$ both inputs essential.

$r \to -\infty$: Leontieff isocost curves $\longrightarrow$ 

The corresponding production function is $f(x_1, x_2) = \left((a_1 x_1)^\rho + (a_2 x_2)^\rho\right)^{1/\rho}$, where ρ and r are "dual": $\frac{1}{\rho} + \frac{1}{r} = 1$. —see Varian p 55-56.

Note that $r = 1 \Leftrightarrow \rho = -\infty$, $r = 0 \Leftrightarrow \rho = 0$, and $r = -\infty \Leftrightarrow \rho = 1$, i.e. linear, C-D and Leontieff cost fns correspond to Leontieff, C-D + linear production fns.

CES stands for constant elasticity of substitution. This elasticity σ is a measure of the curvature of ~~isoquants~~ cost curves - see Varian p 20 - and it turns out that $\sigma = \frac{1}{1-\rho} = r/\rho = 1 - r$. Thus $\sigma = 0$ for linear, $\sigma = 1$ for CD, $\sigma = \infty$ for Leontieff.

The above discussion assumes constant returns to scale. More generally, the production function is $\left((a_1 x_1)^\rho + (a_2 x_2)^\rho\right)^{\alpha/\rho}$, where $\alpha > 1 \Rightarrow$ increasing RS and $\alpha < 1 \Rightarrow$ decreasing RS.

Extra Credit for the ambitious: find the cost function corresponding to $\alpha \neq 1$.

From π-max FOCs, obtain estimating equation

$\ln(x_1/x_2) = a_0 + \sigma \ln(w_2/w_1)$, where $a_0 = -\frac{\sigma}{\rho}\ln(a_2/a_1)$

B. The short run cost curve

1. In the short run, some of the costs are fixed. That is, the firm is unable to mitigate costs by varying all of its inputs.

   a. In the short run we can't buy new machinery, build new factories, hire new management, or change union contracts.

   b. Those costs are fixed – we represent them as F.

   c. Part of F can be recovered if we halt production (e.g.s). This part is called **avoidable**. the remainder is called **sunk**.

   d. Buried here is a general point: economic costs = opportunity costs, not necessarily cash or accrual costs. Ask yourself: does F change in the SR when w_f increases?

2. Total costs incurred by the firm consist of both variable costs and fixed costs (which are in turn simple to express in the two inputs model).

   a. Variable Cost: c_v

      - $c_v(y) = w_v x_v(w_v, y, x_f)$

      Ex: Variable Cost Curve

   b. Fixed Cost: F

      - $F = w_f x_f$

      Ex: Fixed cost curve

   c. Total Cost: $c_v(y) + F$

      - $c(y) = w_v x_v(w, y, x_f) + w_f x_f$

      Ex: Total cost curve

3. Average Costs

a. There are three main types of average costs that are used in studying firm behavior.

- Average Total Cost: $AC(y) = \frac{c(y)}{y} = \frac{c_v(y)+F}{y}$
  - U-shaped
- Average Variable Cost: $AVC(y) = \frac{c_v(y)}{y}$
  - Eventually rising
- Average Fixed Cost
  - Always falling

b. The three average cost functions are related by a simple equation:

c. $AC(y) = AFC(y) + AVC(y)$

Ex: Average cost curves.

4. Marginal Costs

a. $MC(y) = \frac{\partial c}{\partial y} = \frac{\partial c_v}{\partial y}$

b. Relationship between MC and AC.

- When AC is decreasing, MC must be smaller than AC.
- When AC is increasing, MC must be larger than AC.
- MC intersects AC at the minimum point of the AC.
  - Minimum efficient scale
- The same must be true with AVC!
- Integral of (area under) MC gives VC.

Ex: Suppose TC is $c(y, \mathbf{w}) = 128 + 69y - 14y^2 + y^3$ for some fixed input price vector $\mathbf{w}$. Find FC, MC, VC, AVC, etc. Where appropriate, assume all fixed costs are sunk.

To look ahead a bit, find the short run supply curve by solving $p = MC_+(y)$ for y. That is, obtain $y^*(p, \mathbf{w})$ from the increasing portion MC_+ of the marginal cost curve. (Later we will see that not all of MC_+ is relevant, just the part above the AVC curve.)

C. Long run cost curve

1. As we saw in the last section, decreasing returns set in when a factor is fixed.
2. Eventually, the cost of short run factors force the firm to change long run factors.
3. Relationship between long and short term curves

   a. Suppose costs are given as a function of output and the corresponding demand for the fixed factor, $x_f(y)$: $c(y, x_f(y))$.

   b. Fix y at some level $\bar{y}$ and refer to $\bar{x}_f$ as the optimal amount of the fixed factor for producing $\bar{y}$.

   c. Now lets differentiate $c(\bar{y}, x_f(\bar{y}))$ with respect to y at $\bar{y}$.

   - This gives us the following:

   $$\frac{dc(\bar{y},x_f(\bar{y}))}{dy} = \frac{\partial c(\bar{y},\bar{x}_f)}{\partial y} + \frac{\partial c(\bar{y},\bar{x}_f)}{\partial x_f}\frac{\partial x_f(\bar{y})}{\partial y}$$

   d. **Envelope theorem**

   - The conditional factor demand, x_f is the cost minimizing factor choice for producing $\bar{y}$.
   - Thus $\frac{\partial c(\bar{y},\bar{x}_f)}{\partial x_f} = 0$
   - Plugging this into the math above we get a great expression for the relationship between long run cost and short run cost.

   $$\frac{dc(\bar{y},x_f(\bar{y}))}{dy} = \frac{\partial c(\bar{y},\bar{x}_f)}{\partial y}$$

   - That is the slope of long run cost equals the slope of short run cost! (at the point of tangency)
   - The long term costs are the lower envelope of all of the short run costs at various levels of production.

D. Learning curve

1. Experience may enable a firm to discover better procedures and techniques, avoid waste, etc.
2. First quantified in WWII aircraft and shipbuilding, true in teaching classes, manufacturing memory chips, etc etc.
3. The usual specification is in accumulated output $Y_t = \sum_{s \leq t} y_s$ that AC falls proportionately, $\ln AC_t = AC_0 - b \ln Y_t$.

E. Multiproduct firms.

1. Suppose that the joint cost function (estimated directly) is $c(y_1, y_2; \mathbf{w})$.
2. **Economies of scope** exist if $c(y_1, y_2; \mathbf{w}) < c(y_1, 0; \mathbf{w}) + c(0, y_2; \mathbf{w})$, e.g., because distribution networks can be shared, or R&D, or production facilities.
3. Another reason is the presence of cost complementarities, $\frac{\partial^2 c}{\partial y_1 \partial y_2} < 0$, i.e., increasing the output of one product lowers the MC of the other output. E.g., Big Creek Lumber product 1=redwood siding, product 2 = redwood sawdust.

Ex: Multiple plants – a lead-in to supply curve aggregation.