Outline

1. Introduction
2. Cost Minimization
3. Returns to Scale and the Cost Function
4. Long-Run and Short-Run Costs
5. Fixed, Quasi-Fixed, and Sunk Costs

1 Introduction

In the previous lecture we discussed how a firm would choose inputs and outputs to maximize profits, given the output price, input prices, and technology.

This lecture focuses on a distinct, but related question: how should a firm choose inputs to minimize the costs of producing a given level of output, again taking inputs prices as given (the output price is irrelevant).

Cost-minimization is important for two reasons.

First, in many instances economic entities do not necessarily want to maximize profits, but they should still want to minimize costs. Non-profit institutions, such as universities, charities, some hospitals, and so on, are standard examples. Harvard is not in the business of maximizing profits: it is by design a non-profit that seeks to promote other objectives such as learning, truth, and so on.

Harvard still should want to minimize its costs, however, given the objectives that it pursues, because reducing costs leaves more revenue to promote these objectives.

Second, even when we are focusing on a profit-maximizing entity, any profit maximization solution must also minimize costs.
Why? Assume not. That is, consider a set of choices for the inputs and outputs that allegedly maximizes profits, but does not minimize the cost of producing the chosen level of output.

Then, consider holding output constant, but changing the mix of inputs so that the firm is minimizing costs given that level of output. Costs then decline, so profits go up - thus, the original choices cannot have been profit-maximizing in the first place.

In other words, a necessary condition to be maximizing profits is that the firm be minimizing costs. And, in many contexts, we gain substantial insight into the firm’s behavior by focusing just on cost-minimization.

We therefore spend a couple of lectures characterizing the cost-minimization problem and related issues.

2 Cost Minimization

Suppose a firm has access to a production function that relates two inputs, $x_1$ and $x_2$, to one output, $y$. Assume the firm wants to find the cheapest way to produce a given level of output.

Then the problem to be solved is

\[
\min_{\{x_1,x_2\}} w_1 x_1 + w_2 x_2
\]

subject to \( f(x_1, x_2) = y \).

This is a problem of minimization subject to a constraint. It is therefore more involved mathematically than the profit-maximization problem we did earlier. So, we start with a graphical presentation and then discuss several variations on a calculus approach.

(NB: the fact that this is a minimization rather than a maximization is not the reason for the added difficulty; that difference simply changes the sign of the appropriate second-order condition.)

So, consider the following graph:
Graph: Solution to the Cost-Minimization Problem

\[ 4 = x^{1/2}y^{1/2} \]
\[ y = 8 - x \]

The graph shows the isoquant associated with the given level of output, \( y \).

The graph also shows a “cost line.” This is the graph of the equation

\[ C = w_1 x_1 + w_2 x_2 \]

We can also write the equation as

\[ x_2 = -\frac{w_1}{w_2} x_1 + \frac{1}{w_2} C \]

This makes it clear that cost lines closer to the origin correspond to lower levels of costs.

Given the isoquant, the firm wants to choose a combination of \( x_1 \) and \( x_2 \) that puts it on the cost line with the lowest possible level of costs.

This must occur at the cost line that is tangent to the isoquant, as illustrated. For cost lines that do not touch the isoquant, the set of inputs does not allow the
firm to produce the given level of output. For cost lines that intersect rather than just being tangent, some other cost line involves enough inputs to produce the given level of output but at lower costs. The only cost line that minimizes cost and satisfies the constraint must be tangent (for a differentiable isoquant).

Note that in this problem we are taking the “curved line” as given and choosing the straight line that achieves an optimum. We are also minimizing rather than maximizing, which corresponds to the fact that we want to be on a cost line that is as close to the origin as possible. These issues aside, the problem here is similar in structure to the utility maximization model we examined earlier.

The graphical analysis indicates that, at an optimum, the isoquant and the cost line must be tangent – that is, the slopes must be equal. The slope of the cost line is

$$-\frac{w_1}{w_2}$$

and the slope of the isoquant is

$$-\frac{MP_1(x_1, x_2)}{MP_2(x_1, x_2)} \quad = \quad -\frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}}$$

so the cost-minimizing choice of $x_1$ and $x_2$ must satisfy

$$\frac{MP_1(x_1, x_2)}{MP_2(x_1, x_2)} \quad = \quad \frac{w_1}{w_2}$$

In words, the ratio of the marginal products must equal the ratio of the input prices. Intuitively, if this were not true, the firm could decrease cost by substituting away from the input that was providing the lower level of marginal product per unit of input price.

(Remember that because of diminishing marginal product, substituting away from the input would raise its marginal product and increase the ratio of marginal product to input price.)

Note that this condition is one equation in two unknowns. So, we need another equation to pin down the solutions for $x_1$ and $x_2$; this equation is the production function for the level of output the firm is trying to produce at minimum cost.
The solution of these two equations yields the cost-minimizing values of $x_1$ and $x_2$. These are conditional factor demands. They are conditional on $y$. We often write these as

$$x_1(w_1, w_2, y)$$

$$x_2(w_1, w_2, y)$$

We can plug these into the formula for costs and get what is known as the cost function:

$$c(w_1, w_2, y)$$

This function tells us the minimum cost of producing the output level $y$, given the input prices.

Now let’s look at this using calculus. The problem again is

$$\min_{\{x_1, x_2\}} w_1 x_1 + w_2 x_2$$

subject to $f(x_1, x_2) = y$.

This is an optimization problem subject to a constraint, but we can handle it without using a Lagrangian under certain conditions.

In particular, assume we can solve explicitly for $x_2$ as a function of $x_1$ and $y$:

$$x_2 = g(x_1, y).$$

That is, assume such a $g(\cdot)$ function exists. It does not always exist, but under reasonable assumptions about the production function, we can in principle find the appropriate $g(\cdot)$.

Then we can substitute this formula for $x_2$ into the objective function and consider an unconstrained minimization of one variable:

$$\min_{\{x_1\}} w_1 x_1 + w_2 g(x_1, y)$$

This is not always possible, but when it is, this technique is straightforward.
Let’s look at some examples. Assume first that the production function is

\[ y = ax_1 + bx_2 \]

Then it is trivial to solve for \( x_2 \) as a function of \( y \) and \( x_1 \):

\[ x_2 = (1/b)(y - ax_1) \]

Of course, in this case, employing calculus naively will not get us very far. Consider the cost function once we have substituted in for \( x_2 \):

\[ \min_{\{x_1\}} w_1 x_1 + w_2 (1/b)(y - ax_1) \]

If we differentiate this with respect to \( x_1 \), we get an equation that does not include \( x_1 \), so we cannot solve it for \( x_1 \). But, this example illustrates how one can solve the production function for one of the inputs, and then substitute the result into the original problem. (To solve this problem, consider the corner solutions of using only \( x_1 \) or only \( x_2 \).)

As a second example, assume the production function is

\[ y = x_1^{1/2} x_2^{1/2} \]

Then solving for \( x_2 \) yields

\[ x_2 = y^2 x_1^{-1} \]

and substituting this into the cost equation yields the unconstrained minimization problem

\[ \min_{\{x_1\}} w_1 x_1 + w_2 y^2 x_1^{-1} \]

The FOC for this problem is

\[ w_1 - w_2 y^2 x_1^{-2} = 0 \]

and the solution for \( x_1 \) is

\[ x_1 = \left( \frac{w_2}{w_1} \right)^{1/2} y \]
This solution makes sense at least qualitatively: the demand for $x_1$ increases with $y$, decreases with $w$, and increases with $w_2$. There is, of course, a symmetric solution for $x_2$.

**OPTIONAL MATERIAL:** Still a third way to approach this problem is by defining a Lagrangian. This approach is not required for the course, but it is useful to see it as, in some ways, it is more natural than the substitution method (although the two ways are entirely equivalent).

The Lagrangian method proceeds as follows. First, we introduce a new variable, $\lambda$, that is known as the Lagrange multiplier. (In problems with multiple constraints, there is one multiplier per constraint.) Second, we define the Lagrangian function as follows:

$$\mathcal{L}(x_1, x_2, \lambda) = w_1 x_1 + w_2 x_2 + \lambda(y - f(x_1, x_2))$$

That is, we define a new objective function that equals the original objective function plus the Lagrange multiplier multiplied by the constraint. The constraint term is always written so that its value equals zero when it is satisfied.

Now we minimize the Lagrangian by choice of $x_1$, $x_2$, and $\lambda$. That is, we take three partial derivatives and set these equal to zero:

$$\frac{\partial \mathcal{L}(x_1, x_2, \lambda)}{\partial x_1} = w_1 - \lambda \frac{\partial f(x_1, x_2)}{\partial x_1} = 0$$

$$\frac{\partial \mathcal{L}(x_1, x_2, \lambda)}{\partial x_2} = w_2 - \lambda \frac{\partial f(x_1, x_2)}{\partial x_2} = 0$$

$$\frac{\partial \mathcal{L}(x_1, x_2, \lambda)}{\partial \lambda} = y - f(x_1, x_2) = 0$$

Note that this is three equations in three unknowns. To solve, first divide the first equation by the second to get one equation, and then use this new equation in combination with the constraint. This gives

$$\frac{w_1}{w_2} = \frac{\frac{\partial f(x_1, x_2)}{\partial x_1}}{\frac{\partial f(x_1, x_2)}{\partial x_2}}$$
\[ y = f(x_1, x_2) \]

That is, it gives the conditions we derived graphically - that the ratio of the marginal products must equal the ratio of the input prices, and that the amounts of the inputs must be sufficient to produce the given amount of output.

So, this approach is equivalent to the others. It is just more general and convenient. For example, it works with more than just two inputs or general production functions, assuming appropriate differentiability conditions. END OPTIONAL MATERIAL

3 Returns to Scale and the Cost Function

We earlier discussed economies of scale in relation to the production function. To repeat, a technology is said to have increasing, decreasing, or constant returns to scale as \( f(tx_1, tx_2) \) is greater, less than, or equal to \( tf(x_1, x_2) \) for all \( t > 1 \).

It turns out there is a useful relation between returns to scale of the production function and certain characteristics of the cost function.

Suppose we have constant returns to scale. Assume we have solved the cost-minimization problem for 1 unit of output, so we know

\[ c(w_1, w_2, 1) \]

Given this, how should the firm produce \( y \) units of output?

The answer, under CRS, is by using \( y \) times as much of every input. We know this will produce exactly \( y \) units of output. We also know it must be the cheapest way to produce \( y \) units of output; otherwise, the production of each individual unit was not done using the cost-minimizing method.

Thus, the minimal cost to produce \( y \) units of output would just be

\[ c(w_1, w_2, 1)y \]
So, in this case, the cost function is linear in output, i.e., it equals a specific number multiplied by output, and that number is \( c(w_1, w_2, 1) \).

Now consider the case of decreasing returns to scale. We can still calculate

\[ c(w_1, w_2, 1) \]

That is, we can still calculate the minimum cost of producing one unit.

Now, however, we know that multiplying all inputs by \( y \) will produce less than \( y \) units of output; so, to produce \( y \) units, we must add more inputs, and that means that costs to produce \( y \) units must increase more than one-for-one with \( y \).

Similarly, for an increasing returns to scale technology, costs go up less than one-for-one with \( y \).

We can also state this in terms of the behavior of the average cost function, which is costs per unit of output:

\[ AC(y) = \frac{c(w_1, w_2, y)}{y} \]

If the technology has CRS, then

\[ c(w_1, w_2, y) = c(w_1, w_2, 1)y. \]

This means

\[ AC(y) = \frac{c(w_1, w_2, 1)y}{y} = c(w_1, w_2, 1) \]

For decreasing or increasing returns to scale, \( AC(y) \) is either increasing or decreasing in \( y \).

Note that a cost function can have both regions with increasing average costs and regions with decreasing average costs. We will see this shortly.
4 Long-Run and Short-Run Costs

We often want to distinguish costs conditions between the long run and the short run, as we did with profit maximization.

Suppose that in the short run, factor 2 is fixed at some pre-determined level. Then the short-run cost function is defined by

\[ c_s(y, \bar{x}_2) = \min_{\{x_1\}} w_1 x_1 + w_2 \bar{x}_2 \]

subject to \( f(x_1, \bar{x}_2) = y \)

This is a simple problem: the constraint determines how much \( x_1 \) to use, and then costs are determined by that choice. (If there were three or more factors, the problem would be more complicated, but the key insights below would not change).

We can therefore write the short-run, conditional demands for \( x_1 \) and \( x_2 \) as follows:

\[ x_1 = x_1^s(w_1, w_2, \bar{x}_2, y) \]
\[ x_2 = \bar{x}_2 \]

Then, by definition,

\[ c_s(y, \bar{x}_2) = \omega_1 x_1^s(w_1, w_2, \bar{x}_2, y) + w_2 \bar{x}_2 \]

where \( c_s(y, \bar{x}_2) \) is the short-run cost function. This equation just says that the minimum cost of producing output \( y \) is the cost associated with using the cost-minimizing choice of inputs. This is true by definition, but it turns out to be useful.

The long-run cost function is defined by

\[ c(y) = \min_{\{x_1, x_2\}} w_1 x_1 + w_2 x_2 \]

subject to \( f(x_1, x_2) = y \)

The difference here is that both factors can vary. Note that the long-run cost depends on the output level and the factor prices.
We can write the long-run cost function and the long-run conditional factor demands as

\[ x_1 = x_1(w_1, w_2, y) \]
\[ x_2 = x_2(w_1, w_2, y) \]
\[ c(y) = w_1 x_1(w_1, w_2, y) + w_2 x_2(w_1, w_2, y) \]

The interpretation again is just that costs equal the input prices times the cost-minimizing choices of the factors.

The key thing we want to note is then the following: the long-run cost function can also be written as

\[ c(y) = c_s(y, x_2(y)) \]

This says that the minimum costs of producing the amount \( y \) when all factors are variable is just the minimum cost when factor 2 is held fixed, assuming it is held fixed at the level that minimizes LR costs. Note that it is the long-run conditional factor demand in this expression.

It follows that the LR demand for the variable factor—the cost minimizing choice—is given by

\[ x_1(w_1, w_2, y) = x^*_1(w_1, w_2, x_2(y), y) \]

This equation say that the cost-minimizing amount of the variable factor in the LR is the amount that the firm would choose in the SR—if it happened to have the LR cost-minimizing amount of the fixed factor.

5 Fixed, Quasi-Fixed, and Sunk Costs

It is useful to discuss certain types of costs that can arise in various settings.

Certain costs of producing output are fixed rather than variable: they do not depend on the level of output being produced.
Examples of fixed costs include the following: having one cashier behind the counter of a convenience store (the cost is the same regardless of how much merchandise is sold); the drive to work and back (it takes the same amount of time no matter how hard you work once there); the cost of a license to operate a liquor store (the cost does not depend on how much liquor the store sells).

The key thing about fixed factors is that they must be paid whether or not any output is produced.

A different, but related category is quasi-fixed costs: these are costs that do not vary with output, but they must be paid if the firm decides to produce a positive amount.

An example of a quasi-fixed costs is turning on the lights in a factory or heating a building. If the store or factory stays closed, these costs do not have to be paid. If the store or factory opens, the amount to be paid is not dependent on output.

A third key category is sunk costs: fixed costs one cannot recover.

Costs can be fixed but not sunk. For example, consider a lawn service that buys a lawnmower for $500. The cost is fixed whether it is used to mow 10 lawns or none, but it is not necessarily sunk. If the service can resell the mower for the same price it paid, the costs are recoverable rather than sunk. If the service can re-sell for some fraction of the original price (e.g., the price minus the value of depreciation), then only the difference between the purchase price and the re-sale price is sunk.