Unit 5.2: Costs

Michael Malcolm

June 18, 2011

1 Short-Run Cost Function

Suppose that a firm employs L units of labor and K units of capital and has a production function $q = f(L, K)$. Each unit of labor costs w and each unit of capital costs r. What then is the firm's *cost* of producing q units of output?

A firm's total costs are equal to its labor costs plus its capital costs.

$$Cost = wL + rK$$

In the short-run K is fixed, so we simply solve for the level of L needed to produce q units of output, then substitute back into the definition of costs. As an example, consider a firm whose production function is $q = \sqrt{L}\sqrt{K}$ and has $\bar{K} = 100$ units of capital in the short run. Each unit of capital costs $r = 5$ and each unit of labor costs $r = 2$.

Fixing the amount of capital, we can solve for the amount of labor needed to produce q units of output:

$$\begin{aligned} q &= \sqrt{L}\sqrt{K} \\ q &= \sqrt{L}\sqrt{100} \\ \frac{q}{10} &= \sqrt{L} \\ L &= \frac{q^2}{100} \end{aligned}$$

Then the firm's short run cost of producing q units of output is:

$$\begin{aligned} Cost &= wL + rK \\ &= 2\left(\frac{q}{100}\right) + 5(100) \\ &= \frac{q^2}{50} + 500 \end{aligned}$$

2 Short-Run Cost Relations

Total Fixed Cost (TFC) is the cost of fixed inputs that do not vary with the level of output produced (i.e. capital costs). *Total Variable Cost* (TVC) is the cost of variable inputs, increasing as output rises in the short-run (i.e. labor costs). *Total Cost* (TC) is the sum of both the fixed and variable costs.

Average Fixed Cost (AFC) is the fixed cost per unit of output – $AFC = \frac{TFC}{q}$. *Average Variable Cost* (AVC) is the variable cost per unit of output – $AVC = \frac{TVC}{q}$. *Average Total Cost* (ATC) is the total cost per unit of output – $ATC = \frac{TC}{q}$.

Marginal Cost (MC) is the additional cost of producing one more unit of output. As such, it is the derivative of the total cost function with respect to output – $MC = \frac{\partial TC}{\partial q}$. Notice that marginal cost could also be defined as $MC = \frac{\partial TVC}{\partial q}$. Any change in costs must be attributable to a change in variable costs since, by definition, fixed costs do not change in the short run.

Notice that these definitions pertain to the short run since they distinguish between fixed and variable costs. All inputs can be changed in the long run, so by definition there are no fixed costs in the long run. To illustrate with an example, consider a firm whose short-run cost of producing q units of output is.

$$Cost = 450 + 100q - 4q^2 + 0.2q^3$$

This is the total cost function. Of this, 450 is a fixed cost that has to be paid regardless of q, and so $100q - 4q^2 + 0.2q^3$ is the variable cost. Summarizing

$$\begin{aligned} TFC &= 450 \\ TVC &= 100q - 4q^2 + 0.2q^3 \\ TC &= 450 + 100q - 4q^2 + 0.2q^3 \end{aligned}$$

Average costs are as follows.

$$\begin{aligned} AFC &= \frac{TFC}{q} = \frac{450}{q} \\ AVC &= \frac{TVC}{q} = \frac{100q - 4q^2 + 0.2q^3}{q} = 100 - 4q + 0.2q^2 \\ ATC &= \frac{TC}{q} = \frac{450 + 100q - 4q^2 + 0.2q^3}{q} = \frac{450}{q} + 100 - 4q + 0.2q^2 \end{aligned}$$

The marginal cost is the first derivative.

$$MC = \frac{\partial TC}{\partial q} = 100 - 8q + 0.6q^2$$

The standard diagram of the average variable cost, average total cost and marginal cost curves is shown in figure 1.

3 AFC, AVC and ATC

Average fixed costs declines as output rises. To see why, since $AFC = \frac{TFC}{q}$, TFC is constant as q rises, meaning that AFC falls. Since fixed costs don't change, then the fixed cost *per unit* falls as output rises.

Since $ATC = AFC + AVC$, observe that ATC is always higher than AVC. However, the two get closer together as output rises. Since the distance between AVC and ATC is AFC, this means that AVC approaches ATC as output rises since AFC falls as q falls.

4 Shape of MC and ATC curves

How much extra labor is needed to raise output by one unit? We could express this mathematically as:

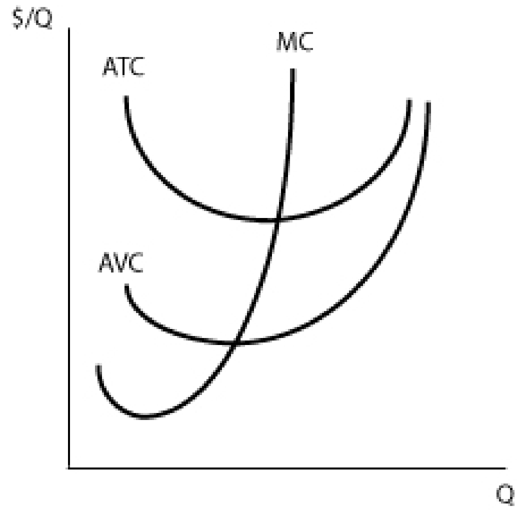

Figure 1: AVC, ATC and MC curves

$$\frac{\partial L}{\partial q} = \frac{1}{\frac{\partial q}{\partial L}} = \frac{1}{MPL}$$

Intuitively, if the marginal product of labor is $MPL = 10$, meaning that each worker produces 10 additional units of output, then the firm needs $\frac{1}{10}$ of a worker in order to produce one more unit of output. Now, since $\frac{1}{MPL}$ workers are needed to produce one extra unit of output and each worker costs w, then the marginal cost of the additional unit of output is

$$MC = w\frac{1}{MPL} = \frac{w}{MPL}$$

As MPL rises, marginal cost falls. This makes sense since, if additional workers are more productive, then the cost of producing each additional unit of output falls. Conversely, as MPL falls, marginal cost rises. But this tells us something important: *diminishing marginal returns is equivalent to rising marginal cost*. If each additional worker produces less output than the previous worker, then the cost per extra unit of that output will be rising as less productive workers are employed. The level of output at which diminishing marginal returns sets in is exactly the point at which the marginal cost curve begins to rise.

Since the law of diminishing marginal returns states that the marginal product of labor will eventually decline, this implies that the marginal cost curve must eventually slope upwards. This explains the typical shape of the MC curve – downward sloping initially but eventually sloping upwards.

As for the shape of the ATC curve, there are two opposing forces. On one hand, AFC falls as output rises, which pulls ATC down since AFC is included in ATC. On the other hand, diminishing marginal returns means that workers are less productive, which tends to bring average costs up. Considering both effects, ATC is typically parabolic. Where ATC switches from sloping downwards to sloping upwards is where the latter effect becomes stronger than the former.

5 Average Costs and Marginal Costs

The MC curve intersects the ATC curve at the minimum point of the ATC curve. The intuitive reason is similar to the reasoning for the relationship between the average product and marginal product functions. If

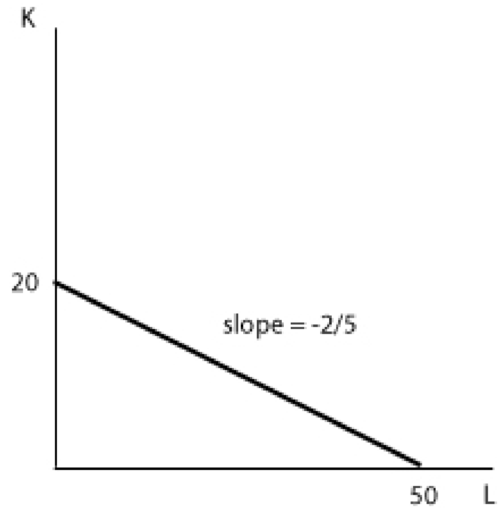

Figure 2: Typical isocost line

$MC < ATC$ then ATC is falling since the cost of an additional unit of output is less than the average cost, which pulls the average down when that additional unit is produced. Similarly, if $MC > ATC$, then ATC is rising since an additional unit costs more than the average unit, which pulls the average up. Combining these two implies that $MC = ATC$ at the minimum point of ATC – exactly where ATC switches from sloping downwards to sloping upwards.

Mathematically, $ATC = \frac{TC}{q}$. The minimum occurs where $\frac{dATC}{dq} = 0$. Computing the derivative using the quotient rule:

$$\begin{aligned}\frac{dATC}{dq} = \frac{\frac{dTC}{dq} \cdot q - TC \cdot 1}{q^2} &= 0\\ MC \cdot q - TC &= 0\\ MC &= \frac{TC}{q}\\ MC &= ATC\end{aligned}$$

In other words, the minimum occurs where $MC = ATC$. Exactly the same argument could be applied to show that MC also intersects AVC at the minimum point of AVC.

6 Long Run Cost Minimization

An *isocost line* gives combinations of K and L that cost the same amount. For example, if each unit of labor costs $w = 2$ and each unit of capital costs $r = 5$, then the isocost line corresponding to $\bar{C} = 100$ is sketched in figure 2.

To derive the endpoints, if the firm employs only labor then it can hire $L = 50$ workers when $w = 2$. Similarly, if the firm employs only capital then it can hire $K = 20$ units of capital at $r = 5$. The slope of the isocost line is $-\frac{w}{r}$, and higher isocost lines represent higher costs. The construction is similar to a budget line in consumer theory.

Given a particular level of output to produce, the firm's objective is to do so using the input basket with the lowest possible cost. In other words, the firm looks for the lowest possible isocost line corresponding to

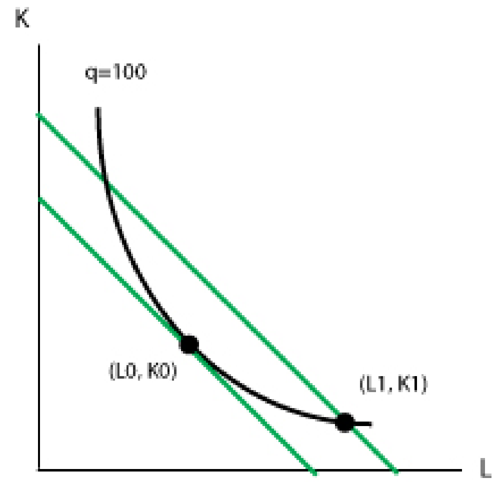

Figure 3: Finding the lowest isocost line corresponding to an isoquant

a fixed isoquant.

Consider figure 3. The isoquant shows many different input bundles that all produce $q = 100$ units of output. The input mix (L_1, K_1) produces 100 units of output, but this is not the cheapest way to do so. The input mix (L_0, K_0) produces 100 units of output at the lowest possible cost. It corresponds to the lowest possible isocost line that achieves the $q = 100$ isoquant.

The optimization is similar to consumer theory, except in consumer theory the primitive is a fixed budget line and we locate the highest indifference curve on the budget line. In firm theory, we locate the lowest possible isocost line that attains a fixed isoquant. In either case, though, the condition for the optimal bundle / input mix is the same. Geometrically, the slope of the isoquant must be equal to the slope of the isocost line. The slope of the isoquant is the negative of the $MRTS$ and the slope of of the budget line is $-\frac{w}{r}$. Thus, *at the optimal input mix*:

$$MRTS = \frac{w}{r}$$

Recall that the marginal rate of technical substitution is defined as $MRTS = \frac{MPL}{MPK}$. Substituting in this definition and rearranging:

$$\frac{MPL}{MPK} = \frac{w}{r}$$
$$\frac{MPL}{w} = \frac{MPK}{r}$$

In words, this condition says that – at the optimal input mix – the extra output per dollar spent on labor should equal the extra output per dollar spent on capital. This is intuitive. If each dollar spent on labor generated more output than each dollar spent on capital, then the firm should spend more on labor and spend less on capital. As it does this, the MPL falls and the MPK rises. The firm should continue this reallocation until equality in the ratios $\frac{MPL}{w}$ and $\frac{MPK}{r}$ is achieved. Intuitively, this condition is analogous to the condition from consumer theory that marginal utility per dollar should be equal across goods.

Notice that changes in input prices would change the firm's optimal input mix. If w rose, then the isocost lines would steepen and the firm would employ an input mix with less labor and more capital.

7 Numerical Example

Consider a firm whose production function is $q = f(L, K) = L^{\frac{2}{3}} K^{\frac{1}{3}}$. The firm's objective is to find the input combination L and K that produces q units of output at the lowest possible cost. Formally, the problem is:

$$\min wL + rK \qquad s.t. \quad L^{\frac{2}{3}} K^{\frac{1}{3}} = q$$

The Lagrangian is:

$$\mathcal{L} = wL + rK + \lambda(q - L^{\frac{2}{3}} K^{\frac{1}{3}})$$

The first order conditions are:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial L} &= w - \frac{2}{3}\lambda L^{-\frac{1}{3}} K^{\frac{1}{3}} = 0 \\
\frac{\partial \mathcal{L}}{\partial K} &= r - \frac{1}{3}\lambda L^{\frac{2}{3}} K^{-\frac{2}{3}} = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= q - L^{\frac{2}{3}} K^{\frac{1}{3}} = 0
\end{aligned}$$

Solving the first two first order conditions for λ:

$$\begin{aligned}
w = \frac{2}{3}\lambda L^{-\frac{1}{3}} K^{\frac{1}{3}} &\Rightarrow \lambda = \frac{3w}{2L^{-\frac{1}{3}} K^{\frac{1}{3}}} \\
r = \frac{1}{3}\lambda L^{\frac{2}{3}} K^{-\frac{2}{3}} &\Rightarrow \lambda = \frac{3r}{L^{\frac{2}{3}} K^{-\frac{2}{3}}}
\end{aligned}$$

Equating the expressions for λ:

$$\begin{aligned}
\frac{3w}{2L^{-\frac{1}{3}} K^{\frac{1}{3}}} &= \frac{3r}{L^{\frac{2}{3}} K^{-\frac{2}{3}}} \\
2rL^{-\frac{1}{3}} K^{\frac{1}{3}} &= wL^{\frac{2}{3}} K^{-\frac{2}{3}} \\
2rL^{-1} K^{1} &= w \\
K &= \frac{wL}{2r}
\end{aligned}$$

Substituting this back into the constraint:

$$\begin{aligned}
L^{\frac{2}{3}} K^{\frac{1}{3}} &= q \\
L^{\frac{2}{3}} \left(\frac{wL}{2r}\right)^{\frac{1}{3}} &= q \\
L^{\frac{2}{3}} L^{\frac{1}{3}} \left(\frac{w}{2r}\right)^{\frac{1}{3}} &= q \\
L &= \left(\frac{w}{2r}\right)^{-\frac{1}{3}} q \\
L &= \left(\frac{2r}{w}\right)^{\frac{1}{3}} q
\end{aligned}$$

Substitute this back into our expression for K:

$$\begin{aligned} K &= \frac{wL}{2r} \\ &= \frac{w\left(\frac{2r}{w}\right)^{\frac{1}{3}} q}{2r} \\ &= \frac{w(2r)^{\frac{1}{3}} q}{w^{\frac{1}{3}}(2r)} \\ &= w^{\frac{2}{3}}(2r)^{-\frac{2}{3}} q \\ &= \left(\frac{w}{2r}\right)^{\frac{2}{3}} q \end{aligned}$$

We have now solved for the *input demand functions*. These give the combination of K and L that the firm should use to minimize cost while producing q units of output.

$$\begin{aligned} L &= \left(\frac{2r}{w}\right)^{\frac{1}{3}} q \\ K &= \left(\frac{w}{2r}\right)^{\frac{2}{3}} q \end{aligned}$$

One could plug in any w, r and q to these equations in order to find the firm's optimal input usage in producing q units of output. As q rises, both L and K rise. On the other hand, as the cost of capital r rises, L rises but K falls – intuitively, the firm substitutes labor in place of capital.

Now, the actual level of cost that the firm experiences is:

$$\begin{aligned} LRC &= wL + rK \\ &= w\left(\frac{2r}{w}\right)^{\frac{1}{3}} q + r\left(\frac{w}{2r}\right)^{\frac{2}{3}} q \\ &= qw^{\frac{2}{3}} r^{\frac{1}{3}} \left(2^{\frac{1}{3}} + 2^{-\frac{2}{3}}\right) \end{aligned}$$

This is a long run cost, by definition, since the firm is choosing L and K both. Notice that cost rises as q rises, w rises and r rises.

8 Long Run Average Cost

The long run average cost function is $\frac{LRC}{q}$. This is the production cost per unit *in the long run* after capital and labor have been adjusted to their optimal levels.

A cost function exhibits *economies of scale* if $LRAC$ falls as output rises. This is related to returns to scale in the production function. If the production function displays increasing returns to scale, then the cost function obviously displays economies of scale – if doubling all production inputs (which doubles costs) more than doubles output, then the cost *per unit* of output is going to fall. Actually, economies of scale holds *only* when the production function displays increasing returns to scale. Stated differently, increasing returns to scale in the production function is a necessary and sufficient condition for economies of scale in the cost function.

A cost function exhibits *diseconomies of scale* if $LRAC$ rises as output rises; this is equivalent to decreasing returns to scale in the production function. A cost function exhibits *no economies of scale* if $LRAC$ is constant as output rises; this is equivalent to constant returns to scale in the production function.

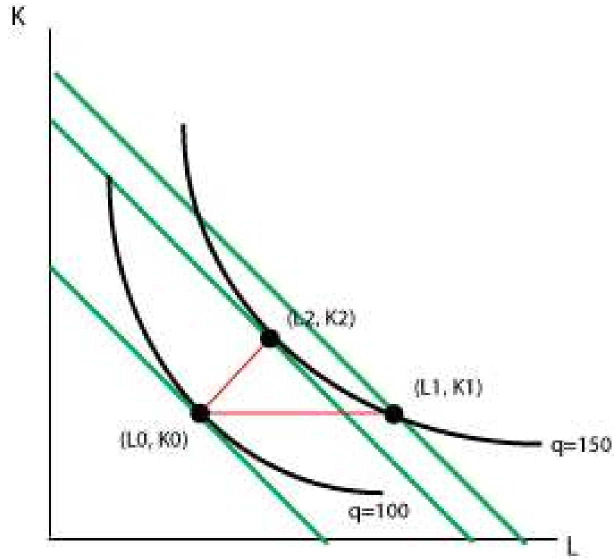

Figure 4: Long-run costs are lower than short run costs

Informally, larger firms are more efficient in industries with economies of scale since cost per unit of output falls as the firm expands. In industries with diseconomies of scale, smaller firms are more efficient. Firm size is irrelevant in industries with no economies of scale. Large and small firms are equally efficient, in the productive sense.

9 Long-Run Costs and Short-Run Costs

Long-run costs are lower than short-run costs. The firm can only change output in the short-run by changing labor input L. However, the firm can change output in the long-run by changing both K and L. If a firm wants to expand its output in the long run, it *could* do it by expanding only labor, so long-run costs certainly can't be any higher than short-run costs. However, if it is cheaper to expand output by expanding both labor and capital, then the firm's long-run costs will be lower than its short-run costs. The firm has more options in the long-run than in the short-run.

It is easy to see this graphically. Consider figure 4. Suppose that the firm initially produces $q = 100$ units of output using the input bundle (L_0, K_0). If the firm wants to expand to $q = 150$ units of output (on a higher isoquant) in the short-run, then its only choice is to increase usage of labor, employing input bundle (L_1, K_1). The *short-run expansion path* connects the two – in the short-run, the firm expands output by adding labor only.

However, this is not the cheapest way to produce $q = 150$ units of output. In the long-run, the firm will add both capital and labor and produce $q = 150$ units of output on a lower isocost line, using the input bundle (L_2, K_2). The *long-run expansion path* connects (L_0, K_0) with (L_2, K_2) – it reflects the long-run expansion of output by adding both capital and labor.

Summarizing, in the long-run, the firm can expand its output from $q = 100$ to $q = 150$ and achieve a lower isocost line in the long-run than it can in the short-run.

Another way to think about this is the following. In the short-run, the firm minimizes $wL + rK$ subject to $K = \bar{K}$ fixed. In the long-run, the firm minimizes $wL + rK$ by choosing both L and K. It could choose $K = \bar{K}$ if it wants to, but it may be able to achieve a lower cost with a different level of capital. Basically, the firm solves the same cost minimization problem in the short-run and in the long-run, but the firm has more constraints in the short-run since it cannot alter its capital.

A firm has many scales (levels of capital) at which it can operate. There are many ways to produce a particular level of output. In the short run, the firm is stuck using the way that corresponds to its current scale. However, each scale of production has its own short-run average total cost curve, and – over the long run – the firm will pick the scale that gives it the *lowest* cost per unit for whatever level of output it wants to produce. Mathematically, we say that the long-run average cost curve is the *lower envelope* of all the short-run average cost curves.