Chapter 7: The Cost of Production

Having described the firm's production process in the previous chapter, it is now time to turn this information into a description of the firm's cost structure. As you probably remember from Econ 201, just about everything the firm does is dictated by its cost structure, so a good understanding of production costs is critical to understanding the behavior of firms.

Types of Cost

We must first distinguish among several different "types" of cost that a firm might consider. First of all, remember that there economic cost is not the same as accounting cost. Accounting cost, calculated by accountants, normally considers actual expenses (cash flows), and is perfectly appropriate if that's what you're interested in. Economic cost, on the other hand, focuses on the opportunity cost of production. Recall that opportunity cost, generally, is defined as the value of the next best alternative. When applied to production, it describes what the firm is actually giving up to be able to produce output. Many of these opportunity costs would not be included in accounting cost.

There are some "costs" that are actually irrelevant for economic decision making because once made, they cannot be recovered. Such costs are called sunk costs. A typical example of a sunk cost would be specialized machinery that can be used only to produce the firm's particular product. Because this machinery has no value to anyone else, it cannot be resold, and whatever funds the firm invested in this machinery is completely and irrecoverably gone. Since the firm has no option to recoup its investment here, such as sunk cost is irrelevant to the firm's production decisions.

Fixed Costs (FC) refers to costs that do not vary with the level of output. Leases on equipment, rent for office space, insurance premiums etc. would all be examples of a fixed costs since the firm must pay these expenses no matter how much it produces. Note that fixed costs are not the same as sunk costs. If the firm decides to go out of business, it can sell its equipment, its office space, etc. and recover these costs.

Variable Costs (VC) refers to costs that do vary with the level of output. Wages and raw material costs are common examples of variable costs. For instance, a baker needs more flour if he is going to produce more bread. Thus, the cost of flour is a variable cost for the baker.

Normally, we will assume that costs related to labor (wages) are variable, while expenses related to capital are fixed. There could be some exceptions to this. For instance, the salaries paid to executives often don't depend on output, while the depreciation cost of some machinery does. We will gloss over these distinctions to keep the analysis simple, though.

Total Cost (TC) = FC + VC

Marginal Cost (MC) is the increase in cost resulting from the production of one more unit of output. MC = $\Delta TC/\Delta Q = \Delta VC/\Delta Q$

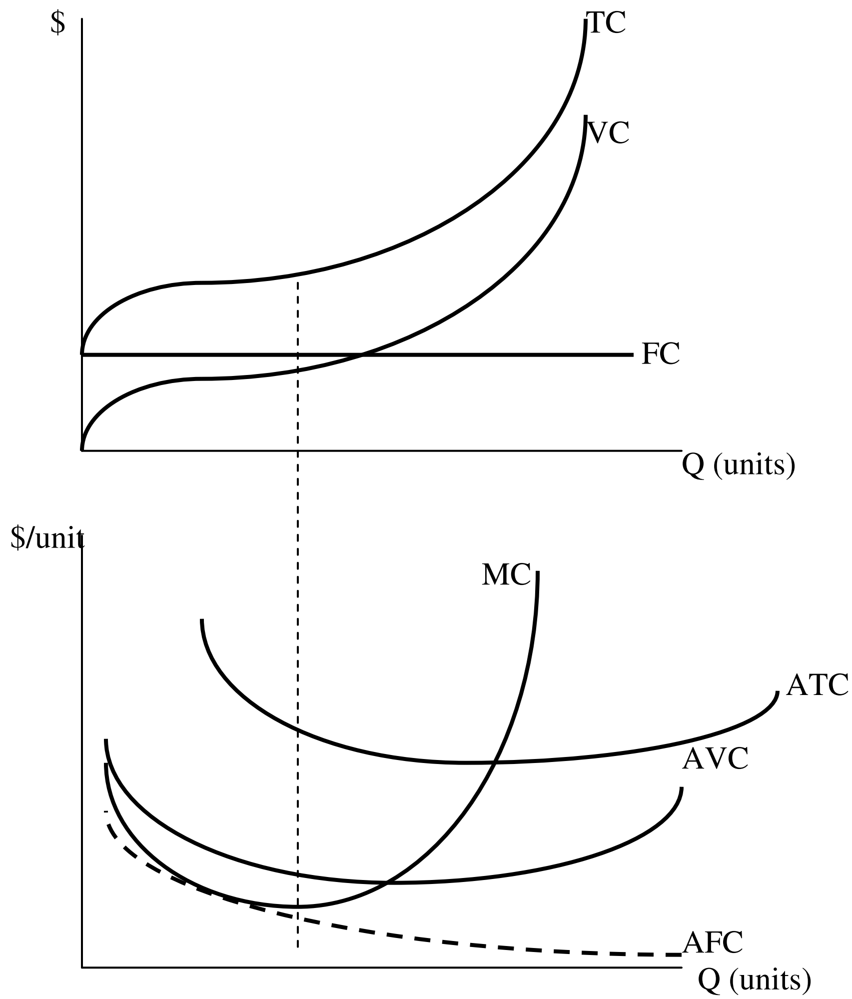
Average Total Cost (ATC) = TC/Q. Notice that if we define Average Variable Cost (AVC) as VC/Q and Average Fixed Cost (AFC) as FC/Q, then ATC = AVC + AFC.

Cost in the Short Run

The following table (taken from P&R) describes a typical cost structure of a firm with a $50 fixed cost.

Q	FC($)	VC($)	TC($)	MC($/unit)	AFC($/unit)	AVC($/unit)	ATC($/unit)
0	50	0	50	--	--	--	--
1	50	50	100	50	50	50	100
2	50	78	128	28	25	39	64
3	50	98	148	20	16.7	32.7	49.3
4	50	112	162	14	12.5	28	40.5
5	50	130	180	18	10	26	36
6	50	150	200	20	8.3	25	33.3
7	50	175	225	25	7.1	25	32.1
8	50	204	254	29	6.3	25.5	31.8
9	50	242	292	38	5.6	26.9	32.4
10	50	300	350	58	5	30	35
11	50	385	435	85	4.5	35	39.5

We can basically make up numbers for FC and VC, and then everything else on the table follows from our definitions. If we graph these costs, they will look something like this:



Some description of each of these curves is in order:

1) Note that FC is simply a horizontal line since it doesn't depend on Q.

2) VC rises at a slower and slower rate at first, presumably because extra units of labor are more and more productive at first (MP_L is initially increasing). Therefore, if the firm wants to produce more output, it doesn't need much extra labor, so VC rises at a *decreasing* rate at first. Eventually, though, diminishing marginal returns kick in. Then, the firm must keep adding labor at a faster and faster rate to produce extra output because MP_L is falling. Thus, VC will eventually start to rise at an *increasing* rate.

3) Since TC = VC + FC, the TC curve looks exactly like the VC curve, only it is shifted up by FC.

4) AFC = FC/Q is strictly downward-sloping and gets close to zero as Q gets large.

5) Just as MP was the slope of TP, MC is the slope of TC. Therefore, since TC is initially rising at a decreasing rate, MC is initially positive and decreasing. Eventually, TC starts to rise at an increasing rate, so MC "bottoms out" and begins to rise.

To make this more clear, suppose that the firm has exactly one variable input, namely labor, and that each unit of labor costs the wage w. In that case,

$$VC = wL \quad \Rightarrow \quad \Delta VC = w\Delta L$$
$$\Rightarrow \quad MC = \Delta VC/\Delta Q = (w\Delta L)/(\Delta Q)$$

but $\Delta Q/\Delta L$ is just MP_L, so that $\Delta L/\Delta Q$ is $1/MP_L$ so

$$\Rightarrow \quad MC = w/MP_L$$

Since we know what happens to MP_L as we increase output, we know what happens to MC. At low levels of output, $\uparrow Q \Rightarrow \uparrow MP_L \Rightarrow \downarrow MC$, but at higher levels of output, $\uparrow Q \Rightarrow \downarrow MP_L \Rightarrow \uparrow MC$. Thus, MC will be "J-shaped" as shown on the graph.

6) Just as AP was the slope of a ray from the origin to TP, AVC (and ATC) is the slope of a ray from the origin to VC (or TC). Notice that the slope of such a ray falls initially (since we are adding proportionally less labor to produce more output, bringing down AVC) but then rises, since with diminishing returns we will eventually be adding proportionally more labor to produce more output. Thus, AVC is "U-shaped."

7) ATC is also "U-shaped" like AVC, except that it is shifted up by AFC.

8) Notice also the relationship between MC and AVC (and ATC). When MC < AVC, this means that each extra unit of output costs "less than average," so that producing one more unit of output brings our AVC down. Thus, if MC < AVC, AVC must be falling (the same relationship holds for ATC)

Furthermore, when MC > AVC, this means that one more unit of output adds more than average to VC, so that if MC > AVC, AVC must be rising (the same relationship holds for ATC).

What this means is that AVC falls as long as MC < AVC, and then AVC begins to rise. Thus, MC intersects AVC at min(AVC). Likewise, MC intersects ATC at min(ATC). Such a phenomenon occurs whenever we have U-shaped ATC and AVC curves and is shown on the graph.

These graphs more or less completely describe the firm's cost structure in the short run, when at least one input is held constant. It would be worthwhile to spend some time studying the diagrams above and making sure you understand why the curves look the way they do. We will see some situations later in which the firms MC, AVC and ATC curves do not look like those drawn above, and it will be much easier to follow this material if you understand the logic behind these shapes as opposed to simply memorizing them.

Cost in the Long Run

In the long run, the firm is able to change its level of capital (assuming that capital was our fixed short run input). We knew that the cost of labor was simply the wage, w. How can we measure the cost of capital?

We normally use a measure called the user cost of capital, which is the annual cost of owning and using a capital asset. It includes two components:

1) The depreciation rate. As you use machinery, it wears down. The cost of this wear and tear should be included when calculating the cost of owning and using capital.

2) Forgone interest. If the firm borrows to buy capital (as firms often do), then it must pay interest on the loan. This interest obviously represents part of the cost of owning capital. However, even if the firm does not borrow, and pays for its capital with cash, it still faces the exact same opportunity cost of forgone interest.

We will use r to denote the per-dollar user cost of capital. It is equal to:

$$r = \text{depreciation rate} + \text{interest rate}$$

Isoquants and Isocost Lines

In the last chapter, we showed how we can use isoquants to describe the firm's production technology. Now we want to combine this understanding with the firm's long run cost structure to see how the firm can minimize the cost of producing a given level of output.

Notice, incidentally, that cost minimization is the "flip side" of profit maximization. If a firm is producing $Q = Q_0$ and is NOT producing that quantity at the minimum cost, then it cannot possibly be maximizing profits, since it could lower its cost, receive the same revenue, and increase profits simply by altering its production process. Thus, any profit-maximizing firm must also be a cost-minimizing firm.

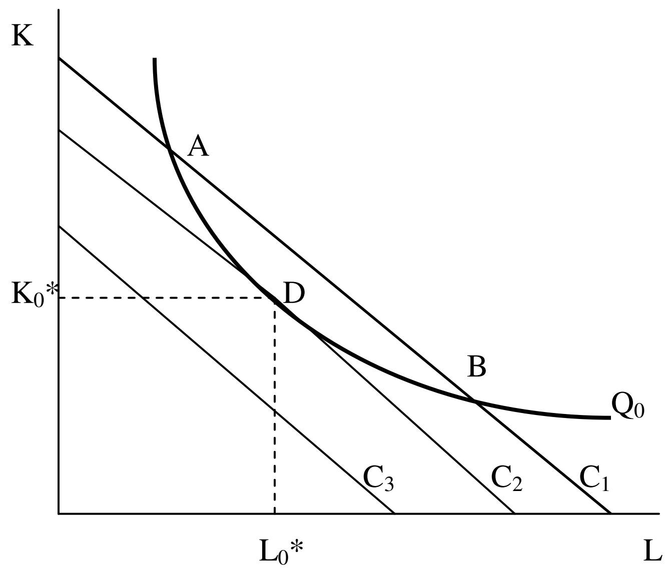
An <u>isocost line</u> shows all possible combinations of capital and labor that can be purchased for a given total cost. (Note that the prefix "iso-" means "same." Thus an isoquant shows input combinations that produce the *same* quantity, and an isocost line shows input combinations that have the *same* cost. We could have called an indifference curve an iso-utility curve by the same logic). The isocost line is described by the following statement:

$$C = wL + rK$$

We can rearrange this and write in a more familiar linear form

$$K = (C/r) - (w/r)L$$

This shows that the isocost line is a straight line having slope $\Delta K/\Delta L = -(w/r)$. We use isocost lines to find the cost-minimizing way to produce a given level of output.



Suppose that this firm wants to produce Q_0. (For now we don't know why this firm would particularly have its heart set on $Q = Q_0$. Later chapters will deal with the issue of the profit-maximizing output level. For now, we will simply assume that Q_0 is best for the firm, and later we'll see how this is determined). The firm could produce Q_0 at points A or B, since those points lie on the isoquant $Q = Q_0$. However, it is clear that the firm can find a point on the same isoquant with a lower total cost (on a lower isocost line). The firm can continue to decrease cost by moving to lower and lower isocost lines, until it gets to point D. At this point, the isocost line corresponding to D is just tangent to the isoquant. If the firm reduces its costs any more (to isocost C_3 for example), there isn't any point on that isocost line that would allow it to product Q_0. Thus, C_2 is the minimum total cost that allows the firm to produce Q_0.

Note that this minimum occurs at a point of tangency (except for corner solutions, which we will ignore). Recall that the slope of the isoquant is the MRTS = $(MP_L)/(MP_K)$, so setting this equal to the slope of the isocost line tells us that costs are minimized when

$$(MP_L)/(MP_K) = w/r$$

or, rewriting this slightly,

$$(MP_L)/w = (MP_K)/r$$

This condition should make some intuitive sense. It says that the additional output gained for an extra dollar's worth of labor must equal the additional output gained from an extra dollar's worth of capital. Suppose that this were not the case. Specifically, let's say that $(MP_L)/w > (MP_K)/r$. This would mean that if the firm spent one more dollar on labor and one less dollar on capital (thus holding costs constant), the additional output gained from the extra labor would outweigh the output lost from having less capital. Thus, costs stay the same and output goes up. This would be inconsistent with the firm having already minimized its costs, because the firm can hold cost constant and produce more, or produce the same amount and reduce cost. (Note that it is also inconsistent with profit maximization, since the firm can hold cost constant and produce more, generating more revenue). Thus, if $(MP_L)/w > (MP_K)/r$ the firm cannot possibly be minimizing cost or maximizing profit, and a similar story holds for $(MP_L)/w < (MP_K)/r$. Only when these to quantities are equal are costs minimized.

e.g. Suppose that at Q_0, $MP_L = 25$, $MP_K = 30$, $w = 5$ and $r = 5$. Then

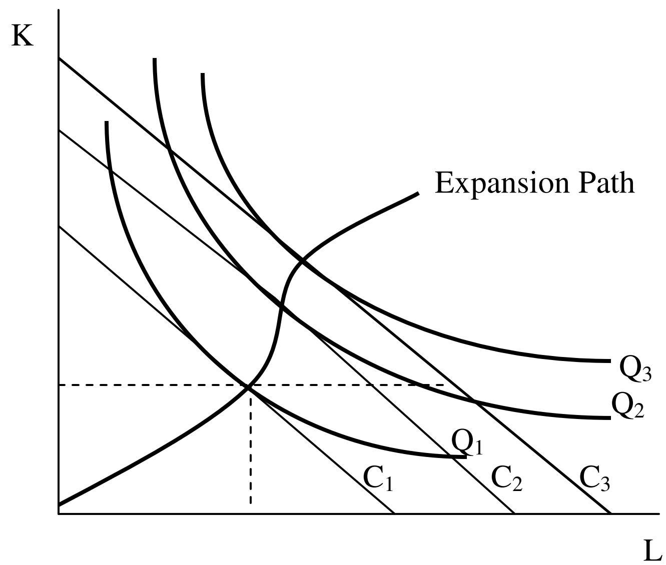
$$(MP_L)/w = 5 < 6 = (MP_K)/r$$

If the firm spends one dollar less on labor, it loses 5 units of output. But if it puts that dollar into capital, it gains 6 units of output. Thus, its cost stays the same but output rises by one unit. Obviously, the firm was using too much labor and too little capital initially.

Also notice in this situation that as the firm adds more capital and uses less labor, MP_L goes up, increasing the LHS quantity, and MP_K goes down, decreasing the RHS quantity. Thus, as the firm rearranges its inputs in the suggested way, we are driven toward equality of $(MP_L)/w$ and $(MP_K)/r$.

Long Run Cost Curves

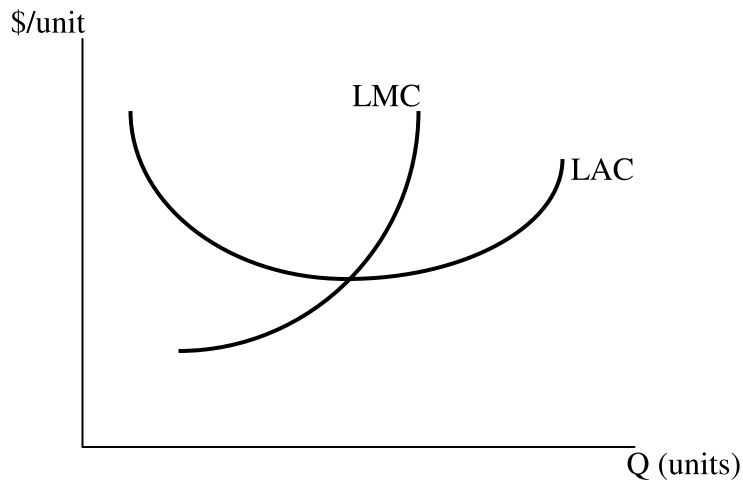
We now know the cost minimizing way to produce Q_0. Notice that the cost of producing Q_0 is C_2 in the above diagram, so we are able to associate a long-run cost to that output level. If we repeat this procedure for all possible output levels, we will have found the firm's long run cost structure.



This diagram shows the cost-minimizing mix of inputs for three different output levels. Notice how the firm is able to reduce its cost in the long run as compared to the short run. For instance, if the firm was initially producing Q_1 at minimum cost and then, in the short run, decided to start producing Q_2, it would have to hold its capital level constant. The resulting input mix places the firm on a substantially higher isocost line than the cost-minimizing mix for Q_2. (Graph not drawn well to show this – be careful to note that K is held constant in the short run, putting the firm at a non-cost-minimizing point on Q_2 if the firm steps up production).

The line connecting the various cost-minimizing input bundles is called the firm's <u>long run expansion path</u> and shows how the firm changes its input usage as it increases output. As we move along the expansion path, we see the firm's output (from the associated isoquant) and its long run total cost (from the associated isocost line). Thus, we are able to plot this information to get the firm's long-run cost curves.

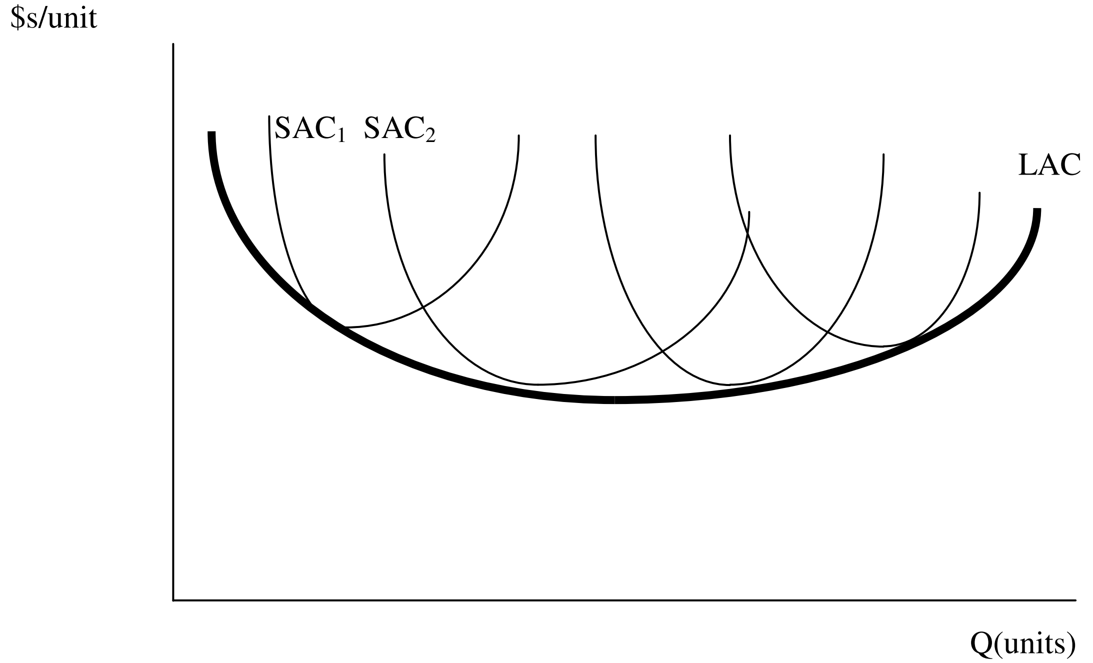
The typical firm is normally assumed to have a U-shaped long run average cost (LAC) curve, as follows:



We assume that LAC is U-shaped because we often assume increasing returns to scale for small levels of production and decreasing returns to scale for larger levels of production. If we have increasing returns to scale, remember, we can scale up all our inputs (thus scaling up our costs) by a given factor, but output goes up more than that factor (i.e. we double inputs and cost, but output more than doubles). Thus, LAC would fall. As we move into a region of decreasing returns, our costs rise by a larger percentage than output as we scale up our input usage, so LAC rises.

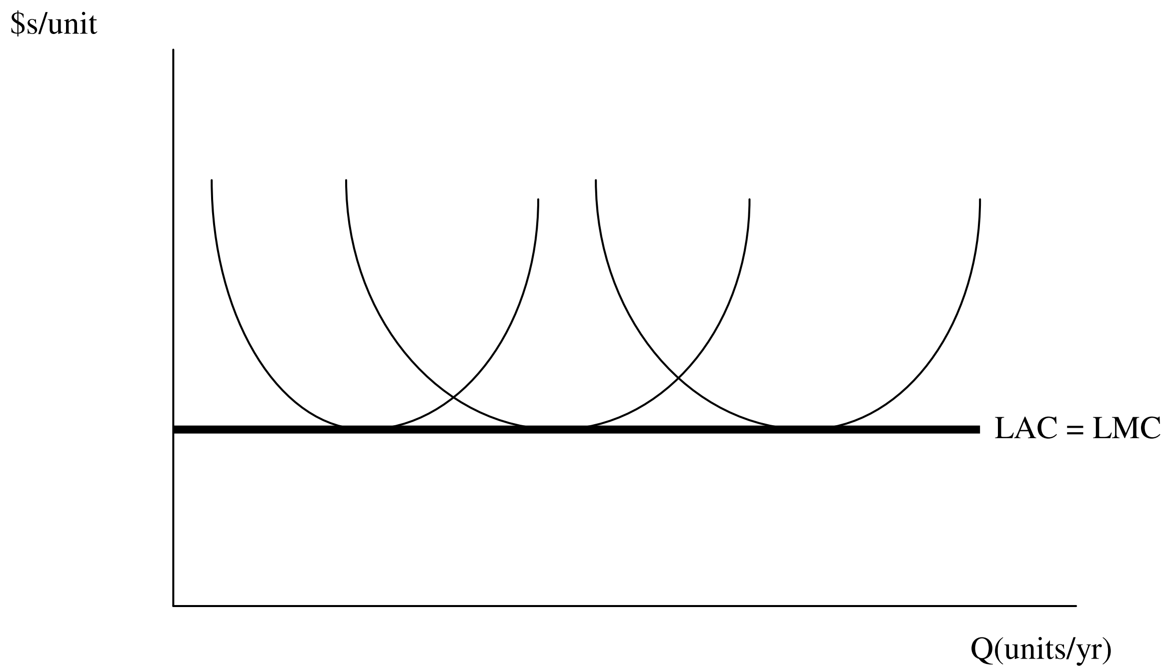
We get our long run marginal cost (LMC) curve from our LAC curve. As long as LAC is falling, LMC must lie below it. When LAC rises, LMC must be above it. Thus, LMC is upward-sloping and intersects LAC at its minimum.

Please notice that these are long-run cost curves. They should represent the lower envelope of the firm's short-run cost curves given various levels of capital.



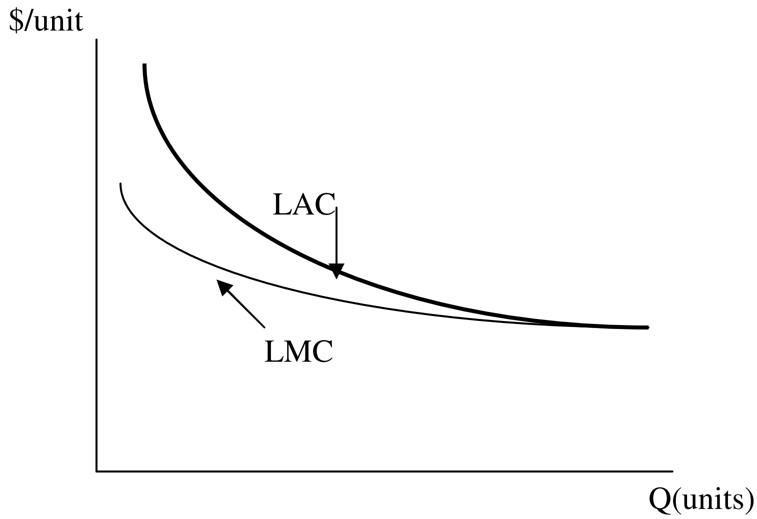
Here is a U-shaped LAC curve and some of the possible short run average cost (SAC) curves that go along with it. Along SAC_1, the level of capital is fixed at K_1, along SAC_2, the level of capital is fixed at K_2, and so on. Notice that the firm has U-shaped SAC curves because of diminishing marginal returns, but that it need not have a U-shaped LAC if it has strictly increasing, constant, or decreasing returns to scale.

For instance, this diagram shows a firm with constant returns to scale.



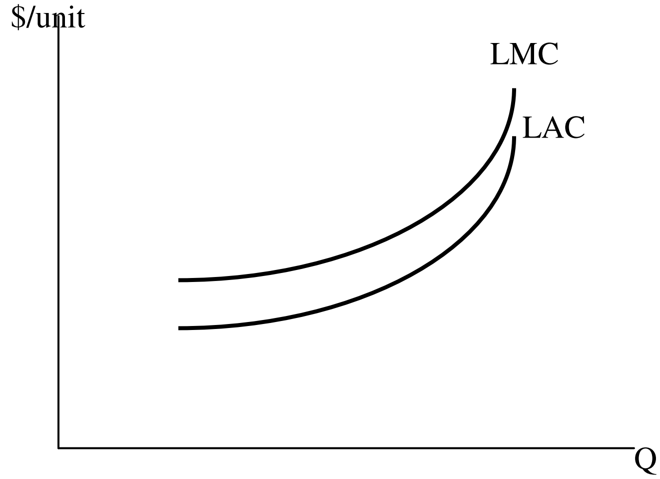
Note that LAC = LMC if the firm has constant returns to scale. CRS means that if the firm increases all inputs (and thus cost) by λ, output increases by λ, so that AC is constant. If LMC either > or < LAC, then LAC would be either rising or falling depending on the inequality, and this is inconsistent with CRS. Furthermore, notice that the firm has U-shaped SAC curves (implying diminishing marginal returns in the short run) but has constant returns to scale in the long run. There is nothing inconsistent about this.

Likewise, a firm could have increasing returns to scale in the long run as shown here:



Here, AC is falling as we produce more output. This occurs because as we increase our inputs (and costs) by λ, output rises more than λ, so AC falls. Note that LMC must therefore lie everywhere below LAC. Here it is drawn downward-sloping, but it could be a straight line to which LAC asymptotes.

Finally, here is a diagram of a firm with decreasing returns to scale (or diseconomies of scale):



LAC rises since as we scale up output by λ, we must scale up inputs (and cost) by more than λ, so that AC rises. LMC must therefore lie everywhere above LAC.