These notes essentially correspond to chapter 7 of the text.

1 Costs

When discussing firms our ultimate goal is to determine how much profit the firm makes. In the chapter 6 notes we discussed production functions, which are the physical methods by which firms produce goods. In this section we will discuss costs. In the next few sections we will discuss pricing policies in order to determine profits.

1.1 Economic vs. Accounting Costs

Although we will rarely distinguish between the two measures of costs (mainly because we will always assume that we are calculating economic cost), there is an important difference between economic and accounting costs. Accounting costs are called explicit costs, which are payments to factors of production. Wages, rent, taxes, advertising costs, costs for materials – these all appear on an accounting statement as costs. With economic costs we add implicit costs to the explicit costs. An implicit cost is an opportunity cost of a resource owned. For instance, suppose a firm brings in total revenue of $60,000 for the year. The explicit costs are $40,000, so the owner brings in an accounting profit of $20,000. However, suppose the owner does not pay himself throughout the year and only keeps the profit at the end of the year. If the owner's opportunity cost of his time is $25,000 (the amount he could earn at another job if he did not run his business), then the implicit cost of his time is also $25,000. We need to subtract this implicit cost from the accounting profit to find that the owner now makes an economic profit of $(-\$5,000)$.

2 Short-run Costs

There are 7 different costs that we need to know about in the short-run. They are broken down into groups below.

2.1 Total costs

Short-run costs are the costs of production when one input is fixed. There are two basic types of short-run costs, fixed costs and variable costs. Fixed costs are costs that do NOT vary with the output level, while variable costs are costs that do vary with the output level. We will say that Total Fixed Cost (TFC) is the total amount of the fixed costs, while Total Variable Cost (TVC) is the total amount of the variable costs. To find the firm's Total Cost (TC) for producing an specific amount of output we need to add the TFC and TVC. Thus,

$$TC = TFC + TVC$$

2.2 Average costs

There are three average cost measures you need to be familiar with: Average Total Cost (ATC), Average Variable Cost (AVC), and Average Fixed Cost (AFC). An average cost is simply a measure of cost per unit produced. To find Average Total Cost, we use:

$$ATC = \frac{TC}{q}$$

To find Average Variable Cost, we use:

$$AVC = \frac{TVC}{q}$$

To find Average Fixed Cost, we use:

$$AFC = \frac{TFC}{q}$$

Another useful relationship is:

$$ATC = AFC + AVC$$

To get that equation simply divide both sides of $TC = TFC + TVC$ by q.

2.3 Marginal Cost

Marginal cost (MC) is the cost of producing one additional unit of output. We can find marginal cost mathematically by:

$$MC = \frac{\Delta TC}{\Delta q}$$

In addition, we can also find MC using the following relationship:

$$MC = \frac{\Delta TVC}{\Delta q}$$

Although it may look odd that marginal cost can be found from either equation, remember that marginal cost is how much total cost changes when we produce one additional unit. Since the TFC does not change when we produce additional units of the good, we know that it will have no impact on marginal cost – therefore, we can ignore it when calculating MC.

2.3.1 Graphs

The graphs of the various total cost functions look as follows:
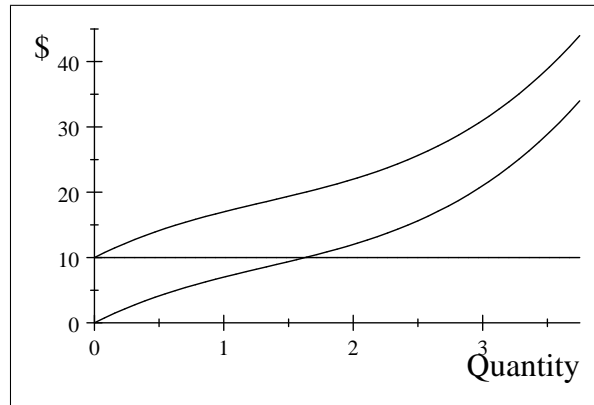
TFC is the flat line at $10 – since the cost does not change when additional output is produced it is always constant. TVC is the curved line that starts at the origin – it starts at the origin because if you produce 0 you do not need to pay for any variable resources. TC is the curved line that starts from 10. It is simply the addition of the TFC and the TVC at every output level. Thus, since the TFC is $10 in this example, the TC at any output level is exactly $10 more than the TVC at that same output level. The TFC is fairly intuitive, but the TC and TVC are not. We will discuss why they look the way they do in a moment. The actual equations that I used to graph these total costs functions are:
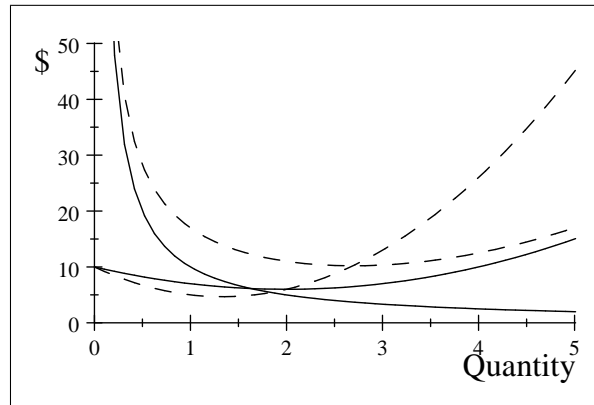
$$TC = 10 + 10q - 4q^2 + q^3$$

$$TVC = 10q - 4q^2 + q^3$$

$$TFC = 10$$

Notice that TFC is just the part of TC that does not depend on q and that TVC is the remaining part of TC that does depend on q.

The graph of the various average cost functions and the marginal cost function look like the one below.

The curves are as follows:

The AFC is the solid line that comes out of the top of the graph (up around $50). Notice that it is always decreasing. The AFC function that is plotted above is:

$$AFC = \frac{10}{q}$$

The ATC is the dotted line that comes out of the top of the graph (up around $50). Notice that it is U-shaped. The ATC function that is plotted above is:

$$ATC = \frac{10}{q} + 10 - 4q + q^2$$

The AVC is the solid line that comes out of the middle of the graph (at $10). Notice that it is U-shaped. The AVC function that is plotted above is:

$$AVC = 10 - 4q + q^2$$

The MC is the dotted line that comes out of the middle of the graph (at $10). Notice that it is U-shaped. The MC function that is plotted above is:

$$MC = 10 - 8q + 3q^2$$

A few key points about the graphs:

1. MC crosses ATC and AVC at their respective minimums.

2. If MC is less than ATC then ATC is decreasing; if MC is greater than ATC then ATC is increasing.

3. If MC is less than AVC then AVC is decreasing; if MC is greater than AVC then AVC is increasing.

4

4. The difference between the ATC and the AVC gets smaller as we increase the quantity produced. Recall that $ATC = AFC + AVC$, so the difference between the ATC and the AVC is just the AFC, or, in equation form, $ATC - AVC = AFC$. Since the AFC is decreasing, the difference between the ATC and the AVC becomes smaller, meaning the two curves get closer together.

2.3.2 Why the graphs have their shapes in the SR

The AFC graph is intuitive – as the firm produces more, the per-unit fixed cost must decline since the total fixed cost always remains constant. The reason the MC is U-shaped is because of the law of diminishing marginal returns. Eventually, when the firm employs an extra worker the output will begin increasing at a decreasing rate. When this occurs, it will cost the firm more to produce an additional unit of output, thus the MC curve must begin to increase. As for the ATC, it is initially high due to a high AFC, and then it begins to decrease as we produce more. It begins increasing due to a high MC of production at high levels of output. The AVC is U-shaped because it is the difference between the ATC and the AFC.

3 LR Costs

In the long-run the firm has the ability to adjust (or vary) all of its inputs. Because of this we are only concerned with 3 costs in the long-run (as opposed to 7 costs in the SR). The costs we are concerned about are: TC, ATC, and MC. We are not concerned with TFC and AFC because there are no fixed costs in the long-run – technically they are both 0. We are not concerned with TVC and AVC because the variable cost in the long-run is exactly the same thing as the total cost since all inputs can be varied. Thus, we will only be concerned with TC, ATC, and MC. The firm's ultimate goal when making LR decisions is to pick the input amounts that minimize the cost of producing a specific level of output. We will call this the cost minimization process.

3.1 Isocosts

We will assume that the firm only uses two inputs in its production process, capital (K) and labor (L). We will let r be the rental rate of capital and w be the wage rate of labor. Thus, the firm's total cost will be:

$$TC = rK + wL$$

This should look very familiar to you – it should look very similar to setting up a budget constraint for a consumer. Now, suppose that we fix the level of total cost at \$100, and we let $r = \$10$ and $w = \$5$. We know that we could spend the entire \$100 on capital and use 10 units of capital, or we could spend the entire \$100 on labor and use 20 units of labor. We could also spend the

$100 on combinations of capital and labor, such as 18 units of labor and 1 unit of capital, or 6 units of labor and 7 units of capital, etc. What we want to find is a function that shows us the trade-off between purchasing capital and labor. If we solve for K in the total cost function above we get:

$$K = \frac{TC}{r} - \frac{w}{r}L$$

Notice that if we know TC, w, and r then we have the equation of a line, just like we did when we created the consumer's budget constraint in chapter 4. Assuming that $TC = \$100$, $r = \$10$ and $w = \$5$, we have:

$$K = 10 - \frac{1}{2}L$$

If we plot this on a capital-labor graph we get:
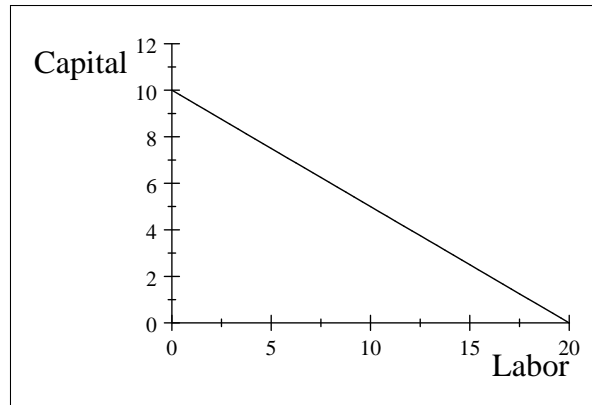
This graph shows the isocost line for the chosen parameters ($TC = \$100$, $r = \$10$ and $w = \$5$). Every combination of inputs along the line has the same total cost of $100.

3.1.1 Slope of the isocost

Note that the slope of the isocost line is $\left(-\frac{w}{r}\right)$. As we have seen before, slopes of isoquants, budget constraints, indifference curves, and now isocosts will be important when finding an interior solution.

3.2 Cost minimization

Suppose the firm wants to produce a specific quantity. If the firm wishes to maximize the profit of producing that specific quantity, then it must minimize the costs of producing that specific quantity. How does the firm do this?

Recall the concept of isoquant from chapter 6. Every combination of inputs along a given isoquant produces the same amount of total output. Thus, we

will fix a specific isoquant (in this case the quantity level is 60 for the production function $q = 10K^{1/2}L^{1/2}$), as in the picture below:[1]
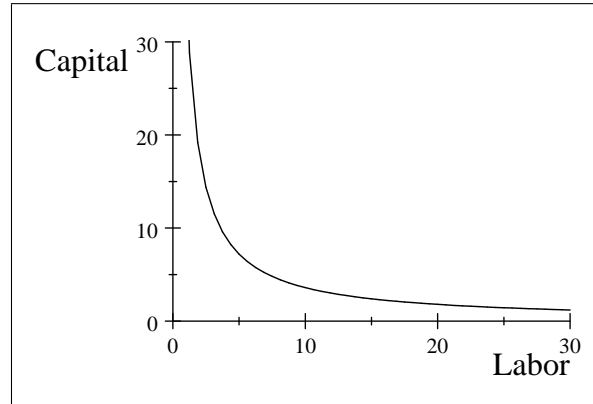
Now, our goal will be to determine the cost-minimizing bundle of inputs that can produce 60 units of output. Suppose that $r = \$10$ and $w = \$5$. We will let TC be 3 different levels: $\$50$, $\$84.85$, and $\$100$. Note that the lowest isocost line will be the one for $\$50$ and the highest isocost line will be the one for $\$100$.
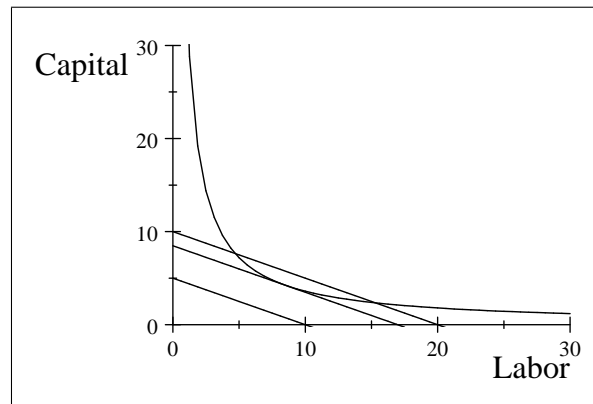
Notice that all the isoquants are parallel. This is because they all have the same slope, $-\frac{w}{r}$ or $\left(-\frac{1}{2}\right)$ in this example. The isocost for $\$50$ does not touch the firm's isoquant for 60 units – this means that the firm must spend more money in order to produce 60 units (much like a point outside a consumer's budget constraint cannot be purchased by that consumer, a point outside the isocost cannot be produced by the firm). The isocost for $\$100$ cuts the isoquant for 60 units twice. If the isocost cuts through the isoquant (and we are not at

---

[1] The actual equation used for graphing this isoquant comes from the production function $q = 10K^{1/2}L^{1/2}$ when q is set equal to 60. If you want to graph the isoquant, substitute 60 in for q and solve for K. This gives you: $K = \left(\frac{60}{10L^{1/2}}\right)^2$. If you graph that function you get the isoquant in the picture.

a corner), then there must be some lower cost combination of inputs that the firm could use to produce 60 units. So we shift the isocost down until it is just tangent (touches one time) to the isoquant. This is the isoquant for \$84.85. It is tangent to the isoquant at the input combination of $\frac{6}{\sqrt{2}} = 4.242641$ units of capital and $6\sqrt{2} = 8.485281$ units of labor. This is the lowest cost combination of inputs the firm could use to produce 60 units of the good.[2]

3.2.1 Cost-minimization rule (interior solution)

If we minimize costs at an interior solution then we know that the slope of the isocost must be equal to the slope of the isoquant at the point of tangency. We know that the slope of the isocost is $\left(-\frac{w}{r}\right)$. We know that the slope of the isoquant is the $MRTS$. We also know that:

$$MRTS = -\frac{MP_L}{MP_K}$$

Now setting the slope of the isocost equal to the slope of the isoquant we get:

$$-\frac{w}{r} = -\frac{MP_L}{MP_K}$$

Or:

$$\frac{MP_K}{r} = \frac{MP_L}{w}$$

Thus, if the firm is minimizing costs at an interior solution, the marginal product of capital per last dollar spent on capital must be equal to the marginal product of labor per last dollar spent on labor. If it were not (suppose that $\frac{MP_K}{r} > \frac{MP_L}{w}$), the firm could take some of the money it is spending on labor and spend that money on capital and produce a higher level of output. Or, the firm could take some of the money spent on labor, buy additional capital with some of the funds and save the rest of the funds to keep output at the same level.

3.2.2 Change in the relative prices of inputs

Suppose that the price of capital in the example increases to \$15 per unit, while the wage stays at \$5 per worker. The old isocost line for \$84.85 (the one that was tangent to the isoquant for 60 units) is no longer relevant. The new isocost line for \$84.85 is now the flatter of the two lines on the graph below.

---

[2] I'll explain how to find this input combination for those interested in a special subsection a few paragraphs down. You do NOT need to know how to do this for the exam.
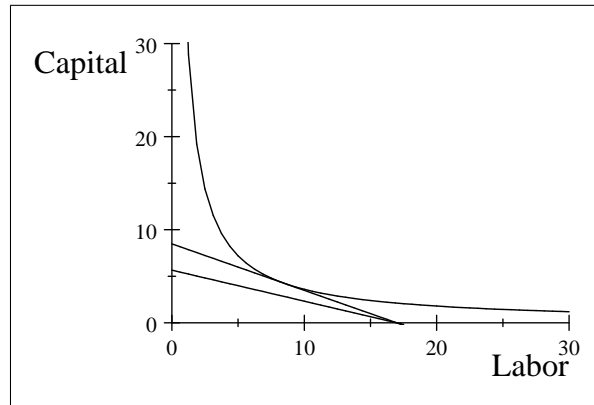
Notice that the firm cannot produce 60 units by spending $84.85 now. What will the firm do? It will find the new isocost line that is tangent to the isoquant of 60 units. The new isocost (based on $r = \$15$ and $w = \$5$) that is tangent to the isoquant for 60 units is found when $TC = \$103.92$. Plotting this new isocost line we see:
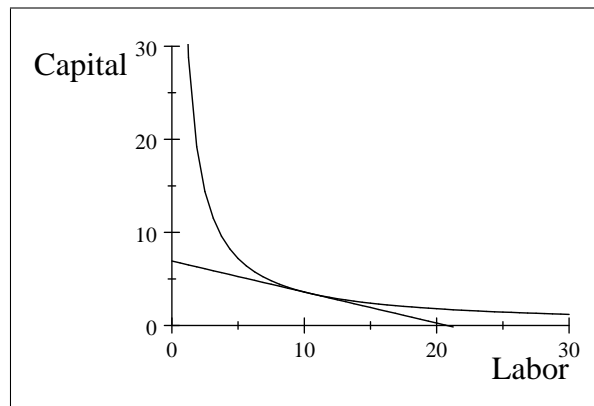
The input combination that minimizes the cost for producing 60 units of the good is $6\sqrt{3} = 10.3923$ units of labor and $\frac{6}{\sqrt{3}} = 3.464102$ units of capital. I will now go into a short digression on how to find the minimum total cost and the corresponding input combination for those interested. Since it involves calculus you will not be tested on this.

3.2.3 ***Short (well, not that short) digression

We know that at an interior solution the isocost line must be tangent to the isoquant, which means the slopes of the two lines must be equal. The slope of the isocost line is easy, as it is just $\left(-\frac{w}{r}\right)$. The slope of the isoquant is a little more difficult. What you need to do is to take the production function and solve for K. Thus, instead of having the production function as $q(K, L)$ you will have the isoquant as $K(q, L)$, although we will not consider q as a variable.

Then, the slope of the isoquant is given by the derivative of capital with respect to labor, or $\frac{dK}{dL}$. In our example, the production function is a Cobb-Douglas, of the form:

$$q = AK^{\alpha}L^{\beta}$$

The parameters are set as $A = 10$, $\alpha = \frac{1}{2}$, and $\beta = \frac{1}{2}$. If we solve for K we get the isoquant, which is:

$$K = \left(\frac{q}{AL^{\beta}}\right)^{1/\alpha}$$

Now, we need to find $\frac{dK}{dL}$. This is a bit messy, but:

$$\frac{dK}{dL} = \left(\frac{q}{A}\right)^{1/\alpha}\left(-\frac{\beta}{\alpha}\right)L^{(-(\beta/\alpha)-1)}$$

Like I said, a little bit messy. Plugging in the parameter values and the fact that $q = 60$ (since this is the quantity the firm wants to produce), we get:

$$\frac{dK}{dL} = \left(\frac{60}{10}\right)^{1/(1/2)}\left(-\frac{1/2}{1/2}\right)L^{(-((1/2)/(1/2))-1)}$$

Simplifying:

$$\frac{dK}{dL} = (6)^2\left(-1\right)L^{(-2)}$$

Simplifying again:

$$\frac{dK}{dL} = -\frac{36}{L^2}$$

Now, we know that the slope of the isocost (which is $-\frac{w}{r}$) equals the slope of the isoquant (which is $\frac{dK}{dL}$) at the cost-minimizing point. Letting $w = \$5$ and $r = \$10$ (this is the first example, with $TC = \$84.85$ at the cost-minimizing input combination), we get:

$$-\frac{1}{2} = -\frac{36}{L^2}$$

Or, solving for L:

$$L = \sqrt{72} = 6\sqrt{2}$$

We have now found one piece of the puzzle, which is the amount of labor used at the cost-minimizing point. A second piece of the puzzle, the amount of capital used at the cost minimizing point, can be found from the production function, which is of the form $q = AK^{\alpha}L^{\beta}$. If we know the parameters, the quantity we want to produce, and the amount of labor we will use to produce that quantity then we must be able to figure out the amount of capital. Solving for K gives us the isoquant, which is:

$$K = \left(\frac{q}{AL^\beta}\right)^{1/\alpha}$$

Now, plugging in all the numbers gives us:

$$K = \left(\frac{60}{10\left(6\sqrt{2}\right)^{1/2}}\right)^{1/(1/2)}$$

Simplifying:

$$K = \left(\frac{6}{\left(6\sqrt{2}\right)^{1/2}}\right)^2$$

Simplifying again:

$$K = \frac{36}{\left(6\sqrt{2}\right)}$$

Finally:

$$K = \frac{6}{\sqrt{2}}$$

The final piece of the puzzle is to calculate the total cost of our cost min-imizing bundle of inputs. Since $TC = rK + wL$, we know that $TC = 10 * \frac{6}{\sqrt{2}} + 5 * 6\sqrt{2} = 84.852\,81$.

Wasn't that fun? Hopefully this digression takes away some of the mystery as to where the numbers come from.

3.3   Graphing the LR curves

To graph the LR curves we first need to discuss the concept of the LR expansion path. The LR expansion path is found by holding the wage and rental rate of capital constant and then connecting the set of cost-minimizing bundles for various output levels. The picture below shows one such possibility for the expansion path.
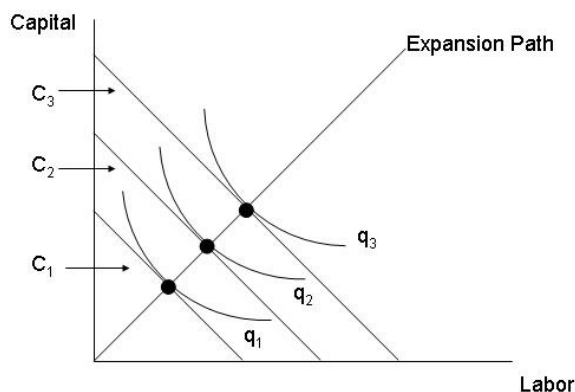
The isocosts are labelled as C_1, C_2, and C_3, with C_1 being the lowest isocost and C_3 being the highest isocost. The isoquants are labelled as q_1, q_2, and q_3, with q_1 the lowest and q_3 the highest. We can translate this picture into a picture of the LR total cost in the following manner. We know that if the firm produces q_1 and minimizes costs given the current wage and rental rate of capital it will cost the firm C_1. We also know that it will cost the firm C_2 if it wishes to produce q_2 and C_3 if the firm wishes to produce q_3. Thus, we can begin to build our LR total cost curve by plotting these points. If we found the cost minimizing amount for every output level then we would connect the dots and the resulting graph would be the firm's LR total cost. For the very special case of a LR expansion path that is a straight line that cuts through the origin, we also get that the LR total cost curve is a straight line that cuts through the origin. The graph is below:
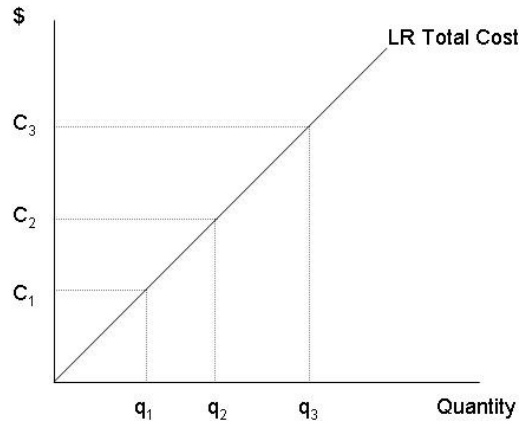
A perfectly linear total cost function is of the form $LRTC = mq + b$, where m is the slope of the cost function and b is the y-intercept. However, since the cost minimizing level of producing zero in the LR is \$0 (why employ inputs if you do not wish to produce anything?), b will always equal 0. So the linear $LRTC$ function will always be of the form $LRTC = mq$. Now, suppose that $LRTC = mq$. What is the $LRATC$ function for this firm? Since $LRATC = \frac{LRTC}{q}$, and $LRTC = mq$, this means that:

$$LRATC = \frac{mq}{q} = m$$

So the per-unit cost of production is a constant (specifically, the slope of the $LRTC$) IF the $LRTC$ is a perfectly straight line. What is the $LRMC$ if $LRTC = mq$? There are a few ways to think about this. First, we could use calculus. Since marginal cost is the derivative of the total cost function with respect to quantity, we get that $LRMC = m$. This means that the $LRMC$ is also a constant, equal to the slope of the $LRTC$. Another method of finding the $LRMC$ would be to use the change in formula. We know that $LRMC = \frac{\Delta LRTC}{\Delta q}$. Suppose we pick two quantity levels, q_1 and q_2, with $q_2 > q_1$. What is the marginal cost between these two quantity levels?
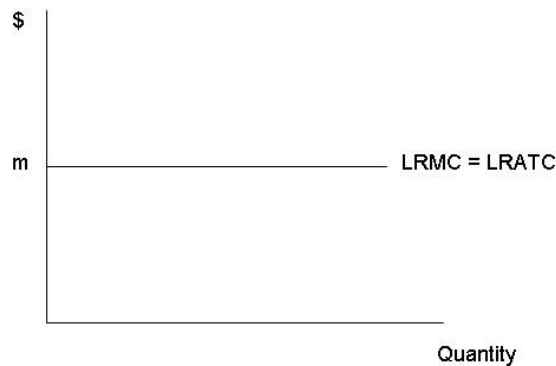
$$\Delta LRTC = mq_2 - mq_1 = m(q_2 - q_1)$$

$$\Delta q = (q_2 - q_1)$$

So:

$$LRMC = \frac{\Delta LRTC}{\Delta q} = \frac{m(q_2 - q_1)}{(q_2 - q_1)} = m$$

Same answer, different method. I will graph the $LRATC$ and $LRMC$ "curves" and then discuss a third, intuitive method of finding the $LRMC$ when $LRATC$ is a constant..


From our discussion of the relationship between marginal costs and average total costs we know that marginal cost intersects average total cost at the minimum of average total cost. However, in the picture above the $LRATC$ is always at its minimum level since it never changes. Thus the $LRMC$ must "intersect" the $LRATC$ at every single point since the $LRATC$ is always at its minimum.

3.3.1 Typical shape of the LR cost curves

The example above is a special case when the $LRTC$ is a perfectly straight line. However, it is typically the case that the $LRTC$ is NOT a perfectly straight line – rather, it looks something like the SR TVC in that it starts from the origin and curves. When this is the case the $LRATC$ and $LRMC$ are NOT constant, but are U-shaped as they were in the short run. However, the reasons that the curves are U-shaped in the LR are different than the reasons the curves are U-shaped in the SR. In the SR, MC was U-shaped because it reflected the law of diminishing marginal returns. ATC was U-shaped because of high AFC for low output levels and high MC for high output levels. However, in

the LR neither of those reasons need to apply, as diminishing marginal returns assumes that we have a fixed factor (which we do not have in the LR) and a high AFC also assumes that we have a fixed factor. In the LR we say that the ATC is U-shaped due to economies and diseconomies of scale. The downward-sloping portion of the LRATC is referred to as the portion with economies of scale and the upward-sloping portion of the LRATC is referred to as the portion with diseconomies of scale. In addition, the minimum of the LRATC may not necessarily be a point but a flat section of the LRATC like the one in the picture shown below. The minimum point (or points if there is a flat section) is called constant returns to scale.
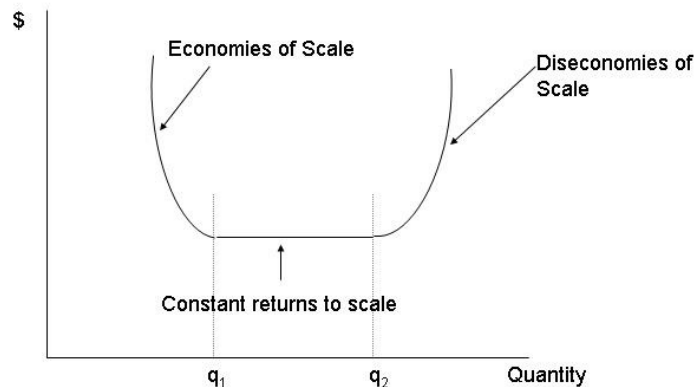

In the picture, there are economies of scale from 0 to q_1. There are constant returns to scale from q_1 to q_2. There are diseconomies of scale from q_2 to infinity. As you can probably tell, economies and diseconomies of scale are tied to the portion of the production function that exhibit increasing, constant, and decreasing returns to scale. If the production function has increasing returns to scale, then the LRATC will have economies of scale. If the production function has constant returns to scale, then the LRATC will be flat.[3] If the production function has decreasing returns to scale, then the LRATC will have diseconomies of scale.

---

[3] In the example where the $LRTC = mq$ and the LRATC was constant at m, the production function was $q = K + L$, which has constant returns to scale over all output levels. Just another reason why the $LRATC$ was constant.

3.3.2 Another method for deriving LRATC

There is one other method for deriving LRATC that is more closely associated with the term "scale". If a firm is entering an industry it will usually have a range of plant sizes from which it will choose. Each of those plant sizes will have a SRATC curve associated with it, as once the firm chooses a plant size there will be fixed costs involved. Suppose that there are 3 plant sizes from which the firm can choose. If we graph their respective SRATCs on one picture, it may look like:
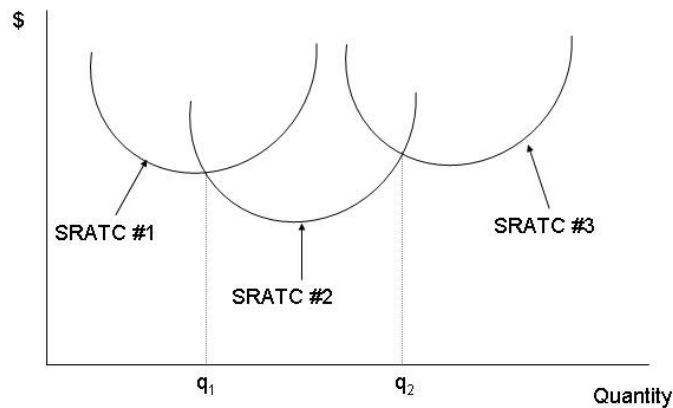
It is important to note that the plant size associated with SRATC #1 is the smallest, while the plant size associated with SRATC #3 is the largest. Now, a firm will wish to minimize the per-unit cost of production in the LR, so it will want to choose the plant size that has the lowest per-unit cost for a specific quantity. Looking at the picture, for any quantity level less than q_1 SRATC #1 is the lowest, so the firm would want to choose that plant. For any quantity between q_1 and q_2, SRATC #2 is the lowest, so the firm would want to choose that plant. For any quantity greater than q_2, SRATC #3 is the lowest, so the firm would want to choose that plant. The LRATC can then be found by looking only at the SRATCs for those quantity levels over which the firm would choose that plant size, as is shown in the figure below.

The entire curve is the firm's LRATC.
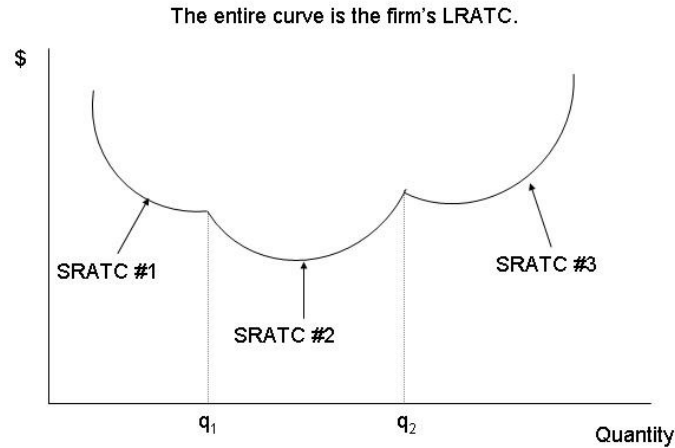
Thus the LRATC is mapped out by the SRATC curves of the various plants. If we could vary plant size by very small amounts (1 square inch or 1 square foot), then we would have many SRATCs on the graph and the LRATC would be a very smooth curve, much as it is depicted in the "Typical shape of the LR cost curves" section above. Hopefully you can see where the terms economies and diseconomies of scale come from – as the firm chooses a bigger plant size (or a bigger scale of production), it has economies of scale if the LRATC decreases. Once choosing a bigger plant size causes LRATC to increase, then the firm has diseconomies of scale.

3.3.3 Reasons for economies and diseconomies of scale

Since economies and diseconomies of scale are tied to increasing and decreasing returns to scale, most of the reasons for economies and diseconomies of scale are the same as those for increasing and decreasing returns to scale.

Economies of scale

1. Bigger plant sizes allow for more specialization among workers.

2. Bigger plant sizes allow firms to utilize mass production methods.

3. Bigger plant sizes allow firms to gain from learning-by-doing. Learning-by-doing is the concept that the more a firm produces, the more it learns about the production process. As it learns more about the production process, it is able to produce more efficiently, thereby lowering per-unit costs.

Diseconomies of scale

1. Bigger plant sizes mean a larger workforce. A larger workforce means that it is not as easy to make sure that workers are working up to their capabilities. Thus, costs may rise because people are shirking.

2. In order to motivate workers and make sure that they work, firms hire supervisors to be in control of groups of workers. Then they hire managers to watch groups of supervisors. Then they hire supervising managers who watch the regular managers. All of these additional hires may add little (depending on the industry) in the way of actual physical output of the product. It also increases the amount of bureaucracy and red tape within the firm. The best example of this that I can think of is the movie Office Space ("I'm telling you Bob, I have 8 managers stopping by my desk asking about my TPS reports").