<u>Chapter 6: Production</u>

This chapter is the first of several (seven, to be precise) that deal with the theory of the firm. As we progress through this section of the course, we will learn more about the firm's cost structure, its profit-maximization problem, and how alternative types of market structures (perfect competition, monopoly, oligopoly) affect the way that firms behave and how they serve consumers.

We begin our analysis by studying the most basic purpose of a firm: production. In the production process, firms transform inputs (labor, capital, materials) into outputs (products that people buy). For simplicity, we will normally assume that firms use labor (workers) and capital (machinery, factories, etc.) to produce goods, but P&R note that other raw materials besides labor (steel, oil, wood, plastic, etc.) are important as well and could be easily incorporated into our conceptualization of the production process.

We describe the firm's production process using a <u>production function</u>, which indicates the maximum amount of output (Q) that a firm can produce for every specified combination of inputs. The production function (with labor and capital as inputs) is written as:

$$Q = F(K,L)$$

Where K denotes capital and L denotes labor. If you tell me how much capital and how much labor a firm is using, F( ) will tell you how much the firm can produce. As far as we are concerned, this production function pretty much completely defines and describes the firm.

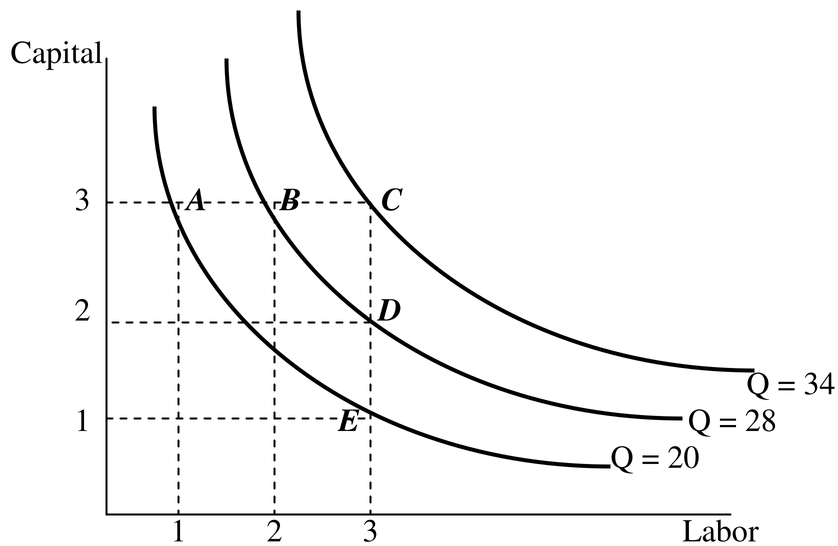
A specific type of production function is the Cobb-Douglas production function, described in the Appendix to Chapter 7. It takes a form similar to the Cobb-Douglas utility function, as follows:

$$Q = AK^{\alpha}L^{\beta} \quad \text{where } 1 > \alpha, \beta > 0$$

This is just an example of one possible form that a firm's production function might hypothetically take. More about the specific properties of this function later. For now, simply note that this function has the property that as we add more of either input, production rises.

Isoquants

We first want to represent the production function graphically, just as we did with the utility function. We can do this using an isoquant, which shows all possible combinations of inputs that yield the same output. (Note how this is analogous to graphing all market baskets that yield the same utility). A typical set of isoquants might look like this:



Points A and E represent points on the same isoquant, Q = 20. The firm can produce 20 units of output using either (K,L) = (3,1) or (1,3). Likewise, the firm could produce 28 units of output using either (K,L) = (3,2) or (2,3) as shown on points B and D. A couple of notes:

1) Isoquants obviously must be downward sloping. If you add labor, for instance, and output remains the same, it must be the case that the level of capital were falling. Presumably, if you added more labor and held K constant, Q would rise, putting the firm on a different isoquant.

2) Isoquants lying to the northeast (or "above" or "to the right of") of others are associated with higher production levels. If the firm uses K=3 and increases its L from 1 to 2 to 3, its production should rise.

The key to this is to recognize that inputs are flexible, and that the firm has a variety of ways to produce the same amount of output. It could use a lot of capital and a little labor, a lot of labor and little capital, or a moderate amount of both inputs.

However, that does not mean that all inputs are always flexible. Particularly in the case of capital, the firm might not be able to change its level of capital on short notice. If a firm is currently producing using a small factory, for example, and decides that it wants a larger factory, it could take years to construct a new production facility. We account for this time-related inflexibility by defining the short run as the period of time in which the amount of one or more output used (usually capital) cannot be changed. The inputs that cannot be changed are said to be fixed inputs; all others (labor usually) are variable inputs. The long run is the time period in which all inputs can be changed (so there are no fixed inputs; all are variable).

Production with One Variable Input

Having made the distinction between the long run and short run, let's analyze the firm's production process in the short run. For simplicity, we'll assume that there is only one variable short run input, namely labor. The following table might summarize a typical production process for such a firm.

L	K	Q	$AP_L(Q/L)$	$MP_L(\Delta Q/\Delta L)$
0	10	0	--	--
1	10	10	10	10
2	10	30	15	20
3	10	60	20	30
4	10	80	20	20
5	10	95	19	15
6	10	108	18	13
7	10	112	16	4
8	10	112	14	0
9	10	108	12	-4
10	10	100	10	-8

The relation between L (holding K constant) and Q is shown in the third column. The fourth column shows the average product of labor (AP_L), measured as output per unit of labor. $AP_L = Q/L$

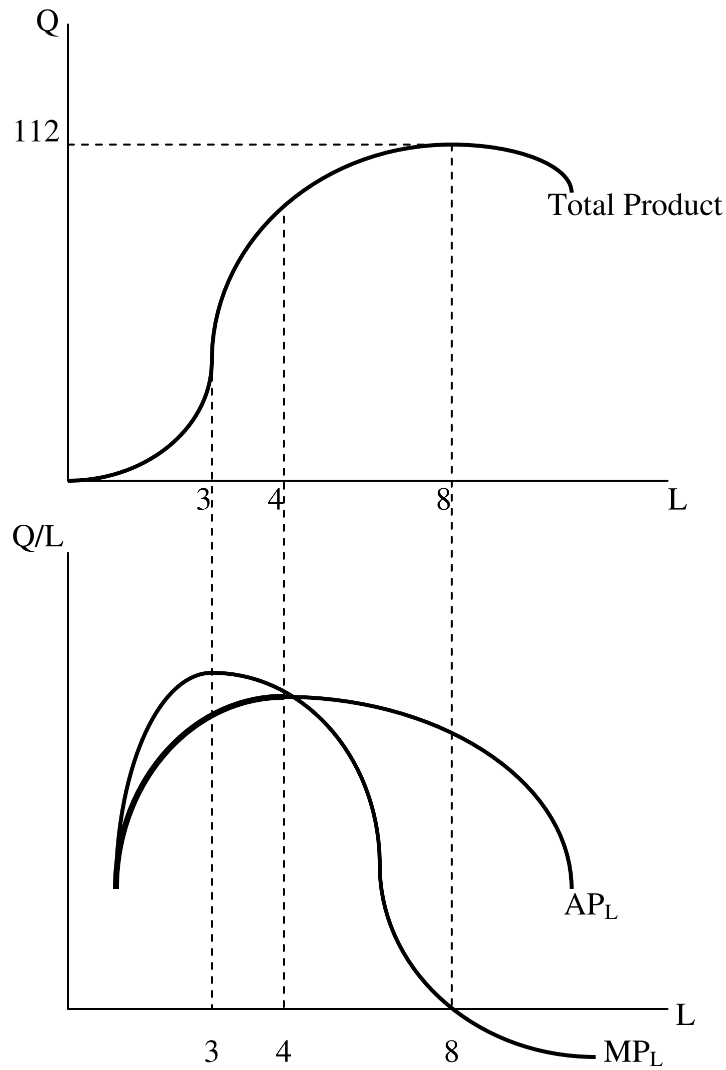
The fifth column shows the marginal product of labor (MP_L), measured as the amount of additional output produced by adding an additional unit of labor. $MP_L = \Delta Q/\Delta L$ or, using calculus, $\partial Q/\partial L$. MP_L gives the rate of change in Q, so if MP_L is positive, Q is rising. If MP_L is rising, Q is rising at an increasing rate. If MP_L is falling, Q is rising at a decreasing rate. If MP_L is negative, Q is falling as we add more labor.

Let's note some of the numeric relationships we see in these data. Notice that, at first, increases in L lead to increases in Q at an increasing rate (with both MP and AP increasing). This is because it may be quite difficult to run a factory efficiently with few workers. One worker with an assembly line all to himself isn't going to get much done. As we add another worker, the first worker is made more productive too, so that MP_L is higher for the second worker than the first.

At some point, though, MP_L should start to fall, and Q will rise at a decreasing rate. Because K is fixed, as we add more workers, each worker will have less and less machinery or other capital to work with. This means that additional workers will add progressively less and less to the firm's output. Note that this occurs NOT because subsequent workers are incompetent or otherwise of lower quality that the first workers hired. We see MP_L fall simply because we are spreading a fixed level of capital across more workers, so that the fixed level of capital becomes an increasingly binding constraint on the firm's production. Decreasing MP_L is very common in real industries, common enough that economists have postulated the law of diminishing marginal returns, which says that as the use of an input increases, with at least one other input held constant, the resulting additions to output will eventually decrease.

Finally, beyond some level, it might make sense to suppose that output falls with additional units of labor. Sooner or later, additional workers just get in the way, leading output to fall. Again, if the firm had a bigger factory this would not be a problem, but since the size of the factory (K=10) is fixed, eventually the firm will rub up against these capacity constraints and exceed them at its own peril.

We can represent a production process like this graphically as follows: (P&R call the graph of Q "total product").



There are several things to note about these diagrams. First, notice that our total product (Q) function looks like what we described. Output at first increases at an increasing rate, then rises at a decreasing rate as diminishing marginal returns kicks in, and finally starts to fall. Of course, no profit-maximizing firm would ever produce at that particular point on the total product curve since they are using *more* of an input to get *less* output.

Second, notice the relationship between Q and MP_L. Because MP is $\partial Q/\partial L$, it is also the slope of the total product function (slope of a line tangent to the total product function at a point). The slope (MP) increases until we get to L = 3, at which point MP reaches its maximum. It then falls, but remains positive until L = 8, at which MP is zero (crossing the horizontal axis on the lower graph). Then it is negative from there on out.

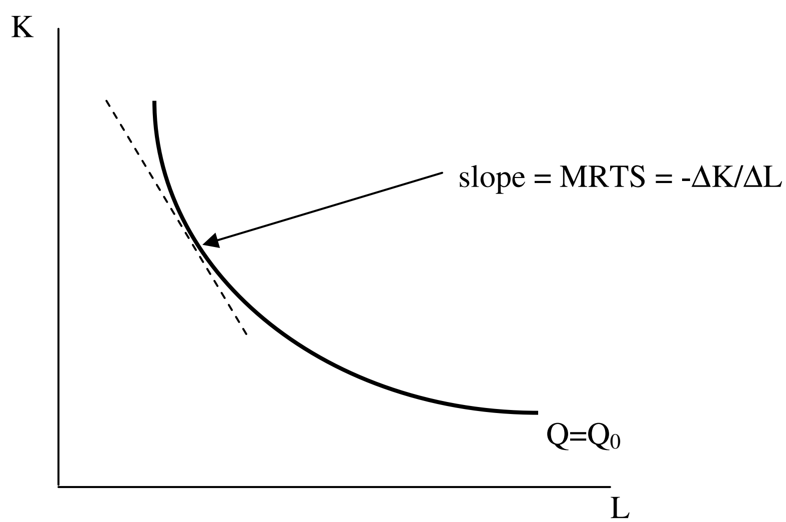
Finally, AP_L can be shown to be the slope of a line drawn from the origin to a point on the total product curve (since the slope of such a line will be simply Q/L). AP is maximized at Q = 4 and then falls continuously. Also notice that whenever MP > AP, AP is rising, since each additional unit of labor adds more output than average, pulling the average up. Furthermore, when MP < AP, each additional worker adds less output than average, dragging the average down. Since AP is rising when it is less than MP and falling when it is greater than MP, it is maximized when it equals MP.

These two graphs pretty much completely describe the firm's short run production possibilities with only one variable input. To describe the firm's long run options, when both capital and labor can be varied, we need to examine . . .

Production with Two Variable Inputs

Once we have two variable inputs, we can use isoquants to analyze production. Notice that in our earlier isoquant example, diminishing marginal returns held for both inputs, holding the other constant, as we would expect (one extra unit of labor, holding K = 3, generated less and less additional output).

The slope of the isoquant, $-\Delta K/\Delta L$ (or $-dK/dL$ using calculus), is called the marginal rate of technical substitution (MRTS). (We insert the minus sign to make this a positive number). This shows the amount by which the quantity of one input can be reduced when another unit of another input is added to keep output constant. In other words, at what rate can the firm trade off capital for labor?



You will notice that the isoquant gets flatter as we go along, meaning that the MRTS falls. This is an assumption we make, much like the assumption we made that indifference curves are convex. Why should the isoquant necessarily be convex, though? To see this, recognize that we can rewrite the MRTS. Start by totally differentiating the production function Q = F(K,L).

$$dQ = (\partial F/\partial K)dK + (\partial F/\partial L)dL$$

But as we move along a given isoquant, Q is held constant, so dQ = 0. Furthermore, $\partial F/\partial K$ and $\partial F/\partial L$ are simply MP_K and MP_L respectively. Thus we have

$$0 = MP_K(dK) + MP_L(dL)$$
$$\Rightarrow dK/dL = -MP_L/MP_K$$
$$\Rightarrow -dK/dL = MP_L/MP_K$$

This tells us that MRTS is simply the ratio of the marginal product of labor and the marginal product of capital. This explains why the isoquant is convex. As we move down along an isoquant, using more labor and less capital, MP_L is falling and MP_K is rising because of the law of diminishing marginal returns. Hence, MRTS must be falling. Hence the convex isoquant.

As we add more inputs, the firm is able to obtain isoquants lying further to the northeast. If the firm is adding more of both inputs, output should always go up (since the firm could simply replicate its current production in another location). How much of an increase in output the firm gets from increases in input levels depends again on the firm's production technology.

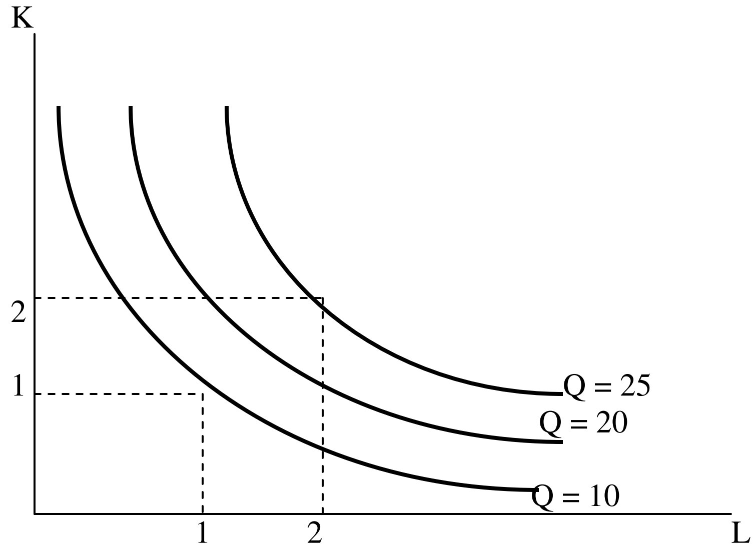
A firm is said to have <u>increasing returns to scale</u> if output rises more than proportionally as all input levels rise. (If all inputs are doubled, output more than doubles).

The firm has <u>constant returns to scale</u> if output rises proportionally as all input levels rise. (If all inputs are doubled, output doubles).

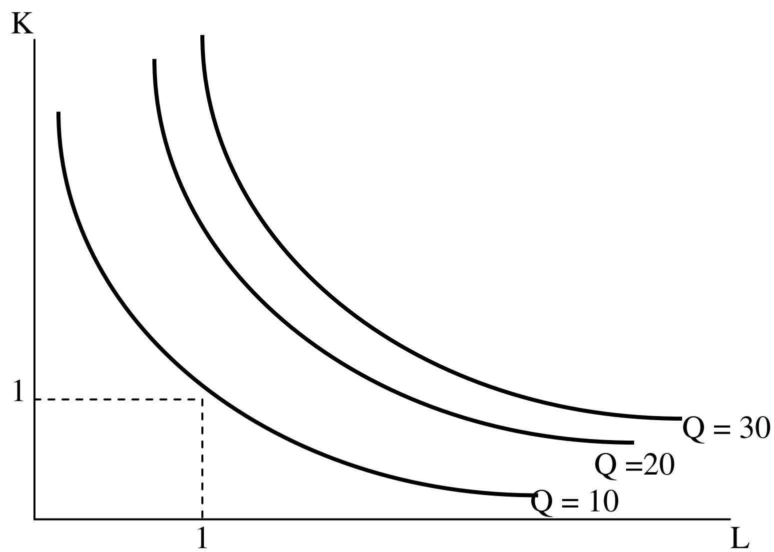
The firm has <u>decreasing returns to scale</u> if output rises less than proportionally as all input levels rise. (If all inputs are doubled, output less than doubles).

Don't confuse returns to scale with diminishing marginal returns. That refers to a situation where one input is being held constant (doubling the amount of labor but holding capital constant). Returns to scale, on the other hand, refers to all inputs being increased simultaneously (doubling the amount of labor *and* capital). They are completely different phenomena.

We can show returns to scale possibilities graphically as well:

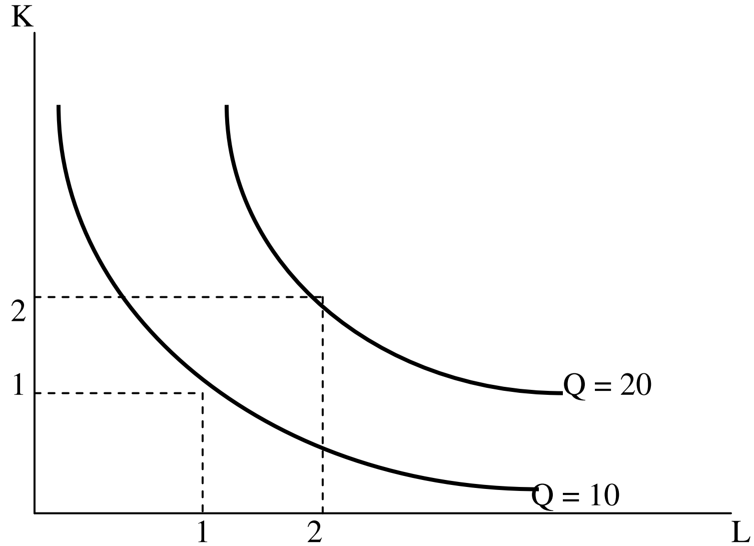


This diagram shows a firm with increasing returns to scale. As inputs are raised by a factor of 2, output rises by a factor of 2.5. Put another way, the firm is able to double output without having to double all input levels. In that case, we can look at isoquants like this:

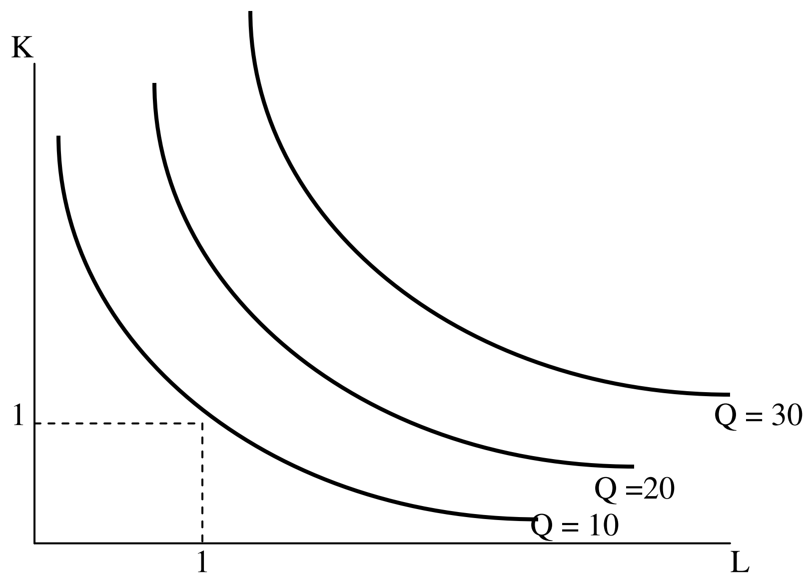


The isoquants get closer together as we increase output, since inputs do not need to be increased as much.

Here we see a firm with constant returns to scale:

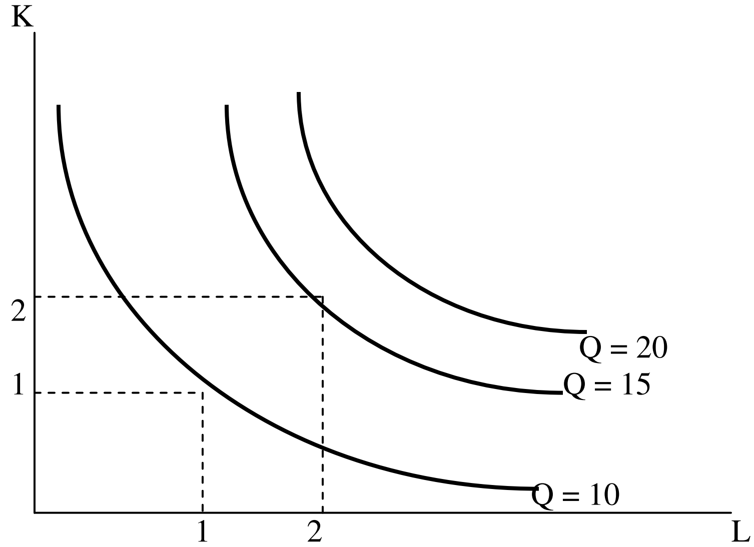


This diagram shows a firm with constant returns to scale. As inputs are raised by a factor of 2, output rises by a factor of 2. Put another way, the firm is able to double output by doubling all input levels. In that case, we can look at isoquants like this:

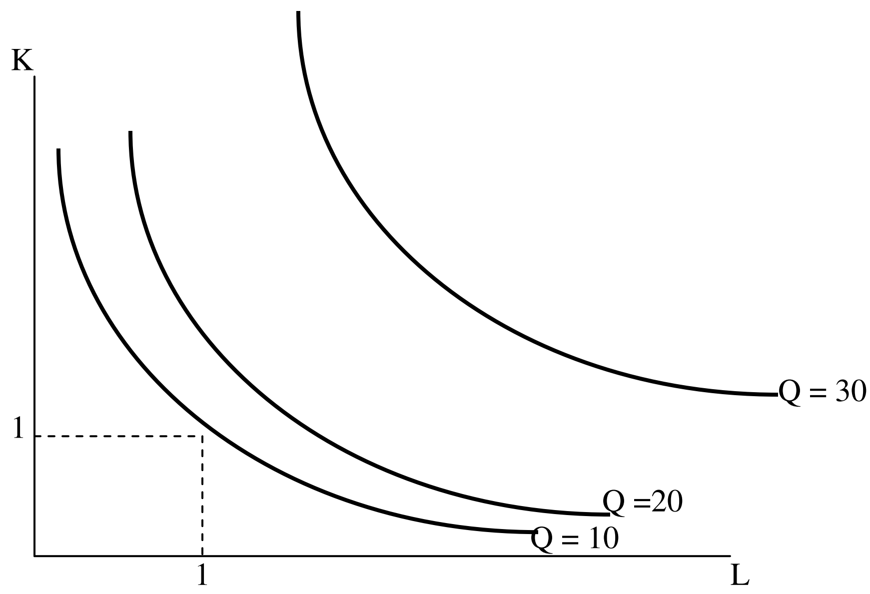


The isoquants get "move" outward proportionally as we increase output, since inputs have to be increased proportionally.

Finally, here is a firm with decreasing returns to scale:



This diagram shows a firm with decreasing returns to scale. As inputs are raised by a factor of 2, output rises by a factor of only 1.5. Put another way, the firm cannot double output without having to more than double all input levels. In that case, we can look at isoquants like this:



The isoquants get further apart as we increase output, since inputs need to be increased as more each time with scale up production.

The Cobb-Douglas Production Function

Notice that the Cobb-Douglas production function described earlier does a good job of demonstrating the diminishing MRTS that we would expect from a production function, as well as being versatile enough to handle increasing, constant, or decreasing returns to scale.

First consider the MRTS of the Cobb-Douglas function. You'll recall that the function takes the form

$$Q = AK^{\alpha}L^{\beta} \quad \text{where } 1 > \alpha,\beta > 0$$

$$\Rightarrow \quad MP_L = \partial Q/\partial L = A\beta K^{\alpha}L^{\beta-1}$$
$$MP_K = \partial Q/\partial K = A\alpha K^{\alpha-1}L^{\beta}$$

$$\Rightarrow \quad MRTS = MP_L/MP_K = (A\beta K^{\alpha}L^{\beta-1})/(A\alpha K^{\alpha-1}L^{\beta}) = (\beta/\alpha)(K/L)$$

In fact, as we increase L and decrease K (moving down along the isoquant), K/L falls, so the MRTS falls, so the isoquant gets flatter and takes the convex shape we expect.

Now let's think about returns to scale with the Cobb-Douglas production function. If we increase K and L by the same factor, we can write that by "scaling up" both inputs by the factor λ. Thus, we scale up K to λK and L to λL. If we double both inputs, $\lambda = 2$ so K becomes 2K and L becomes 2L.

What does this do to production if we have $Q = F(K,L) = AK^{\alpha}L^{\beta}$? Well,

$$\begin{aligned} F(\lambda K,\lambda L) &= A(\lambda K)^{\alpha}(\lambda L)^{\beta} \\ &= A\lambda^{\alpha}K^{\alpha}\lambda^{\beta}L^{\beta} \\ &= \lambda^{(\alpha+\beta)}AK^{\alpha}L^{\beta} \\ &= \lambda^{(\alpha+\beta)}F(K,L) \end{aligned}$$

What does this mean? It means that if we scale up both inputs by the same factor λ, output will rise by the factor $\lambda^{(\alpha+\beta)}$. We can tell whether we have increasing, decreasing or constant returns to scale by observing whether $\lambda^{(\alpha+\beta)}$ > or < or = λ. But that just depends on whether $(\alpha+\beta)$ > or < or = 1. Thus,

$$\alpha + \beta > 1 \quad \Rightarrow \quad \lambda^{(\alpha+\beta)} > \lambda \quad \Rightarrow \quad \text{increasing returns}$$
$$\alpha + \beta = 1 \quad \Rightarrow \quad \lambda^{(\alpha+\beta)} = \lambda \quad \Rightarrow \quad \text{constant returns}$$
$$\alpha + \beta < 1 \quad \Rightarrow \quad \lambda^{(\alpha+\beta)} < \lambda \quad \Rightarrow \quad \text{decreasing returns}$$