Yuriy Gorodnichenko
Econ 101B, Spring 2025

Aggregate Supply, Unemployment and the Natural Rate

These notes will construct our model of aggregate supply. The key building block of the supply side of the economy will be the labor market. We will consider labor supply and labor demand to understand the determinants of the real wage, employment, and unemployment in the economy. Given the labor market, we will be able to derive a relationship between prices and output that ensures equilibrium in the labor market. Later, we will combine this with the aggregate demand framework to close our model of the business cycle.

The Labor Market

We turn first to the labor market. For simplicity, we assume that there is a single market for all of labor within the economy, as well as one level of wages that applies to all workers. We will need to describe both the supply and demand for labor in this market.

Labor Demand

Consider a firm that is trying to maximize profits. This firm produces a good that is sold at price P. To produce this good, it uses capital and labor as inputs. The ***production function*** is given by

$$Y = AK^{\alpha}N^{1-\alpha}$$

where K is the stock of capital and N is the amount of labor employed. A is the level of technology involved in the production process. The coefficient α is assumed to be between 0 and 1. This production function exhibits *diminishing returns to each input* separately (because $0<\alpha<1$) , but *constant returns to scale* overall (because doubling both inputs doubles output). The firm must pay wage W to its workers. We'll treat the capital stock as fixed over the horizon of the firm's maximization problem, i.e. the capital stock is exogenously given. We'll assume the firm faces a (potentially) downward-sloping demand curve

$$P = Y^{-\frac{\mu}{1+\mu}}$$

where $\mu \geq 0$. When $\mu>0$, then to sell more output (higher Y), the firm must charge a lower price. If $\mu=0$, then the price level is independent of output. This is the case under perfect competition when the firm is a price-taker. Thus, $\mu>0$ implies some degree of monopoly power for the firm. ***Profits*** (Π) for the firm are given by the difference between total revenue (price times output) and total costs (number of workers times wages)[1]

$$\Pi = PY - WN$$

[1] Technically we should have $\Pi = PY - WN - RK$, i.e., we should also subtract the rental cost of capital RK. However, we assume that capital is fixed and thus RK is just a constant which we will omit to simplify algebra.

which, if we plug in the demand curve for the price can be rewritten as

$$\Pi = Y^{1-\frac{\mu}{1+\mu}} - WN = Y^{\frac{1}{1+\mu}} - WN$$

The profit-maximization problem is then given by

$$\max_{N} \quad Y^{\frac{1}{1+\mu}} - WN$$

$$subject\ to\ \ Y = AK^{\alpha}N^{1-\alpha}$$

You may recall the optimality condition for how much a firm should produce from microeconomics:

$$MR = MC$$

i.e. the firm should keep producing until the marginal revenue of another unit produced (*MR*) is equal to the marginal cost of that unit (*MC*). Why does this have to be true? Suppose the firm is producing a quantity such that MR>MC. In this case, by increasing production, it will raise revenues more than costs, thereby increasing profits. Hence, MR>MC cannot be an optimal outcome. Similarly, if MR<MC, a firm could decrease production, thereby lowering revenues but by less than the decrease in costs, again increasing profits.

What is marginal revenue here? Total revenue is given by prices times quantity, *P*Y*, which given the demand curve, simplifies to $Y^{\frac{1}{1+\mu}}$. The marginal revenue of an additional unit of output is the derivative of total revenue with respect to output, or

$$MR = \left(\frac{1}{1+\mu}\right)Y^{\frac{1}{1+\mu}-1} = \left(\frac{1}{1+\mu}\right)Y^{\frac{-\mu}{1+\mu}} = \left(\frac{1}{1+\mu}\right)P$$

What is marginal cost? Marginal cost is the cost of producing one more unit. To produce an additional unit, a firm must pay for additional labor, at rate *W*. How much additional labor depends on the marginal product of labor MP_N. The marginal product of labor tells us how many units one additional worker will provide. Given our production function, the marginal product of labor is

$$MP_N = (1-\alpha)AK^{\alpha}N^{-\alpha}$$

How much labor we need to produce one more unit is then given by $1/MP_N$. Thus, marginal cost is

$$MC = \frac{W}{MP_N} = \frac{W}{(1-\alpha)AK^{\alpha}N^{-\alpha}}$$

The optimality condition for production is then:

$$MR = MC$$

$$\Leftrightarrow \left(\frac{1}{1+\mu}\right)P = MC$$

$$\Leftrightarrow P = (1+\mu)MC$$

which says that the optimal price a constant markup over marginal cost. The markup is given by 1+μ. Note that if μ=0, i.e. if the firm is in a perfectly competitive industry in which it is a price-taker, we have the familiar condition that price equals marginal cost. If μ>0, our firm has market power and charges a price in excess of marginal cost. The greater the market power (or monopoly power), the bigger the markup.

Plugging in for marginal costs:

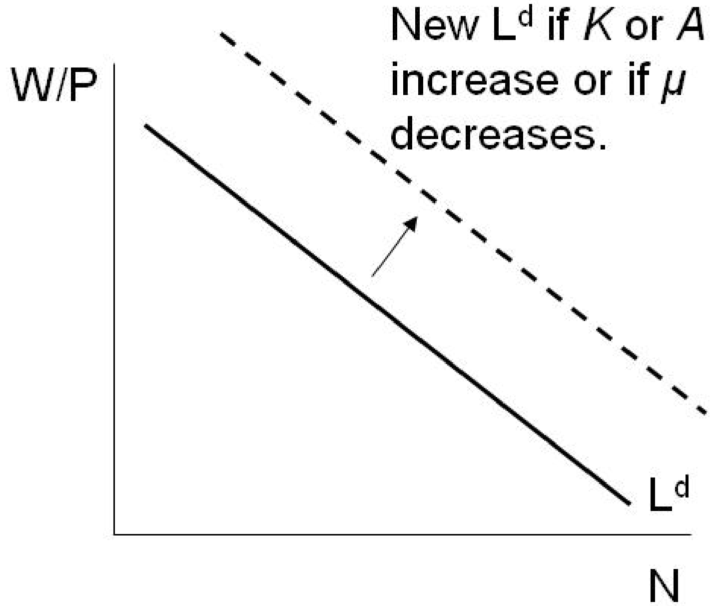
$$P = (1+\mu)\frac{W}{(1-\alpha)AK^{\alpha}N^{-\alpha}}$$

which can be rearranged as

$$\boldsymbol{\frac{W}{P} = \frac{(1-\alpha)}{(1+\mu)}AK^{\alpha}N^{-\alpha}}$$

This is the ***labor demand*** for the firm. W/P is the ***real wage*** that the firm must pay its workers. Note that because the exponent on N is negative, the higher the real wage the fewer workers firms want to hire.

Graphically, this is as shown in the graph below



Note how the exogenous variables shift the demand curve:

1) When μ goes up, the demand curve shifts in. Recall that μ tells us how monopolistic firms are. High values of μ mean that firms have significant monopoly power and industries are less competitive. This leads to lower levels of production and lower demand for labor.
2) When the stock of technology (A) rises, workers become more productive. Thus firms are willing to hire more workers at the same real wage, expanding the demand for labor.
3) When the amount of capital (K) in the economy rises, workers again become more productive. By the same logic as with technology, labor demand expands.

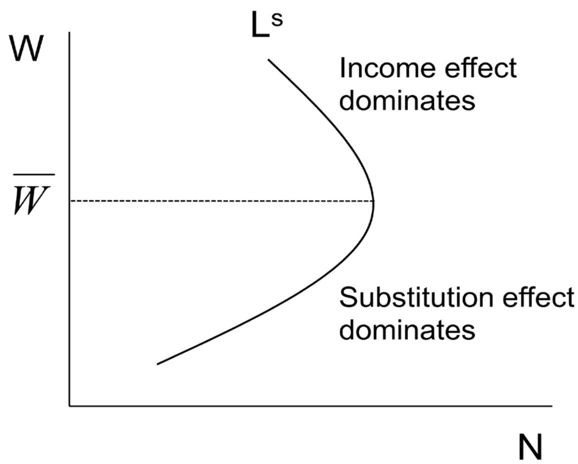
We can express our ***labor demand curve*** more generally as

$$\frac{\boldsymbol{W}}{\boldsymbol{P}} = \boldsymbol{L^d}\left(\overbrace{\boldsymbol{\mu}}^{-}, \overbrace{\boldsymbol{A}}^{+}, \overbrace{\boldsymbol{K}}^{+}, \overbrace{\boldsymbol{N}}^{-}\right)$$

Labor Supply

Consider now workers' labor supply condition: the willingness of agents to provide labor to firms in return for wages. Suppose firms are offering a nominal wage W. How much labor are agents willing to supply? In general, there are two conflicting effects operating at the same time. When wages go up, workers should tend to think that the opportunity cost of leisure time has gone up: i.e. each hour not spent working now entails a higher sacrifice in terms of foregone earnings. This means that with leisure time being more expensive, agents should tend to work more. This is called the *substitution effect*. However, when wages go up, it also means that workers should feel richer. When agents feel richer, they should tend to want to enjoy more leisure time, i.e. work less. This is known as the *income effect*. Which effect is stronger determines whether workers will supply more or less labor as wages rise.

Typically, economists assume that at low levels of income, the substitution effect dominates, so that workers are willing to supply more labor for higher wages. However, once wages are high enough, the income effect tends to dominate, so that workers are willing to supply less labor for higher wages. This results in what's known as the ***"backward-bending" labor supply curve****:*



The critical wage level $\overline{W}$ is the wage level where the income and substitution effects exactly cancel out. Labor economists often find that when wages rise, there is little change in labor supplied by workers. This would seem to indicate that for many workers, the income and substitution effects nearly cancel out. On the other hand, macroeconomists typically argue that during business cycles, employment fluctuates substantially whereas wages move only little, though the two tend to go in the same direction. This would tend to imply that the substitution effect dominates (some would say it dominates strongly).

An important distinction that helps resolve this apparent contradiction is the notion of *intensive **and extensive margins*** of labor supply change. When labor economists find that a given worker will not increase labor supply when wages rise, this is called low elasticity of labor supply along the intensive margin. However, if small changes in wages induce new agents to join the labor force, causing overall labor supply to rise strongly, then we can observe a high elasticity of labor supply along the extensive margin. For the purposes of this class, we will focus on the qualitative result that when wages rise, labor supplied overall goes up as well.

In addition to the wage, a second element that affects labor supply is expected prices. Most wages are predetermined in the sense that workers and firms agree on a wage to be paid over a period of time. However, people typically understand that what matters for their purchasing power is not the nominal wage W but rather the purchasing power of the wage, i.e. the wage relative to prices. If the wage is set contractually for a set period of time, then workers should care about expected prices over the duration of the contract (P^e).

There are, of course, other elements that affect agents' willingness to supply labor given the wage. For example, the generosity of unemployment benefits is important. The higher these benefits, the less willing some agents will be to work at low wages. Economists frequently refer to something known as the ***reservation wage***: this is the income that workers would receive if they are not working. The higher is the reservation wage, the lower labor supplied will be for a given wage since some agents will choose not to work. Unemployment benefits clearly fit this definition, but many others do as well. For example, if a country has a nationally-provided system of health care, this will be part of the reservation wage since workers will receive this regardless of whether they work. In countries like the U.S. where health care is typically received through employment, the reservation wage does not include health insurance. We will group all these elements into a composite term z.

Thus, one could generally write the labor supply condition as

$$W = P^e g(N^S, z)$$

where $g(\,)$ is an (unspecified) increasing function of N^S and z. This says that higher wages lead to higher labor supplied, given price expectations and the reservation wage. In addition, if expected prices rise, workers will demand a higher wage to work the same amount. Finally, if the reservation wage (z) is higher, then workers will demand a higher real wage to supply the same labor. To simplify matters, we will make a slightly more extreme assumption: let's assume that the labor supply elasticity is infinite. In other words, let's assume workers will supply as much labor as is required at the given real wage.[2] This may seem extreme given the discussion of inelastic labor supply, but again, the aggregate labor supply elasticity can be high because of changes along the extensive margin. This is also consistent with workers signing contracts agreeing to provide as much labor as may be necessary for a given wage. We'll also assume that g(z)=z. Thus, the labor supply condition becomes simply:

$$W = P^e z$$

Dividing both sides by P yields the ***labor supply condition*** in terms of the real wage W/P

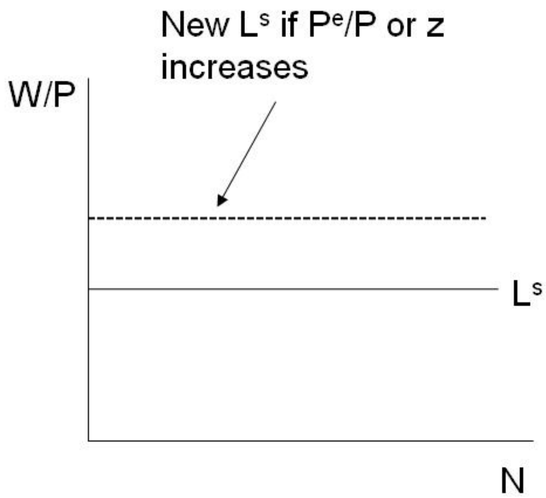
$$\boldsymbol{\frac{W}{P} = \frac{P^e}{P} z}$$

Note the following:

a) If workers expect prices to rise (P^e goes up), they will demand higher wages and this will raise the labor supply curve.
b) If current prices rise, this decreases the real wage and labor supply curve.
c) If the reservation wage (z) rises, then workers will demand higher real wages to work and the labor supply curve will shift up.

[2] Mathematically, this means that $g(N^S, z) = g(z)$ and the wage does not depend on N^s.

Graphically, the labor supply curve would be represented as



Note that if expected prices rise relative to actual prices, then the real wage will rise as the labor supply curve shifts up. Similarly, increases in z (the determinants of the reservation wage) will also shift the labor supply curve up.

Labor Market Equilibrium and the Short-Run Aggregate Supply Curve

As usual, equilibrium in the market requires that quantity supplied and quantity demanded be equal. This occurs, in the case of the labor market, where the labor supply and labor demand curves intersect. This point determines both the equilibrium real wage and equilibrium employment. Formally, we have the following labor supply and demand conditions:

Labor supply: $\boldsymbol{\frac{W}{P} = \frac{P^e}{P} z}$

Labor demand: $\boldsymbol{\frac{W}{P} = \frac{(1-\alpha)}{(1+\mu)} A K^{\alpha} N^{-\alpha}}$

In equilibrium, the real wage implied by the two must be equal. Hence, we have

$$\frac{P^e}{P} z = \frac{(1-\alpha)}{(1+\mu)} A K^{\alpha} N^{-\alpha}$$

Or equivalently

$$\boldsymbol{P = z P^e \frac{(1+\mu)}{(1-\alpha)} \frac{N^{\alpha}}{A K^{\alpha}}}$$

Note that this expression links the price level positively to the level of employment. Because output depends positively on employment, the equilibrium in the labor market implies a positive relationship between prices and employment. To see this more formally, rewrite the last expression solving for N:

$$\boldsymbol{N} = \left[\frac{\boldsymbol{P}}{\boldsymbol{P^e}}\frac{(\boldsymbol{1}-\boldsymbol{\alpha})}{\boldsymbol{z}(\boldsymbol{1}+\boldsymbol{\mu})}\boldsymbol{A}\right]^{\boldsymbol{1/\alpha}}\boldsymbol{K}$$

which is the equilibrium level of employment implied by the labor market, conditional on the price level and the exogenous variables. Plugging this into the production function:

$$Y = AK^{\alpha}N^{1-\alpha}$$

yields

$$Y = AK^{\alpha}\left[\left[\frac{P}{P^e}\frac{(1-\alpha)}{z(1+\mu)}A\right]^{1/\alpha}K\right]^{1-\alpha}$$

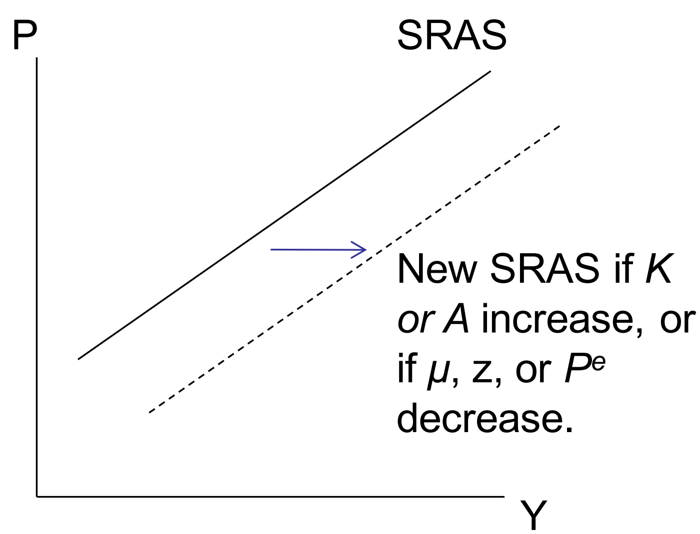
which is equivalent to

$$Y = A^{1/\alpha}K\left[\frac{P}{P^e}\frac{(1-\alpha)}{z(1+\mu)}\right]^{(1-\alpha)/\alpha}$$

Finally, let's rewrite this in terms of the price level:

$$P = P^e\frac{z(1+\mu)}{(1-\alpha)}\left[\frac{Y^{\alpha}}{AK^{\alpha}}\right]^{\frac{1}{(1-\alpha)}}$$

which gives us our ***short-run aggregate supply*** (SRAS) relationship. Graphically in P-Y space, this yields



More generally, we could write the short-run aggregate supply relationship as

$$\boldsymbol{SRAS} = \boldsymbol{SRAS}\left(\overset{-}{\boldsymbol{P^e}}, \overset{+}{\boldsymbol{A}}, \overset{+}{\boldsymbol{K}}, \overset{-}{\boldsymbol{z}}, \overset{-}{\boldsymbol{\mu}}\right)$$

Note that if expectations of prices rise, then the SRAS curve shifts up (or left), reflecting the fact that workers will demand higher wages and force firms to charge higher prices. If A or K rise, then labor productivity will be higher, firms will hire more labor, and output will rise, causing SRAS curve to shift right. If z rises, then workers' reservation wage is higher and they demand higher wages. This causes firms to hire fewer workers and lowers output, therefore shifting the SRAS curve left. Finally, if μ rises, firms have more market power, causing them to want to decrease output to raise prices, shifting the SRAS curve left.

Long-Run Aggregate Supply and the Natural Rate of Output, Employment, and Unemployment

You may have noticed that we labeled the previous relationship "short-run". Why was this distinction added? A key feature of the labor market that we considered is that expectations of prices may differ from actual prices. This assumption is generally considered valid in the short-run, but not in the long-run. In the long-run, agents are assumed to be rational and therefore to be able to accurately adjust their expectations of the price level to the right level. In other words, the *long-run assumption* is that

$$P^e = P$$

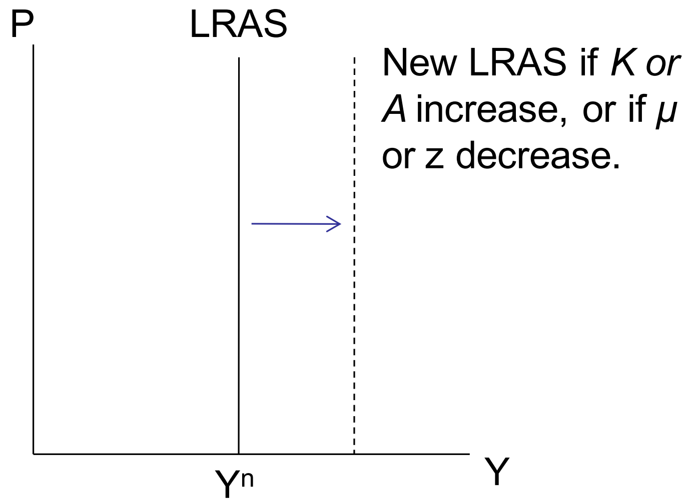
This has crucial implications for the supply side of the economy. Recall that the short-run aggregate supply relationship was given by:

$$P = P^e \frac{z(1+\mu)}{(1-\alpha)} \left[\frac{Y^\alpha}{AK^\alpha}\right]^{\frac{1}{(1-\alpha)}}$$

which, when $P^e = P$, reduces to

$$\boldsymbol{Y^n = A^{1/\alpha} K \left[\frac{1-\alpha}{z(1+\mu)}\right]^{\frac{1-\alpha}{\alpha}}}$$

This expression defines what's called the ***natural level of output (Y_n)***. This is the level of output that would occur if price expectations were equal to actual prices. This is also called the ***long-run aggregate supply*** relationship, as it relies on imposing the long-run assumption that price expectations are equal to actual expectations. The striking thing about this relationship is that it is independent of the price level, i.e. in P-Y space, it is a vertical line at the natural level of output



Similarly, one can define a ***natural level of employment*** as the employment level that occurs when price expectations equal actual prices. Recall that equilibrium in the labor market requires

$$P = zP^e \frac{(1+\mu)}{(1-\alpha)} \frac{N^\alpha}{AK^\alpha}$$

so that, when $P=P^e$, this can be rewritten as

$$\boldsymbol{N^n = \left[\frac{(1-\alpha)}{z(1+\mu)}\right]^{1/\alpha} A^{1/\alpha} K}$$

Finally, we can link this back to the unemployment rate. Recall that

$$U = \frac{L-N}{L} = 1 - \frac{N}{L}$$

defines the unemployment rate via the ration of employment to the labor force. This can be rearranged in terms of employment as

$$N = L(1-U)$$

Plugging this last expression in for the natural level of employment yields

$$\boldsymbol{U^n = 1 - \left[\frac{(1-\alpha)}{z(1+\mu)}\right]^{1/\alpha} A^{1/\alpha} \left(\frac{K}{L}\right)}$$

which defines the ***natural rate of unemployment,*** also called the ***structural rate of unemployment***. The natural rate of unemployment is the unemployment rate that would occur if price expectations were equal to actual prices. In other words, this is the average rate of unemployment around which the economy will bounce around.

Note the key determinants of the natural rate of unemployment:

1. A high reservation wage (z), due for example to high UE benefits, leads to a higher natural unemployment rate. High unemployment rates in Europe since the 1970s are often attributed to labor market policies that discourage workers from working and firms from hiring.
2. More competitive economies, with lower μ, will tend to have lower rates of unemployment. This is because monopoly power causes firms to produce at lower levels of output to keep prices higher. This lower level of output implies lower demand for labor, and hence less employment and higher unemployment.
3. Productivity (A) also matters. Improvements in technology or general productivity can push the natural rate of unemployment down. For example, the low rates of unemployment in the 1990s are often attributed to the "New Economy" and the adoption of computers in the workplace.
4. A high capital-labor ratio (K/L) also tends to lead to lower unemployment rates. Economies with more machines per person will tend to have a higher marginal product of labor, leading to more demand for workers and lower unemployment rates. What affects the capital to labor ratio? A key element is taxes on capital. When capital is heavily taxed relative to labor, the capital to labor ratio will be low as firms will try and avoid using capital.

Extension: Oil Prices and Aggregate Supply

Suppose we want to study the effect of oil prices (or raw materials more generally) on the economy. We need to integrate these into the supply side of our model. Let's allow for a more general production function that implies that firms need oil to produce:

$$Y = AK^{\alpha}N^{\beta}O^{1-\alpha-\beta}$$

where K is the stock of capital, N is the amount of labor employed, and O is the amount of raw materials used in production (O is for oil). A is the level of technology involved in the production process. The coefficients α and β are assumed to be between 0 and 1. This production function exhibits *diminishing returns to each input* separately (because $0<\alpha,\beta<1$) , but *constant returns to scale* overall (because doubling both inputs doubles output). The firm must pay wage W to its workers and P_o is the price for materials. We'll treat the capital stock as fixed over the horizon of the firm's maximization problem, i.e. the capital stock is exogenously given. We'll assume the firm faces a (potentially) downward-sloping demand curve

$$P = Y^{-\frac{\mu}{1+\mu}}$$

Profits for the firm are given by the difference between total revenue and total costs

$$\Pi = PY - WN - P_o O$$

which, if we plug in the demand curve for the price can be rewritten as

$$\Pi = Y^{\frac{1}{1+\mu}} - WN - P_o O$$

The profit-maximization problem is then given by

$$\max_{\mathrm{N,O}} \quad Y^{\frac{1}{1+\mu}} - WN - P_o O$$

$$subject\ to\ \ Y = AK^{\alpha} N^{\beta} O^{1-\alpha-\beta}$$

The optimality condition for inputs is that the firm should keep using more inputs until the extra revenue generated by an input is exactly equal to its marginal cost. If the firm is hiring fewer inputs, then it could increase profits by hiring more since the marginal cost of these inputs is less than the extra revenue they generate. The extra revenue from hiring one more input is known as the *marginal revenue product* of the input, which the extra cost of an additional input is known as the *marginal cost* (mc_i). Thus, the profit-maximizing condition for each input is

$$MRP_i = MC_i$$

where i is the relevant input (N for labor, O for materials, …). The marginal cost of an input is simply the wage for labor (W) or the price of materials (P_o). The marginal revenue product is a more complicated creature: we have to take into account two factors: if we use one more input, by how much does output rise (this is the *marginal product MP*)? Then, if output goes up by that amount, by how much does total revenue rise (this is the *marginal revenue MR*)? The product of these two effects is the marginal revenue product: *MRP=MR*MP*.

For our problem, the marginal product of the input is the derivative of the production function with respect to that input. Thus, for labor and materials respectively we have:

$$MP_N = \beta AK^{\alpha} N^{\beta-1} O^{1-\alpha-\beta}$$

$$MP_O = (1-\alpha-\beta) AK^{\alpha} N^{\beta} O^{-\alpha-\beta}$$

The marginal revenue of an additional unit of output is the derivative of total revenue with respect to output, or

$$MR = \left(\frac{1}{1+\mu}\right) Y^{-\frac{\mu}{1+\mu}}$$

Therefore, the marginal revenue product for each input is given by the marginal product of the input times the marginal revenue, which gives:

$$MRP_N = \left(\frac{1}{1+\mu}\right) Y^{-\frac{\mu}{1+\mu}} \beta AK^{\alpha} N^{\beta-1} O^{1-\alpha-\beta}$$

$$MRP_O = \left(\frac{1}{1+\mu}\right) Y^{-\frac{\mu}{1+\mu}} (1-\alpha-\beta) A K^{\alpha} N^{\beta} O^{-\alpha-\beta}$$

Therefore, the optimality conditions, MRP=MC, are

$$\left(\frac{1}{1+\mu}\right) Y^{-\frac{\mu}{1+\mu}} \beta A K^{\alpha} N^{\beta-1} O^{1-\alpha-\beta} = W$$

$$\left(\frac{1}{1+\mu}\right) Y^{-\frac{\mu}{1+\mu}} (1-\alpha-\beta) A K^{\alpha} N^{\beta} O^{-\alpha-\beta} = P_o$$

OK, so this looks hideous, but this is where the beauty of Cobb-Douglas production functions kicks in. Let's divide the top condition by the lower condition and note how much cancels out:

$$\frac{\beta}{(1-\alpha-\beta)} \frac{O}{N} = \frac{W}{P_o}$$

or equivalently

$$\boldsymbol{\frac{OP_o}{NW} = \frac{(1-\alpha-\beta)}{\beta}}$$

Note that OP_o is the total cost of material inputs while NW is the total cost of labor inputs. The ratio of the two should exactly equal the ratio of the exponents on each in the production function. Note the implications of this result:

a) If the price of materials rises relative to wages (P_o/W goes up), then O/N must fall. This means firms will hire relatively more workers and buy less raw material.
b) If the firm needs to expand production but there has been no change in input prices, then the firm must increase its use of inputs by the same proportion.

Now, we'll use this expression to back out the firm's demand for labor as a function only of exogenous variables and the real wage (i.e. let's get rid of the O term). First, let's rewrite the last expression

$$O = \frac{NW}{P_o} \frac{(1-\alpha-\beta)}{\beta}$$

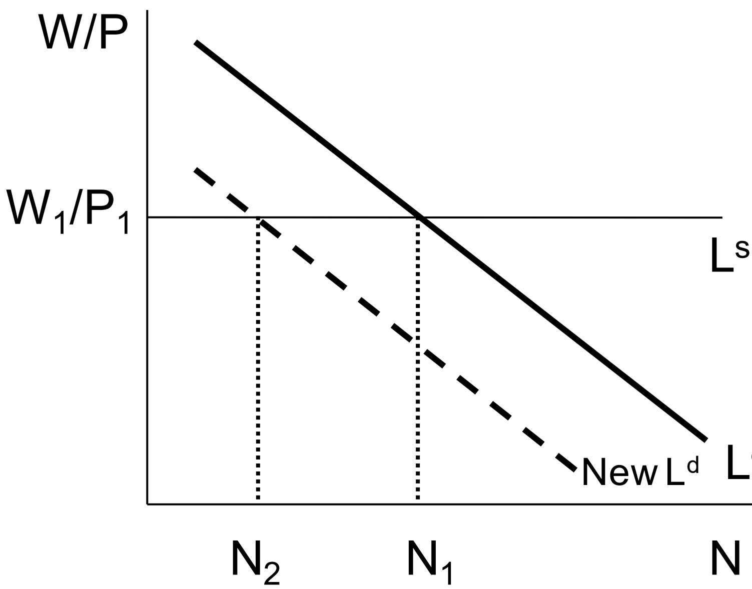
and plug this back into the optimality condition for labor. After extensive rearranging, we get the following ***demand curve for labor***:

$$\boldsymbol{\left(\frac{W}{P}\right)^{\alpha+\beta} = \left(\frac{1}{1+\mu}\right) cAK^{\alpha} N^{\beta-1} \left(\frac{P_o}{P}\right)^{\alpha+\beta-1}}$$

where c is just a constant term (function of α and β).

Note that there is still a negative relationship between labor demand and the real wage. The only important difference is that movements in the relative price of oil shift the demand curve for labor: higher oil prices lower labor demand.

We could go through and derive the aggregate supply curve, but this would be mathematically demanding. Instead, let's just ask how increases in oil prices shift the supply curve. First, recall that our assumption about labor supply implies that the real wage is determined entirely by labor supply (since labor supply is horizontal curve). Thus, when oil prices rise, labor demand goes down but the real wage stays unchanged. See picture below:

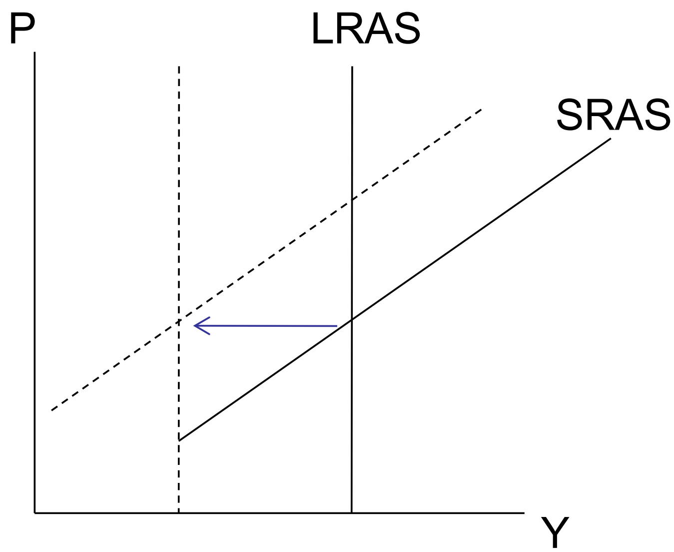


Hence, it is clear that employment must fall. Does this imply that output falls also? Not necessarily since the production function depends on two variable inputs: employment and oil (capital is fixed). But recall the condition regarding the relative demand for oil and labor by firms:

$$O = \frac{NW}{P_o}\frac{(1-\alpha-\beta)}{\beta} = N\left(\frac{W}{P}\right)\left(\frac{P}{P_o}\right)\frac{(1-\alpha-\beta)}{\beta}$$

Note that W/P is unchanged. P/P_o must fall, since P_o is rising. Finally, N is falling by the labor market diagram above. Hence, oil use by firms must also decline. This implies that total production must be falling as well, since both variable inputs – labor and oil – are declining.

Thus, the key implication of allowing for the use of raw materials in production is that when the price of a raw material rises, that shifts aggregate supply back (to the left), as illustrated below:



An increase in oil prices shifts both the long-run and short-run aggregate supply curves back. Not surprisingly, it also raises the natural rate of unemployment (deriving this is a good exercise).