

The Aggregate Production Function

Revised: January 5, 2012

We want to understand why some countries are richer than others, in the sense of having higher GDP per capita. Since rich means they produce more output, the question becomes where the output comes from. Here we describe a tool for answering that question: a *production function* that relates the quantity of output produced to the quantities of inputs and the efficiency or productivity with which they're used. Doing this for an entire economy takes a leap of faith, but the reward is a quantitative summary of the sources of aggregate economic performance.

The production function

Economic organizations transform inputs (factories, office buildings, machines, labor with a variety of skills, intermediate inputs, and so on) into outputs. Boeing, for example, owns factories, hires workers, buys electricity and avionics, and uses them to produce aircraft. American Express's credit card business uses computers, buildings, labor, and small amounts of plastic to produce payment services. Pfizer hires scientists, MBAs, and others to develop, produce, and market drugs. McKinsey takes labor and information technology to produce consulting services.

For an economy as a whole, we might think of all the labor and capital used in the economy as producing real GDP, the total quantity of goods and services. A production function is a mathematical relation between inputs and output that makes this idea concrete:

$$Y = AF(K, L),$$

where Y is output (real GDP), K is the quantity of physical capital (plant and equipment) used in production, L is the quantity of labor, and A is a measure of the productivity of the economy (we call it *total factor productivity*). More on each of these shortly.

The production function tells us how different amounts of capital and labor may be combined to produce output. The critical ingredient here is the function F . Among its properties are

1. More input leads to more output. In economic terms, the marginal products of capital and labor are positive. In mathematical terms, output increases in both K and L :

$$\frac{\partial Y}{\partial K} > 0, \quad \frac{\partial Y}{\partial L} > 0.$$

Consult the Math Review if this seems overly mysterious to you.

2. Diminishing marginal products of capital and labor. Increases in capital and labor lead to increases in output, but they do so at a decreasing rate: the more labor we add, the less additional output we get. You can see this in Figure 1: for a given capital stock \bar{K} , increasing labor by an amount Δ starting from L_1 has a larger effect on output than increasing labor by the same amount starting from $L_2 > L_1$. That is: $AF(\bar{K}, L_1 + \Delta) - AF(\bar{K}, L_1) > AF(\bar{K}, L_2 + \Delta) - AF(\bar{K}, L_2)$. This condition translates into properties of the second derivatives:

$$\frac{\partial^2 Y}{\partial K^2} < 0, \quad \frac{\partial^2 Y}{\partial L^2} < 0.$$

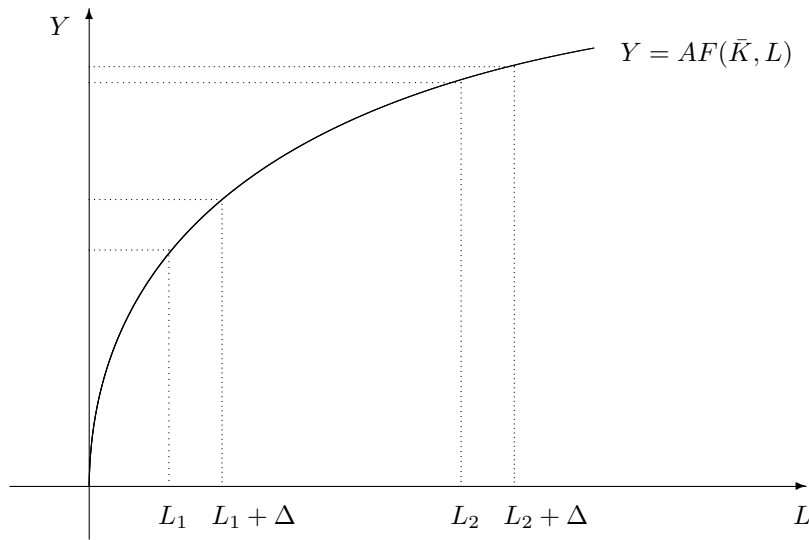


Figure 1: The Production Function.

3. Constant returns to scale. This property says that if we (say) double all the inputs, the output doubles, too. More formally, if we multiply both inputs by the same number $\lambda > 0$, then we multiply output by the same amount:

$$AF(\lambda K, \lambda L) = \lambda AF(K, L).$$

Thus there is no inherent advantage or disadvantage of size.

These properties are more than we need for most purposes, but we mention them because they play a (sometimes hidden) role in the applications that follow.

Our favorite example of a production function is $F(K, L) = K^\alpha L^{1-\alpha}$, which leads to

$$Y = AK^\alpha L^{1-\alpha} \tag{1}$$

for a number (“parameter”) α between zero and one. Circle this equation so you remember it! It’s referred to as the *Cobb-Douglas* version of the production function to commemorate two of the earliest people to use it. (Charles Cobb was a mathematician. Paul Douglas was an economist and later a US senator.) Let’s verify that it satisfies the properties we suggested. First, the marginal products of capital and labor are

$$\begin{aligned}\partial Y/\partial K &= \alpha AK^{\alpha-1}L^{1-\alpha} = \alpha Y/K \\ \partial Y/\partial L &= (1-\alpha)AK^{\alpha}L^{-\alpha} = (1-\alpha)Y/L.\end{aligned}$$

Note that both are positive. Second, the marginal products are both decreasing. We show this by differentiating the first derivatives to get second derivatives:

$$\begin{aligned}\partial^2 Y/\partial K^2 &= \alpha(\alpha-1)AK^{\alpha-2}L^{1-\alpha} \\ \partial^2 Y/\partial L^2 &= -\alpha(1-\alpha)AK^{\alpha}L^{-\alpha-1}.\end{aligned}$$

Note that both are negative. Finally, the function exhibits constant returns to scale. If we multiply both inputs by $\lambda > 0$, the result is

$$A(\lambda K)^{\alpha}(\lambda L)^{1-\alpha} = A\lambda^{\alpha}K^{\alpha}\lambda^{1-\alpha}L^{1-\alpha} = \lambda AK^{\alpha}L^{1-\alpha},$$

as needed. We will typically use $\alpha = 1/3$. If you’d like to know why, see the “Review questions” at the end.

Capital input

The capital input (or capital stock) K is the total quantity of plant and equipment used in production. We value different kinds of capital (machines, office buildings, computers) at their base-year prices, just as we do with real GDP in the National Income and Product Accounts. It’s somewhat heroic to combine so many different kinds of capital into one number, but that’s the kind of people we are.

Fine points:

- How does capital change over time? Typically capital increases with investment (purchased of new plant and equipment) and decreases with depreciation. Mathematically, we might write

$$K_{t+1} = K_t - \delta_t K_t + I_t, \tag{2}$$

where δ_t is the rate of depreciation between t and $t + 1$. On average, the capital stock depreciates about 6% a year, but this is an average of depreciation rates for structures (which depreciate more slowly) and equipment (for example,

computers, which depreciate more quickly). In practice, we use (2) to construct estimates of the capital stock from investment data.

Digression. Note that we've used a different timing convention than financial accountants. Capital at time t is the amount available for production during the period. We use the amount available at the start of the period, which in financial statements would be the end of the previous period. Why do we do this? Because otherwise current production would depend on last period's capital stock, which seems a little strange. Note, too, that for a period like a year, this is a moving target: the amount of capital available in December is likely to be different from the amount available in March. That's not a big deal, because the capital stock is slow to change, so any changes within a period are likely to be small.

- **Quality.** In principle, we want to take into account changes over time in the quality of capital. Computers, for example, are more productive than they were 10 years ago, so a computer today should count as more capital than a computer 10 years ago. Ideally this happens when we construct our real investment series: the national income and product accountants consider changes in quality when they divide investment into price and quantity components. In recent times, the effect of this has been a sharp decrease in the price of investment goods, particularly equipment, so that a given dollar expenditure results in greater additions to capital than in the past.
- **Wars and natural disasters.** Wars can have an impact on the capital stock — natural disasters, too, although they're rarely as big. Experts estimate that the German and Japanese capital stocks declined by about 50% between the start and end of World War II. In modern times, the impact is almost always negligible. September 11 and Hurricane Katrina, for example, had enormous effects on New York City and New Orleans, respectively, but the impact on the US capital stock was tiny in both cases.
- **Does land count?** The short answer: no. In principle maybe it should, but in modern economies land is far less important than plant and equipment. For very poor agricultural economies, land and livestock are important inputs to production, but they're not typically included in our measures of the capital stock.
- **Intangibles.** Capital here consists solely of physical capital. We do not include "investments" in such things as research and development, patents, brands, and databases. These aren't part of traditional measures of capital yet, but there's been some progress on including them.

Labor input

The next component of our production function is labor. The first-order approximation is simply the number of people employed (L), which is a number we can find

for most countries. (It's not as easy as you might think to measure employment, especially in countries with a large informal sector.) In some cases, we also include measures of the quality of labor ("human capital" H) and hours worked (h). If we include both, our measure of labor input becomes hHL .

The starting point for the labor input is, of course, the population. Populations of countries differ not only in quantity, but in their age distribution and its evolution. Right now, for example, China has a relatively young population, but with a low birth rate it is aging rapidly. The US has a younger population than Europe or Japan, the result of a higher birth rate (more young people!) and a higher rate of immigration (immigrants tend to be young, too). These demographic issues are interesting in their own right. They play an important role in government policy — many countries have state-supported pension and health-care systems, for example, so changes in the age distribution can have a significant impact on government budgets. They're also a critical input in product decisions, telling you, for example, whether you should be selling diapers or walkers.

Our focus, however, is on the quantity and quality of labor. There's no question that individuals differ in skill. Derek Jeter's skills earn him \$15m a year as a shortstop for the New York Yankees baseball team, but most of us would be worth far less in the same job. American workers earn more than Mexican and Chinese workers, in part because their skills are better. There are many skills we might want to measure. One that's relatively easy to measure is the level of education of the workforce. In 2000, the average Korean worker had 10.46 years of schooling, and the average Mexican worker had 6.73 years. We know that individuals with more education have higher salaries, on average, so we might guess that Koreans have higher average skills than Mexicans. We call this school-based difference in skill human capital and take it into account by putting it into our production function:

$$Y = AF(K, HL),$$

where H is a measure of human capital.

We won't spend much time on human capital — we don't have time — but there are two common measures we could use, both tied to the number of years of school S of the workforce. The first is to set human capital equal to average years of school:

$$H = S.$$

This seems to be a relatively good approximation for most purposes, but it leads to unreasonably large percentage increases in H at low levels of schooling. For example, workers in India had an average level of schooling of 1.7 years in 1960, so one additional year of school increases H by 59%. Another approach, based on a huge body of evidence, is to credit each year of school with (say) a given percentage increase in skill. Mathematically, we might say

$$H = \exp(\sigma S),$$

where σ is the extra value of a year of school. A good starting point is $\sigma = 0.07$, which means that every year of school increases human capital by 7%.

A second refinement of our measure of labor input focuses on quantity: the number of hours worked. Curiously enough, there are substantial differences in average hours worked across countries. If we use h to represent hours worked, our state-of-the-art modified production function is

$$Y = AF(K, hHL) = AK^\alpha(hHL)^{1-\alpha} \quad (3)$$

We'll generally stick with (1), but will turn to (3) if labor is of particular interest.

Productivity

The letter A in the production function plays a central role in this course — and in the economic performance of countries. We refer to it as total factor productivity or TFP, but what is it? Where does it come from?

The word productivity is commonly used to mean several different things. The most common measure of productivity is the ratio of output to labor input, which we'll call the *average product of labor*. This is typically what government agencies mean when they report productivity data. It differs from the *marginal product of labor* for the same reason that average cost differs from marginal cost. *Total factor productivity* is the letter A in the production function. It measures the overall efficiency of the economy in transforming inputs into outputs.

Mathematically, the three definitions are

$$\begin{aligned} \text{Average Product of Labor} &= Y/L \\ \text{Marginal Product of Labor} &= \partial Y / \partial L \\ \text{Total Factor Productivity} &= Y/F(K, L). \end{aligned}$$

For the Cobb-Douglas production function they are

$$\begin{aligned} \text{Average Product of Labor} &= A(K/L)^\alpha \\ \text{Marginal Product of Labor} &= (1 - \alpha)A(K/L)^\alpha \\ \text{Total Factor Productivity} &= A. \end{aligned}$$

Holding A constant, the first two increase when we increase the ratio of capital to labor. Why? You can be more productive if you have (say) more equipment to work with. TFP is an attempt to measure productivity independently of the amount of capital each worker has. That allows us to tell whether the US is more productive than India because it has more and better capital (higher K) or uses the labor and

capital it has more effectively (higher TFP A). In practice both play a role, and this allows us to tell which effect is larger.

In practice we measure total factor productivity as a residual: We measure A by taking a measure of output (real GDP Y) and comparing it to measures of capital and labor inputs. In the simplest case (without corrections to labor), we solve

$$A = Y/(K^\alpha L^{1-\alpha}).$$

As a result, anything that leads the same inputs to produce more output results in higher TFP. What kinds of things might do this? One is innovation: if we invent the computer chip or a drug that cures cancer, they will clearly increase measured productivity (or one would hope they would). But there are many other examples. One is security: if we establish personal safety and security, then individuals can spend more time working productively, and less time worrying about being robbed or murdered. Another is competition. If the economic system reallocates resources from less productive to more productive firms, that will lead to an increase in country-wide productivity. Capital and labor market laws and regulations play a clear role here. In short, anything that affects the allocation of resources can have an impact on total factor productivity.

Marginal products

In competitive markets, labor and capital are paid their marginal products. We could show that, but for now would prefer to simply take it on faith. That, in turn, tells us where payments to labor and capital come from.

Consider payments to labor. Firms hire workers until the marginal product of an additional unit of labor equals its cost, the wage w . We'll go into this in more detail when we study labor markets, but for now note that this bit of logic can be represented mathematically by

$$w = \text{MPL} \equiv \partial Y / \partial L,$$

where MPL means the marginal product of labor. With our basic Cobb-Douglas production function (1), this becomes

$$w = (1 - \alpha)AK^\alpha L^{-\alpha} = (1 - \alpha)A(K/L)^\alpha.$$

We can now ask ourselves: what do we need to generate high wage rates? The answer: high total factor productivity and/or high capital-labor ratios. In words: workers are more productive, at the margin, if TFP is high and if they have more capital to work with.

Note that high wages are a good thing for an economy: they reflect (for example) high productivity. Often countries with high TFP also have high capital per worker, so

the two terms drive wages in the same direction. It doesn't seem fair, but it happens because the same productivity that makes workers valuable also raises the return on capital, as we see next.

The market return on capital (r , say) equals the marginal product of capital. In this case, there's an additional adjustment for depreciation, so we have

$$r = \text{MPK} = \alpha A (K/L)^{\alpha-1} - \delta.$$

The right-hand side here is the net marginal product of capital — net because we have netted out depreciation. Without that term, we have the gross marginal product of capital, because our measure of output is gross of depreciation (the G in GDP).

In short, the productive value of labor and capital (ie, their marginal products) depends in large part on total factor productivity. To understand this, it's important that you be able to distinguish between total factor productivity (the letter A in the production function) and the marginal products of labor and capital.

Executive summary

1. A production function links output to inputs.
2. Inputs include physical capital (plant and equipment) and labor (possibly adjusted for skill and hours worked).
3. Total Factor Productivity (TFP) is a measure of overall productive efficiency.

Review questions

1. A small country invests a large fraction of GDP in a major infrastructure project, which later turns into a “white elephant” (that is, it's not used). How does this affect the components of the production function?

Answer. The investment will raise the stock of capital K , but since it's not used, we would expect no increase in output Y . We would therefore expect measured productivity to fall.

2. Suppose an economy has the production function

$$Y = AK^{1/4}L^{3/4}.$$

If $Y = 10$, $K = 15$, and $L = 5$, what is total factor productivity A ?

Answer. $A = Y/(K^{1/4}L^{3/4}) = 1.520$.

3. Suppose the production function is

$$Y = 2K^{1/4}L^{3/4}$$

and $K = L = 1$. How much output is produced? If we reduced L by 10%, how much would K need to be increased to produce the same output?

Answer. With $K = L = 1$, $Y = 2$. If L falls to 0.9, $K = 1/0.9^3 = 1.372$ (a 37% increase in K). The reason for the difference between the magnitudes in the changes in K and L is the difference in their exponents in the production function.

4. Worker 1 has 10 years of education, worker 2 has 15. How much more would you expect worker 2 to earn? Why?

Answer. If H = years of education, then one hour of worker 2's time is equivalent to 1.5 (= 15/10) hours of worker 1's time, so we'd expect her to be paid 50% more. A more complex answer is that skill may increase in a more complicated way with years of education, and that types of education may differ in their impact on earning power (an MBA may be worth more in this sense than a PhD in cultural anthropology, however interesting the latter may be).

5. Consider the augmented production function

$$Y = K^{1/3}(HL)^{2/3}.$$

If $K = 10$, $H = 10$, and $L = 5$, what is the average product of labor? How much does the average product increase if H rises to 12?

Answer. Output is $Y = 29.24$ so $Y/L = 5.85$. If H rises to 12, $Y/L = 6.60$.

6. Conditions 2 and 3 seem to contradict each other: one says increases in inputs have a declining impact on output, the other says that proportional increases in capital and labor lead to the same proportional increase in output. What's going on here?

Answer. This is a subtle issue, but the answer is that the conditions are different. Condition 2 concerns increases in one input, *holding constant the other input*. Condition 3 concerns increases in both inputs at the same time. Different concepts, different properties.

7. Why did we set $\alpha = 1/3$?

Answer. If we look at the income side of the National Income and Product Accounts, about two-thirds is paid to labor and one-third to capital. We'll see later that firms will hire labor until its marginal product equals the wage. For our Cobb-Douglas production function,

$$w = \text{MPL} = (1 - \alpha)AK^\alpha L^{-\alpha}.$$

Total payments to labor are the product of the wage and labor:

$$wL = (1 - \alpha)AK^\alpha L^{1-\alpha} = (1 - \alpha)Y.$$

So we set $1 - \alpha = 2/3$, as stated.

If you're looking for more

MBA '11 alum Matthew Cedergren supplies these links to interactive graphs of Cobb-Douglas production functions:

- Manfred Gartner's [eurmacro site](#).
- [Wolfram demonstrations](#).