Lecture 4

Cost and Supply: Production, Supply, and the Demand for Factors

I. Introduction: What this Lecture is About

This lecture deals with the behavior of firms and businesses—the supply side of the market. We begin by asking a fundamental and surprisingly complicated question: What is a firm, and why do firms exist? The central feature of a firm is that managerial decisions serve to organize production and coordinate resources.

We shall assume that the primary goal of the firm is to maximize profits, defined as the difference between revenues and costs. As noted earlier, in economics the cost of any activity is the highest valued alternative action that must be foregone. This leads to a distinction between accounting profits and economic profits, which take into account the full opportunity cost of all the firm’s resources. The prospect of earning economic profits determines how resources are allocated among the various goods and services that might be produced in a modern economy. Paradoxically, we will find that in competitive markets the prospect of earning positive economic profits causes them to be eliminated. Under competition, the tendency is always toward a zero profit equilibrium.

Before getting on with the details of firms’ decision making, it is worth pondering two preliminary questions: (1) What is a firm?; and (2) What is the goal of firms’ decisions?

II. The Nature of Firms

The Boundaries of the Firm

Production occurs when inputs such as raw materials, labor and machines are transformed into something useful.

Of course not all production occurs in traditional business firms. Governments hire teachers to provide educational services and soldiers to provide defense. Families engage
in “home production” by preparing meals, cleaning, teaching children and so on. Economists view all these activities as productive because they yield things that people want. But for present purposes we will be concerned with production that takes place within business firms. Broadly speaking, a firm is an organization of one or more people devoted to producing a good or service. Within firms, resources are allocated by managerial decisions. For example, within an automobile manufacturing company like General Motors or Toyota workers don’t simply choose which tasks to perform and when they will do them. Rather, they are assigned to particular tasks by managers, who allocate resources and coordinate the productive activities of thousands of workers.

The fact that management decisions allocate resources within a firm may seem like a pretty obvious point. But why are there firms at all? At a firm like GM, many activities that could be managed within the firm are instead accomplished by arms-length market transactions. GM could produce its own tires, but instead chooses to buy them on a market. On the other hand, GM could buy spark plugs, car bodies and engines from separate producers, but instead GM circumvents these possible market transactions and produces these parts itself. Conceivably, every activity in GM might be accomplished by market transactions: car designs could be purchased from independent designers; the assembly line worker who installs the steering wheel could buy a partially assembled car from the previous worker on the line, and then sell it to the next person, who might install the radio. If all of these thousands of activities were accomplished by separate market transactions, there would be no separate management function, and thus no firm. All would be accomplished by markets and prices.

In a market economy the coordination and allocation of resources is accomplished by a combination of managerial decisions, taken within firms, and market transactions. Then a firm is a place where potential market transactions are instead accomplished by managerial decisions. Why is managerial authority sometimes substituted for the market? That is, why are there firms? Ronald Coase (a Chicago guy) won part of his Nobel Prize for pondering this question. There are two good reasons.

The first is that it is often cheaper to coordinate resources by managerial authority. The workers at GM could engage in a set of separate transactions as each partially-assembled car rolled down the assembly line. Separate contracts would have to be negotiated for each of these transactions. These contracts must include the negotiated price, conditions for delivery, characteristics of the product, and legally enforceable penalties in case the contract is breached. The costs of all these separate contracts would be enormous. In this case it is cheaper to delegate authority to managers, who direct workers and divisions of the firm in what to do, and then pay them for doing it. So one reason for the existence of firms is that there are high transactions costs of using markets.

Example 1: Gun Manufacturing in the 19th Century
Small arms manufacturing was an important industry in the English city of Birmingham in the 1860s. The process was organized by a master gunmaker, who owned a warehouse but did no manufacturing himself. Instead he acquired parts from specialists, who made gun barrels, sights, triggers, and so on. He then
contracted out the assembly to a long succession of “setters up”, each of whom specialized in particular step in the assembly of a gun: barrel and lock assemblers, breach assemblers, barrel strippers, hardeners, borers, and lock-freers—who performed final adjustments to make the parts work together. Each craftsman acted as an individual “firm.”

One of the drawbacks of this system is that the parts produced by separate craftsmen were not easily conformable, so it took a lot of work to make them fit. Eli Whitney revolutionized the assembly of guns in America by bringing all these functions within one firm, where he standardized the manufacture of parts that were meant to fit together. Whitney avoided the high transactions costs of obtaining conformable parts by overseeing the production of these parts within his own firm.

**Example 2: GM and Fisher Body**

In the early 20th century the exclusive supplier of auto bodies to General Motors was a company called Fisher Body. Since car bodies are large and heavy, it was efficient for Fisher to locate its factories near GM’s assembly plants. But because doing so would enhance GM’s bargaining power in subsequent negotiations—once a factory was built GM could offer a low price for Fisher’s bodies and Fisher would have no alternative use for the specialized factory. Fisher wasn’t stupid. Knowing that GM would do this, Fisher refused to locate efficiently—the costs of reaching an efficient agreement between the two firms were too high. How did they get around the impasse? GM bought Fisher and produced its own bodies in a specialized plant, replacing market transactions with internal decisions.

A second reason for the existence of firms is team production. Consider the task of moving a grand piano, which may require the efforts of several people to lift it. Since it is nearly impossible to distinguish the separate contributions of each person to the lifting of the piano, it is similarly hard to write an enforceable contract specifying what each person’s product (effort) should be. As above, this means that it is costly to write contracts that specify payment for a deliverable product from each person. Further, since effort is both difficult to measure and unpleasant for each person to provide, each has an incentive to shirk by reducing his own efforts somewhat—it is easy to look like one is working hard without actually doing so. Again, the most effective means of production in this case may be to delegate authority to a manager, who monitors the efforts of individual workers and directs them in which tasks to perform.

**Example 3: Monitoring Effort on the Yangtze**

Long ago, a Westerner was traveling up the Yangtze by boat. At a narrowing of the river where the water ran fast, it was impossible for boats to proceed under their own power. They were pulled upstream by teams of coolies, who trudged along the shore with harnesses around their necks. An overseer beat the coolies to elicit more effort from them.
The Westerner was horrified by the cruel treatment of the coolies. But then he learned that the coolies owned the rights to pull boats over this stretch of water. Because each coolie had an incentive to shirk on the team effort of pulling boats, they had hired the overseer to monitor their efforts and provide “incentives” to work hard. [Related by S.N.S. Cheung in “The Costs of Alternative Economic Organizations,” Canadian Journal of Economics, 8/75].

**The Goals of the Firm**

When we analyzed individual behavior in Lecture 3, we assumed that consumers make decisions with a particular goal in mind: maximizing happiness, or “utility.” To build a theory of how firms make decisions, we also assume that firms pursue a goal. But what goal shall we assume? As firms are owned by people who are interested in being happy, it would seem logical to assume that firms are run so as to maximize the utility of owners. Yet we will assume that the firm’s goal is to maximize profit, the difference between its revenues and its costs. Why? By maximizing profits the owners of the firm also maximize their utilities.

Look at Figure 4.1a, which shows Robinson alone on his desert island, producing and consuming fish and coconuts along his production possibilities frontier. We recognize by now that the “best” use of Robinson’s efforts is to produce and consume at point α, where his marginal value of a fish is equal to the marginal cost of obtaining one. Robinson maximizes his utility by producing \( F_\alpha \) fish and \( C_\alpha \) coconuts.

![Figure 4.1a](image)

**Lonely Robinson**

One day Robinson awakes and, miraculously, finds himself in the middle of an archipelago inhabited by other Robinsons, one on each island. Using canoes, they
establish a fish and coconut market on a central island, which results in a market price for fish of $p_F$ coconuts per fish. There are so many participants that Robinson is a price taker—he can buy or sell as many fish as he wants at price $p_F$. The situation is shown in Figure 4.1b.

One result of the emergence of a market is to open new consumption opportunities for Robinson. Suppose Robinson continues to produce the combination $\alpha = (F_\alpha, C_\alpha)$. Using his $C_\alpha$ coconuts, he can purchase fish by moving to the southeast along the budget line $B_\alpha$, which has slope $-p_F$. Or he could sell fish for coconuts and move to the northwest along $B_\alpha$. As we have drawn things, Robinson prefers to consume more fish than he produces, consuming at point $\beta = (F_\beta, C_\beta)$. He consumes $F_\beta - F_\alpha$ more fish than he catches, and pays for them by consuming $C_\alpha - C_\beta$ fewer coconuts than he harvests. He exports (sells) coconuts and imports (buys) fish from other islands, and he is better off than when there were no opportunities to trade.

But Robinson isn’t done. He didn’t have to produce combination $\alpha = (F_\alpha, C_\alpha)$, and Robinson notices that his budget constraint would offer even greater consumption opportunities if he produced more coconuts and fewer fish. Indeed, by producing the combination $\gamma = (F_\gamma, C_\gamma)$ his budget constraint, $B_\gamma$, is farther from the origin than for any other possible combination on his PPF. In other words, Robinson’s income is maximized when he produces combination $\gamma$, which then allows him to maximize utility by consuming combination $\delta$. Why does $\gamma$ maximize Robinson’s income? At point $\alpha$, Robinson’s marginal cost of a fish is the slope of his PPF at that point. The price of fish, $p_F$, is the slope of budget constraint $B_\alpha$. As the slope of the PPF is steeper at that point,
we know that \( P_F < MC_F \) --Robinson’s cost of producing fish is greater than the price of acquiring them on the market. His income will rise if he produces fewer fish and more coconuts. His income is maximized at point \( \gamma = (F_F, C_G) \), where the marginal cost of a fish (the slope of the PPF) is equal to the price of a fish (the slope of \( B_f \)). He maximizes his income by producing where \( p_F = MC_F \). And by maximizing his income (maximizing his profit) he also maximizes his consumption opportunities.

Notice that Robinson’s profit-maximizing decision to produce combination \( \gamma \) is completely independent of his preferences. We drew the picture so that Robinson, having produced at \( \gamma \), uses his coconuts to buy more fish, ultimately consuming combination \( \delta \). If Robinson’s preferences had been different—say he just loves coconuts—he might have consumed fewer fish and more coconuts than he produces at \( \gamma \). The shape of his indifference curves (his tastes) doesn’t matter: To maximize his utility, Robinson should first maximize his profits.

This fact underlies our assumption that firms in a market economy desire to maximize profits. (Whether they do so is a more complicated question, and a matter for another class). By doing so, they maximize the income of the firm’s owners, which is the owners’ first step to maximizing utility. So firms’ owners do maximize utility, but the best way to do that is maximize profits. The particular tastes of the owners do not affect the way the firm is run.

**Costs, Profits, and other Misunderstood Things**

We have already emphasized that all costs are opportunity costs—the highest valued foregone alternative for the resources in question. So when we think of a firm’s “costs” we have to include all the foregone alternatives for the firm’s resources. This means that economic costs and profit have different meanings than accounting costs and profits. My friend Red runs a hot dog stand (really). If Red’s Hot Dog Stand takes in revenues of $200,000 per year, with costs of wieners, mustard, hired labor and so on of $150,000, then Red’s accounting profit would be $50,000. But if Red’s next best use of his time is to paint houses for $50,000 per year, that $50,000 is also a cost of running the hot dog stand. Red’s economic profits are zero—he and all the other resources of the firm are earning exactly as much selling hot dogs as they would earn in their best alternative activities.

**Definition: Economic Profit**

Economic Profit is the difference between a firm’s revenues and the opportunity costs of the resources used by the firm.

More generally, in order for a firm to attract and retain productive resources, owners of those resources must earn an amount that is at least equal to what they could earn elsewhere. This is obvious for inputs like workers, who must be paid as much as what other employers are willing to pay for their services, or raw materials, which must be purchased at a market price that (by definition) others are willing to pay. But some
inputs, like Red or a firm’s investors, earn their compensation from an ownership claim on the accounting profits of the firm. If investors can earn a 10 percent annual return by putting their money in a mutual fund, then the accounting profits of a firm hoping to attract investors must be large enough to yield an equivalent risk-adjusted return on investment. The level of accounting profits necessary to pay the firm’s owners what they could earn elsewhere is sometimes called normal profits.

Now suppose that Red discovers a new and better way of making great hot dogs—people are willing to pay more for his dogs than for others’—and his revenues rise to $225,000 per year. But Red did not become a better house painter, so he can still get $50,000 from his next best alternative. Now Red earns an economic profit of $25,000 per year—he gets $25,000 more per year making hot dogs than he could earn in his next best alternative.

How we think of that profit depends on how other people can respond to it. If others cannot copy what Red has done, so Red continues to earn more from making dogs than from his best alternative, we will say that Red earns economic rents from making hot dogs. An economic rent is a payment to a factor (here, Red) that is greater than the amount necessary to bring the resource into its current use (here, making dogs). So a resource that earns economic profits is also earning rents. [So Lebron James earns lots of rents from playing basketball: his basketball salary from the Miami Heat is about $15 million, but one would guess that his alternatives are less attractive.] If others can imitate Red’s methods, however, his ability to charge higher prices for his dogs will soon dissipate, and Red’s profits will eventually disappear. Then we refer to Red’s profits as quasi rents—they don’t affect the use of Red’s time, but they are competed away as time passes.

As we discussed in Lecture 2, economic profits and losses are both a signal and an incentive to resource owners to change the allocation of resources. Profits signal that buyers are willing to pay more than the cost of production in order to obtain a good. They are also the incentive that attracts new resources, which causes an industry to expand and profits to fall. Losses have the opposite effect. These forces clearly tend toward zero economic profit as the “normal” state of affairs, so long as resources are mobile among alternative uses.

**A Taxonomy of Costs**

It is useful to categorize costs as avoidable or non-avoidable. **Avoidable** costs are costs that are, well, avoidable. The hot dog business is slow in August, so Red often paints houses instead. He doesn’t have to pay for a month’s worth of wieners, mustard, or buns. Those costs are avoidable or, some would say, variable because they vary with the amount of hot dogs that Red sells. If Red leases his grilling equipment on an annual basis, however, those costs are **unavoidable**, or fixed, because they don’t depend on how many hot dogs he sells. He must pay them whether he sells 10,000 hot dogs or none at all. Of course if Red exits the hot dog business he doesn’t have to renew his lease on the
equipment, so these costs that are fixed for one period of time (the term of the lease) are variable over a longer period—Red can avoid them.

Last, suppose Red contracted to have his stand built in the shape of a giant hot dog with an expensive, smiling statue of him on top. The stand has no other use—much like Fisher’s body plant above—and we say that the costs of building the stand are sunk or non-salvageable because no one will pay anything for them. In contrast, salvageable costs are those that can be recovered, say by selling the asset in question.

Economically rational decision making requires that the only costs that should matter for deciding on some action are those that can be avoided if the action is not undertaken. In deciding whether to paint houses in August, Red compares his revenues to the avoidable costs of selling dogs. The lease payment on the grilling equipment is fixed—he has to pay it anyway—so it doesn’t matter to his calculation of whether to open the stand for the month. Similarly, in deciding whether to stay in business for another year, the non-salvageable costs of the stand are irrelevant to that decision.

Costs and Decisions: In deciding on some course of action A, the only costs that matter to the decision are those that can be avoided if A is not undertaken.

The Long Run and the Short Run

An input or factor of production is something that is used by the firm to produce output. So professors’ time and chalk are inputs to the production of educational services, as are buildings, energy, chairs and so on.

Suppose that Booth wanted to increase its output of MBAs. With current faculty and facilities, we could increase teaching loads, fit a few more bodies in classrooms, and schedule classes at night. A substantial increase in output would be difficult because it is very costly to quickly expand the resources that Boot uses to educate students. To reflect this type of situation, the term short run is meant to define a time period during which it is very costly to change the quantities of some inputs. To keep things simple, we can think of the short run as a period of time in which some of the firm’s inputs are fixed, that is, they can’t be varied at all.

Definition: Short Run and Long Run

The short run is a period of time during which some factors of production are fixed.

The long run is a period of time during which all factors of production can be varied.

We typically think of “fixed” factors as representing durable inputs like plant, equipment and machinery, which are gathered under the term capital.

Definition: Capital

Capital refers to durable assets that yield a stream of services over time.
Thus a machine, a desk or a house or building are forms of capital. A robot used to assemble automobiles yields productive services this year and, subject to some depreciation, it will yield services next year and the year after. The salient feature of capital for our discussion in this Lecture is that it is typically more costly to quickly vary the amount of capital employed in production than to vary the amount of labor or raw materials—it’s hard to build a factory in a week. Thus capital is often assumed to be fixed in the short run. At the other extreme, inputs like materials and supplies, energy, and (often) labor are commonly characterized as “variable” factors because they can be obtained quickly by paying a market price. Even labor, however, often has attributes of a fixed factor. For example, in many countries regulations make it costly to terminate workers, which also make a forward-looking firm reluctant to hire them in the first place. More generally, in order to expand, a firm must often train workers in specialized techniques. These investments in human capital may take a long time to reach fruition, so in some cases a firm’s labor force may be considered “fixed” in the short run.

Like many other concepts in economics it is important to understand that “short run” and “long run” are abstractions that facilitate thinking about certain problems. Inputs are rarely fixed in a literal sense. But it is often costly to quickly obtain additional amounts of certain inputs, like plant and equipment or trained labor, so we think of them as fixed during some loosely-defined “short” period. There might be many “short runs” corresponding to the costs of adjusting various inputs. In the shortest of short runs, capital and the number of trained workers might be fixed, so greater output is achieved by working overtime and using existing equipment more intensively. Over a longer period, the firm might train more workers and add a second shift to work on existing equipment. And over the long run, when all factors are variable, the firm can add both new capital and new labor to its productive capacity. The point is that some types of decisions can be implemented quickly, while others take longer.

III. Costs and Output

Firms pay for the inputs they use, and these payments to factors constitute the firm’s costs. In turn, the usage of the inputs produces output. So the relationship between costs and output depends directly on the relationship between inputs and output. We consider the relationship between inputs and output, and the implied relation between costs and output. We will consider these issues in the context of a particular numerical example (Bob’s farm), and then generalize.

Given a technology for producing output, the quantifiable relationship between inputs and output is called a production function. More precisely:

**Definition: Production Function**

For any combination of inputs \( x = (x_1, x_2, \ldots, x_n) \), a firm’s production function tells us the maximum amount of output, \( Q \), that the firm can produce with those inputs. We write the production function as:

\[
Q = F(x)
\]
To illustrate the relationships between inputs, output and costs, consider Bob’s farm. Bob uses his own managerial skills, hired labor, land, and machinery to produce wheat. With the combination of inputs labeled ‘B’ in Table 4.1, the largest amount of wheat that Bob can produce is 10,000 bushels per year. Now suppose we replicate the combination of inputs used in B, doubling the amount of every input (including a clone of Bob). How much output will be produced? It stands to reason that if the combination in B is exactly replicated with the same technology, then output must double to 20,000 bushels. We made an exact replica of the original farm, so it must produce the same amount as shown by combination C in the table. Following this logic, adding proportionally larger amounts of all inputs must yield equal proportional increases in output. This is illustrated by combinations D and E in the table.
Table 4.1
Production and Costs with Constant Returns to Scale: Bob’s Farm

Inputs
(rental prices per unit in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Land ($400)</th>
<th>Workers ($10000)</th>
<th>Tractors ($2000)</th>
<th>Farmers ($20000)</th>
<th>Output Q</th>
<th>Total Cost C(Q)</th>
<th>Average Cost AC(Q)= C(Q)/Q</th>
<th>Marginal Cost MC(Q)= ΔC(Q)/ΔQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>10,000</td>
<td>100,000</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>40</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>20,000</td>
<td>200,000</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>60</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td>30,000</td>
<td>300,000</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>E</td>
<td>80</td>
<td>28</td>
<td>4</td>
<td>4</td>
<td>40,000</td>
<td>400,000</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Constant Returns to Scale

The input-output combinations in Table 4.1 are an example of what economists call constant returns to scale in production. This term means that a given percentage increase in the utilization of all inputs must yield an equal percentage increase in output. So doubling all inputs doubles output. Here is a formal statement:

**Definition: Constant Returns to Scale**

A production function $Q = F(x)$ exhibits constant returns to scale (CRS) if a given proportional increase in all inputs yields the same proportional increase in output. Formally, this says that:

$$F(kx_1, kx_2, ..., kx_n) = kF(x_1, x_2, ..., x_n)$$

**Example: The Cobb-Douglas Production Function**

In the 1930’s a Chicago economist named Paul Douglas (who later became a U.S. Senator) wanted to find a production function that would “fit” data on US national output at date $t$ ($Q_t$), the labor force ($L_t$), and measures of the capital stock ($K_t$). He wanted the function to have constant returns to scale, because he noticed that areas with larger amounts of inputs had proportionally larger amounts of output. But he also wanted to allow for factors to become more productive over time—a source of economic growth. So he visited his friend Professor Cobb in the math department, and they came up with this:

$$Q_t = A_t L_t^\alpha K_t^{1-\alpha}$$

which came to be known as the **Cobb-Douglas Production Function**. This function has constant returns:
so a 10 percent increase in all inputs \((k=1.1)\) raises output by 10 percent. The term \(A\) is called total factor productivity because an increase in \(A\) makes all factors more productive—it captures technical progress that allows us to get more output from given inputs.

Despite its bare-bones simplicity, it turns out that the Cobb-Douglas production function does a remarkably good job of fitting data on national output and the contributions of labor, capital, and technical progress to economic growth. We will return to it later, when we use it to compare sources of economic growth across countries.

Returning to our discussion of Bob’s farm in Table 4.1, we concluded that a literal doubling of all inputs must double output. So our conclusion is that when an exact replication of the activities of a firm is possible, the firm’s production function must have constant returns to scale. Whether exact replication is actually possible is a key issue, to which we return in a moment. But first we show the implications of constant returns for the firm’s costs.

Bob’s total cost of production for any rate of output, \(C(Q)\), is derived by multiplying each input quantity by the price of the input, and adding the components. So the total cost of producing 10,000 bushels of wheat is $100,000. Since there are constant returns, a doubling of inputs (including Bob) doubles output, so it should be obvious that the total cost of producing 20,000 bushels is $200,000, and so on. So with CRS, total costs rise proportionally with output, as in Figure 4.2a.

Table 4.1 also shows two measures of cost per unit of output. The first is average cost (AC), which is simply total cost divided by output:
At $Q=10,000$ bushels average cost is $100,000/(10,000 \text{ bushels})=\$10/\text{bushel}$. With CRS the numerator and denominator of $AC$ change by the same proportional amount when output rises, so average cost is a constant, independent of the rate of output.

The second measure of unit cost is marginal cost (MC) which we already know is the cost of producing an additional unit. A more formal definition is:

**Definition: Marginal Cost**

For a small increase in output, $\Delta Q > 0$, marginal cost at output rate $Q$ is defined as

$$MC(Q) = \frac{\Delta C(Q)}{\Delta Q}$$

The last column of Table 4.1 shows calculated values of marginal cost for our CRS farming technology. At all levels of output $Q$, $MC(Q)=\$10$, which is again an implication of CRS. A unit change in the rate of output always causes total cost to change by the same fixed amount, no matter what the original level of output may be. So with constant returns to scale, marginal cost is constant and equal to average cost. This result is shown in Figure 4.2b.

**Figure 4.2b**

*Average Cost and Marginal Cost With CRS*

$C(Q) = \lambda Q$

$AC = MC = \lambda$
Diminishing Returns and the Firm’s Choices

If literally all inputs can be varied proportionally, we have concluded that the production function must exhibit constant returns to scale, and that marginal cost (and average cost) is flat. Does this mean that all production is CRS, with flat average and marginal cost curves? No.

The key point is that, even in the long run, a firm that wants to double its output cannot double each and every one of its inputs. Think again of Bob who manages the operation of his farm, deciding how and when all his inputs will be used. In the long run he can acquire twice the land, hire twice as much labor, and rent another tractor. But he can’t replicate himself, and his managerial talent is an input to the production process. Because there is only one entrepreneur to manage production, at least one of the inputs to the production process—call it “managerial capacity”—is fixed.

It is easiest to see what happens in this case if we allow only one input to vary, and hold the others fixed. So suppose that in the short run all of Bob’s factors of production are fixed, except labor. Specifically suppose that Bob’s “fixed” inputs are 20 acres of land, one tractor, and Bob. Then the only way to produce more wheat is by hiring more workers. So we can write Bob’s production function as

\[ Q = F(L) \]

where \( L \) is the amount of labor employed, and it is understood that other inputs are used in production but cannot be varied (the are subsumed in the function \( F(.) \)). We continue to assume that Bob is a price taker in the market for the inputs he uses, and for the output he sells. Assume that the annual rental price for land is $400/acre and that tractors rent for $2000/year. Bob’s own opportunity cost of working in farming is $20,000 per year, so his annual fixed cost is

\[ C_0 = 2000 \times (1 \text{ tractor}) + 400 \times (20 \text{ acres}) + 20000 \times (1 \text{ Bob}) = 30000 \]

Workers can be hired for annual wage \( W \), so Bob’s variable costs of producing wheat are simply \( WL \). Each unit of output can be sold for price \( P \), so Bob’s annual profits from running the farm with \( L \) workers are:

\[ \Pi(L) = PF(L) - WL - C_0 \]

Bob’s goal is to choose \( L \) so as to make \( \Pi \) as big as possible.

Some hypothetical production and cost data for Bob’s farm are recorded in Table 4.2. We can use these data to demonstrate most of what we need to know about production, cost, and profit maximization. Columns (1) and (2) of the table show rates of wheat production (\( Q \)) when various quantities of labor (\( L \)) are hired, holding other inputs fixed at the levels shown at the top of the table. We assume that output cannot be
Table 4.2
Short Run Production, Costs, and Profits for Bob’s Farm

Other inputs are 20 acres of land, 1 tractor, and Bob
Rental Price of Land = $400/acre, Rental Price of Tractor = $2000
Bob’s Opportunity Cost = $20000, Annual Wage of Workers = $10000
Price of Output = $15/bushel

Fixed Cost is $C_0 = 400 \times 20 + 2000 \times 1 + 20000 \times 1 = 30000$

<table>
<thead>
<tr>
<th>(1) Labor</th>
<th>(2) Output</th>
<th>(3) Marginal Product of Labor</th>
<th>(4) Average Product of Labor</th>
<th>(5) Value of Marginal Product of Labor</th>
<th>(6) Variable Cost (V(Q))</th>
<th>(7) Total Cost (TC)</th>
<th>(8) Marginal Cost (MC)</th>
<th>(9) Average Variable Cost (AVC)</th>
<th>(10) Average Total Cost (ATC)</th>
<th>(11) Revenue (R)</th>
<th>(12) Profit (Π)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>30000</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0</td>
<td>-30000</td>
</tr>
<tr>
<td>1</td>
<td>500</td>
<td>500</td>
<td>7500</td>
<td>10000</td>
<td>40000</td>
<td>20.00</td>
<td>20.00</td>
<td>80</td>
<td>7500</td>
<td>-32500</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1800</td>
<td>900</td>
<td>19500</td>
<td>20000</td>
<td>50000</td>
<td>7.69</td>
<td>11.11</td>
<td>16.67</td>
<td>54000</td>
<td>-23000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3600</td>
<td>1200</td>
<td>27000</td>
<td>30000</td>
<td>60000</td>
<td>5.56</td>
<td>8.33</td>
<td>16.67</td>
<td>54000</td>
<td>-6000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5600</td>
<td>1400</td>
<td>30000</td>
<td>40000</td>
<td>70000</td>
<td>5.00</td>
<td>7.14</td>
<td>12.50</td>
<td>84000</td>
<td>14000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7600</td>
<td>1520</td>
<td>30000</td>
<td>50000</td>
<td>80000</td>
<td>5.00</td>
<td>6.58</td>
<td>10.53</td>
<td>114000</td>
<td>34000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9600</td>
<td>1500</td>
<td>21000</td>
<td>60000</td>
<td>90000</td>
<td>7.14</td>
<td>6.67</td>
<td>10.00</td>
<td>135000</td>
<td>45000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>10000</td>
<td>1429</td>
<td>15000</td>
<td>70000</td>
<td>100000</td>
<td>10.00</td>
<td>7.00</td>
<td>10.00</td>
<td>150000</td>
<td>50000</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10833</td>
<td>1354</td>
<td>12495</td>
<td>80000</td>
<td>110000</td>
<td>12.00</td>
<td>7.38</td>
<td>10.15</td>
<td>162495</td>
<td>52495</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>11500</td>
<td>667</td>
<td>1278</td>
<td>10000</td>
<td>90000</td>
<td>15.00</td>
<td>7.83</td>
<td>10.44</td>
<td>172495</td>
<td>52495</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>12050</td>
<td>550</td>
<td>1205</td>
<td>8250</td>
<td>100000</td>
<td>18.18</td>
<td>8.30</td>
<td>10.79</td>
<td>180745</td>
<td>50745</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>12500</td>
<td>1136</td>
<td>6750</td>
<td>110000</td>
<td>140000</td>
<td>22.22</td>
<td>8.80</td>
<td>11.20</td>
<td>187495</td>
<td>47495</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12900</td>
<td>1075</td>
<td>6000</td>
<td>120000</td>
<td>150000</td>
<td>25.00</td>
<td>9.30</td>
<td>11.63</td>
<td>193495</td>
<td>43495</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13250</td>
<td>1019</td>
<td>5250</td>
<td>130000</td>
<td>160000</td>
<td>28.57</td>
<td>9.81</td>
<td>12.08</td>
<td>198745</td>
<td>38745</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>13550</td>
<td>968</td>
<td>4500</td>
<td>140000</td>
<td>170000</td>
<td>33.33</td>
<td>10.33</td>
<td>12.55</td>
<td>203245</td>
<td>33245</td>
<td></td>
</tr>
</tbody>
</table>
produced without workers, so total output is zero in the first row. One worker, when combined with the other inputs, can produce 500 bushels per year. In general, we expect that adding more workers will raise output. The question is, how much does each additional worker add to the total output of the farm?

Suppose employment is doubled from one worker to two. It is easy to imagine that two identical workers could produce more than twice the output of one. A person working alone must both drive the tractor and load the harvest, each as a separate task. With two workers, one can specialize in driving the tractor while the other loads. It is possible, even plausible, that addition of a second worker more than doubles output, to 1800 bushels, as shown in column (3) of the table. The marginal product of a second worker is the change in total output when a second worker is added, or 1800-500 = 1300 bushels. More generally:

**Definition: Marginal Product**

The marginal product of input $x_i$ is the additional output produced by using an additional unit of $x_i$, holding constant the use of all other inputs. Formally,

$$MP_i = \frac{\Delta Q}{\Delta x_i} = \frac{\Delta F(x_1, x_2, \ldots, x_n)}{\Delta x_i}$$

We refer to the marginal product of additional workers as the marginal product of labor, $MP_L$. In our example, specialization and cooperation among workers makes the marginal product of a second worker (1300) larger than the marginal product of the first (500), as shown in column (3). These effects continue as we add a third worker with marginal product of 1800 bushels and a fourth with marginal product of 2000 bushels.

With fixed quantities of land and other inputs, the marginal product of labor cannot continue to increase forever. (If it did, all the farmworkers in the world should work on one acre of land, which would feed everyone). Only one person is needed to drive the tractor, and with a limited number of tasks to be performed additional workers will be assigned to successively less productive tasks. The marginal product of labor must start to decline, which in this example occurs with the fifth worker. After that, additional workers add successively smaller amounts to total output. It is even possible that an additional worker just gets in the way, reducing total output, though we do not go that far in the table.

The tendency for the marginal product of an input to decline as larger quantities of the input are used was first noted by David Ricardo, an English economist, in the 19th century. Ricardo observed that the supply of agricultural land in England is essentially fixed, and he postulated that additional units of inputs used with land—labor and capital in his analysis—would yield smaller and smaller additions to agricultural output. This observation is important enough that it is termed a **law**:
The Law of Diminishing Returns: Holding constant the quantities of all other inputs, the marginal product of successively larger amounts of an input will eventually decline.

Figures 4.3a and 4.3b graph the total and marginal products of labor, using the data from Table 4.2. Graphically, the marginal product of labor in Figure 4.3b (column (3) of Table 4.2) is simply the slope of the production function shown in Figure 4.3a. The range of diminishing returns is that part of the marginal product of labor curve where $MP_L$ slopes down or, equivalently where the slope of $F(L)$ in 4.3a is diminishing with $L$. 

![Figure 4.3a](Production Function on Bob's Farm)
Figure 4.3b also graphs the average product of labor, $AP_L$—calculated in column (4) of Table 4.2—defined as total output divided by the number of workers employed. More generally,

**Definition: Average Product**

The average product of input $x_i$ is total product divided by the number of units of the input employed:

$$AP_i = \frac{Q}{x_i} = \frac{F(x_1, x_2, ..., x_n)}{x_i}$$

Notice in Figure 4.3b that $MP_L$ exceeds $AP_L$ when $AP_L$ is rising, and $MP_L$ is below $AP_L$ when $AP_L$ is falling. Why is this? Think of your test scores in a class. If your average on 3 tests is 90, and on the 4th test you score an 85, your average will fall. The marginal score (85) pulls the average down because the marginal is less than the average. But if you had scored 95 your average would rise. More generally, if the average value of some variable $X$ is rising, the marginal must be above it, pulling it up. If the average is falling, the marginal must be below it, pulling it down. This implies that $MP_L$ must equal $AP_L$ at the maximum of $AP_L$. That is exactly what Figure 4.3b shows for the relationship between average and marginal productivity—when average product is falling, marginal product is below it.

**Example: Productivity and Wage Growth in the United States**
We will find that the concept of marginal productivity is an important tool for characterizing a firm’s decisions about how much of an input to use and how much of output to produce. Average productivity is less useful for these purposes, but it is much easier for economists, statisticians and government agencies to measure—to compute the average product of labor all we need to know is total output divided by the number of workers and hours used to produce it. Indeed, the average product of labor is the concept used in calculating productivity statistics that are reported for developed economies. So when you read that “productivity in the US grew at an annual rate of 5.1 percent in the third quarter of 2003,” the statement refers to the average product of labor in the United States.

What causes measured (labor) productivity to grow? Take the simple example of a Cobb-Douglas production function introduced above, \( Q_t = A_t L_t^\alpha K_t^{1-\alpha} \). The average product of labor simply divides both sides by \( L \), the amount of labor employed:

\[
AP_L = \frac{Q}{L} = A \left( \frac{K}{L} \right)^{1-\alpha}
\]

Average productivity increases when, ‘\( A \)—total factor productivity—rises. That simply means that the state of technical knowledge has grown, so we can get more output from given inputs. The average product of labor also rises when the capital-labor ratio, \( \frac{K}{L} \), increases. Labor is more productive when it has more “capital” (machines, computers, buildings and so on) to work with. That is, in the Cobb-Douglas production function adding more capital raises the average productivity of labor. More generally, over time in a market economy we expect improvements in the state of technical knowledge (\( A \)) and greater accumulation of factors of production that “work with” labor, so that productivity will typically grow over time. In fact, we will see later that these forces are the main sources of economic growth and rising living standards in both developed and developing countries.
To illustrate these forces at work, I went to The Economic Report of the President, available on the Web\(^1\), which reports hundreds of statistics on the performance of the US economy. Using data on output, total hours worked, and compensation of employees, Figure 4.4 shows the correspondence between the growth rates of hourly productivity and wages in the US over since 1964.\(^2\) Over an extended period, firms can’t pay workers more than what they produce, and firms that try to pay less will see workers bid away by other employers. (Things are a bit more complicated than this, but not much). Thus we expect that economy-wide increases in labor productivity will parallel changes in average earnings and benefits of American workers. The Figure shows that in the early 1960’s productivity grew rapidly, reflecting rising total factor productivity and the accumulation of capital. Wages also grew rapidly during this period. But productivity growth slowed through the early 1980s—a period widely referred to as the “productivity slowdown”—and the rate of wage growth slowed as well. After 1983 productivity and wages began to grow more rapidly, especially in the productivity “boom” of the late 1990s. The figure shows that wage growth closely follows productivity growth over long periods: when productivity growth slows, so does wage growth, and conversely. In short, productivity improvements find their way to workers’ pockets, and are the driving force behind rising wages and living standards.

\(^1\) [http://www.gpoaccess.gov/eop/tables10.html](http://www.gpoaccess.gov/eop/tables10.html)

\(^2\) To smooth things out a bit, the figure shows five-year moving average of growth rates. So, for example, the recorded rate for 1964 is the average of the rates of annual growth for 1960-1964.
IV. The Demand for a Factor of Production: How Many Workers Should be Hired?

When an additional worker is hired, she produces additional output that we called the “marginal product of labor,” tabulated in column (3) of Table 4.2. An additional concept asks: “What is the output of an additional worker worth to the firm?” A price-taking firm sells the output for whatever price, $P$, the market dictates. So the value of a worker’s additional output is simply the marginal product of labor times the price of output, called the value of the marginal product of labor $VMP_L$. More generally:

**Definition: Value of the Marginal Product of an Input:**

The value of the marginal product of input $x_i$ is simply the revenue produced by an additional unit of the input, or price times marginal product:

$$VMP_i = P \times MP_i$$

This concept is central to understanding the demand for factors of production. Recall that Bob’s profits are

$$\Pi(L) = PF(L) - WL - C_0$$

Increasing employment by $\Delta L$ changes profits by

$$\frac{\Delta \Pi(L)}{\Delta L} = P \frac{\Delta F(L)}{\Delta L} - W$$

$$= P \times MP_L - W$$

$$= VMP_L - W$$

per additional worker hired. Profits rise when an additional worker is hired if $VMP_L > W$; if the worker adds more to revenues ($VMP_L$) than what she is paid ($W$). In terms of the data in Table 4.2, hiring another worker costs Bob $W =$10,000, the market wage for a year’s work. Each unit of output sells for $P =$15. If the worker adds more to revenue than she costs—$VMP_L > W$—then it is worthwhile to hire her because doing so increases Bob’s profits. Following this rule, if Bob has six employees he should hire a seventh worker because $VMP_L(7) = $15 \times 1000 = $15000$, so hiring a fifth worker increases his profits by $15000-10000 = $5000. He should also hire an eighth because $VMP_L(8) = $15 \times 833 = $12495$, which adds another $2495 to his profits. If he hires a ninth worker we have $VMP_L(9) = $15 \times 667 = $10000$, which is just equal to the wage Bob has to pay, so a 9th worker adds zero profit. To avoid problems of discreteness, we assume that Bob is willing to hire so long as $VMP_L \geq W$, so hiring a 9th employee is (just) worthwhile because Bob earns zero marginal profit by doing so.
Column (12) of Table 4.2 confirms that Bob’s profits are maximized when he hires \( L^* = 9 \) workers and produces \( Q^* = 11,500 \) bushels per year. His total revenues are \( PQ^* = $172,495 \), and his costs are $120,000, leaving a profit of \( \Pi(L^*) = $52,495 \). We can summarize this profit-maximizing decision as a rule:

**The Profit Maximizing Employment of a Factor of Production:**

If the value of the marginal product of input \( x_i \) exceeds its price, \( W_i \), then it is profitable to employ another unit of the input. So with diminishing marginal productivity, the condition for profit maximizing use of an input is

\[
VMP_i = W_i
\]

This decision is shown graphically in Figure 4.5, which shows the value of the marginal product of labor on Bob’s farm along with the market wage of \( W = $10,000 \). He hires an additional worker so long as \( VMP_L \geq W \), which leads him to hire 9 employees.

**Question:** There are two places in the graph where \( VMP_L = W \), one on the upward sloping part of \( VMP_L \) and one on the downward sloping part. Why did profit-maximizing Bob choose the one on the downward sloping part?

Notice that if the market wage of farmworkers fell from $10,000 to $6,000, our decision rule says that it would be profitable to hire a 10\(^{th}\) worker because \( VMP_L(10) = $8250 \). Indeed, with \( W = $6000 \) Bob would increase employment from 9 to 12 workers. Graphically, this is simply a move down the \( VMP_L \) curve when the wage—the price of labor—falls. So the \( VMP_L \) “curve” is Bob’s demand curve for labor. This should make sense in light of our previous discussion of demand as willingness to pay or “marginal value.” Since a 10\(^{th}\) worker adds \( VMP_L(10) = $8250 \) (the height of the curve) to Bob’s revenues, he would be willing to pay up to $8250 to obtain another worker. Since he only has to pay $6000, the benefits outweigh the costs and so he does it.

**Question:** Technical improvements in farming raise productivity on Bob’s farm by 20 percent at all levels of employment. Assuming that wages and the price of wheat are unchanged, use Figure 4.5 to show how this affects employment. Does employment unambiguously rise? Why, if it now takes fewer workers to produce a given level of output, would Bob increase employment?

**Question:** The government of Bob’s state raises the minimum wage from $5 per hour ($10,000 per year) to $7.50. How will this affect the optimal level of employment on Bob’s farm? What profit was Bob earning before the new minimum took affect? What is the largest level of profit that he can earn after the new minimum takes effect? Assuming that Bob has leased his land and tractor for the year, and cannot take another job, should he farm or shut down his operations? Why? If another job paying $20000 per year becomes available, then what should he do?
Our conclusion that the $VMP_L$ curve in Figure 4.5 is Bob’s demand curve for labor has one caveat. Suppose the wage were to rise from $10,000 to $25,000 per year. Then the best that Bob can do by hiring workers is to employ 5 of them (Why?), which yields output of 7600 bushels. Bob’s total revenue is $PQ = $114,000, but his labor costs alone are $WL = $25,000 \times 5 = $125,000. Even at the “optimal” level of employment, he can’t cover his variable costs of hiring workers. In this case, Bob should shut down the farm, He incurs his fixed cost of $30,000, but at least he doesn’t lose more money selling wheat for less than it costs to produce it. This is an application of the principle stated above: only avoidable costs (here, labor costs) should affect the decision to produce, and Bob’s fixed costs cannot be avoided.

More generally, Bob should produce so long as revenues ($PQ$) exceed his variable costs ($WL$), which requires:

\[
\frac{PQ}{L} \geq W
\]

\[
P \times AP_L \geq W
\]

\[
P \times AP_L \geq VMP_L
\]

So Bob only hires workers when the average product of labor exceeds the marginal product of labor. This means that Bob’s labor demand curve is the downward sloping portion of $VMP_L$ that lies below the curve $P \times AP_L$ in Figure 4.5—only for wages in
this range is it worthwhile for Bob to produce, because then he covers his variable costs of production.

V. Supply: Shapes of Cost Curves, and the Profit Maximizing Rate of Output

**Figure 4.5** and the concept of value marginal product allowed us to characterize Bob’s profit maximizing choice of how many workers to hire. Of course, if he hires 9 workers, who work with Bob’s other fixed inputs, he gets a determinate rate of output which Table 4.2 shows to be 11500 bushels per year. In other words, by choosing the profit maximizing amount of *inputs* to use, Bob also chose the profit maximizing rate of *output* to produce.

Looking at the firm’s decisions from the input side of things is often useful—for example when we want to know how a tax on payrolls would affect employment, or how a subsidy to investment might impact the use of capital. In those cases we are interested in a firm’s demand for factors of production, so we build on the framework set out above. To understand other phenomena, such as how Bob will respond to a tax on wheat sales or a government subsidy to wheat production, it is more useful to study his decision of how much output to produce, relegating his input choices to the “background.” To do this, we characterize the relationship between costs and output.
Figure 4.6a graphs three concepts of “costs per unit” that are recorded in Table 4.2. The first is marginal cost, which the data indicates is ‘U-shaped’—first falling with output and then rising. Why does this happen? Diminishing returns. Indeed the range of output where marginal cost is rising (falling) corresponds exactly to the range of input \((L)\) where the marginal product of labor is falling (rising). So marginal cost must eventually rise because marginal productivity of inputs must eventually fall. They are the same thing.

Figure 4.6a also graphs average variable cost, which is simply variable cost \((C(Q)=WL)\) divided by output:

\[
AVC(Q) = \frac{C(Q)}{Q}
\]

Notice that \(AVC\) is falling when \(MC\) is below it, “pulling it down,” and \(AVC\) is rising when \(MC\) is above it, “pulling it up.” This relationship between “marginals” and “averages” means that the marginal cost curve must intersect average variable cost at the minimum of \(AVC\). So \(AVC\) is also ‘U-shaped’ because of the law of diminishing returns.

The final curve in Figure 4.6a is average total cost, defined as total cost (fixed cost plus variable cost) divided by the rate of output:

\[
ATC = \frac{C_0 + C(Q)}{Q} = \frac{C_0}{Q} + AVC
\]
So the average total cost curve is simply the $AVC$ curve shifted up at each quantity by the amount $C_0/Q$. This latter amount (sometimes called “average fixed cost”) gets smaller and smaller as $Q$ gets big, because the numerator is a constant. Graphically, this means that $ATC$ lies everywhere above $AVC$, but the vertical distance between them gets smaller and smaller as $Q$ rises. That’s what the graph shows.

Bob’s goal is to maximize his profits by choosing an appropriate rate of output, $Q$. So we now write his profits as depending on $Q$:

$$\Pi(Q) = PQ - C(Q) - C_0$$

We can use Figure 4.6b to solve this problem in terms of marginal decisions. Figure 4.6b differs from 4.6a in only one respect; it shows the price of output, $P$= $15 per bushel. The price of output is the increment to Bob’s revenue whenever he produces and sells another bushel, called his marginal revenue. The increment to his costs from producing and selling another unit is (by definition) marginal cost. Producing and selling another unit raises Bob’s profits so long as marginal revenue (price) exceeds marginal cost, so his decision rule is to produce more so long as $P > MC(Q)$. With rising marginal cost, these marginal additions to profits are exhausted at the rate of output $Q^*$ where $P = MC(Q^*)$. In the Figure and in Table 4.2, this occurs at $Q^* = 11500$ bushels per year.

Bob’s maximized profits can also be shown graphically. His profits are:

$$\Pi(Q^*) = PQ^* - C(Q^*) - C_0$$

$$= \left( P - \frac{C(Q^*) + C_0}{Q^*} \right) Q^*$$

$$= \left( P - ATC(Q^*) \right) Q^*$$

Bob’s profits are the difference between price and average cost, times the number of units he produces. This is simply the area of a rectangle in Figure 4.6b. The height of the rectangle is $P - ATC(Q^*) = $15-$10.44 = $4.56, which is the vertical distance from $P$ to the $ATC$ curve at $Q^* = 11500$ bushels. The base is $Q^*$, so the area is $4.56 \times 11500 = $52495.

If the price of wheat were to rise from $15/bushel to, say, $20, Bob would increase his rate of output, moving out along his marginal cost curve. If the price were to fall to $12.50 he would reduce his output, moving back down $MC$. This is the proof of our earlier assertion that the supply curve of a competitive firm is the firm’s marginal cost curve. But there is a caveat, and it’s the same one we had in defining Bob’s demand curve for labor.
Suppose price were to fall to $10/bushel. At this price, Bob could set $P=MC$ and produce $Q=10000$ bushels per year. This price also happens to equal the minimum level of $ATC$, so Bob is operating at the point where $MC$ intersects $ATC$. As $P-ATC = 0$, Bob earns zero economic profit. We know this is worthwhile, however, because he is able to cover all of the opportunity costs of the resources he employs, including himself. If price were to fall even more, to $7.14$, he could cut his production to 9000 bushels, but then he loses money: $P-ATC = $7.14-$10 = -$2.86, so his profits are $2.86x9000 = -$25740. Even so, he should continue to operate because he at least covers his variable costs of production. If he shuts down the farm he must still incur his fixed costs of $30000, but by continuing to farm he at least covers the variable costs of farming. Indeed, he will continue to farm so long as the price of wheat exceeds the minimum of his average variable costs, for then he can at least minimize his losses. In our example, this minimum occurs when Bob hires 5 farmworkers and produces 7600 bushels per year, where $AVC = $6.58. If price falls below that, he should quit. This means that Bob’s supply curve is his marginal cost curve in the range of output where $MC$ lies above the minimum of $AVC$. This is worth a formal statement:

**The Shutdown Condition and the Supply Curve of a Competitive Firm:**
A competitive firm should produce the rate of output where $P=MC$, so long as $P \geq \min AVC$. If $P < \min AVC$ the firm should cease operations. This means that the supply curve of a competitive firm is the marginal cost curve in the range where $MC$ lies above $AVC$.

So Bob should cease farming if price falls below $6.58, otherwise he should produce according to $P=MC(Q)$. What happens in the long run? In the long run, all costs are variable: Bob doesn’t have to renew his lease on the land or on the tractor, and he can move to the city and become an economist (for $20000/year, the going rate). Then the $ATC$ curve is the same as $AVC$, and he should only produce if the price exceeds $10$, the minimum value of $ATC$ that allows him to earn zero economic profit. So Bob would only enter the business if he expected $P \geq 10$; operating at lower price means that price unexpectedly fell.

*Choosing Inputs or Choosing Output: It’s the Same Decision*

Figure 4.5 represented the firm’s profit maximizing decision in terms of hiring workers. We found that profits are largest when

$$P \times MP_L(L^*) = W$$

Figure 4.6b purported to represent the same decision, but measured in terms of output. Then we found that profits are maximized when

$$P = MC(Q^*)$$
These appear to be different conditions, but they really can’t be because if profits are maximized it had better be true that \( Q^* = F(L^*) \); that is, profit maximizing employment produces the profit maximizing rate of output. We can see that this is true by rewriting the employment condition as

\[
P = \frac{W}{MP_L(L^*)}
\]

The numerator of this expression is the increase in the firm’s costs if it hires one more worker: \( \Delta C = W \). The denominator is the additional output the firm gets if it hires one more worker: \( \Delta Q = MP_L \). So the ratio is

\[
P = \frac{W}{MP_L(L^*)} = \frac{\Delta C(Q^*)}{\Delta Q} = MC(Q^*)
\]

In other words, a firm that hires labor to the point where value marginal product equals the wage must also—by definition—be producing output at the quantity where price equals marginal cost. You can check this in Table 4.2: marginal cost, calculated as \( CQ/Q \), is equal to the ratio \( W/MP_L \) at every level of employment. This re-emphasizes the point made above: sometimes it is useful to look at the firm from the input side, and sometimes it is useful to look at the output side. But we should not lose sight of the fact that the underlying behavior is identical.

VI. Multiple Factors of Production: The Long(er) Run

We have characterized how a firm chooses a single variable input, labor, by comparing the value of what an additional worker produces to the cost of hiring her. This framework “works” no matter how many inputs the firm uses in production.

We now drop our numerical example of Bob’s farm and consider the more general problem of choosing multiple inputs. To keep things simple, think of a firm that uses two inputs, capital, \( K \), and labor, \( L \). Recall that capital refers to durable assets that yield a stream of services over many periods, like a machine. To square our treatment of capital with the timing of production and input use, we think of the stream of capital services as being proportional to the amount of capital employed. We will want to contrast two extreme situations, one when capital is fixed (the short run) and the other when capital is freely variable (the long run). Our input prices also have to be expressed as payments per unit of time, say per year, so we will denote the rental price of capital by \( R \). Then \( R \) is the amount that a firm must pay in order to use one unit of capital for one period.
With those preliminaries, output is given by
\[ Q = F(L, K) \]
and profits are
\[ \Pi(L, K) = PF(L, K) - WL - RK \]
where \( R \) is the rental price per unit of capital and we assume for the moment that all costs are variable—there are no fixed costs. How does the firm simultaneously choose \( L \) and \( K \) to maximize its profits?

To understand how the solution works, suppose we knew that the optimal amount of capital was \( K^* \), because God told us. If we fix \( K = K^* \), we can think of using various quantities of labor, just as we did on Bob’s farm. With capital fixed, we can trace out the value marginal product of labor curve shown in Figure 4.7a, defined by
\[ VMP_L(L, K^*) = P \times MP_L(L, K^*) \]
where \( MP_L(L, K^*) \) is the marginal product of various quantities of labor when capital is held fixed at \( K^* \). From our previous analysis, \( VMP_L(L, K^*) \) measures willingness to pay for another unit of labor, so if \( VMP_L(L, K^*) > W \) it is profitable to hire more workers, and so on. It follows that the optimal quantity of labor to hire, given \( K = K^* \), satisfies:
\[ P \times MP_L(L^*, K^*) = W \]
which is shown in Figure 4.7a. This says that when inputs have been chosen appropriately, the value of the marginal product of labor must equal the cost of hiring another worker—just as above.
Now think of fixing labor at its optimal level $L = L^*$ and varying capital. This experiment traces out the value marginal product of capital curve shown in Figure 4.7b, defined by

$$VMP_L(L^*, K) = P \times MP_L(L^*, K)$$

where $MP_L(L^*, K)$ is the marginal product of various quantities of capital when labor is fixed at its optimal quantity. Since $VMP_K(L^*, K)$ measures willingness to pay for an additional unit of capital, if $P \times MP_K(L^*, K) > R$ it pays to employ more capital. It follows that the optimal quantity of capital to employ, given $L = L^*$, satisfies:

$$P \times MP_K(L^*, K^*) = R$$

which solution is shown in Figure 4.7b. So when inputs have been chosen appropriately, the value of the marginal product of a unit of capital must equal the price of capital—otherwise the firm has not maximized its profits.

The 2-input solution shown in Figure 4.7 is just a graphical representation of some math: we had two equations in two unknowns, $K$ and $L$. The simultaneous solution of the two equations gives the optimal quantities of capital and labor to use.
We can generalize these findings to an arbitrary set of inputs $x_i$ (land, energy, MBAs, accountants, hot dog buns, wiener ...) as follows:

**The Profit Maximizing Use of Factors:**

Let production be $Q = F(X)$, where $X = (x_1, x_2, \ldots, x_n)$ are inputs with market prices $W_1, W_2, \ldots W_n$. Then a price taking firm maximizes its profits when the value of the marginal product of each input is equal to the input’s price. That is, the optimal input quantities $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$ satisfy

$$P \times MP_1(x^*) = W_1$$
$$P \times MP_2(x^*) = W_2$$
$$\vdots$$
$$P \times MP_n(x^*) = W_n$$

Why do I make you learn this in such generality? Because I want you to understand that for all inputs, prices reflect the value of what the input produces at the margin. So if MBAs in consulting firms earn $120000 per year, that is the value of the marginal product of an MBA—if another were hired, that’s about the value of what she would produce. If one died, that’s the value of output that would be lost. If a law firm rents space for an annual price of $500/square foot, then the value of output attributable to a 200 square foot office is $100000 per year. And so on.

These ideas are the foundation for what economists call the “marginal productivity theory of income distribution,” which says, in a nutshell, that everybody and every thing gets paid the value of its marginal product. Then a household’s income is determined by what resources it owns (human capital, physical capital, etc.) and the productivity of those resources. We will return to this idea in Lecture 5, when we examine the evolution of wages and wage inequality in the United States.

**Question:** True, False, or Uncertain. Oil is an input to its own production—it takes oil to produce oil. Therefore, the marginal product of oil in an oil producing firm must be 1.0.

**Cost Minimization**

In our analysis of Bob’s farm, we showed that the problem of choosing how much output to produce was the twin of choosing how much labor to hire, because $Q=F(L)$ on the farm. With many inputs, the combination that produces a fixed level of output, say $Q=10000$ hot dogs, is not unique—we could produce this quantity with lots of labor and a little capital, or with more capital and less labor. In other words, (we assume) it is possible to substitute capital for labor while holding output fixed, just as we assumed that consumers were willing to substitute one good for another at a given level of utility.
Since the combination of inputs that produces a given output is not unique, we can think of the firm’s decisions in two steps. First, the firm chooses the combination of inputs that minimizes the cost of producing each level of output \( Q \), resulting in \( C(Q) \). Then it chooses the profit maximizing rate of output \( Q^* \) where \( P = MC(Q^*) \).

We could spend way too much time on the details of how the firm does this, but we already have the tools to characterize all that we need to know. Consider the solution to the firm’s problem of choosing capital and labor to maximize profits, given by

\[
P \times MP_K(L^*, K^*) = R
\]

\[
P \times MP_L(L^*, K^*) = W
\]

If we divide the first equation by the first, we get an expression that is independent of \( P \):

\[
\frac{MP_K(L^*, K^*)}{MP_K(L^*, K^*)} = \frac{R}{W}
\]

This is a condition for cost minimization at any rate of output: the ratio of the marginal products of any two inputs must equal the ratio of their prices. This is more intuitive with one more rearrangement

\[
\frac{W}{MP_L(L^*, K^*)} = \frac{R}{MP_K(L^*, K^*)}
\]

As we noted in our discussion of decisions on Bob’s farm, the first ratio, \( W / MP_L \) is the marginal cost of additional output when slightly more labor is hired. The numerator is the change in the firm’s cost when it hires another worker (\( \Delta C = W \)) and the denominator is the change in output when it hires another worker (\( \Delta Q = MP_L \))—so the ratio is marginal cost. Since this notion of marginal cost is associated with hiring more labor, we might awkwardly refer to it as \( MC_L \). Similarly, the numerator of the second ratio is the change in cost from acquiring another unit of capital, \( R \), and the denominator is the additional output that the capital would produce. So the ratio of the two is another measure of marginal cost, where changes in cost and output are due to using a little more capital. Call this \( MC_K \). So the condition states that \( MC_L = MC_K \): No matter how we increase output, using capital or labor, the marginal cost must be the same.

To see why this condition must hold, suppose it didn’t. Suppose

\[
\frac{W}{MP_L} > \frac{R}{MP_K} \quad [\text{means } MC_L > MC_K]
\]

This says that it is cheaper to produce the marginal unit of output using a bit more capital than by using a bit more labor. So the firm’s costs would be lower if it substituted away
from labor and toward capital, holding output fixed. As it uses more capital the marginal product of capital falls, so \( \frac{R}{MP_k} \) rises. (Why?) As it uses less labor the marginal product of labor falls, so \( \frac{W}{MP_L} \) falls. This process of substitution continues until equality is restored. The upshot is that when we talk about “marginal cost” at any level of output, this notion reflects the cost-minimizing choice of all the firm’s inputs. We can summarize this fact as follows:

**The Condition for Cost Minimization:**

For any rate of output, \( Q \), a firm minimizes its cost when the marginal cost of an additional unit is the same no matter how that unit is produced:

\[
MC(Q) = \frac{W_1}{MP_1} = \frac{W_2}{MP_2} = \cdots = \frac{W_n}{MP_n}
\]

**Question:** Suppose that firms in the energy industry employ twice as many college graduates as high school graduates. The wage rate of skilled workers is \( W_s = $60000 \) per year and the wage of high school graduates is \( W_h = $40000 \). True or False: A law mandating that firms use equal numbers of high school and college graduates will reduce costs, because high school graduates are cheaper.

**The Demand for Factors in the Short Run and the Long Run**

I said earlier that the short run is a window of time in which we may think of some factors of production as being fixed. To represent this idea, let’s fix \( K = K^* \), the optimal value from Figure 4.7. If the firm can’t vary its capital, how does this affect its demand for labor? In particular, how does this “fixity” affect the sensitivity of the quantity of labor demanded to a change in the wage? That is, is the demand for labor more or less elastic when capital cannot be varied?

The analysis is shown in Figure 4.8. We start from the optimal quantities \((L^*, K^*)\) that are chosen when prices are \((W^0, R^0)\). Now suppose the wage increases to \( W^1 \), so labor is more expensive. With capital fixed, we know the firm will move up the \( VMP_L(L, K^*) \) curve, hiring \( L^1 < L^* \) workers. As \( K \) was fixed for this experiment, we can think of the reduction in employment from \( L^* \) to \( L^1 \) as the short run response of labor demand to a \( W^1 - W^0 \) increase in the wage.
Now consider the long run, where capital can also be varied in response to the change in the wage. We can think of things sequentially. First, with capital fixed at $K^*$, the firm reduces employment to $L^1$. As time passes, it is able to change $K$ as well. Should it increase $K$, decrease $K$, or do nothing at all? Which of these occurs depends on how labor and capital interact in the production process. In particular, we need to know whether having fewer workers changes the marginal product of capital. If it does, then the firm will change the amount of capital it uses.

There are two possibilities. First, a reduction (increase) in labor might reduce (increase) the marginal product of capital. So, for example, hiring another worker makes capital more productive, so $MP_k$ rises with $L$. If this is true (you’ll have to trust me on this) then it also must be true that acquiring an additional unit of capital makes labor more productive, so $MP_L$ rises with $K$. In this case we say that capital and labor are productive complements—more of one makes the other more productive. For example, power drills and hammers (capital) make carpenters (labor) more productive, so hammers and carpenters are complements.

But capital and labor don’t have to be complements. Think of robots on the assembly line of an automobile plant, which (let’s say) can do exactly the same things that humans do—tightening bolts, installing hubcaps, and so on. Then having more robots for a fixed number of workers is exactly like having more workers—the marginal
product of labor must fall. In this case we say that capital and labor are productive substitutes, because more of one reduces the productivity of the other.

**Definition: Productive Complements and Substitutes**

If an increase in input $x_i$ raises the marginal product of $x_j$, then we say that $x_i$ and $x_j$ are **productive complements**. If an increase in input $x_i$ reduces the marginal product of $x_j$, then we say that $x_i$ and $x_j$ are **productive substitutes**.

With these definitions behind us let’s go back to Figure 4.8 and assume that capital and labor are complements. Then a reduction in $L$ from $L^*$ to $L'$ must reduce the marginal product of capital. In Figure 4.8b, this means that the value of the marginal product of capital shifts down to $VMP_K(L', K)$. Since $VMP_K$ is lower, the firm reduces the amount of capital it desires from $K^*$ to $K'$. So, when labor and capital are complements, an increase in the price of one reduces the demand for the other. If $K$ and $L$ had been productive substitutes, a reduction in $L$ would cause $VMP_K$ to shift up, so the demand for capital would rise. So when capital and labor are substitutes, and increase in the price of one increases the demand for the other.

**Fact: Cross-Price Effects on the Demand for Factors of Production:**

If $x_i$ and $x_j$ are complements, an increase in the price of one reduces the demand for the other. If $x_i$ and $x_j$ are substitutes, an increase in the price of one raises the demand for the other.

Now we have a problem. The firm’s decision to reduce employment to $L'$ was predicated on the assumption that capital was fixed at $K^*$. But when the firm employs $L'$ workers, and labor and capital are complements, the firm chooses $K' < K^*$ units of capital. So $L'$ is not the right number of workers to employ when the firm can adjust capital from $K^*$ to $K'$. We can find the correct number to employ with $K'$ units of capital by noting that fewer units of capital must reduce the marginal product of labor—because we assumed they are complements—so $VMP_L$ must shift down to $VMP_L(L, K')$. Then the firm chooses to employ $L'' < L'$ when capital is fixed at $K'$.

But now $K'$ isn’t the right amount of capital, because it was predicated on employing $L'$ workers, so $VMP_K$ shifts down again, so less capital is employed, and dot, dot, dot. We can do this forever. But all we really need to understand was in the first iteration, when the firm reduced its demand for capital from $K^*$ to $K'$. This caused the demand for labor to fall to $L'' < L'$. Therefore, when capital can be varied, the demand for labor was more responsive to a change in the wage than when capital was fixed. So we conclude that the demand for labor is more elastic in the long run, when all factors are variable, than in the short run, when some factors are fixed.
**Exercise:** We assumed that capital and labor are complements, and we found that the demand for labor is more elastic in the long run. Convince yourself (draw the pictures) that the same is true when $K$ and $L$ are substitutes. That is, a higher wage reduces $L$, which raises the demand for capital, which further reduces the demand for labor …..

**Fact: The Elasticity of Demand for a Factor in the Long and Short Runs**

The own-price elasticity of demand for a factor of production is larger in the long run, when all factors are variable, than in the short run, when some factors are fixed.

This is an extremely useful rule. For example, it says that an increase in the minimum wage will have a larger impact on the demand for teenage workers in the long run, when employers can adjust other factors, than in the short run. In the short run, fast-food restaurants are stuck with the capital (hamburger grills, cash registers, etc.) they have. But in the long run they can reduce the scale of operations, choose more automated ways of doing things, and so on, to respond to the higher costs of labor. Our analysis tells us that these adjustments will reduce teenage employment by more at the end of 2 years than at the end of 2 months or 2 weeks. So we cannot judge the overall impact of the law by looking at employment only a few months after the minimum increased. Similarly, when the price of oil spiked in the 1980’s, firms were stuck with their current stocks of capital and methods of production. Energy use fell, but it fell by much more in the long run, when firms were able to adjust other factors of production.

The rule stated above is the first of what are called “Marshall’s Laws” for the elasticity of demand for factors of production (same Marshall, he was a smart guy). There are three others. I give an intuitive summary, and then a current example.

Marshall’s second law has to do with the ease of substitution between inputs. In our conditions for cost minimization discussed above, an increase in the wage would cause the firm to substitute capital for labor at each level of output. If capital and labor are good substitutes, then labor falls by a lot at each level of output—demand is more elastic. If capital and labor are poor substitutes, labor falls by less at each level of output—demand is less elastic. More generally, the demand for a factor of production is more elastic the greater the ease of substitution between it and other factors of production.

The third “law” has to do with the elasticity of demand for the good being produced. When we increased the wage in our analysis above, we held the price of the good fixed. This might be true if the wage increase applied to only one seller, but something like an increase in the minimum wage applies to all sellers in a given industry or market. So a higher minimum wage raises the marginal cost of selling hamburgers, which must raise the price of hamburgers. When the price rises, people eat fewer hamburgers, so employment of hamburger makers must fall. How much does it fall? It depends on the elasticity of demand for hamburgers—when it’s big employment falls by more, and when it’s small employment falls by less. More generally the demand for a
The fourth “law” is related to the third. Suppose that labor accounts for 75% of the marginal cost of producing hamburgers, and that making a hamburger requires a fixed amount of labor. Then (assuming all workers earn the minimum wage) a 10% increase in the minimum wage will increase marginal cost by about 0.75X0.10=0.075, or 7.5%. As price equals marginal cost in a competitive industry, price will rise by about 7.5% as well. If the elasticity of demand for hamburgers is 1.0, then the quantity of hamburgers eaten must also fall by 7.5% and, because each burger takes a fixed amount of labor, employment of teenagers will fall by 7.5%. On the other hand, if labor accounts for only 25% of marginal cost of a hamburger, then employment will fall by only 2.5%. In other words, the demand for a factor of production is more elastic the greater the share of the factor in marginal cost. Factors that have a small cost share are less elastically demanded, because they don’t have much impact on the price of the final good.

Marshall’s Laws for the Demand for Factors of Production

The elasticity of demand for a factor of production is larger:
1. In the long run than in the short run.
2. The greater the ease of substitution between it and other factors.
3. The more elastic is the final demand for the good being produced.
4. The greater the share of the factor in marginal cost.

[The fourth “law” isn’t always right, but for practical purposes we can take it to be true].

Example: Antitrust and the elasticity of demand for Microsoft Windows

Marshall’s second, third and fourth laws played an important role in the Department of Justice’s recent antitrust suit against Microsoft Corporation, which began in 1998 and concluded in 2002. Putting aside the legal issues and outcomes, DOJ alleged that Microsoft charged a monopoly price for its Windows operating system (OS)—a price that was well above what would occur in a competitive market. Applying the second law, DOJ (and the judge) pointed out that there are no good substitutes for an OS—PCs need to have one in order to work—which makes the demand for OSs highly inelastic. Reinforcing this, they also applied the fourth law: the OS accounts for only a small portion of the overall cost of a PC, which is another factor causing the demand for an OS to be small. Microsoft supplied over 90 percent of OSs for PCs, so the inelastic demand allowed them to raise price a lot without reducing quantity demanded.

Microsoft turned the arguments around, arguing that they could not possibly be charging a monopoly price, even if their share of OS sales was huge—they were forced by competition to charge a lower one. Their economics experts did the following calculation to obtain the elasticity of demand for operating systems (OSs), which are inputs to the building of a complete PCs.

First, using the third law, they presented evidence that the elasticity of demand for personal computers (the final product) was about 2.0. Then, they showed the prices
charged to computer makers (Dell, Compaq, etc.) for copies of Windows installed on their machines, which came to about $60 per copy. In the 1990s, this price accounted for about 2.5% of the cost of a typical computer. As computer manufacturing is highly competitive, the price of PCs equals marginal cost and an increase in the price of the OS would raise the price of a PC by the same amount. So, using Marshall’s fourth law, the elasticity of demand for operating systems of personal computers was about 2x(0.025)=.05, which is a tiny demand elasticity. As we know from Lecture 3, if the elasticity of demand for Windows was .05 (much smaller than 1.0), then Microsoft’s price was far below the price that would maximize its profits (Why?).

Microsoft argued that their pricing behavior belied the allegation that they controlled the market for operating systems. Windows is a brand of OS, but the elasticity of demand for one firm’s OS was far greater than the elasticity of demand for OSs in general, just as the elasticity of demand for Tropicana orange juice is greater than the elasticity of demand for orange juice in general. In other words, there are very good substitutes for Windows (Marshall’s second law). This fact forced them to price low, they argued. Indeed, if they really had monopoly power, as the DOJ alleged, they calculated that the profit maximizing price of Windows would have been about $1000, or nearly 20 times higher than it was.

How is it that Microsoft sells 95% of all OSs for personal computers, but still prices low? The argument above is one explanation. We will take up others when we discuss monopoly pricing, and other details of the case, later in the quarter.

VII: The Long Run Elasticity of Supply

We have demonstrated that factors of production are more elastically demanded in the long run. The same type of reasoning implies that supply is more elastic in the long run, when all factors can be varied. This is demonstrated in Figure 4.9.

The Figure shows a long run average cost curve, $LAC$, drawn under the assumption that a firm can choose any quantities of labor and capital that it wants. For example, when output is $Q_0$ the firm employs $L_0$ workers and $K_0$ units of capital. When output is $Q_1$, the firm would use $L_1$ workers and $K_1$ units of capital, and so on. It isn’t important to us what the actual input quantities are, only that the firm got to choose them to minimize the cost of production, as above. Associated with $LAC$ is a long-run marginal cost curve, $LMC$, which also assumes that capital and labor are both variable.
Assume that the price of output is $P_0$, which is equal to the minimum of long run average cost, so the firm is earning zero economic profit at output $Q_0$. To do this, it uses $L_0$ workers and $K_0$ units of capital. Now suppose that price rises from $P_0$ to $P_1$. In the long run, when both capital and labor can be varied, we know that the firm will choose to produce $Q_1$ units, where $P=LMC$. But what about in the short run, when capital is fixed at its original level $K_0$?

To figure this out, suppose we told a firm that was producing at $Q_0$: “You have to increase your output to $Q_1$, but you must keep your capital fixed at $K_0$. That is, all of the increased production must be achieved by using more labor.” Would it be more costly to produce the additional units? Less costly? What do the short run average and marginal cost curves for this experiment look like?

When capital can be varied, the average cost of producing $Q_1$ units is given by the height of $LAC$ at that quantity. If it could, the firm would use $K_1$ units of capital, but we have told it to use $K_0$. In minimizing its costs the firm could have chosen $K_0$, but didn’t. So the cost of producing $Q_1$ units while forcing the firm to use $K_0$ units of capital must be higher than when it uses its best choice $K_1$. This means that short run average cost $SAC$ must lie above $LAC$ for all quantities greater than $Q_0$, and it just touches $LAC$ at

![Figure 4.9](image)

The Elasticity of Supply is Greater in the Long Run than in the Short Run
output $Q_0$, where $K_0$ is the optimal quantity of capital to use. The same applies to quantities smaller than $Q_0$: the firm cannot do better when it is prohibited from varying its capital, so $SAC$ lies above $LAC$ in this range too.

What about marginal cost? The cost of producing $Q_1 - Q_0$ additional units is given by the area under marginal cost between $Q_0$ and $Q_1$: the sum of all the marginal costs on each additional unit produced. The costs of producing these units cannot be lower when the firm is forced to produce them with $K_0$ units of capital, instead of the amount that minimizes cost. So the short run marginal cost curve, $SMC$, must rise more rapidly than long run marginal cost, $LMC$. Since $SMC$ is the firm’s short run supply curve, we have established:

**Fact: The Elasticity of Supply in the Short Run and the Long Run**

The elasticity of supply of a competitive firm is greater in the long run, when all factors are variable, than in the short run, when some factors are fixed.

Again, “short run” and “long run” are abstractions meant to convey the idea that it is more costly to adjust to new circumstances in a short period of time. The real message is that supply responses of firms will be larger as time passes and firms can adjust in the most cost effective way.

**VIII. Summary: The Supply of Output and the Demands for Inputs**

This Lecture has developed the basic tools for analyzing the profit-maximizing decisions of a competitive business firm. We have characterized supply decisions that follow from the incremental (marginal) revenues and costs from producing and selling output, as well as the decisions of which inputs to use in production, and in what quantities to use them. Lecture 5 extends and applies this framework to help us understand two important phenomena. First, why do economies grow, and what are the implications of growth for human welfare? Second, what determines the distribution of incomes in a market economy, and why does income inequality change over time. The tools developed in this lecture are the main weapons economists use to study these issues.