Lesson-20

Theory of the Firm

For developing the supply and demand approach to economics, the economists first worked out the basis of the demand curve. By treating the demand for a product or service as a rational decision by a (primarily) self-interested individual or family, the economists were able to understand the relation of the demand for one product or service to the demands for other products and services and to many other forms of economic activity. It was natural to apply the same approach to supply. As a first step, one needs to think about the decision-makers in supplying goods and services and what a "rational decision" to supply goods and services would mean. In economics, this is often called the "theory of the firm."

A firm is a unit that does business on its own account. Firm is from the Italian word, *firma*, a signature and the idea is that a firm can commit itself to a contract. Thus, a firm is the decision-maker in supplying goods and services.

There are three main kinds of firms in modern market economies which are as follows:

Proprietorships

A proprietorship (or proprietary business) is a business owned by an individual known as the "proprietor." Many "mom and pop stores" and other "mom and pop" businesses are proprietorships. Some of the proprietorships are too small even to employ a person full time. Craftsmen, such as plumbers and painters, may have "day jobs" and work as self-employed proprietors part time after hours. The computer programmers and others may also do that. At the other extreme, some proprietary businesses employ many hundreds of workers in a wide range of specializations. In a proprietorship, a proprietor is almost always the decision-maker for the business.

Partnerships

A partnership is a business jointly owned by two or more persons. In most of the partnerships, each partner is legally liable for debts and agreements made by any partner. Ofcourse, this requires a great deal of trust and thus, partners generally know one another well enough to have that sort of trust. Family partnerships are very common for that very reason. There are now a few "limited partnerships" in which some of the partners are protected from legal liability for the agreements made by others, beyond some limits. In many cases, one partner is designated as the managing partner and is the main decision-maker for the business.

Corporations

A corporation has two characteristics that distinguish it from most of the proprietorships and partnerships which are as follows:

- Limited liability
- Anonymous ownership

Limited liability means that the owner of shares in a corporation cannot lose more than a certain amount if the company fails. Usually, an amount is the money paid to buy the shares. Anonymous ownership means that the owner of the shares can sell them without getting the permission of anyone other than the buyer. By contrast, in most of the partnerships, no single partner can sell without getting the agreement of the other partners. In such a case, the continuing partners will, ofcourse, want to know about the new partner. And he will not be an "anonymous owner." In a typical corporation, the shareholders formally elect a board of directors who, in turn, select the officers of the company. One of these officers, often called the "president," will be the principal decision-maker for the firm but he will be expected to make decisions in the interest of the shareholders.

While there are millions of proprietorships, typically very small, the biggest businesses are corporate and corporations are particularly important because of their size.

Objectives

Malthus did not have firms in mind when he formulated the law of diminishing returns. But this law has applications which Malthus did not envision. In the *Reasonable Dialog of economics* in the nineteenth century, the development of these ideas was a bit indirect. In circa eighteen-seventies, the economists were rethinking the theory of consumer demand. They applied a version of "diminishing returns" and the Equimarginal Principle to determine how a consumer would divide his/her spending among different consumer goods. That worked pretty well and, therefore, some other economists, especially the American economist, John Bates Clark, tried using the same approach in the theory of the firm. These innovations were the beginning of neoclassical economics.

By following the neoclassical approach, one may interpret "rational decisions to supply goods and services" to mean decisions that maximize something. What does a supplier maximize? The operations of a firm will, ofcourse, depend on its objectives. One of the objectives that all the three kinds of firms share is profits and it seems that profits are the primary objective in most of the cases. We will follow the neoclassical tradition by assuming that the firms aim at maximizing their profits.

There are two reasons for this assumption which are as follows:

1. Despite the growing importance of non-profit organizations and the frequent calls for corporate social responsibility, profits still seem to be the most important single objective of producers in our market economy. Thus, it is the right place to start.

2. A good deal of the controversy in the *Reasonable Dialog of economics* has centered on the implications of profit motivation. Is it true, as Adam Smith held, that the "invisible

hand" leads profit-seeking businessmen to promote the general good? To assess that question, one needs to understand the implications of profit maximization.

Profit

Profit is defined as revenue minus cost. However, one needs to be a little careful in interpreting that. Remember, the economists understand cost as opportunity cost, i.e. the value of the opportunity given up. Thus, when we say that businesses maximize profit, it is important to include all costs whether they are expressed in money terms or not.

For example, a cab-driver, the self-employed proprietor of an independent cab service, says, "I'm making a 'profit,' but I can't take home enough to support my family, so I'm going to have to close down and get a job." The proprietor is ignoring the opportunity cost of his own labor. When those opportunity costs are taken into account, one will find that he is not really making a profit after all.

Let us say that the cab driver makes $500 a week driving his cab, after taking out all the expenses (gasoline, maintenance, etc). Suppose, he can get a wage (including tips) of $800 driving for someone else with hours no longer and about the same conditions otherwise. Then, $800 is the opportunity cost of his labor. After deducting the opportunity cost from his $500 net as an independent cabbie, he is actually losing $300 per week.

This is one of the most important reasons for using the opportunity cost concept. It helps us to understand the circumstances that lead people to get into and out of business.

Because the accountants traditionally considered only money costs, the net of money revenue minus money cost is called "accounting profit." Actually, modern accountants are well aware of opportunity cost and use the concept for special purposes. The economist's concept is sometimes called as "economic profit." If there will be some doubt as to which concept of profit one means, one will sometimes use the terms "economic profit" or "accounting profit" to make it clear which is intended.

The John Bates Clark Model

Like any other unit, a firm is also limited by the technology available. Thus, it can increase its outputs only by increasing its inputs. As usual, this will be expressed by a production function. The output that a firm can produce depends on the land, labor and capital the firm puts to work.

In formulating the neoclassical theory of a firm, John Bates Clark took over the classical categories of land, labor and capital and simplified them in two ways. These are as follows:

1. He assumed that all labor is homogenous-- One labor hour is a perfect substitute for any other labor hour.

2. He ignored the distinction between land and capital, grouping together both kinds of non-human inputs under the general term "capital." And he assumed that this broadened "capital" is homogenous.

Ofcourse, the simplifying assumptions are not true. John Bates Clark's conception of a firm is highly simplified, like a map at a very large scale. In more advanced economics, one can get rid of the simplifying assumptions and deal with a much more realistic "map" of a business firm.

In the John Bates Clark model, there are some important differences between labor and capital and they relate to the long and short-run.

Short and Long-Run

A key distinction between the short and long-run is as follows:

Some inputs can be varied flexibly in a relatively short period of time. One conventionally thinks of labor and raw materials as "variable inputs" in this sense. Other inputs require a commitment over a longer period of time. Capital goods are thought of as "fixed inputs" in this sense. A capital good represents a relatively large expenditure at a particular time with the expectation that the investment will be repaid and any profit paid by producing goods and services for sale over the useful life of the capital good. In this sense, a capital investment is a long-term commitment. So, capital is thought of as being variable only in the long-run but fixed in the short-run.

Thus, one can distinguish between the short-run and the long-run as follows:

In the perspective of the short-run, the number and equipment of firms operating in each industry is fixed. In the perspective of the long-run, all inputs are variable and firms can come into existence or cease to exist. So, the number of firms is also variable.

Assumptions

The John Bates Clark model of a firm is very simple. Think of a business that uses two inputs, homogenous labor and homogenous capital and produces a single homogenous kind of output. The output could be a product or service but, in any case, it is measured in physical (not money) units such as bushels of wheat, tons of steel or minutes of local telephone calls. In the short-run, in addition, the capital input is treated as a given "fixed input." Also, one can identify the price of labor with the wage in the John Bates Clark model. In a modern business firm, one has to include benefits as well as take-home wages. The technical term for the total of wages and benefits is "employee compensation."

The two more simplifying assumptions are as follows:

- The price of output is a given constant

- The wage (the price of labor per labor hour) is a given constant

By putting all of them together, just two kinds of input and one kind of output, one kind of output fixed in the short-run and given output price and wage, it seems to be a lot of simplifying assumptions and it is. But these are actually not arbitrary simplifying assumptions. They are the assumptions that fit best into many applications and the starting point for still others.

Once, you have simplified your conception of the firm to this extent, nothing much is left for the director of a firm to decide.

The Firm's Decision

In the short-run, there are only two things that are not given in the John Bates Clark model of the firm. These two things are as follows:

- Output produced
- Labor (variable) input

And that is not actually two decisions but just one, since labor input and output are linked by the "production function." Either the output is decided and the labor input will have to be just enough to produce that output or the labor input is decided and the output is whatever that quantity of labor can produce.

Thus, a firm's objective is to choose the labor input and corresponding output that will maximize profit.

Let us continue with the numerical example given in a previous lesson. Suppose, a firm is producing with the production function as shown there in the short-run. Also presume that the price of the output is $100 and the wage per labor-week is $500. How much labor the firm would use and how much output it would produce in order to maximize profits?

The relationship between labor input and profits is as follows:

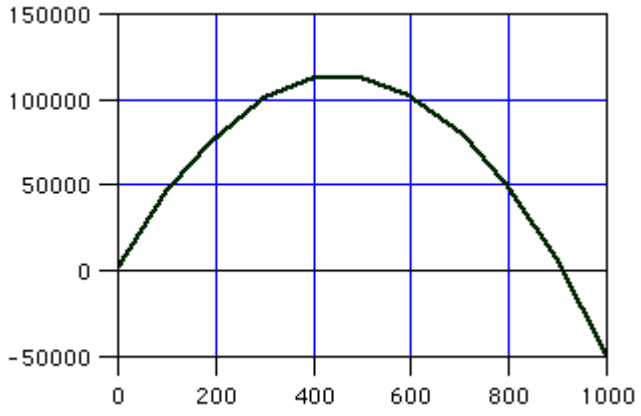

Figure 3.1 Labor input and profits

In the figure 3.1, the green curve shows the profits rising and then falling and the labor input increases. Ofcourse, the eventual fall-off of profits is a result of "diminishing returns" and the problem a firm faces is to balance "diminishing returns" against the demand for the product. The objective is to get to the top of the profit hill. One can see that this means hiring something in the range of four to five hundred workers for the week. But just how many?

The way to approach this problem is to take a bug's-eye view. Think of yourself as a bug climbing up that profit hill. How will you know when you are at the top?

The Marginal Approach

The bug's-eye view is the marginal approach. However, much labor is being employed at any given time, the really relevant question is, "Suppose one more unit of labor is hired, will profits be increased or decreased?" "If one unit of labor is eliminated, will profits increase or decrease?" In other words, what does one additional labor unit add to profits? What would elimination of one labor unit subtract from profits?

We can break that question down. Profit is the difference of revenue minus cost. Ask, "What does one additional labor unit add to cost? What does one additional labor unit add to revenue?"

The first question is relatively easy. What one additional labor unit will add to cost is the wage paid to recruit the one additional unit.

The second question is a little trickier. It is easier to answer a related question, "What does one additional labor unit add to production?" By definition, that is the marginal product. The marginal product of labor is defined as the additional output as a result of increasing the labor input by one unit. But we need a measurement that is comparable with revenue and profits, i.e. a measurement in money terms. Since the price is given, the measurement we need is the "value of the marginal product."

Value of the Marginal Product (VMP)

The value of the marginal product of an input is the increase in the money value of a firm's output when it adds one more unit of the input, keeping the quantities of other inputs fixed. One can say that an input's marginal benefit is the value of its marginal product.

For the purpose of reviewing, add one more unit to the labor input.

Increase in revenue =	Value of marginal product
Increase in cost =	Wage

1. What will one additional labor unit add to profits?

The VMP minus the wage.

2. What will the elimination of one labor unit add to profits?

The wage minus the VMP of labor.

And, in either case, the "addition to profits" may be a negative number-- either building up the workforce or cutting it down can drag down profits rather than increasing them.

So, by again taking the bug's-eye view, one may ask "Is the VMP greater than the wage or less?" If greater, increase the labor input because it will increase profits by the difference, i.e. VMP-wage. If less, cut the labor input because it will increase profits by the difference, i.e. wage-VMP. One should continue doing this until the answer is "neither." Then, there will be no further scope to increase profits by changing the labor input as the stage of maximum profits would already have been achieved.

Let us go back to the numerical example from a previous lesson. Assume that the price of output is $100 per unit and the wage is $500. In figure 3.2, the VMP is $100*MP.

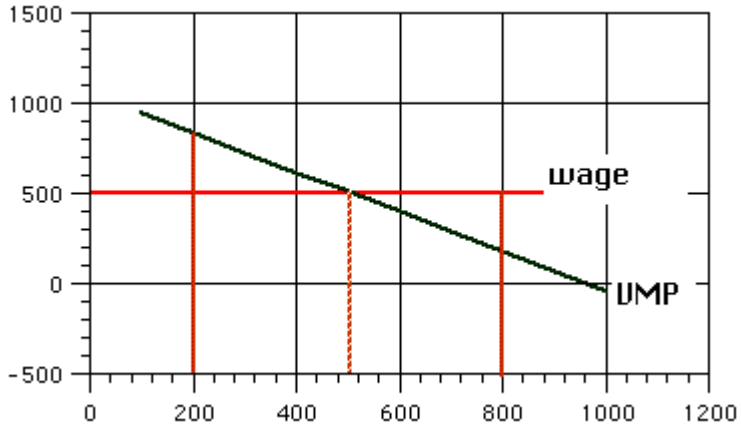

Figure 3.2

Suppose, the firm begins by using just 200 units of labor as shown by the orange line. The manager asks herself, "If I were to increase the labor input to 201, it would increase both costs and revenue but by how much?"

The VMP is 850. So, an additional worker will add $850 to revenue. Since the wage is $500, the additional worker will add just $500 to cost for a net gain of $350. It is a good idea to "upsize" and add one more worker.

On the other hand, suppose the firm is using 800 units of labor as shown by the other orange line. The manager asks herself, "If I were to cut the labor input to 799, it would cut both costs and revenue but by how much?"

The VMP is 200. So, the additional worker will add just $200 to revenues. Since the wage is $500, the additional worker will add just $500 to cost for a net loss of $300. It is the time to "downsize" and cut the labor force.

In each case, there is an unrealized potential. The amount of unrealized potential is the difference between the VMP and the wage. The firm's profit potential will not be 100% realized until the VMP is equal to the wage. That is again according to the "Equimarginal Principle."

The Equimarginal Principle

By taking the marginal approach, i.e. the bug's-eye view, we have discovered the diagnostic rule for maximum profits. The way to maximize profits is to hire enough labor so that

$$VMP=wage,$$

where p is the price of output and VMP = p*MP, i.e. the marginal productivity of labor in money terms.

This is another instance of the Equimarginal Principle. According to its rule, profits are not maximized until the labor input is adjusted so that the marginal product in labor, in dollar terms, is equal to the wage. Since the wage is the amount that the additional (marginal) unit of labor adds to cost, one could think of the wage as the "marginal cost" of labor and express the rule as "value of marginal product of labor equal to marginal cost."

Profit Maximization

In the numerical example, assume that the price of output is $100 per unit and the wage is $500 per worker per period. Then, the p*MP, wage and profits will be as follows:

Labor	Marginal Productivity	p*MP	Wage	Accounting Profit
0	9.45	945	500	0
100			500	44500

	8.35	835		
200			500	78000
	7.25	725		
300			500	100500
	6.15	615		
400			500	112000
	5.05	505		
500			500	112500
	3.95	395		
600			500	102000
	2.85	285		
700			500	80500
	1.75	175		
800			500	48000
	0.65	65		
900			500	4500
	-0.45	-55		
1000			500	-50000

Table 3.1

In the table 3.1, the transition from 400 to 500 units of labor gives p*MP=505 which is near to the VMP=wage. And this is the highest profit. So, the profit-maximizing labor force is about 500 units.

Following is the figure of the profit maximizing hiring in this example:

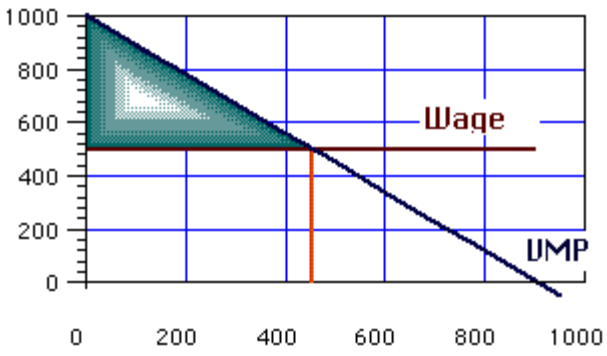

Figure 3.3 Maximizing profit

Figure 3.3 suggests that the exact amount is a bit less than 500 units of labor. The exact number is 454.54545454545 ... units of labor-- a repeating decimal fraction.

The shaded area between the VMP curve and the price (wage) line should be noticed. In the figure 3.3, the area of the shaded triangle is the total amount of payments for profits, interest and rent, i.e. everything that the firm pays out for factors of production other than labor. The rectangular area below the wage line and left of the labor=454. The line represents the wage bill. Thus, the John Bates Clark model provides a visualization of the division of income between labor and property.

Figure 3.3 can also be used to understand why VMP=wage is the diagnostic that tells us the profit is at a maximum. Suppose, the labor input is less than 500. It is 200. Then, an additional labor-day of labor will add about 7.8 units to output and about $780 to the firm's sales revenue. But it will only $500 to the firm's costs, adding roughly $220 to profits. So, it is profitable to increase the labor input from 200, or by the same reasoning, from any labor input less than $500.

This difference between the VMP and the wage is the increase or decrease in profits from adding or subtracting one unit of labor. It is sometimes called the marginal profit. The absolute value of the marginal profits is a measure of unrealized potential profits. This is the reason why a businessman wants to adjust the labor input so that VMP-wage=0.

Suppose the labor input is 800 labor-days per week. If the firm "downsizes" to 799 labor-days, it will reduce its output by just about 1.2 units and its sales revenue by about $120. But, it will reduce its labor cost by $500 and increase its profits by about $380. Thus, a movement toward the VMP=wage again increases profits by realizing some unrealized potential profit.

The formula VMP=wage is a diagnostic for maximum profits because it tells that there is no further potential to increase the profits by adjusting the labor input or marginal profit is zero.

The marginal productivity rule is the key to maximization of profits in the short-run. But now, let us take a look at the long-run perspective.

Increasing Returns to Scale and the Long-run

In microeconomics, diminishing returns is a short-run thing. In the long-run, all inputs can be increased or decreased in proportions. A reduction in the marginal productivity of labor due to an increase in the labor input can be offset by increasing the tools and equipment the workers have to work with. How will that come out on net?

In the long-run, there can be three possible cases which are as follows:

1. Decreasing returns to scale-- If an increase in all the inputs in the same proportion k leads to an increase of output in a proportion less than k, there are decreasing returns to scale. For example, if the inputs to a dairy farm have increased (cows, land, barns, feed, labor etc) by 50% and milk output by only 40%, there are decreasing returns to scale in dairy farming. This is also known as the "diseconomies of scale," since production is less cheap when the scale is larger.

2. Constant returns to scale-- If an increase in all the inputs in the same proportion k leads to an increase in output in the same proportion k, there are constant returns to scale. For example, if the number of machinists and machine tools have increased by 50% each and the number of standard pieces produced by 50%, there are constant returns in machinery production.

3. Increasing returns to scale-- If an increase in all the inputs in the same proportion k leads to an increase of output of a proportion greater than k, there are increasing returns to scale. For example, if the inputs to a software engineering firm have increased by 50% and output by 60%, there are increasing returns to scale in software engineering. This might occur because in the larger workforce, some of the programmers can concentrate more on particular kinds of programming and get better at them. This is also known as the "economies of scale," since production is cheaper when the scale is larger.

In introductory economics, these long-run tendencies are usually discussed in the context of cost analysis rather than marginal productivity analysis. However, increasing returns to scale, in particular, creates some complications for the application of marginal productivity thinking. Thus, there may be something to gain by exploring how increasing returns to scale goes well with marginal productivity. To keep it as simple as possible, let us consider a numerical example of a two-person labor market and a fictitious product that is produced with increasing returns to scale. The economists often like to talk about the production of "widgets," therefore, this fictitious industry is the widget-tying industry.

Example of production with increasing returns to scale is as follows:

Assumptions

Since this is a long-run analysis, there is no fixed input. Indeed, for simplicity, there is only one input. Labor is the only input and is variable.

There are three people in a small economy. These three people are as follows:

- Bob (worker)
- John (worker)
- Gordon, an entrepreneur-- He organizes a business if and only if it is profitable to do so

Quantum of the work they perform is as follows:

- Bob can alone produce output worth 2000 per week
- Bob's opportunity cost is 2100 per week
- John can alone produce 2000 per week
- John's opportunity cost is 2800 per week

If Bob and John work together, they can produce 5500 per week. Suppose, for example, Gordon sets up a widget-tying business and hires Bob and, later, John to do the work. This is an example of "increasing returns to scale," since input increases by 100% when the second worker is hired and output increases by 175% as a result.

The output would increase more than in proportion to inputs because of the following reasons:

1. Simply having four hands may increase productivity as the two men can simultaneously do different parts of the job.
2. Each may concentrate on some part of the work, getting better at it with more practice but leaving the other part to the other worker who also gains practice and skill in that part. (These were the kinds of advantages Adam Smith particularly stressed)
3. Each may concentrate on the tasks for which he has a greater inborn talent.

It should be noticed that the two-person widget-tying operation uses resources with an opportunity cost of 2800+2100=4900 and produces output worth 5500 for a net increase in production of 600. Evidently, it is a good thing that such a team be organized.

The Dark Side of the Force

Increasing returns to scale are a powerful force for increasing productivity but the problem of organizing them efficiently is "the dark side of the force." According to the above-mentioned example, an enterprise that yields a net gain of 600 to society cannot be organized without producing a loss. The market system cannot take advantage of the potentiality for gain through division of labor and increasing returns to scale in this case.

This possibility was discovered by an early 20th century British economist, Arthur Charles Pigou. Despite after 80 years of discussion, this analysis is not at all widely understood, even among professional economists. Pigou thought it might be a good idea for the government to subsidize enterprises with increasing returns to scale. In this case, a subsidy of 150 would make the widget-tying enterprise profitable and produce a gain of 600 in national product.

There may be another solution. Since the widget-tying enterprise adds 600 to national output but loses atleast 100, what happens to the difference of 700? Then, Bob will get it. Bob is paid atleast 2800 but his opportunity cost is only 2100, accounting for the difference of 700. Suppose, Bob and John were not paid the same wage and each was paid his opportunity cost plus 100. Then, the wage bill would be 2200+3000=5200 and Gordon would finish with a profit of 300. Thus, wage discrimination may make it possible for the widget-tying enterprise to exist when it cannot exist so long as each worker is paid the same wage for the same work.

Although the conclusions are surprising and controversial yet the number of people support them. Some people believe that each person should be paid according to her or his contribution and interpret “marginal productivity” as a person’s contribution. However, this may impossible when there are increasing returns to scale as there may not be enough output to pay everyone on that basis. Some people also say that each person ought to be paid in proportion to her or his contribution, so that people are paid equally for the same work. That, too, may be impossible.

Discrimination or subsidy may be necessary to allow some socially useful activities to exist. There may be no simple system of payment (such as, supply and demand or equal pay for equal work) that will allow a socially useful enterprise with increasing returns to scale to exist.

This is the reason for the existence of organizations. If there were no increasing returns to scale, there would be a little reason for any business to employ more than one person. There would instead have an economy consisting of self-employed individuals like a yeoman agricultural system. Instead, there would be an economic system consisting in part of large, complicated organizations with internal arrangements and payments systems that have little to do with contributions or marginal productivity and may be discriminatory. From an abstract point of view, they may waste resources by not paying at the marginal productivity. But the benefits of increasing returns to scale are so large that even falling for short of potential efficiency, they can still be very productive.

People naturally avoid complexity and organizations sometimes try to set up simple, market-like internal payment and fund transfer systems. They do it with a hope that this will increase efficiency. But this can fail badly in the context of increasing returns to scale.

One can refer to an experience at Drexel. A few years ago, we went over to “revenue centered budgeting.” The idea was to let the colleges retain a high proportion of the

revenue they produce through tuition, grants, contracts and so on. We felt that this would give the deans and college faculties more "incentives" to set up new programs and initiatives. However, it was not possible to let the colleges keep 100%, since some money was needed to run shared services like the computer center, student-life activities and the library, not to mention the salaries of high administrators. But it could not be made to work. If the proportion kept by the colleges was high enough to make it profitable for them to set up new programs and initiatives, there was not enough for the purposes of the central administration. If the proportion taken by the central administration was enough to do its job, the colleges were losing money on their new programs and initiatives. So, Drexel moved away from "revenue centered budgeting" in practice, although there is still some work being done to work out a "revenue centered budgeting" system that will work.

Following is a prediction based on the theory of increasing returns to scale:

A revenue-centered budgeting system probably can be made to work, but it will be just as complex and frustrating than the centralized budgeting traditionally has been. That complexity and frustration (and large organizations) are the price which one pays for the benefits of increasing returns to scale.

Summary

The concept of marginal productivity and the Law of Diminishing Returns play a central roles in both the efficient allocation of resources in general and in profit maximization in the John Bates Clark model of the business firm.

The John Bates Clark model and the principle of diminishing marginal productivity provide a good start on a theory of the firm and of supply. In applying the marginal approach and the Equimarginal Principle to profit maximization, it extends the understanding of the principles of efficient resource allocation. Some key points of the lesson are as follows:

- The distinction between marginal productivity and average productivity
- The "Law of Diminishing Marginal Productivity"
- The rule for division of a resource between two units producing the same product, i.e. equal marginal productivities
- The diagnostic formula, i.e. VMP=wage, that tells us the input and output are adjusted to maximize profits in the business firm in the short-run
- In the long-run, there may be increasing, decreasing, or constant returns to scale
- Increasing returns to scale complicates things somewhat for the marginal productivity approach