

ECON 600
Lecture 3: Profit Maximization

I. The Concept of Profit Maximization

Profit is defined as total revenue minus total cost.

$$\Pi = TR - TC$$

(We use Π to stand for profit because we use P for something else: price.)

Total revenue simply means the total amount of money that the firm receives from sales of its product or other sources.

Total cost means the cost of all factors of production. But – and this is crucial – we have to think in terms of opportunity cost, not just explicit monetary payments. If the owner of the business also works there, we must include the value of his time. If the firm owns machines or land, we must include the payments those factors could have earned if the firm had chosen to rent them out instead of using them.

If only explicit monetary costs are considered, we get accounting profit. But to find economic profit, we need to take into account the opportunity cost, implicit or explicit of all resources employed.

The main constraints faced by the firm are:

- technology, as summarized in the cost curves of the last lecture;
- the prices of factors of production, also taken into account by the cost curves; and
- the demand for its product.

II. Demand Curve Facing the Firm

The firm's demand curve tells how much consumers will buy at each price from a particular firm. (This is distinguished from other kinds of demand curve, such as the market demand curve, which shows how much consumers will buy at each price from all firms put together.)

The shape of the firm's demand curve is related to the degree of competition in the market. Loosely speaking, more competition causes the firm's demand curve to be more elastic (flatter), because consumers can respond to price increases by shifting their purchases to other firms. Less competition, on the other hand, implies a more inelastic (steeper) demand curve. Perfect competition and monopoly turn out to be the extreme ends of the spectrum:

- a perfectly competitive firm faces a perfectly horizontal demand curve;
- a monopoly faces the whole market demand curve.

III. Total and Marginal Revenue

Total revenue (TR) is the total amount of money the firm collects in sales. Thus,

$$TR = pq$$

if p is the price and q is the quantity the firm sells. Notice that I'm using a small q , because this is just one firm (Q is reserved for the market as a whole).

If the firm faces a downward-sloping demand curve, picking a quantity q automatically implies picking a price p . Why? Because the firm must be operating on the demand curve. Any chosen price corresponds to a specific quantity that consumers will buy at that price, and any chosen quantity corresponds to a specific price that will induce consumers to buy the chosen quantity. To sell more, the firm must lower its price.

Graphically, TR is represented the rectangle created by p and q .

Marginal revenue is the change in total revenue from increasing quantity by one unit. That is, $MR = \Delta TR / \Delta q$, where the change in q is usually one.

Now, you might think that MR must be equal to the price, p , because that's how much you get paid for selling one more unit. But this is not true in general. Why not? Because if you face a downward-sloping demand curve, you have to lower your price to sell more. So if you increase your quantity, you're also lowering your price for all the previous units of the good.

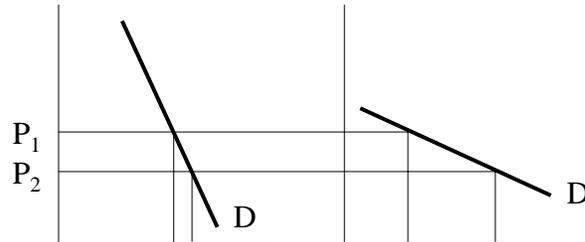
Example: Suppose you're currently selling 10 units at \$20 each. To sell an 11th unit, you'll have to lower the price to \$19. So, you gain \$19 from the additional unit. But you also lose \$1 each on the previous ten units that you could have sold at \$20 each. So your marginal revenue from the 11th unit is not \$19, but rather \$9. You can see this from looking at the total revenue: before, $TR = 20(10) = 200$, but now $TR = 19(11) = 209$. Thus, TR rose by only \$9.

This analysis leads to the following general conclusion: that MR is always below the demand curve. Why? At any quantity, the demand curve tells us the price corresponding to that quantity. But we've just shown that the MR must be less than the price, and hence below the demand curve. The only place MR and the demand curve are equal is the point where the cross the vertical axis.

In general, any time a firm lowers its price, there are two effects: the "people buy more units" effect, and the "people pay less per unit" effect. Which effect is larger determines whether MR is positive or negative. It turns out this is closely related to the price elasticity of demand.

IV. A Digression: Price Elasticity of Demand

Elasticity refers to the degree of responsiveness of one variable to another. It's not enough to say, for instance, that a rise in price leads to a fall in quantity demanded (the Law of Demand); we want to know how much quantity changes in response to price.



A simple way to see the degree of responsiveness is simply to look at the slope. A flatter demand curve represents a greater degree of responsiveness (for a supply or demand curve), as shown in the above graphs: the flatter demand curve produces a larger change in quantity for the same change in price.

Using just the slope is the quick-and-dirty way to think about elasticity. The extremes are easy to remember: A perfectly elastic demand curve is horizontal, because an infinitely small change in price corresponds to an infinitely large change in quantity; the graph looks like the letter E for elastic. A perfectly inelastic demand curve is vertical, because quantity will never change regardless of the change in price; the graph looks like the letter I (for inelastic).

But using the slope can be misleading, because it doesn't tell us the significance of the quantities. Suppose a \$1 dollar increase in price leads to almost everyone choosing not to buy the good. That would not surprise us for gumballs, but it would certainly surprise us for televisions. The point is that a \$1 increase is not much relative to the total price of TVs, but it is huge relative to the total price of gumballs. This is why we use elasticity instead of just the slope.

Definition of price elasticity of demand: the percentage change in quantity demanded divided by the percentage change in price. That is,

$$E_d = |\% \Delta Q_d / \% \Delta P|$$

How do you find the percentage change in something? You find out how much it changed, and divide by the initial value. For example, suppose your income rises from \$400 a week to \$500 a week. The change is \$100, so the percentage change is $\$100/\$400 = .25$ or up 25%. N.B.: The percentage change depends on the direction you're going. If your income went from \$500 to \$400, the percentage change would be $-\$100/\$500 = -.2$ or down 20%. [Note: For this reason, some textbooks use a slightly different formula. When calculating the percentage change in a variable, instead of dividing by the original

point, they divide by the average of the two endpoints. In this class, we will use the original point as described above.]

Thus, we can also write

$$E_d = |(\Delta Q_d / Q_d) / (\Delta P / P)|$$

Example: Suppose that at a price of \$10, consumers will buy 80 units. If the price rises to \$10.50, the consumers will reduce their purchases to 72 units. The percent change in quantity is $(80 - 72)/80 = .1$ or 10%. The percent change in price is $(10.50 - 10)/10 = .05$ or 5%. So the price elasticity of demand on this section of the demand curve is $10\%/5\% = 2$. In words, this means that a 1% increase (decrease) in price corresponds to a 2% decrease (increase) in quantity.

Now, we can rearrange the formula for elasticity like so:

$$E_d = |(\Delta Q_d / \Delta P) \times (P / Q_d)|$$

Look at the first term, change in Q over change in P. This is basically the slope of the curve. I say “basically” because when we talk about the slope of a line, we usually measure the rise (vertical distance) over the run (horizontal distance). In a supply and demand graph, we usually measure price vertically. So actually, the slope is change in P over change in Q. What we have here is the inverse of the slope. If we use m to stand for the slope, we have:

$$E_d = |(1/m) \times (P/Q_d)|$$

This is the easiest formula to use when you have a straight line for a demand curve.

Example: You have the demand curve $P = 50 - .1Q_d$. The slope is -1 . Using our formula, the slope at the point (100, 40) is given by $E_d = |(-1/.1)(40/100)| = 4$. The slope at the point (200, 30) is given by $E_d = |(-1/.1)(30/200)| = 1.5$. Notice that the slope is not the same at every point on a line.

Note: The previous method for calculating elasticity, in which we considered a discrete change in price and quantity, gives us “arc elasticity.” With arc elasticity, you are finding the elasticity over a section of the demand curve. The method just demonstrated, on the other hand, gives us “point elasticity.” Point elasticity tells you the elasticity at a single point – specifically, at the (P, Q_d) point you plug in. Think of point elasticity as the elasticity for an interval (change in price) that is very, very small.

We have the following definitions:

When $E_d > 1$, we say the curve is elastic at that point.

When $E_d < 1$, we say the curve is inelastic at that point.

When $E_d = 1$, we say the curve is unit elastic at that point.

[Note: Textbooks differ on whether to take the absolute value or not. If you don't take the absolute value, you'll get a negative elasticity, which means that the demand curve is downward sloping. But to keep things positive, we'll always take the absolute value.]

In general, any demand curve will have one point that is unit elastic, which means that a one percent change in price corresponds to a one percent change in quantity.

Example: continued from above. We can find the unit elastic point by setting

$$\begin{aligned}
 E_d &= 1 \\
 |(-1/.1)(50 - .1Q)/Q| &= 1 \\
 (50 - .1Q)/Q &= .1 \\
 50 - .1Q &= .1Q \\
 50 &= .2Q \\
 Q &= 250
 \end{aligned}$$

So the point $Q = 250$, $P = 25$ is the unit elastic point.

How does all of this relate to the firm's MR curve, which is what we were talking about? It turns out that whether the MR is positive or not depends on whether the demand curve is elastic or not, as shown in the following grid.

elastic demand	1%Δ in price → more than 1%Δ in quantity	“people buy more” outweighs “people pay less”	MR > 0 (above horizontal axis)
unit elastic demand	1%Δ in price → 1%Δ in quantity	effects exactly cancel out	MR = 0 (crossing horizontal axis)
inelastic demand	1%Δ in price → less than 1%Δ in quantity	“people pay less” outweighs “people pay more”	MR < 0 (below horizontal axis)

When the demand curve is a straight line, the MR always has a slope exactly twice that of the demand curve, so that it crosses the horizontal axis exactly half-way between the origin and the point where the demand curve does crosses the axis. Why? Because it turns out that a linear demand curve is always unit elastic exactly half-way down its length, and therefore MR crosses the horizontal axis at a quantity exactly half-way down the demand curve.

So, if I give you a demand curve $P = b - mq$, what's the MR? $MR = b - 2mq$.

V. The Profit-Maximizing Decision

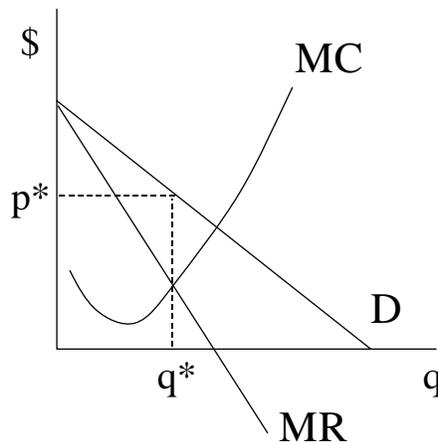
The firm uses MR and MC to decide how much to produce.

- Suppose increasing output by one unit will bring in more additional revenues than it costs to produce. That is, $MR > MC$. Then it makes sense to produce the unit, because doing so will create more added benefit than added cost.
- On the other hand, suppose your last unit cost more to produce than it brought in additional revenue. Then $MR < MC$. And it makes sense to cut back your production

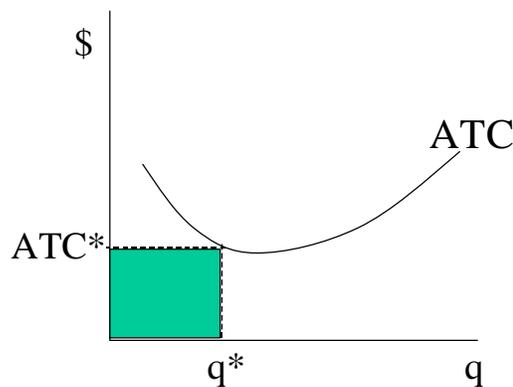
(not produce that unit after all), because the last unit created more added cost than added benefit.

- So the only point at which you're satisfied is where $MR = MC$. This is the general rule for profit maximization, which we'll use with any firm, whether a perfect competitor, monopolist, or anything in between.

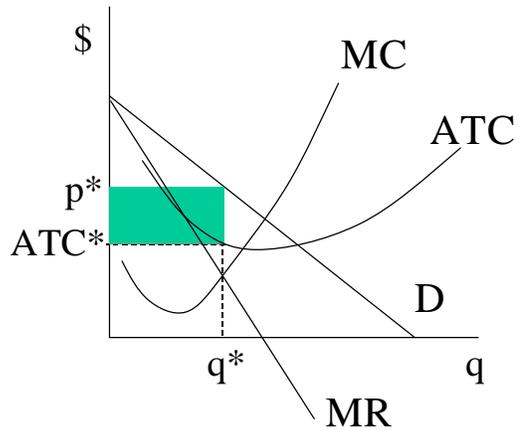
Graphically, we find the quantity q^* where the MR and MC curves cross. What price corresponds to this quantity? We find it by extending q^* up to the demand curve to get p^* . No other price will work: a higher price will reduce sales below q^* , and a lower price will increase sales above q^* .



How much profit does the firm actually make? To find this out, we need more information about the firm's costs. We can get all the information we need from ATC. Recall that $ATC = TC/Q$. Therefore, $TC = ATC \times Q$. Graphically, we can see this as the rectangle created by q^* and ATC^* on the ATC curve.



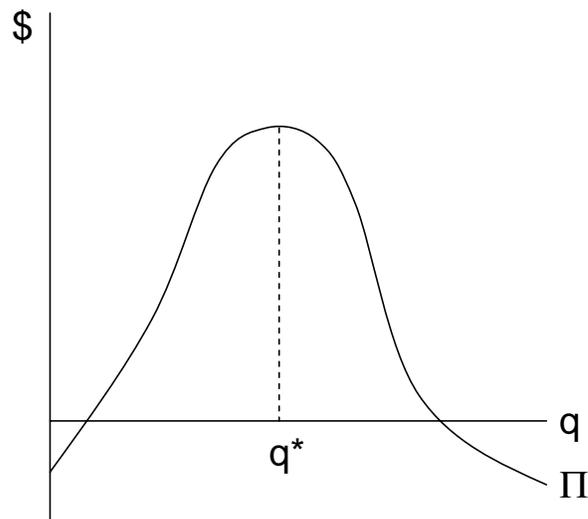
Profit is defined as the difference between TR and TC. TR, as we saw before, is the rectangle p^*q^* . TC is the rectangle ATC^*q^* . So the difference in the areas of these rectangles is profit, as shown below.



As drawn, the firm is making profits, but it needn't be such a happy outcome. If the ATC curve had been higher, we could have gotten at ATC^* greater than the price. When this happens, the firm is making negative profits, also known as losses. This can happen even though the firm is doing the best it can, because any choice other than q^* would make it lose even more (with the possible exception of shutting down entirely – something we'll take up later).

VI. The Profit Function

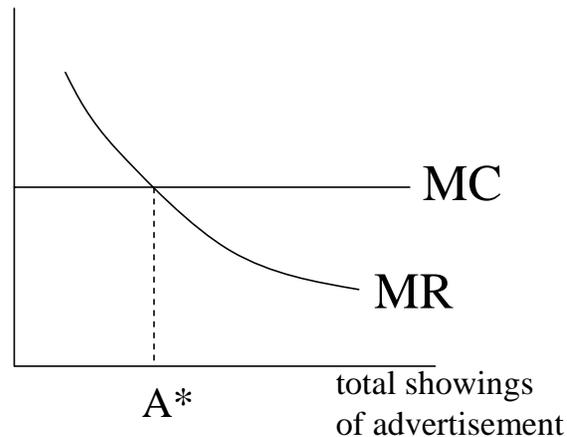
The MC-MR approach yields the same optimal quantity that you would get if you looked directly at a profit function. A profit function simply shows how much profit you make at each level of output.



The very highest point on this curve occurs at q^* , which is also exactly the quantity at which the MC and MR cross.

VII. Applications of $MC = MR$

The $MC = MR$ rule is a very versatile one, which can be applied to many other decisions by firms other than how much to produce. It can be applied to hours of operation: stay open as long as the added revenue from the additional hour exceeds the cost of staying open another hour. Or it can be applied to advertising: increase the number of times you run your TV commercial as long as the added revenue from running it one more time outweighs the added cost of running it one more time. Assuming there is a constant price per 30-second commercial, the situation might look like this:



This is just one of the many ways in which $MR = MC$ rule can be used to make business decisions other than choosing the profit-maximizing quantity.

Also, notice that average cost did not play any role in the choice of output. All that matters is the marginal effect on profits of increasing or decreasing your output. If a firm fails to follow the $MC = MR$ rule and uses average cost instead, it will lose profits.

Example: The case of Continental Airlines. An airline has to decide how many flights to run during any given period of time. In the early 1960s and before, airlines typically made this decision by asking whether the additional revenue from a flight (the MR) was greater than the per-flight cost of a flight. In other words, they used the rule $MR = TC/q$. But Continental broke from convention and started running flights even when the added revenues were below average cost. The other airlines said Continental was crazy – but Continental made money hand over fist. Eventually, the other airlines followed suit. What made Continental’s strategy superior? The per-flight cost consists of variable costs, including jet fuel and pilot salaries, and those are very relevant to the decision about whether to run another flight. But the per-flight cost also includes expenditures like rental of terminal space, general and administrative costs, and so on. These costs do not change with the number of flights, and therefore are irrelevant to that choice.

In order to put $MC-MR$ thinking to its greatest use, we need to adopt a broader notion of fixed and variable inputs. So far, we’ve thought about them in terms of time: variable inputs can be changed in a given period of time, while fixed inputs can’t. But a better way of thinking about them is in terms of “the meeting you’re at” – that is, what kind of a

decision are you trying to make? Some inputs are fixed with respect to some decisions, while they are variable with respect to others. For whatever kind of decision you're making, only the inputs that are variable with respect to that decision are relevant.

Example: Airlines again. When the airline is deciding whether to run an additional flight, the costs of G&A, reservation systems, etc., are fixed. The costs of jet fuel, pilots' wages, meals, flight attendants' wages, etc., are variable. But now suppose the airline has already decided to run a particular flight, and now they're deciding how many seats to sell (that is, what price to charge per seat). For this decision, the costs jet fuel and pilots' wages are fixed, while only the costs of meals and flight attendants' wages (assuming more customers require more flight attendants) are variable.

Example: If you're deciding how many hours your employees should work, their wages are variable costs, but their fixed benefits packages (e.g., health insurance) are fixed costs. If you're deciding whether to hire a new employee, both wages and benefits are variable costs.

VIII. The Shut-Down Decision

There is one exception to the rule that average costs are irrelevant in deciding how much to produce: the decision about whether to produce at all.

You should think about the $MC = MR$ rule as the rule you should follow if you've decided to produce at all. But it might make sense to shut down the firm instead. How should the firm make this choice? You might think that the firm should shut down if it's making losses, but that's not quite true. Why? Because even if the firm shuts down, it will still have to pay its fixed costs (in the short run). So what really matters is whether the firm's revenues cover its variable costs. If the answer is yes, the firm should stay open even though it's making losses – because its revenues still partially offset the fixed costs. But if the answer is no, the firm should shut down – because the choice to produce is just adding even more losses on top of the fixed costs.

In other words, it must be the case the firm's total revenues (TR) exceed the firm's total variable costs (TVC).

$$\begin{aligned} TR &\geq TVC \\ p \times q &\geq AVC \times q \\ p &\geq AVC \end{aligned}$$

This shows that the firm should stay open if the price (at its optimally chosen quantity) is greater or equal to its average variable cost. But if the price is below the average variable cost, the firm should shut down.

Note that this is a short-run shut-down decision. In the long run, there are no fixed costs, only variable costs. That means the firm can avoid making any losses at all by shutting down, so the firm should shut down any time it's making economic losses -- unless there

are changes the firm can make to its scale that allow it to escape those losses. This may be possible because the firm can shift to a different SRATC curve via a different capital choice.