Introduction

Traditional distinction between microeconomics (small) and macroeconomics (big):

Microeconomics is usually subdivided in the study of the consumer and the study of the producer: so we study individual *economic agents behaviour*.

How do we think about the consumer then?

To us a consumer does not have needs, we think of a consumers as somebody who is rational, has a bunch of desires which are possibly limitless and has *a bunch of (un sacco)* resources (illimited).

So, the problem of the consumer is to allocate scarce resources, for instance his income, towards different ends: it's a problem of choice. Our consumers choose.

On the other hand our enterprises also choose: they are institutions designed to produce goods and services.

There is somebody whom we call an entrepreneur, who organizes factors of production

and the task of the entrepreneur is to organize factors of production in such proportion as to be able

to generate any given level of output at the possible lower cost.

So, the difference between what the entrepreneur spends to produce and the income he makes

from selling the goods and services that are produced we call a profit.

In economy people produce for a profit, so consumers maximize their own satisfaction, producers maximize profits.

In **macroeconomics** we study the great aggregates of the economy: unemployment, inflation, international trade. We don't concern ourselves with individuals behaviour. We want to know what is the relationship between interest rates and investment. We want to know why some countries export some goods and other countries export other goods instead. The macroeconomies we compare across countries: why do countries differ, why there is some rate of unemployment in the US and a different rate of unemployment in Europe at this stage.

Supply and Demand

OVERVIEW SUPPLY AND DEMAND

Basic supply and demand graph referred to the market for pork: in this pork market, you have a demand curve and a supply curve.

Figure 2-1: Pork market equilibrium

Each point in the demand curve represents the price consumers are willing to pay for that quantity and it's downward sloping, because as the price goes up, consumers are willing to buy less of a good. So their willingness to pay changes. The supply curve represents how much they're going to charge for a given quantity of the good and is upward sloping because as the price rises, they're willing to supply more. As the price rises, producers produce more. So as the price rises, consumers demand less.

And equilibrium is that point where supply equals demand. Equilibrium equals happiness. It's the point where suppliers and demanders are both happy, because at a point such as **e**, the amount that consumers demand at that price is equal to the amount suppliers are willing to supply at that price.

You've reached a point where consumers want a certain amount at a certain price. At that price, producers say, great, you want **e**, I'm happy to produce **e** at that price.

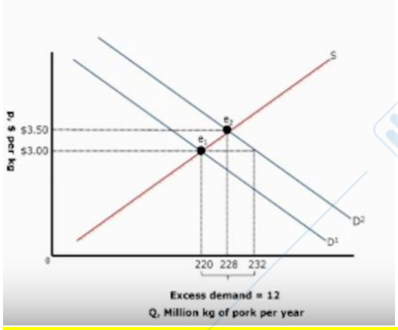
IMPACT OF A DEMAND SHIFT

Let's talk about what happens when we shock that equilibrium. This is the market for pork, imagine that the price of beef rises. When the price of beef (manzo) rises, the effect on people's demand for pork is its increase because they're substitutes. What's determine these demand curves is the substitutability across goods.

So here we have a situation where the substitute for pork has gotten more expensive: as a result, people want more pork.

That is a shift out/shift up in the demand curve: a shift up or out, depending on it's out into space or up vertically in the demand curve.
So what happens here is that folks shift to consuming pork, so they want more of it.

Figure 2-2: Impact of a demand shift

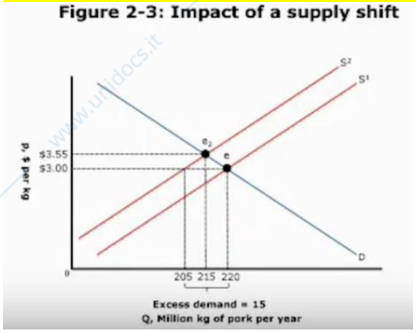


So we're initial equilibrium at 220 millions of kilograms of pork a year and a price of $3 a kilogram. That was the initial equilibrium, the point where demanders and suppliers were happy.
Now the price of beef has gone up and people want more pork. That shifts the demand curve out.
Initially, if the price remained at $3 people would now want 232 millions of kilograms of pork a year, a lot more pork.
So what you're going to have initially is excess demand, because at that price of $3, suppliers are only willing to supply 220 million kilograms and they were happy at point e1.
They're not happy to provide more at that price: if we want more, they are going to have to charge a higher price.
So they say, if you want more pork, we're going to have to produce more. So we are going to slide up the supply curve and we're going to charge you a higher price if you want more.
Well, as the price goes up, consumers say, well, wait a second. If the price is going up, I don't want quite as much more. And you meet at the new equilibrium e2.
Consumers go back up the demand curve and don't quite want as much at that higher price and the new equilibrium is e2. That's where consumers and producers are now happier with that outward shift in demand curve.
But the key point is consumers are not getting as much as they originally wanted at the original price, because the original price will not hold. The price is going to change. Given the price is going to change, the quantity is going to change. And you end up with both a higher price and a higher quantity for pork. The price of beef rising raised the price of pork. It didn't just increase the demand for pork, but through increasing the demand, it raised the price.
So price in one market affects the price in another market, not just the quantity but the price in another market.
And that's our new equilibrium. That's the shift in the demand curve. They're happy at this new point e2, this new equilibrium, because given the new demand curve, at a price of $3.50 a kilogram, producers are happy to produce the 228 million kilograms of pork a year that consumers want.
And at that price, that's what consumers want.

IMPACT OF A SUPPLY SHIFT

Let's imagine there's a pig disease so suddenly, it becomes more expensive to produce pork, because there's fewer pigs, it's more expensive to produce it. It causes a shift in in the supply curve, inward or upward of the supply curve.

Figure 2-3: Impact of a supply shift



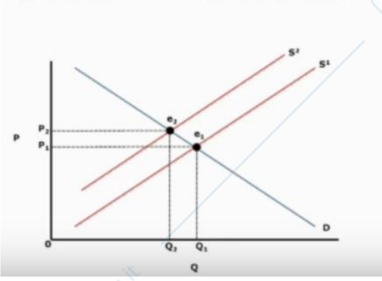
The fact that it's more expensive to produce pork, the suppliers need to get paid a higher price.
Initially, if the price stayed constant at $3.00 we would not begin with excess demand, because then consumers would want 220 million kilograms of pork, which is what they wanted before. Nothing's changed from their perspective.
But producers can only now produce 205, because supply curve shifted up.
So once again, you have to move to a new equilibrium, where suppliers and consumers are happy.
That will happen at a point like e2, where given the higher price of producing pork, producers will now sell 215 million kilograms at $3.55 and consumers don't want quite as much as they wanted before but only 215. And they're happy.
We had two very different phenomena: we had a demand shift and a supply shift, both led to a higher price.
Demand shift led to higher demand, to higher quantity sold in the market, the price went up, and the quantity in the market went up from 220 to 228.
Supply shift led to higher price almost exactly the same amount, but here, the quantity fell. *So you can't tell from a price increase what happened.* If the price of pork goes up, you can't tell me whether that was a demand or supply shift. You need to know both the price and quantity to be able to tell me that.
Both changes led to the same outcome in terms of prices.

THE ELASTICITY OF SUPPLY AND DEMAND

What determines the shapes of supply and demand curves, how we think about supply and demand interacting in a market and what determines how responsive individuals and firms are to prices.

The different effects on quantity and the size of shifts on supply and demand curves come from the shapes of the supply and demand curve. We'll talk both theoretically about what determines these shapes and empirically about how economists go about figuring out the shapes of supply and demand curves. So, to think about this, let's start with Figure 3-1, which is a standard market diagram.

Figure 3-1: Market equilibrium with supply shift



With an initial equilibrium at point E1, with an initial price P1 and a quantity Q1. That's a stable equilibrium, because at that price P1, consumers demand Q1 units, and suppliers are willing to provide Q1 units.

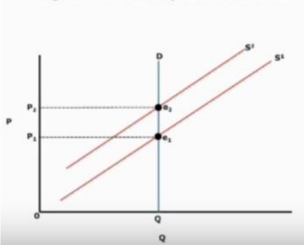
Now we have some supply shift. So the supply curve rises to S2. At that new price, initially, you would have excess demand. But quickly the price increases to shut off that excess demand. And you end up with a new equilibrium with a higher price, P2, and a lower quantity Q2, and new equilibrium point E2.

What determines the size of that shift from Q1 to Q2 and that price increase from P1 to P2? What's going to determine it is the elasticity of supply and demand. **The elasticity** of supply and demand is how much do supply and demand respond, the quantities supplied and the quantities demanded respond, when the price changes.

When we say, how elastic is demand, what we mean is how sensitive to price is the quantity demanded. Or, alternatively, what is the slope of that demand curve? So the slope of the demand curve will be the sensitivity of quantity demanded to the price consumers face. And that will determine the market responsiveness.

We look at the extremes that don't exist in the real world. So let's think about one extreme case in Figure 3-2. Let's think about the case of **perfectly inelastic demand**: there's no elasticity of demand, demand for a good is unchanged regardless of the price. So perfectly inelastic demand is a case where demand for the good is unchanged regardless of the price. That would lead you to have a vertical demand curve at a given quantity. What this says is regardless of the price, people always demand Q.

Figure 3-2: Perfectly inelastic demand

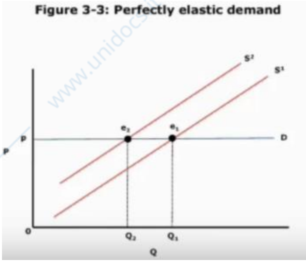


When there's no substitutes, when there's nowhere to go, it doesn't matter what the price is. When there's no substitutes, demand will be perfectly inelastic, because you have to have Q.

Now, what happens with inelastic demand when there's a supply shock? When supply increases, what happens? In that case, there can never be excess demand, because demand doesn't change. So all that happens is price just increases and no change in quantity.

Now, let's consider the opposite. Let's look at Figure 3-3 and think about perfectly elastic demand.

Figure 3-3: Perfectly elastic demand



Perfectly elastic demand is demand where consumers, essentially, don't care about the quantity, they just care about the price. A perfectly elastically demanded good would be one where there are, essentially, perfect substitutes. Technically, if a good is perfectly elastically demanded, then you are completely indifferent between that good and a substitute. If you're completely indifferent, then if the price changed at all, you would immediately switch and so the price can't change. So if there's a supply shock to a provider that's facing a perfectly elastic demand curve, they cannot raise their price, because people will just switch, so quantity will fall a lot.

So with perfectly inelastic demand, the quantity didn't change. With perfectly elastic demand, we saw a big quantity change. So, more generally, *what determines the quantity change in response to a price change is the elasticity*. More generally, we're between these two cases of perfectly elastic and perfectly inelastic. And what's going to determine the price change is the price elasticity of demand that is the percentage change in quantity for each percentage change in price or, in calculus terms, $\frac{dQ}{dP}$.

$$\boldsymbol{\varepsilon = \frac{\Delta Q/Q}{\Delta P/P}}$$

So the price elasticity of demand will typically be between 0 and negative infinity. And the larger it is the more quantity will change when prices change.

Consumer Theory

INTRODUCTION TO CONSUMER THEOR

We're gonna talk about where the demand curve comes from. In economics all consumer behavior comes from utility maximization. That's the basic building block of consumer behavior and an overview is composed by:

- We posit consumer preferences: what consumers would like.
- We posit some budget constraint: what resources consumers have to get what they'd like.

And then we do a constrained maximization problem that says, given your preferences, given what you'd like, subject to the resources you have available, what choices will you make? And in particular, the term we'll use a lot, what bundle of goods makes you the best off? Given your preference, given your constraints, what bundle of goods?

So think about consumers choosing across a set of goods. Typically, we'll think about two goods because graphs are easier to think about two dimensions than more. So we'll typically think *about trading off two goods*.

So think about *consumers with preferences across two goods*, some budget they can allocate, and how they make those choices. But this basic framework applies to the multiplicity of choices we all make along many, many dimensions. So doing two dimensions to make as just one of the simplifying assumptions.

So basically we we're going go through three steps:

1. What assumptions we make about preferences (preference assumptions): the axioms that underlie how economists model consumer preferences.
2. How we translate these preferences assumptions into mathematical tractability through the use of the utility function, which is basically a mathematical representation of underlying consumer preferences. So we'll talk about how we basically take these preferences and translate them into something that we can work with here at MIT by making it mathematical, by making a utility function.
3. Budget constraints.

And armed with these three things, we'll then be able to model how consumers make decisions.

CONSUMER PREFERENCE ASSUMPTIONS

To model consumers' preferences across goods we're going to impose three preference assumptions:

1. Completeness: when comparing two bundles of goods, you prefer one or the other. You can't say, I'm not sure.
2. Transitivity: if you prefer x to y, and y to z, you've got to prefer x to z.
3. Assume non-satiation: more is always better, you never would turn down having more.

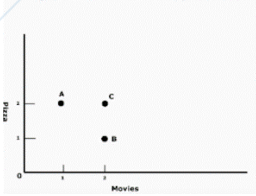
PROPERTIES OF INDIFFERENCE CURVES

Indifference curves are our name for what you could also think of as preference maps: are the graphical representation of people's preferences which we do through graphics that we call indifference curves.

So now let's go to the example of a decision you have to make: imagine your parents gave you some money and you had to decide whether to buy pizza or see movies. That's your decision. That's the trade-off you're making.

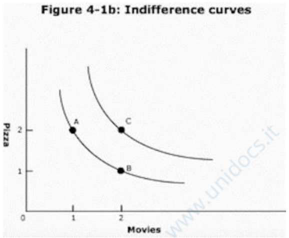
We're in a world with only two goods, pizza and movies and you're deciding how to allocate the money your parents gave you over pizza and movies.

Figure 4-1a: Pizza/movie preferences



Consider three choices of pizza and movies in figure 4-1a.
You could have two pizzas and one movie, that's point A.
You could have one pizza and two movies, that's point B.
Or you can have two of both, that's point C.
That's just three choices you're facing. Once again, we're ignoring paying for them, not considering budget constraints. Now we're just saying I'm giving these three choices, how do you feel about them? Well let's assume that you're indifferent between two pizzas and one movie, and one pizza and two movies. But clearly you like two pizzas and two movies better than either the first two combinations. Then what we can do is we can draw what we call indifference curves.

Figure 4-1b: Indifference curves

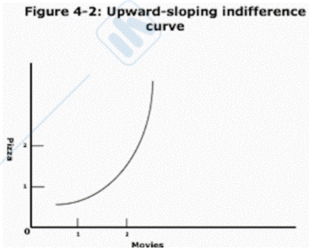


These are maps of your preferences: an indifference curve is the curve showing all combinations of consumption along which the individual is indifferent. So you have an indifference curve between A and B.

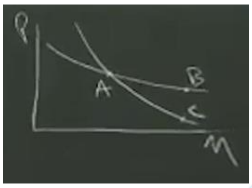
That means that all combinations along this curve are indifferent, you're equally happy getting two pizzas and one movie or one pizza and two movies. But point C, which is two pizzas and two movies is on a different indifference curve.

You're not indifferent between point C and points A and B, you clearly like two pizzas and two movies better than one of one and two of the other.
So armed with those assumptions, there are four key properties of indifference curves that we have to keep track of:

- The first is that consumers prefer higher indifference curves: the further out indifference curve from the origin, the more you prefer it and this comes naturally from the non-satiation assumption. Given that we've assumed non-satiation, you must always prefer an indifference curve that's further from the origin because it's more, and more is better.
- The second point is that indifference curves are always downward sloping. And that, once again, comes from non-satiation. To see this, let's look at the an upward sloping indifference curve. Upward sloping indifference curve violate non-satiation because you're indifferent to getting more. Because this would say you're indifferent between (1,1) and (2,2) that would say you're indifferent between getting one pizza and one movie or two pizzas and two movies. You can't be because that violates more is better. So indifference curves can't be upward sloping, they've got to be downward sloping by the non-satiation assumption.



- The third property of indifference curves is indifference curves cannot cross. Because imagining a situation where you have your pizza and your movies and two indifference curves that look like these. Transitivity says I must then be indifferent between B and C through the logic you just laid out. But I can't be indifferent between B and C because B dominates C. B has a basically the same number of movies, but more pizza, so I must like B better. So by the combination of transitivity and non-satiation indifference curves can't cross.



- And finally completeness that simply means you can't have more than one indifference curve through a point. So basically, every possible bundle has one indifference curve, you can't have two indifference curves through it and you are not sure which indifference curve I'm on. There's one indifference curve through every bundle. There's not two indifference curves through a bundle.

UTILITY FUNCTIONS

Now everything you need to know about preferences is represented in those indifference maps.
The problem is they're pretty awkward to work with when we need to actually prove theorems and solve and understand how people make decisions. That's a lot easier if we have a mathematical representation of those preference maps. And that's the utility function: *the utility function is a mathematical representation of people's underlying preferences*. And the key thing is that we assume individuals have these well-defined utility functions, and by maximizing those utility functions we can tell what choices they're going to make.
So for example, suppose that I said that your utility function over pizza and movies was the square root of pizza times movies. $u = \sqrt{p * m}$
It's a mathematical representation of your preferences. It tells us about your preferences, your preferences can be represented.
Figure 4-1b tells us your preferences because you're indifferent between two pizzas and one movie and one pizza and two movies, both give a utility square root two. But you prefer two pizzas and two movies because that gives a utility of two. So this is a mathematical representation consistent with those utility indifference curves.
This is a mathematical representation of your tastes, utility means nothing in the sense that it is not a cardinal concept, it's only an ordinal concept. An utility of 2 doesn't mean anything it just means that you get more than from one pizza and one movie. And we can even get the ratio that you get square root of two more, than you get from one pizza and two movies. We can do ranking and ordinality, but we can't assign cardinality. I can't say how happy you are in some abstract absolute sense from two pizzas and one movie. I can't give a cardinal form preference, but this is an ordinal ranking of preferences. That's why utility function is a representation of indifference maps.

MARGINAL UTILITY

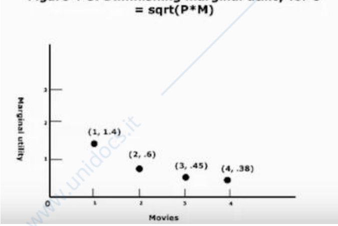
The key concept for consumer theory for understanding how consumers make decisions is the concept of marginal utility. *That is how your utility changes with each additional unit of the good, or the derivative of the utility function.*

If you want to do it in calculus terms, marginal utility is the derivative of your utility function with respect to one of the inputs. But if you don't want to put it in calculus terms, it's as you add each unit of one of the elements of the utility function, how does utility change.

So to see this, let's do an example of marginal utility: imagine for a moment that you have two pizzas, p=2. How does your utility change as you see additional movies? Figure 4.4 is showing how your marginal utility for movies evolves as you get more movies. Given that you have two pizzas, this is the evolution of your utility as you get more movies. So each additional movie increases your utility. The slope is positive, more is better, it still improves your utility, but at a diminishing rate. And that's the key is that we assume diminishing marginal utility. We assume that each additional movie increases your utility, but at an ever diminishing rate.

Figure 4-4: Diminishing marginal utility

Figure 4-5: Diminishing marginal utility for U = sqrt(P*M)



So basically, we can actually graph your marginal utility in figure 4-5.

So the very first movie gives you marginal utility of 1.4 because you go from 0 to square root of 2.

From square root of 2 to 2, you get 0.6 the next movie.

From 2 to square root of 6, you get 0.45 for the third movie.

For square root of 6 to square root of 8, you only get 0.38 from the fourth movie, and so on.

So the key point is that these marginal utilities are ever decreasing: each additional movie gives you less incremental utility.

It is intuitive: think about the movies you want to see right now. Presumably whichever you ranked first would give you more utility to see than whichever you ranked second. By the time you get to the fourth movie, you're not getting much utility from it at all.

You're getting a lot of utility from that first movie you see, but each additional one is giving you less and less.

Likewise with pizzas, if you haven't eaten all day, that first pizza can give you a very high marginal utility. The enjoyment you get from eating that first pizza can be very large, but the second pizza, not so much.

BUDGET CONSTRAINTS AND THE MARGINAL RATE OF TRANSFORMATION

Today we're going to continue our discussion of consumer choice. If you remember the set-up from last time, the main motivation is you're trying to understand what underlies demand curves, how consumers ultimately decide to trade off price and quantity of goods. We said that ultimately that came from the principle of utility maximization, and that utility is maximized when individuals maximize the utility function, which is this mathematical representation of preferences. And last time we talked about how if individuals were unconstrained how they choose what they want, they would just like more of everything, and their ranking across different bundles would depend on that underlying utility function.

Now, of course, what's stopping individuals from consuming everything they want is their budget constraints.

We assume that your income equals your budget, that is, you spend your entire income, ignoring the possibility of savings.

Let's image that you've got some budget saved from your parents, called y. These are your money you have to spend. And let's imagine that you have to allocate that budget only across two goods, pizza and movies. So that gives you your budget constraint. So how do you allocate that?

You can buy movies, the number of movies you can get, plus the number of pizzas and how many of each you can get depends on their price. In particular, budget constraint is the number of movies times the price per movie plus the number of pizzas times the price for pizza. $Y = M * Pm + P * Pp$

Figure 5-1: Budget constraint

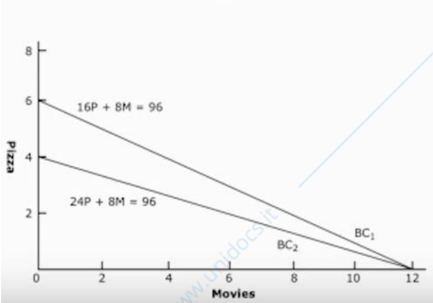


Figure 5-1 is a graphical illustration of a budget constraint.

The x-axis is how many movies you could have if all you did with your income was consume movies.

If all you did with your income was consume movies, you could have $M = \frac{Y}{Pm}$ movies.

If instead you decided to devote all your income to pizza, then you could have $P = \frac{Y}{Pp}$ pizzas. So the y-axis is going to be the point where you consume zero movies and all pizza.

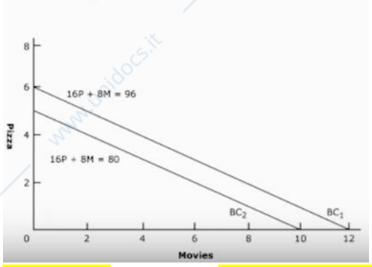
And then there'll be some combination in between, which is our budget line. Which is the combinations of pizzas and movies you can consume given your total income y. So basically the slope of that line is going to be the price ratio. Or the negative of the price ratio. The slope of that line is: $Budget\ Constrain\ Slope = MRT = -\frac{Pm}{Pp}$

So basically it's the negative of the price ratio: minus the price of movies over the price of pizzas because they're in the denominators as you said, because as the price goes up, the quantity goes down.
Imagine that income equals $96. Imagine that the price of movies is $8 and the price of a pizza is $16. This means that with your income of $96, you could either get 8 pizzas or 12 movies. So that means that the price ratio of the slope of your budget constraint is - 1/2.
The name for this slope is the marginal rate of transformation MRT*: it is the marginal rate at which you can transform pizzas into movies*. But the market essentially is giving you a rate at which you can transform a pizzas into movies. Given that you have a certain amount of money, $96, and given the prices that you face in the market, you could transform pizzas into movies by trading one pizza for 1/2 a movie. That's the trade-off that you face when you're trying to transform one to the other. So effectively, it's the same as if you're trading them for each other and that's because of the concept of opportunity cost. The opportunity cost *is the value of the forgone alternative, the opportunity cost of consuming one good instead of another*. Basically it means that if you decide to forgo a pizza, that's the same as forgoing two movies. So the opportunity cost of a movie, what essentially the movie is costing you, is 1/2 a pizza. And the reason is because you have a fixed budget. If you had an infinite budget, there'd be no opportunity cost. But because you have a fixed budget and you have to allocate that budget, thus there's an opportunity cost.

SHOCKING THE BUDGET CONSTRAINT

Let's talk about what happens when we shock the budget constraint.
Let's imagine the price of pizzas rose from $16 to $24, becoming really expensive.
Figure 5-2 shows our new budget constraint, instead of being 16P + 8M = 96, it's now 24P + 8M = 96. Or, more relevantly, the slope of the budget constraint has flattened from - 1/2 to - 1/3.
Now, forget utility for a second. Just looking at this, your opportunity set has been restricted by this price change. Your opportunity set is the set of choices you can make given your budget. Before, you could make choices all the way up to the upper line. Now your set of choices that are available have just fallen.
Now you're no poorer; it's not like your parents have cut you off, they still give you the $96. But you're effectively poorer, because the set of things you could afford with that $96 has just been restricted because the price has increased.
A price increase makes you worse off. It restricts your opportunity set, because with the same amount of income, you can now afford fewer goods.
Likewise, now let's talk about what happens when your income falls.

Figure 5-3: Decrease in income



Now, let's suppose your parents are pissed at you and they cut you down to $80.
The slope of the budget constraint has not changed, because what determines the slope of the budget constraint is prices, and no prices have changed. But your opportunity set is once again restricted because you now have lower income, so you can now afford fewer pizzas and movies. So now, instead of being able to afford up to 6 pizzas and up to 12 movies, you can now only afford up to 5 pizzas and up to 10 movies because your income has fallen. Your opportunity set has contracted.
So your opportunity set will contract whenever income falls or whenever price increases. And how it affects the graph will depend on whether it affects prices, which affects the slope, or just income, which affects the intercepts.

CONSTRAINED UTILITY MAXIMIZATION: GRAPHICAL ANALYSIS

We know what your preferences are, we've mathematically represented those by utility function. We know what your budget set is, we've mathematically represented that based on your income and prices. Now let's put them together and talk about how you make choices.
What's the highest utility you can achieve given the your budget constraints put on you? You want understand this intuitively, graphically, and mathematically.

- Intuitively the idea is just what's the most you can have given the constraints that are placed on you?
- Graphically, we represent that as asking, what is the furthest out indifference curve you can achieve? Because more is better. Indifference curves that are further out make you happier. So what's the furthest out indifference curve that you can achieve given your budget constraint?

Let's imagine your utility is the square root of pizza times movies. So your preferences are mathematically represented by $u = \sqrt{p * m}$. The same budget constraint and prices: income is \$96, price of movies is \$8, price of pizza is \$16.

Figure 5-4: Budget constraint and indifference curves

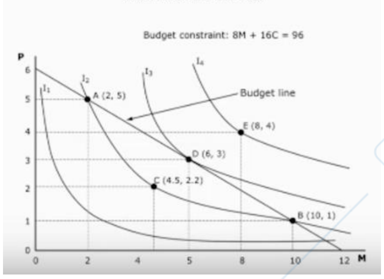


Figure 5-4 puts together our indifference curve analysis with our budget constraint analysis.

The straight line running from a y-intercept of 6 to an x-intercept of 12 is your budget constraint. Then we have here a series of indifference curves that are a mathematical representative of this utility function. And what we see is that the best you can do is to choose point D. Point D, with six movies and three pizzas Has an utility of the square root of 6 times 3, which is the square root of 18.

This is the best point for you, compared to point A that is on a lower indifference curve: utility's a lower value, it's only square root of 10.

So point A is dominated by point D; the same for point C because you could afford more.

So basically, *the point is that the point which will make you happiest is the point at which your indifference curve is tangent to the budget constraint.* Because that is the point of the farthest out indifference curve that you can reach given your budget constraint. The tangency of the indifference curve and the budget constraint is the point which makes you best off given your available budget and the available prices*. And that's the point where the slope of the indifference curve equals the slope of the budget constraint.* That's the graphic intuition.

CONSTRAINED UTILITY MAXIMIZATION: MATHEMATICAL DERIVATION

Let's talk about what it means that these slopes are equal.

The slope of the indifference curve is the marginal rate of substitution. In particular, it's the negative of the marginal utility of movies over the marginal utility of pizza. Remember, it's the negative of the marginal utility of the x-axis over the marginal utility of the y-axis.

$$Slope\ indifference\ curve = MRS = -\frac{mu_x}{mu_y}$$

So marginal rate of substitution is the rate at which you're willing to substitute between movies and pizza, which is a function of your marginal utilities. If your marginal utility for movies is very high, then you need a lot of pizzas to trade for it. If your marginal utility of movies is very low, you'd be happy to give up a movie even for a small fraction of a pizza.

At the same time, we're saying that that marginal rate of substitution (the slope of indifference curve) is equal to the slope of the budget constraint.

We call the slope of the budget constraint as the marginal rate of transformation, which is the negative of the price ratio.

$$Slope\ of\ budget\ constrain = MRT = -\frac{P_x}{P_y}$$

So preferences give us this, the marginal rate of substitution.

The mechanics of the market give us the marginal rate of transformation.

And utility maximization gives us that those are equal, because they're equal at that tangency.

$$\boldsymbol{Maximization: MRS = MRT} \rightarrow\ Marginal\ Benefit = \text{Marginal Cost}$$

Think about the ratio of the marginal utilities as the marginal benefit of another movie in terms of pizza. It's how much you like that next movie relative to how much you like that next pizza.

The marginal rate of transformation is the price of that next movie relative to the price of that next pizza or the cost to you of that next movie in terms of pizza.

At the maximization we're setting benefits equal to the costs. In particular, we're setting marginal benefits equal to marginal cost.

For now, the price is constant, but the marginal utilities are not constant. Marginal utilities are obviously changing the more moves you see.

In the equilibrium, at the optimum the marginal utility of movies over the price of movies equals the marginal utility of pizza over the price of pizza.

$$\boldsymbol{Equilibrium: \frac{mu_x}{P_x} = \frac{mu_y}{P_y}}$$

For each dollar of movie expenditure, what's it buying? What's that next dollar of movie expenditure buying you?

This is saying, what's that next dollar of pizza expenditure buying you and they've got to be equal.

If the next dollar of movie expenditure buys you a lot more happiness than the next dollar of pizza expenditure, then you're not at the right place. You should shift your money and spend more on movies and less on pizza. If the next dollar of pizza expenditure buys you a lot more happiness than the next dollar of movie expenditure, then you're not in the right place either. You should see fewer movies and buy more pizza.

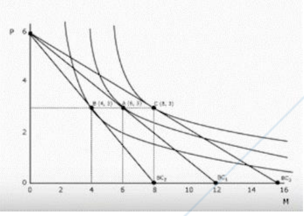
DRAWING DEMAND CURVES

Where demand curves come from?

Demand curves come from underlying utility maximization.

Return to your example, your parents gave you $96; you could buy movies at $8 a pop or pizzas at $16 a pop. If given your utility function, U equals square root of P times M, you would choose a point like A, where you're consuming six movies and three pizzas.

Figure 6-2: Drawing a demand curve



That steepens the budget constraint, moves it inward. Because now think about your opportunity set, for the same income of $96 you can buy the same number of pizzas you could have before, but now you're buying fewer movies. Same number of pizzas you could have bought before, but now you can buy fewer movies.

With a steeper budget constraint, the slope, instead of being - 1/2, is - 3/4.

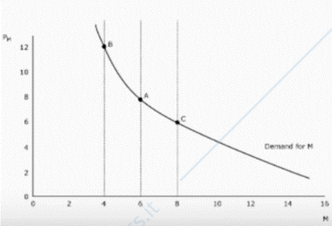
And given the preference I wrote, U equals square root of C times M, you should be able to show yourself that you'd now choose a point like B, where you have three pizzas but now only four movies. So you reduce the number of movies, you keep the number of pizzas constant, still spending the total budget.

The marginal rate of substitution you can compute if you write it down from that utility function, will be - 3/4, which is the same as the marginal rate of transformation with this new price. So you will choose a point like point B.

Now, let's say instead the price of movies fell from $8 to $6, your budget constraint would flatten. It would move outwards and your opportunity set would expand in that case, because effectively you're richer. Your opportunity set expands. You move to BC_3 and given those preferences I wrote down, U equals square root of P times M. You end up choosing point C, with the same 3 pizzas but now 8 movies.

Once again, how do we know that's right? First of all the marginal rate of substitution will equal the new marginal rate of transformation and also you can see you spend your entire $96 income. You're still roughly splitting it with $48 on movies and $48 on pizza, exactly splitting it.

All we're doing in this top diagram is saying, given your utility is U equals square root of P times M and given your income and the prices, these are the choices you would make as prices change.



Now armed with that, we can draw a demand curve, because we've just given you three different prices for movies

and three different quantities of movies you choose.

When the price of movies is $8, that's point A in the middle, you choose 6 movies.

When the price of movies rises to $12, your demand for movies falls, only choosing 4 movies.

When the price of movies falls to $6, you choose 8 movies.

The demand curves just come from constrained utility maximization. You just take your utility function, you maximize it, given the budget constrain.

Producer theory

FIRM PRODUCTION FUNCTIONS

Now let's come to supply curves. Now on the one hand, this will be much easier than demand curves, because a lot of the logic is the same. The analysis basically has the same kind of tangency of curves with straight lines that yield supply curves as for consumer theory. On the other hand, supply curves are a ton harder, because now we don't just have the price as given to consumers, but the suppliers actually make up the price. With consumers, you're given a price and you choose what to buy at those prices. Who set those prices are the producers and that's what determines the underlying supply curve.

Just as we had consumers making decisions, we thought of a consumer as somebody choosing across a bundle of goods, pizza versus movies, now we're going to think of a producer, a firm very simply as a black box, where inputs go in and outputs come out. Inputs go in and outputs come out; the firm has a simple goal which is to maximize their utility, producers have a simple goal which is to maximize their profits. $Profit = \pi = Revenues - Costs$

And the key to maximizing profits is efficient production, is going to produce goods as efficiently as possible. So profit maximization requires *production efficiency*.
Now we'll focus on production functions, that's essentially the technology by which a firm takes inputs, or what we call factors of production, and turns them into outputs, is through this production function. So just like your utility function is a tool for which we take bundles of goods and turn them into happiness, *a production function is a tool for which we take bundles of inputs and turn them into outputs.*
We're going to think about two different kinds of inputs that firms are going to use to make life easy: we're going to use labor and capital. Labor is just hours of labor, hours of work in production. Capital is everything else that goes into production, the machines, the buildings, the land, everything. So basically, when you produce stuff, you produce it with workers working with stuff. The output produced by the firm is q (units of production).
So basically, we can think of a production function is q, is some function of L and K: $q = f(Q, K)$.
We use little **q** to represent a firm's output, and big **Q** to represent market output.
So little q is some function of the amount of workers you have and the amount of capital you use.
Now, the important distinction we're going to make here is between variable versus fixed inputs. Variable inputs are inputs that are easily changed, like how many hours somebody works. In principle, you could just have some work five hours one day, one hour the next day, ten hours the day after that. It's easy to change hours of work. Fixed inputs are things which are harder to change quickly, like the size of the building that the workers are building in. Once that building's built, it's pretty hard to change it.
And this will lead to a critical distinction for production theory, which is the short run versus the long run.
The long run is the period over which all inputs are variable. The short run is a period over which some inputs are fixed.
I can tell you that, clearly, tomorrow is the short run. Clearly, you can't vary all your inputs over one day. And probably next month is the short run. And probably even next year is the short run. There's a lot of inputs that it takes more than a month to change, or a year to change. On the other hand, 10 years from now is almost certainly the long run. There's very few inputs to production you can't change over a 10-year period. So we know a day is the short run. And we know 10 years is the long run. We don't really know where the transition is, but in substance that doesn't matter for you. It's a theoretical concept: the break between the short run and the long run is the break between when some inputs are fixed and all inputs are variable.
A lot of times, economists will talk about quasi-fixed factors of production, which are things which could change in between the short run and the long run. So for instance, take labor as a variable unit of production you can change in the short run, but in fact, we know in practice you can't.
We're typically on some kind of reasonably set work schedule. Now that work schedule can evolve, but it can't change day by day. We're all going to have jobs with a fairly smooth distribution. There'll be peaks and valleys, but a fairly smooth distribution of our labor effort. So really, truly, there's very few inputs which are truly perfectly variable.
Just like there are no inputs which are truly perfectly fixed, but for the purposes of this model, let's think about labor as a variable input. Let's think about labor as being like hourly labor, like hourly construction. And let's think about capital as being a fixed input, like a building that you can't pare down or rebuild overnight, but over a 10-year period you can.

SHORT RUN PRODUCTION AND DIMINISHING MARGINAL PRODUCT

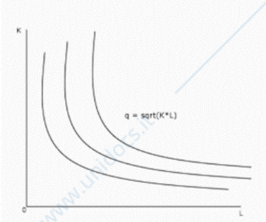
How firms make short-run production decisions. So let's start by considering the short run, and considering that period of time over which labor is variable but capital is fixed. That is, you have a given plant, but you can adjust how many workers you use every day in that plant. And now the firm has to decide, given that that plant exists, how many workers should I hire to produce my good? How many workers should I hire to produce my good?
And the key concept that's going to determine that is something we'll call the **marginal product of labor**, *which is the change in total output resulting from the next unit of labor used* : $MP_L = \frac{\Delta Q}{\Delta L}\Big|_{\bar{K}}$
It's going to be the change in total output from another unit of labor, once again holding capital constant because that's fixed: at a given k bar, but that's implicit in the fact it's the short run.
We're typically going to assume this like marginal utility: marginal product is like marginal utility.
Just as the marginal utility was your utility from another unit of one good, holding the other good fixed, marginal product is the marginal production from another unit of an input, holding the other input fixed.
And just as we've assumed and discussed the intuition for diminishing marginal utility, we're going to typically assume *diminishing marginal product*. That is, from a given level of labor, the next worker you add increases your total product by less than the previous one.
Now once again, just like we have a non-satiation rule in utility, we're not going to say the next worker doesn't help.

Every worker helps. But every worker helps less and less and less, just like every pizza means less and less and less to you. Why shouldn't the next worker do less than the first? The key is that we're holding capital constant; the reason each worker does less is because they only have the same amount of stuff to work with. And the classic example we use here is the example of digging a hole. You go to dig a hole and your capital's a shovel (and let's say for some reason the shovels are out, so you can't get another shovel for a while).
There's one shovel, so you have one worker digging a hole, then the next worker comes along and that's where he can help because he can spell the first worker. And maybe the next worker's just as good because he can work more hours, but probably it's a little bit less good. But certainly by the time you add a fourth and a fifth and a sixth person, with one shovel, they probably each help because they can rotate and rest each other. But certainly the sixth person is not going to help dig the hole as much as that second person did, or the third, or fourth, or fifth person, because only one shovel. So they can share a little bit and share the burden a little bit. But at some point each additional worker helps less because they have to share the same shovel.
So that's what diminishing marginal product is, because capital's fixed. With a certain amount of capital to work with, each additional worker just can't help as much as the one before.
Now eventually, you can add more shovels. And you could have wheelbarrows. So some people could run the wheelbarrows, some people could run the shovels. But that's the long run.
In the short run, there's the one shovel, so each additional worker does less good.

LONG RUN PRODUCTION AND THE MARGINAL RATE OF TECHNICAL SUBSTITUTION

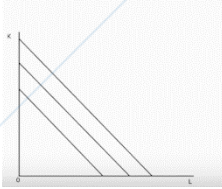
Now in the long run, all inputs are variable. So now a firm doesn't just choose how many workers to hire, or how many hours of labor to buy. It chooses both L and K, and has to trade them off, just like you chose both pizzas and movies and had to trade them off. So the long-run production theory is the same, basically the same mechanics, as utility theory. There's a production function; you have two inputs; you have two goods to consume and you trade them off just like you're a consumer. When you decide how to trade them off as a consumer, you're given a budget constraint, while for ultimately production is going to self: *the difference with production is the budget constraint is going to be itself determined by the same system.* So you're not only going to develop your production function, but you're going to develop your budget constraint and, you're going to decide both.
For now, let's just think about the parallels to consumer theory, and think about a *production function*: $q = \sqrt{K * L}$.
(That same functional form I used with pizza and movies, where utility was the square root of pizza times movies) Now I'm going to say what you produce of your good is the square root of K times L.

Figure 8-3: Isoquants



In figure 8-3 we can see here the trade off between K and L and deciding to produce, then you get what's called isoquants.
Isoquants are the parallel to indifference curves. Just as there were sets of goods across which you were indifferent, two pizzas and one movie versus one pizza and two movies. *Isoquants are sets of inputs along which production is the same. So along a given isoquant, q is fixed.* Each of those isoquants is a different level of q, but they show how you can vary K and L to get the same amount of q.
So producing q equals square root of 2, I can use two units of capital and one unit of labor, or one unit of capital and two units of labor. So I can choose lots of combinations of K and L along that isoquantto produce a given amount of output.
And isoquants have all the same features as in indifference curves:
- The further out the better because you're producing more.
- They can't cross.
- They slope downwards because there's a trade off between capital and labor.
  What's going to determine the slope of an isoquant? Likewise, the substitutability between labor and capital will determine the slope of these isoquants.

Figure 8-4a: Perfectly substitutable inputs



In that case, you would have a linear isoquant, that would mean you don't care if you have three capital and one labor, or three labor and one capital, or two labor and two capital, or three labor and one capital, as long as you get a total of four, that is the only thing that matters. They're perfectly substitutable inputs, which would say that it would be something like $q = K + L$. You don't care if it's K or L, you just care about the total.

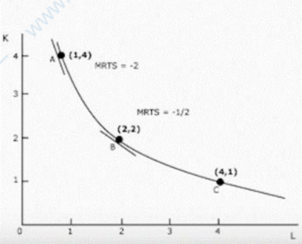
Figure 8-4b: Non-substitutable inputs

On the other hand, let's think about goods which are not at all substitutable, like cereal and cereal boxes: the cereal wouldn't be any good unless you have a box to put it in; the box doesn't do anything unless you have cereal to put in it.
8-4b would show you non-substitutable inputs where basically, given the amount of one input, it doesn't matter how much you have of the other.
Leontief production function: $q = MIN(K, L)$
So given how much K you have, given you have 10 cereal boxes, once you have 10 chunks of cereal, it doesn't matter if you have 10, 11, 12, 1 million. You only have 10 cereal boxes. So given an amount of K, then it doesn't matter how much L you have, and vice versa.
So substitutability determine the slope of the isoquants and of this production function.
Now in general, we'll be in between these cases. There'll be some substitutability, goods won't be perfectly substitutable. And more generally, just as the slope of the indifference curve is the marginal rate of substitution, *the slope of the isoquant we will call the* **marginal rate of technical substitution**. *The marginal rate of technical substitution is the rate at which you can substitute one input for another in a production function, that* we'll define as:

$MRTS = \frac{\Delta K}{\Delta L}\Big|_{\bar{q}}$, the rate at which you can trade off K for L to hold q bar fixed.

Now, as with marginal rate of substitution, the marginal rate of technical substitution will change along the isoquant.

Figure 8-5: Isoquants and MRTS



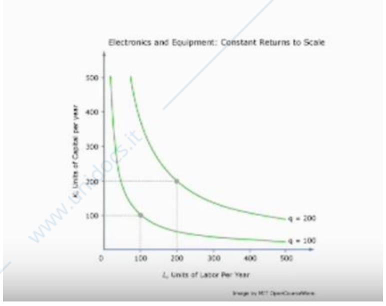
So in 8-5, we've drawn a typical isoquant for the production function $q = \sqrt{K * L}$.
This is the isoquant of all combinations which produce two units: q=2.
Now, unlike utility where u=2 was meaningless, utility was an ordinal concept and not a cardinal concept, here, quantity is meaningful. If you produce four, you would have only produced twice as much as if you produced two. We can care about both the ordinality and the cardinality of these outcomes. So we can say, what are the combinations of inputs which lead you to produce two units? So whereas you can have one unit of labor and four units of capital, two of each, or four units of labor and one unit of capital, all will produce two units of the output. And what you can see is that the marginal rate of technical substitution varies. So for instance, when we start with four units of capital and one unit of labor, and we think about adding a second unit of labor, then the marginal rate of technical substitution is minus 2. That is, one unit of labor is worth two units of capital. In other words, we can produce the same amount of widgets of q, but if we replace two units of capital with one unit of labor, so at that point we're very capital-intensive, and that unit of labor is very valuable.
On the other hand, now if you imagine we're down at 4-1, at the point C, now if you'd be willing to give up two units of labor just to get one unit of capital, that's the marginal rate of technical substitution is now minus 1/2. When you're very labor-intensive, you'd be happy to give up a lot of labor to get a little capital. Once again, the principle of diminishing marginal product, just like the principle of diminishing marginal utility, implies that the marginal rate of technical substitution is going to be falling as you go down the isoquant.
Just like the marginal rate of substitution fell as you went down the indifference curve*, the marginal rate of technical substitution is going to fall as you go down the isoquant*. Once again it's because of this diminishing marginal productivity: as you add more and more labor, given capital, each unit of labor can do less and less. And likewise, as you add more and more capital, given an amount of workers to use it, each unit of capital can do less and less.
So it's a very different approach, but gets you the same answer, which is with consumption, each piece does less and less for you, but we can see that as consumers. Here's the notions of labor, each worker does less and less for you, holding capital fixed. And each machine by the same logic, if you have one guy in the hole, it doesn't matter how many shovels you throw there. He still is only one guy. So each machines is doing less and less by the same logic.
And those diminishing marginal products lead to this decreasing marginal rate of technical substitution as you move along the isoquant.

RETURNS TO SCALE

One other concept, which is very important for thinking about production theory and comes back to these assumptions that we make which might be a little bit unrealistic, which is the concept of returns to scale.
So here the question is, what happens if we increase all inputs proportionally?
By scale means what happens if we just double everything, twice as much labor and twice as much capital. That's an increase in scale, increasing or decrease all inputs equal proportionally, twice as much labor, twice as much capital, half as much labor, half as much capital. So a change in scale is an equal increase or decrease in all inputs?
The answer is, it depends on the production process:

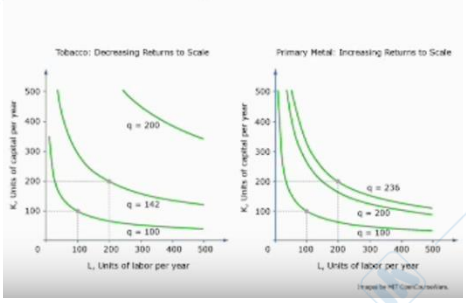
- Some production processes will exhibit a constant returns to scale: $CRS: f(2L, 2K) = 2f(L, K) = 2q$
  That is, if it's constant, you can just pull the 2 out. Doubling the inputs leads to doubling the outputs, which equals to 2q. So if I have twice as much labor and capital, that's the same as twice producing what I had with the original labor and capital which will get me twice the original production. So every time I double my firm, I get exactly twice as much stuff out.
- Increasing returns to scale: $IRS: f(2L, 2K) > 2f(L, K) > 2q$
  That is, when I double my firm, I produce more than twice as much stuff.
- Decreasing returns to scale: $DRS: f(2L, 2K) < 2f(L, K) < 2q$
  When I double my firm I get less than twice as much stuff.

Figure 8-6a: Isoquants and constant returns to scale



Let's start with a constant returns to scale: when we doubled the inputs from 100 labor, 100 capital, to 200 labor, 200 capital, you doubled output. Those are constant returns to scale isoquants.

Figure 8-6b: Isoquants with increasing and decreasing returns to scale



Decreasing and increasing returns to scale in figure 8,6b: in tobacco example when you double the inputs from 100 to 200, your output only goes from 100 to 142. And if you want to double the 200 units, you've got to go way the heck up there on inputs. On the other hand, primary metal production is something where you can have increasing returns to scale, because there's such high costs of building a plant that once it's built, the second digger of the hole might actually make it more productive. Once that plant's built, there's one worker banging around the plant, a second worker adds a huge amount to productivity because they can specialize. Basically, *increasing returns to scale is going to come from specialization*. So if you build this big plant to build primary, to build steel, one worker in that plant is no good, because he's got to pour the steel, then run over and cool it down, et cetera. Once you have two, they can specialize, and three, they can specialize more.
And that's illustrated here, where doubling the inputs actually leads to more than doubling the output.

PRODUCTIVITY

The form of the production function as $q = f(Q, K)$, really more generally, can be written as $Q = A * f(Q, K)$q, where A is aggregate productivity.
Now let's think of aggregate quantity Q for society or else we wouldn't talk about a specific firm. But if we think about aggregate product, aggregate quantity produced in society, it's a function of the aggregate capital and labor of the society, but also a function of productivity. It's also a function of the fact that we use our inputs more effectively over time.
So while K if it's defined as land may be fixed, and L therefore there's diminishing marginal product of a given production function, the production function itself is improving over time because of productivity improvements.
Productivity, the arability of land, disease-resistant seeds, and other things are making that given quantity of land more productive over time.
So effectively, in the long-run if A goes up faster than the marginal product of labor diminishes, then overall quantity can increase even though K, the underlying level of land, is fixed.

INTRODUCTION TO COSTS AND SHORT RUN COSTS

The firm has to maximize profits, which is revenues minus cost. So we have to ask what are costs if we're going to make this profit maximizing decision. Costs are going to have two major components: fixed costs and variable costs.
Fixed costs *are the costs of inputs that cannot be varied in the short-run* (a period over time which only some inputs can vary): fixed costs are the costs of those inputs that can't vary in the short-run.
Variable costs *are the cost of goods that can vary in the short run*, that's like labor.

$$\text{Total costs} = \text{fixed cost} + \text{variable cost}$$

Marginal cost is the change in cost with a change in output: $\boldsymbol{Marginal\ Cost} = \frac{\Delta C}{\Delta q}$

The change in a firm's cost with the change in the firm's output is marginal cost.

Average cost is just C over q, it's just the average. $\textbf{Average Cost} = \frac{C}{q}$

So the difference between marginal and average cost is basically: average costs is the average over the whole set of goods produced; marginal cost is the cost of that next unit of production.

How do we get costs? The answer is we get them from the production function. Once we do a production function, we can derive costs.

So if we have some production function $q = f(L, K)$, then we can say the cost of producing q is $c(q) = f(wL + rK)$. Where w is the wage rate, or the rate you pay per unit of labor, and r is the rental rate, or the rate you pay per unit of capital. It's easy to think the cost of an hour of labor, it's the wage you pay for an hour. It's harder to think about the cost of a unit of capital because we buy the machines, but for now imagine all machines are rented and think of r as the rental price of that unit of capital. The key point is, the reason we have to do this is the wage is a flow measure, every hour I pay you a new wage. If I use the cost of buying the machine, that be a stock measure, so you couldn't really compare it to wages. So we want to use a flow measure, the flow measures is what we have to pay every period to rent the machine.

Now, in the short-run capital is fixed, so our fixed costs are $FC = r\overline{K}$, the rental rate times the fixed amount of capital in the short-run. And our variable costs are $VC = w * L(q)$, the more you produce the more labor you use in the short-run.

$$Short - run\ term\ Total\ Costs = r\overline{K} + w * L(q).$$

K is not a function of q because K's fixed in the short-run, but the amount of labor used is a function of how much you produce. This implies that the marginal cost, which is the derivative of total costs function with respect to quantity, is:

$$Marginal\ Costs = \frac{\Delta C}{\Delta q} = w * \frac{\Delta L}{\Delta q}$$

So the marginal cost of producing the next unit is going to be how much labor I have to produce to produce the next unit, times the wage I pay per unit of labor.

So marginal cost is the wage over the marginal product of labor. $Marginal\ Costs = \frac{\Delta C}{\Delta q} = w * \frac{\Delta L}{\Delta q} = \frac{w}{MP_L}$

Where marginal product labor is: $MP_L = \frac{\Delta q}{\Delta L}$

The cost of the next unit of production is declining with the marginal product of labor. The more productive is a worker, the less expensive is producing the next unit. The less productive is the next worker, the more expensive is producing the next unit. So it's an inverse relationship between the marginal cost and the marginal product where the wage is the constant that scales that relationship. So basically, when workers are very, very high marginal product, then it's going to be cheap to produce the next unit. When workers have a low marginal product, it's going to be expensive to produce the next unit, and that's going to depend on what you actually have to pay the worker.

LONG RUN COSTS

Firms get to choose their input mix to maximize their production efficiency. So *input mix is chosen to maximize production efficiency which equates to minimizing costs.*

Isoquants are combinations of labor and capital that delivered the same output, just like indifference curves are combinations of pizza and movies that deliver the same utility.

The key point is that, technologically, any choice of labor and capital produces the same q, so there's nothing that tells you technologically which of those to use. We just know, technologically, is a variety of a set of choices which deliver the same q. How do we tell which to use?

We want to choose the one which is minimizing costs and to do that, we're going to introduce the cost of those inputs. Just like there's a set of pizza and movies, all of which leave you indifferent, how do you decide which pizza and movies to choose? You introduce the relative price of pizza and movies.

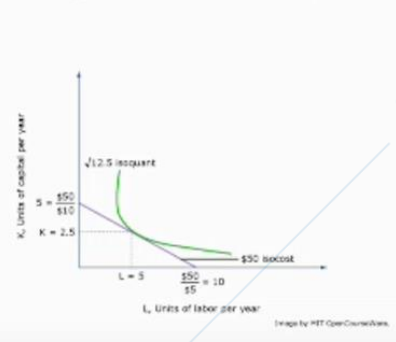
Here, we're going to bring in the relative price of capital and labor to determine how we choose between capital and labor. So to do that, we're going to draw isocost lines which are going to be just like our old budget constraints.

Isocost lines which *represent the cost of different combinations of inputs*, just like our old budget constraint represented the cost of different consumption goods.

Figure 9.1: Isocost Lines

In figure 9-1 we have isocost curves, and we're going to assume here that the wage is w=\$5 an hour, and the rental rate is r=\$10 per unit of capital. So, in other words, the \$50 isocost line in figure 9-1 shows all combinations of labor and capital that cost \$50. So you could spend \$50 in production if you had 10 units of labor and no units of capital. Or five units of capital and no units of labor, or any combination in between.
Likewise, the \$100 isocost is all combinations of labor and capital that cost \$100.
So each of these isocosts give you the combination of inputs that cost a certain amount. Just like a budget constraint gave you the combination of pizza and movies on which you spent your income.
The difference with consumers is we knew their income so we knew what their budget constraint is, but here we don't know whether to choose the \$50 cost, the \$100 cost, \$150. We don't know what the total amount is.
For now, let's just say there's a set of trade-offs that the firm can choose from, and a set of isoquants that they have.
The slope of the isocost is the negative of the wage rental ratio: $\boldsymbol{Isocost\ Slope = -\frac{w}{r}}$.
It's basically the trade-off between labor and capital's going to be determined by the relative prices of those inputs. So basically, how many units of capital do you have to give up to get the next unit of labor? What this isocost tells you is you have to give up 5/10= 1/2 a unit of capital to get a unit of labor, so the slope is - 1/2.
Likewise, you could say you have to give up two units of labor to get one unit of capital.
Budget constraints are about opportunity costs. How much labor do you have to give up to get another unit of capital? Or how much capital do you have to give up to get another unit of labor?
Now, armed with isoquants, which are like indifference curves, and these isocosts which are like budget constraints, we can then figure out what is the economically efficient combination of inputs for the firm to use.
The economically efficient combination of inputs for a given level of output *is determined by the tangency of the isoquant with the isocost*, as you see in figure 9-2.

Figure 9-2: Cost Minimization



$Isoquant: q = \sqrt{K * L}$
The efficient, if you want to produce, a given amount q, is at the tangency of that isoquant with the isocost.
And you're going to say that the efficient way to produce that is going to be to use K=2,5 units and L=5 units.
It's going to say look, given the relative prices that are given to us by this budget constraint, the production technology is given to us by this production function from which we derived isoquants last time. And that will produce basically square root of 12,5, that is basically $q = \sqrt{2{,}5 * 5}$.
Let's think about for a second the mathematics. We know that the slope of the isoquant at any given point is $\boldsymbol{MRTS = \frac{MP_L}{MP_K} = \frac{w}{r}}$
The efficient thing to do is to set the marginal rate of technical substitution equal to the price ratio, that's what happens when the slopes are equal.
Rewriting this equation as the marginal product of labor over the wage, equals the marginal product of capital over the rental rate $\boldsymbol{MRTS = \frac{MP_L}{w} = \frac{MP_K}{r}}$.
What this is telling us is the efficient place is where essentially for every dollar you spent on workers, you're getting the same return as a dollar spent on machines.
The marginal product of labor over the wage is sort of the bang for buck of workers. What are you getting for your next dollar of wage? The marginal product of capital over r is the bang for the buck of machines. What are you getting for your next dollar of rent? And the efficient point is where these are equal. If they're not equal, then you have too much of one and not enough of the other.

Competition

INTRODUCTION TO PERFECT COMPETITION

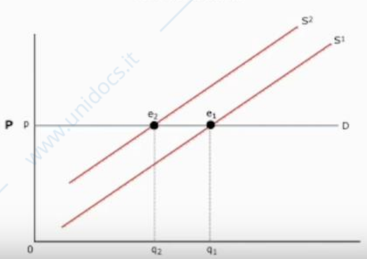
So, basically, the level of production for a given firm, q, will be derived from how firms behave in different market settings and we have 3 main market settings:

- Perfect competition: this is basically a case where many firms are selling goods to many consumers.
- Monopoly is where one firm sells to many consumers.
- Oligopoly when several firms sell to a large market, which is probably the most realistic setting of all.

Technically, perfect competition *exists whenever firms are price takers on both the output and input markets*. That is no action that they take can affect either the price at which they sell their goods or the price that they pay for their inputs. Technically this would be true if a firm faced *perfectly elastic demand for their goods*, and if they had *perfectly elastic supply of inputs.*

So let's focus on the first of those conditions which is perfectly elastic demand.

Figure 10-1: Supply shift with perfectly elastic demand



Let's take a look at Figure 10-1, where we have firm (q) facing a perfectly elastic demand, that means that the firm's quantity is pegged by their supply curve. Or, in other words, *the firm cannot change the price from that level P*. So, in other words, if this is a supply shift, the price of the firm's inputs go up, the firm doesn't get to charge any more for their goods. They just sell fewer goods from q1 to q2 with the supply shift from S1 to S2. They're going to reduce the quantity they sell, but they cannot change the price.

So when does this make sense as a description of the world? It makes sense as a description of the world under four conditions:

- Identical products: in a perfectly competitive market, the firms in that market sell identical products. They don't have to literally be identical. They have to be perceived by consumers as identical, for purposes of their demand across firms. Because if products aren't identical, then firms will be able to charge different prices from each other because they have something different to sell.
- Consumers have to have full information on all prices.

The next two conditions, because they're related.

- Low transaction or shopping costs.

Conditions 2 and 3 are correlated, because consumers are going to shop across firms selling identical goods and they're going to buy from the cheapest one. And if there's any failure of either of these, the consumers might not know if you're the cheapest. And, therefore, you might be able to charge extra. Perfectly elastic demand is going to require that consumers know all the prices and can cautiously shop across all the options, otherwise, firms might have some opportunity to charge different prices.

- It needs to be free entry and exit of firms.

Now, in reality, no perfectly competitive market exists. There's never been a perfectly competitive market.

FIRM DEMAND VS. MARKET DEMAND

Distinction between firm demand and market demand. *Even if a given firm faces perfectly elastic demand, it doesn't necessarily mean that market demand is perfectly elastic.*

That is the overall demand for little, fake Statues of Liberty around Port Authority in New York is not perfectly elastic. As the price goes up, fewer people will buy them. As the prices goes down, more people will buy them. But for any given vendor selling them, it is perfectly elastic because there's always someplace next door you can go.

So it's very important to distinguish between the demand facing the firm being perfectly elastic and the demand facing the market not being perfectly elastic.

And the way to think about this is to think about the concept of residual demand.

We have a demand function for market D(p), which is that as the price goes up demand goes down.

Now the demand function for a given firm, we'll call the residual demand, is equal to: $D(q)^r = D(q) - S^0(p)$

It's equal to the demand for the market minus the supply that all other firms in the market provide.

So the demand for my product as a firm is my residual demand. It's the market demand minus what other firms supply.

If you differentiate this with respect to price: $\frac{dD^r}{dp} = \frac{dD}{dp} - \frac{dS^0}{dp}$

This first one is the market demand curve: $\frac{dD}{dp}$, that's a negative number, because demand curves slope down.

But the second is a positive number: supply curves slope up. The amount that other firms in the market will supply as the prices goes up is positive: supply curves slope up.

So $\frac{dD^r}{dp}$ is a very negative number: $\frac{dD^r}{dp} = \frac{dD}{dp} - \frac{dS^0}{dp}$

The firm's residual demand responds more to price than the market's demand does because the firm's residual demand is after all the supply of other firms. So we can rewrite this in terms of elasticities.

So let's assume, for a second, that all firms are identical: $q = \frac{Q}{n}$. And so, therefore, the amount produced by other firms, $Q^0 = (n-1) * q$.

The elasticity of demand facing a given firm is n times the elasticity of demand for the entire market ε minus (n - 1) times the elasticity of supply for the market η. $\varepsilon_i = n * \varepsilon + (n-1) * \eta$

So, for example, let's say you've got a market with 100 firms in it and the elasticity of demand for this market equals minus 1, so it's in between elastic and inelastic, and the elasticity of supply is 1. For a given firm, if you used this formula, the elasticity of demand facing a given firm is -199=100*(-1)+(100-1)*1.

It's a huge negative number, even though the market demand is modestly elastic, -1.

The point is that even if a market does not have super elastic demand, a given firm can face very elastic demand. And that's what can lead to perfect competition.

When we talk about demand, think about demand at the firm level versus demand at the market level. Demand at the market level, that's about substitutability with other goods and the things we've talked about deriving demand curves. When we derive demand curves, we're not deriving firm demand curves. We're deriving market demand curves. And so the demand curve was a function of elasticities and substitutability across goods. The firm demand curve is a function of all that, but also how many firms are in the market. If there are a lot of firms in the market, it's going to be very elastic in a perfectly competitive market.

SHORT RUN PROFIT MAXIMIZATION IN A COMPETITIVE MARKET

How does a firm maximize profits?

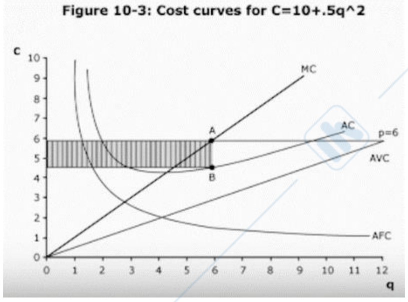
We say that profits is a function of quantity produced, that is revenues, as a function of quantity produced, minus cost as a function of quantity produced. Our goal is to figure out what little q a firm chooses and it is dictated by maximizing this equation: $\pi(q) = R(q) - C(q)$

So a firm will choose its quantity, little q, such that: $\frac{dR}{dq} = \frac{dC}{dq}$, that's the profit maximizing equation. Or, in economic terms, where marginal revenue equals marginal costs: $MR = MC$. Marginal revenue is the revenue made from selling the next unit; marginal cost is the cost incurred by making the next unit. In a competitive market, we know that $\frac{dR}{dq}$ is given to the firm by the market and it is equal to the price: $\frac{dR}{dq} = P$.

The price is given to the firm, so in a competitive market, the **profit maximizing equation** is: $\mathbf{P = MC = \frac{dC}{dq}}$

You will produce until the marginal cost of producing the next unit is equal to the price you can sell that unit for in the market.

Figure 10-3: Cost curves for C=10+.5q^2



In figure 10,3 we have an example with cost curves for a cost function = 10 + 0.5q^2.

Average costs is that line in the middle there and is the average cost of all the units you've made..

You have an average variable cost that's a line, that's linear, an average variable cost that has a slope of 1.

You have an average fixed cost that's everywhere declining, because your fixed costs of 10 is everywhere declining, as you produce more and more, your fixed costs are declining.

You have a marginal cost of q, that is the cost of the next unit, of 1 unit is 1; your marginal cost of 2 units is 2, etcetera. So this draws out the cost curves that correspond to that cost function.

The firm is facing a perfectly elastic demand curve, so it is horizontal at price equals 6: it's a perfectly competitive market. The firm chooses to produce where marginal cost equals price.

So if I make 6 units, the profit I make on that 6th unit is 1+1/3: 8/6 (tot profit/q), but I make those profits on all 6 units I sell. So that means is, in total, I'm going to make a profit of 8 that is the area of this rectangle: 6*(1+1/3)=8. You cannot choose a production level that produces a bigger rectangle than this. So if you produce 7, your rectangle will be longer, but the gap between price and average cost will be smaller and your total rectangle size would fall. The largest rectangle is produced at a production level of 6 that is the most efficient use of your resources producing at a point where marginal cost equals price, that causes the maximum gap between price and average cost. AB max is 1+1/3.

SHORT RUN SHUTDOWN DECISIONS

In the long run we also have to decide whether or not we want to shut the firm down.
One condition is to set price equal to marginal cost; but actually, short run profit maximization has a second condition: check whether the firm wants to shut down. A firm might want to shut down if it actually loses money by continuing to produce. Firms may lose money but not shut down or firms may lose so much money they shut down.
Imagine the price in this market suddenly fell from 6 to 3 per unit you sell. If the price fell to 3, the firm would choose to produce 3 units. You would still have marginal cost equals price: price is 3, and marginal cost is q, so q is 3. If you produce 3 units, its costs are 10 + 4.5= 14.5, while at a price of 3 revenues is 9. So its profits are - 5.5. It would lose money from this production. If a price changed, it's not like you change which equation you follow. You always follow the maximization equation. The efficient production level is always marginal cost equals price regardless of what the price is. So if the price is 3, the efficient thing to do would be to produce 3 units and lose money. If that goes negative, wouldn't they just shut down? The answer in the short run is no, because *in the short run, the fixed costs that you paid to produce are sunk*. They're unchangeable in the short run. In the long run, they're changeable, you can just leave. But in the short run, you've invested fixed costs of 10 to produce in this market, so you will not exit unless you lose more than $10.
You will not shut down unless you're losing so much money that you can't cover your fixed costs, because you've paid those fixed cost, they're sunk.
So unless you're actually losing more than your fixed costs, you will not shut down your firm.

RECAP: Cost function of the firm= $10 + 0.5q^2$.
The key condition for profit maximization with a perfectly competitive firm is that $P = MC$.
If you differentiate this with respect to q, you get that that means that $P = q$: it is the profit maximizing condition for this firm. It sets the price equal to quantity it's going to sell with this particular functional form of the cost function.
In the short run, a firm might not shut down even if it's losing money, because the firm has already paid its fixed costs. It's already paid 10, so even if it's losing money, it might still not shut down. The profits to shuts down is -10. So if profit are higher than -10 it actually makes more money by staying in business than shutting down. So, because it's going to pay the 10 anyway, as long as it's going to lose less than 10, it might as well stay in business.
As long as they cover its fixed costs, that means a firm will stay in business as long as its revenues are greater than or equal to its variable costs. $R \geq VC \rightarrow P * q \geq VC$
And that means that it will stay in business at long as its price is greater than or equal to variable costs over quantity: $R \geq VC \rightarrow P \geq \frac{VC}{q}$
Or as long as price is greater than or equal to average variable cost: $\boldsymbol{R \geq VC \rightarrow P \geq AVC}$
The variable costs for our firm are: $VC= 0.5 * q^2$.
We know that in equilibrium, if it's profit maximizing, it will produce where q =P, so we can replace the q with the p: $VC= = 0.5 * P^2$ and average variable costs are: $AVC= 0.5 * P$.
By definition, P is always greater than 0.5*P, because our firm will never go out of business in the short run if $P \geq AVC$.
So, more generally, when we think about a short run supply decision for a firm, there's two steps:

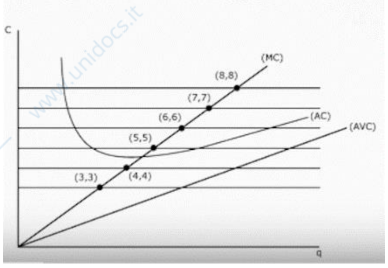
1. The first step is set price equal to marginal cost to figure out what the firm is going to produce. And that will give you the firm's q*, that will give you what the firm is going to produce. $P = MC \rightarrow q^*$
2. The second step is check that price is greater than or equal to average variable costs, because you may solve for an optimal quantity that turns out to be a money loser for the firm. $P \geq AVC$

You've got to first solve for the optimal quantity that the firm is going to produce, but then you've got to make sure that the firm actually makes money on that quantity, or it won't produce at all. And that's how we do the profit maximization decision in the short run for the firm. You've got to produce at the efficient point and make sure the firm actually makes some money.

DETERMINING SHORT RUN MARKET EQUILIBRIUM

Now, armed with these rules, we can now, finally, derive the supply curve.
We derived the demand curve by getting the tangency at different price ratios with the indifference curves.
To derive the firm's supply function, we need to define, at different prices, how much will the firm produce.

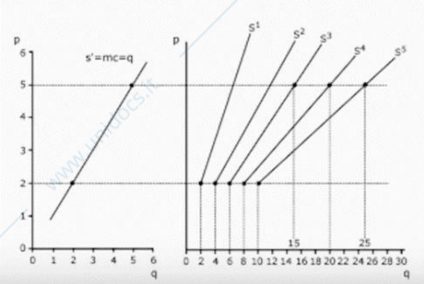
Figure 11-1: The firm's supply decision



In Figure 11-1 we can see the supply curve for this firm.
What we see is that at a price of 3, it will produce 3 units. At a price of 4, it will produce 4 units, etcetera.
The supply curve is the marginal cost curve. **Firm's short run supply curve is marginal cost curve above the point where price equals average variable cost, $P \geq AVC$.**
In our case P is always greater than average variable cost, so the second condition is irrelevant: the firm's supply curve is just literally that marginal cost curve.
Where do market supply curves come from? We now know where firm supply curves come from.

- The first step of the short run is you're going to enter this market and to enter this market, you're going to have an amount of capital you're going to pick. So each firm is going to have some cost function which involves picking some amount of capital or fixed costs. It's going to say, I want to build a building this big, having built that building, we're going to get the firm's supply curve which is $P = MC$.
- The second step is we're going to add up the firm's supply curves to get a market supply curve.

Figure 11-2: Short-run market supply curve



So, for example, suppose that there's five firms in the market; each firm has a marginal cost curve. We use the same cost function used before where $P = MC$, so the supply curve is $P = q$.
So each firm has that supply curve you see in the first panel.
The second panel gives you the market supply curve, as you add more firms. So, if there's only one firm in the market, the market supply curve would be S1. Now, if there were two firms in the market, the market supply curve is S2, that is, at a price of 2, you're now producing 4 units in the market. If there's three firms, the curve is S3, four firms, S4, and so on. *As you add more firms, that market supply curve shifts out and becomes flatter.*
Remember firms are identical here and we're adding identical firms, that is an assumption of perfect competition.
You can see that market supply curve is shifting out and becoming flatter, that is the supply of goods is becoming more elastic as there are more firms.

- The third step is we intersect market supply with market demand to get the equilibrium price.

There's some market supply, which we've derived and there's some market demand: that will give us the equilibrium price.
So, for example, in our case, there's five firms in the market. The total market supply $Q = 5q$, because there's five identical firms in the market.

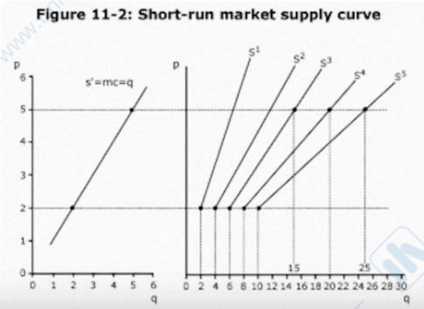


Figure 11-2: Short-run market supply curve

We know, from the marginal cost condition, that's the same as saying $Q = 5q = 5p$.

So, our market supply curve, which is actually S^5 on Figure 11-2, is $Q = 5q$, because when $P = 2, Q = 10$; when $P = 5, Q = 25$.

So, you can see that S^5 is the market supply curve; and the quantity supplied is $Q^S = 5P$.

The demand function, the quantity demanded is $Q^D = 30 - P$: we have a downward sloping demand curve with a slope of -1.

To get equilibrium, we set these equal: $Q^S = 5P = Q^D = 30 - P$ and we get that $30 - P = 5P \rightarrow P = 5$, that's the equilibrium price. At $P = 5$, $Q^D = 25$, so at a price of 5 the market wants 25 of these things, whatever the heck it's producing.

- Then, the final step in solving for equilibrium is that each firm then decides how much to produce. Each firm is going to produce $P = 5$ and there are 5 firms and is produced 25 that is exactly what people want.

Through these four steps, we've gotten equilibrium which is defined as the quantity supplied equals the quantity demanded.

To find the short run equilibrium, you need a demand function, a cost function, and a number of firms. You have to be given a number firms, because there's no entry and exit in the short run, remember.

LONG RUN MARKET EQUILIBRIUM: FIRM ENTRY AND EXIT

In long run, the only one condition to worry about is $P = MC$. And you're only going to be in a situation, in the long run, where you're only going to be making either 0 or positive profits. If you're making negative profits, you'll be gone. Now, the key difference in the long run, is now we can't take the number of firms as given, we need to derive the number of firms. The way we do that is by thinking about entry and exit. What's going to determine entry and exit is if in the market, as it stands today with some number of firms, there's profit to be made. If in the market, as it stands today with some number of firms, there's losses being made, some firms will leave: no one stays in making losses anymore. That continues until you reach a situation where all firms make zero profit.

In a perfectly competitive **long run equilibrium, all firms make zero profit.** Obviously, there's no place that works like this in the world, this is an extreme.

In a perfectly competitive long run equilibrium, all firms make zero profit, because if there's any profit to be made, a new firm will enter and take it away. And if there's any unprofitable industry, a firm will exit until the profits go back to zero. So profits will always be zero in the long run equilibrium.

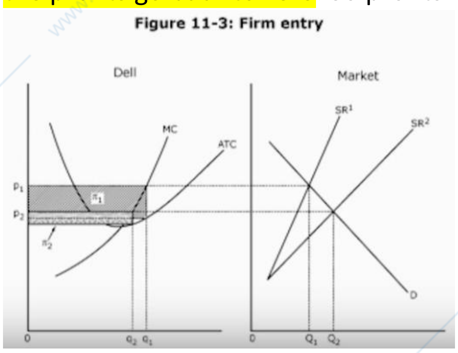


Figure 11-3: Firm entry

This is the market for PCs.

The supply curve was pretty steep because there weren't many firms making PCs. So the market price was P1.

On the right, you have the market. On the left, you have Dell.

In the market, in initial equilibrium, there's a price of P1 with Q1 being sold. So Q1 PCs are being sold at a high price of P1.

It's the novel technology, people want it, but not many firms are doing it.

What happens with Dell? Dell firm is producing where that price equals their marginal cost. So $P_1 = MC$ at q1: Dell's producing q1. But its average costs at that point are all the way down, it's where that vertical line for Q1 intersects average total cost. That's where their costs are. So, in each unit, they're making the height between marginal cost and average total cost at that Q1. They're making that vertical bar. So they make that entire rectangle of profit: Dell makes a big profit, because not many firms are in this business and yet demand for PCs is high. This is a profitable business, we can make PCs and it's not that expensive.

Well, what happens when a new firm comes in? The market supply curve flattens, because now, at any price, you're producing more. So the market supply curve flattens to the point SR2. Maybe you get a bunch of entrants until you get the market supply curve SR2, but then SR2 intersects demand at a new higher market quantity Q2. That higher market quantity going to the left, there's now no longer profits to be made, because at that market quantity, Dell is going to produce little q2. It's where the marginal cost curve intersects the average total cost curve, q2 is exactly at the minimum of the average total cost curve. So Dell no longer makes profits, the entry of some firms into the PC business has removed the profit from the PC business. Market quantity has gone up: $Q_2 \geq Q_1$; but Dell's quantity has gone down: $P_2 \leq P_1$ because more firms are in the market producing. So, *as more firms come in, total market quantity goes up, but any given firm is going to produce less and that will continue until profits go to zero.*

That is how firm entry wipes out profits: in the long run, firms make zero profit because, first of all, *entry drives price down to average cost* $(P \rightarrow AC)$ *and when price equals average cost, profits are zero*.

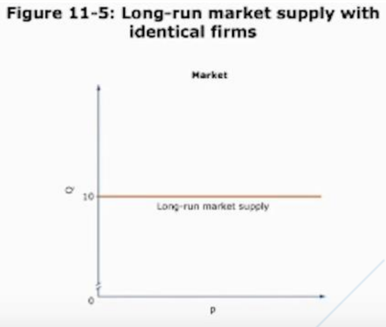
Because $\pi = P * q - C$, and if you divide by profits by $q, \frac{\pi}{q} = P - AC$.

So if $\boldsymbol{P = AC \rightarrow \pi = 0}$.

So entry drives profits to zero. It drives price to equal average cost.

Since $P = MC$, it's the point where $MC = AC$.

LONG RUN MARKET SUPPLY CURVE WITH PERFECT COMPETITION



In figure 11-5, we see that in the long run, firms always supply not on a single curve, but at a single point. *In the long run, with a perfectly competitive market, for a given firm, there is no longer even meaningfully a supply curve to a firm, there's just literally a supply point*.

Every firm produces at exactly the point where $\boldsymbol{MC = AC}$.

For a given firm, in the long run, they literally choose one production point which is technologically given. For a given firm, the market doesn't matter, for a given firm in a perfectly competitive market, we don't need to know anything about demand: all we need to know is the firm's production function. We don't even need to know anything about costs. All we need to know is their cost function. And then all we need to do is derive where $MC = AC$.

This is the power of the perfectly competitive equilibrium. In the long run, it's easy.

The firm will produce where $MC = AC$ and the price will be where $MC = AC$.

We can define the P and the Q in equilibrium just if we have a cost function.

Looking at the graph, the point where marginal costs equals average costs is the point of cost minimization in the long run. In the long run of perfectly competitive equilibrium firms will, by definition, minimize their costs, they will produce as efficiently as possible through the power of the market.

If you start a firm and you aren't minimizing in the short run, you might make money even if you aren't cost minimizing but, in long run, you'll get driven out of business. Because if there's someone else who can produce more cheaply than you, they'll be able to charge a lower price and drive you out of business. Your price will end up above the long run equilibrium price if you're not cost minimizing. So any firm that is not cost minimizing will get driven out of business. And the equilibrium will be a market where all firms are producing at the cost minimizing level.

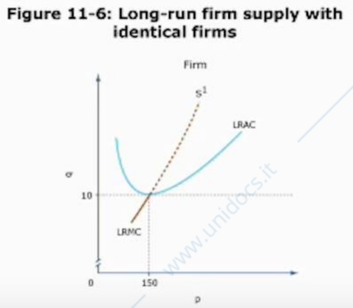


Figure 11-6: the long run market supply curve is perfectly elastic.

What determines perfect competition are two things: the demand curve to the firm was perfectly elastic, and the supply curve to the market is perfectly elastic.

And we talked last time about why the demand curve to the firm is perfectly elastic, because with lots of firms, any given firm has a perfectly elastic demand.

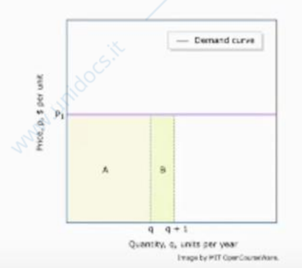
Now we've just derived why the market supply curve is perfectly elastic: it's perfectly elastic at the cost minimizing point.
If the price suddenly rises above that cost minimizing point, firms enter and drive the price back down. If the price ever drops below that cost minimizing point, firms exit, and the price goes back up. So through the power of firm entry and exit, in the long run, you end up with a horizontal or perfectly elastic supply curve.

Monopoly

MARGINAL REVENUE FOR A MONOPOLISTIC FIRM

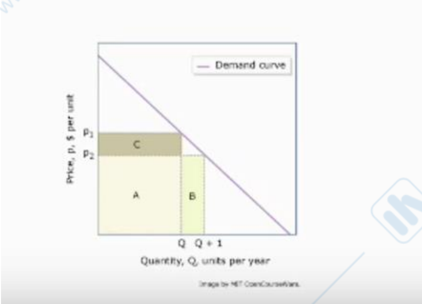
And today we're going to move beyond the unrealistic case of perfect competition to the somewhat more realistic case of monopoly. We've been discussing perfect competition thus far as a form of market organization, and that makes sense in some context like fast food and other things. But in most contexts, we think perfect composition is not the way the world works, there's some limits on competition: many markets, many of the goods we consume have only a few firms. So the most realistic model of markets would be one which accounts for the fact that there's less than an infinite number of firms, there's only a few firms.
We started with one extreme, which is a competitive market where there's an infinite number of firms. Now we're going to reverse field and talk about the other extreme, monopoly, which is only one firm. Then we'll talk about oligopoly, that middle case, which is multiple firms.
Monopoly is a market where there's only one firm and it is no longer price taker, it is now price maker.
Competitive firms are price takers: they were given a price by the market and they reacted to that and in the long run, that price is settled at the minimum of average cost. So basically, they were given the price which dictated production efficiency. You got a flat long run supply curve.
In a monopoly market, we don't meet the conditions for perfect competition. In particular, one condition was that consumers had perfect substitutes between your good and other goods they could buy; that's not true in a monopoly market.
The profit maximizing condition is that $MR = MC$ in perfect competition. And I decide to produce the next unit as long as the money I make off that unit exceeds what it cost to produce that unit. For a competitive firm, we said marginal revenue was just price, so the rule was set price equal to marginal cost: $P = MC$. But that was a particular case of marginal revenue. So for example, to see that, let's look at Figure 14-1, just a way to think about how marginal revenue is priced.

Figure 14-1: Average and marginal revenue for a competitive firm



A perfectly competitive firm faces a perfectly elastic demand curve, so they have to think about what the implications are of the marginal unit they sell. If they sell little q units, their revenue is A. If they sell one more unit, their revenue is B. And the marginal revenue is the height of that rectangle B times the base; the base is 1 because it goes from q to q + 1. The height is p, so the marginal revenue is price because marginal revenue is price for the perfectly competitive firm.
Looking at the monopoly case, the difference with a monopolist, as you see in Figure 14-2, is *they no longer face a perfectly elastic demand curve: they now face a downward-sloping demand curve.*
Remember, that for perfectly competitive case the graph is for little q, because a perfectly competitive case was not that we said the demand for the entire good was perfectly elastic, but just the residual demand facing one firm was perfectly elastic.
Now, a monopolist is the only firm in the market, so their residual demand equals total demand: they face the entire market demand curve. So as long as total market demand is downward-sloping, as we typically think it is, then they'll face a downward-sloping demand curve.

Figure 14-2: Average and marginal revenue for a monopolistic firm



Now we're talking about big Q's not little q's anymore, or we could say $q = Q$.

So now the firm reacts not to its residual firm demand curve, which is flat, but the downward-sloping market demand curve.

Now, we're going to make one assumption here that's very important: we're going to assume that the monopolist can only charge one price for their good to all consumers, we're going to assume a non-price discriminating monopolist.

Let's think about his decision to produce another unit: he's originally producing at Q at a price P1. If he wants to sell one more unit, he's going to have to lower the price to P2, because he now faces a downward-sloping demand curve.

So if he wants to sell one more unit, he's going to have to lower the price to P2. So on the one hand he's going to get the rectangle B, but on the other hand, on all the units he was selling at P1, he now gets a lower price P2 so he loses the rectangle C. So the marginal revenue for this monopolist is equal to the rectangle B minus the rectangle C. Or alternatively, $\boldsymbol{MR = P_1 - (P_1 - P_2)Q_1}$.

We could rewrite this as P1 plus how much the price changes when you change the quantity, times the original quantity Q1: $MR = P_1 + \frac{\Delta P}{\Delta Q} * Q_1$.

If revenue equals p times q, and q is a function of p, $R = P * q(P)$.

And then differentiating that, marginal revenue is: $MR = P + \frac{dP}{dQ} * Q$.

So marginal revenue is the price plus the change in price from selling another unit times the initial quantity. P is positive, it is always greater than 0; but the second term is negative because demand curves slope down:

$$MR = \underbrace{P}_{+} + \underbrace{\frac{dP}{dQ} * Q}_{-}.$$

So there's now two effects: there's a positive effect, which is if I sell another unit, I make money on that other unit; but there's also a negative effect, which is to sell that other unit, I have to lower the price because I face downward-sloping demand. So there's two effects a monopolist as he thinks about wanting to sell another unit. There's the money from that unit, but the lower willingness to pay for all previous units, that makes a monopolist a little more interesting.

We basically think of the monopolist as basically having to work down the demand curve. With a perfectly competitive firm, they don't have to work down the demand curve, the demand's flat to them. They can sell as much as they want at that price because they don't affect the price.

If the monopolist wants to sell more, he has to face the wrath of the market: if he wants to sell more, he have to lower the price to do so. What the monopolist is going to want to do is draw a marginal revenue curve. With the perfectly competitive firm, marginal revenue curve was just a price, it was given to them.

There was no marginal revenue curve. For a monopolist, there is a marginal revenue curve. So here I have a demand curve.

ELASTICITY AND MARGINAL REVENUE

There's a very important relationship between marginal revenue and the elasticity of demand.

So let's take our marginal revenue equation and put it back in change terms: : $MR = P + \frac{\Delta P}{\Delta Q} * Q$ and let's multiply and divide by P: : $MR = P + P(\frac{\Delta P}{\Delta Q}) * \frac{Q}{P}$

The reason to rewrite this is because now looks like the inverse of elasticity demand: $\frac{1}{\varepsilon} = \frac{\Delta P}{\Delta Q} * \frac{Q}{P}$

So we can rewrite this as : $\boldsymbol{MR = P * \left(1 + \frac{1}{\varepsilon}\right)}$.

The elasticity of demand facing a perfectly competitive firm is negative infinity, perfectly elastic: $MR = P$.

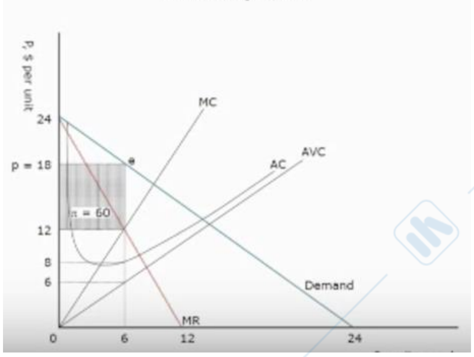
Now, instead, if we took a firm where the elasticity of demand was -1, $MR = 0$.

That is saying: if you're a monopolist facing an elasticity of demand of -1, then you make no money by selling the next unit, because these two effects exactly cancel. Exactly what you make by selling one more unit is offset by how much you have to lower the price on all your previous units.
As the elasticity of demand gets below -1, as it approaches 0 from below.
As the elasticity of demand approaches 0 from below, then the marginal revenue is going to become negative. So for example, if the elasticity of demand equals -0.5, then $MR = -P$, you lose money.
So as that elasticity of demand approaches 0, you're going to have a negative marginal revenue from selling the next unit. With a very inelastic good, you have to push the price down so much to sell the next unit that you lose money. Think about a very elastic versus very inelastic good:

- With a very elastically demanded good, to sell another unit you don't have to change the price much, because the demand curve's very flat, $\frac{\Delta P}{\Delta Q}$ is small, or $\frac{\Delta Q}{\Delta P}$ is big.
- With a very inelastically demanded good, to sell one more unit you're going to lower the price a ton, which is going to poison the revenues you get from selling that extra unit. So that's why marginal revenue will be higher, or will be a larger fraction of P as this elasticity becomes more negative.

PROFIT MAXIMIZATION AND SHUTDOWN CONDITIONS

Figure 14-4: Profit maximization for a monopolist



Profit Maximization for a Monopolist.
This cost function is: $C = 12 + q^2$ and the demand function is: $Q = 24 - P$.
The profit is maximized where $MR = MC$.
We know marginal revenue: $MR = 24 - 2Q \rightarrow 12 * 2 + \frac{dq^2}{dq}$
Marginal cost is the differentiation of the cost equation: MC =2Q.
So the optimization term for a monopolist is where $MR = 24 - 2Q = MC = 2Q$, which is $Q = 6$.
That's going to be the optimal production level for the monopolist.
So we can see that graphically: marginal cost curve hits the marginal revenue curve where the sales are 6 units.
What's the price? We might say marginal cost and marginal revenue intersect at 6 that means the price is going to be 12, but it can not be the price because it's not on the demand curve. The monopolist still has to respect the demand curve, so monopolists in setting their quantity, gets the intersection of MR and MC, but then in setting the price, they still have to read off the demand curve. They can't change consumer tastes. So they charge a price of 18.
You set marginal revenue equal marginal cost to derive Q, but then to get P, you've got to go back and plug that into the demand curve.
To get that P, I've got to go back and plug this in the demand function $P = 24 - Q = 24 - 6 = 18$.
The monopolist picks both price and quantity, but he has to pick them such that you get a point on the demand curve, respecting the demand curve.
In the short run, we still have another condition for profit maximization, which is the shutdown rule.
Remember the shutdown rule we talked about perfectly competitive firms in the short run, which is even if profits are negative, you might not shut down unless price is less than average variable cost: $P \leq AVC$.
In this case, the monopolist profits is 60, that's graphically the box, the rectangle, that's the difference between the average cost curve and the price they get.
Now once again, marginal revenue is gone, marginal revenue isn't something that actually exists in the market; it's just something the monopolist draws to pick what they're going to do, but then it disappears. What the monopolist cares about then is price, they're charging 18. Their average cost for that unit is only 8. So they're making a profit of 10 per unit on 6 units, so they're making a profit of 60.

If the monopoly sold that seventh unit, they would lose money on the seventh unit. Because if they sold that seventh unit, their price would have to be 17, so they'd sell one more unit at 17 that'd be good. But they'd lose $1 on the previous 6 units, which is bad. So how much revenues would they make?
The marginal revenue would be they make 17 minus the 6 poisoning effect, so $MR = 11$; their $MC = 2Q = 14, so$ they lose money.
The marginal cost of that next unit is only 14, they sell it for 17: they should do it. But we are missing that by selling it for 17, they've lost the dollar extra they make on each of the previous 6 units and that poisoning effect makes it unprofitable to do this. And that's why the monopolists stop short of what would be the perfectly competitive outcome.
The perfectly competitive firm would set marginal cost equal to demand and they would end up producing where marginal cost equals demand. $MC = Q \rightarrow 2Q = 24 - P$. So they would end up producing where marginal cost equals demand at a much higher level charging a slightly lower price.
So what you see is the monopolist ends up selling fewer units at a higher price.

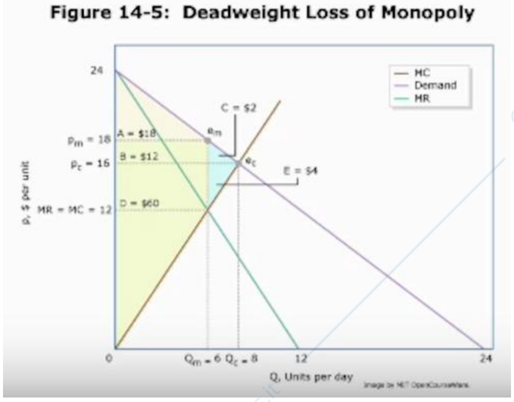
MARKET POWER

What monopolists have, is market power. Market power *is the ability to charge price above marginal cost*: $P > MC$.
The summary statistic of how much power a monopolist has is how much they can drive their price above marginal cost.
Remember the condition for profit maximization: $MR = P * \left(1 + \frac{1}{\varepsilon}\right) = MC$
So we can rewrite this as: $\frac{MC}{P} = \left(1 + \frac{1}{\varepsilon}\right)$
Let's define the mark-up *as price minus marginal cost, how much money you make on the next unit:* $Markup = P - MC$ *or the* $\%\ Markup = \frac{(P-MC)}{P}$.
Then you can see that, for a monopoly firm, $\boldsymbol{Markup} = \frac{(P-MC)}{P} = -\frac{1}{\varepsilon}$
The lower its elasticity, the more the monopolists can mark up their price, the monopolist will charge an incredible price. They'll still lose a lot of money if they try to raise that price, if they try to sell one more unit, but the first initial price they'll set will be incredibly high. So, basically at some point, there is some elasticity because there's a market demand curve, and basically what's going to determine how much market power the monopolist has is going to be how elastic it is, how close the substitutes are for that good. If there's close substitutes, the monopolist won't be able to charge a very high mark-up. If there's not close substitutes the monopolist can charge a very high mark-up and become very, very rich.

WELFARE EFFECTS OF MONOPOLY

So what is the welfare effects of monopoly?

Figure 14-5: Deadweight Loss of Monopoly



We know that the competitive firm maximizes welfare, we know that the best you can do is to sell 8 units at a price of 16.
What happens when you sell 6 units at a price of 18?
What happens is consumer surplus falls from A + B + C with perfect competition, to a consumer surplus equals to A with a monopoly. So you lose B + C with monopoly.
Producer surplus under perfect competition was the area D + E; now, under a monopolist, the producer surplus is equal to D + E + B. The monopolist , in this case, gained the rectangle B, but gave up the rectangle E. The consumer lost the rectangle B, that was a transfer to the monopolist. So there was a transfer of the rectangle B from the consumer to the monopolist, but C + E have disappeared. They're a deadweight loss, because in the perfectly competitive equilibrium these are trades that would have made both parties better

off. That is, these are trades which socially should happen; they are trades where the value to the consumer exceeds the cost of producing that unit.
Those seventh unit worth 17.
We can read that off the demand curve, that's a willingness to pay curve. People are willing to pay 17 for that seventh unit. And it costs 14 to be produced. So you have a unit which people want more than it costs to produce, yet it's not getting sold. That's deadweight loss. So monopolists induce deadweight loss because units that people value above their marginal cost doesn't get sold.

Oligopoly

TYPES OF OLIGOPOLIES

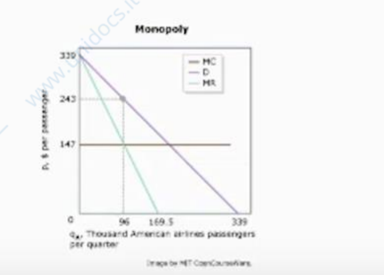
Most markets are better described as oligopolies, *these are markets where there's more than one market player, yet where each firm is large enough to actually affect the price*. So an oligopoly market is where there'll be a small number of firms in the market with *substantial barriers to entry from additional firms*.
So the classic example of an oligopoly industry is the auto industry: it's a market with a small number of dominant players; there's been some entry and exit over time, obviously, but it moves pretty slowly.
Firms in this market not behave like perfect competition where they can lazily take a price out of the market and just produce based on that price, but it's also not the same as monopoly where they can just get to set the price and not worry about what other people do. They're in this in-between situation where they have price setting power. They have some market power, but in a context where they have to worry about competitors. And so in this context there are two different ways firms can behave: they can behave cooperatively or non-cooperatively.
If they behave cooperatively to determine the outcome, we say that they form a cartel. The classic example, of course, here being OPEC, which is a cartel that drives the price of oil. Those countries cooperate in how much oil they produce to move the price up or down according to what the group desires. *And what cartels do is essentially turn oligopolies into monopolies, the cooperative equilibrium behave as if we're one big monopoly by cooperating*. And therefore, if you cooperate you can get all the wonderful things monopolies get: huge market power, huge profits, etcetera. But it turns out to be pretty hard to get a cooperative oligopoly, there's lots of reasons why it might fall apart and that's why in most oligopolistic markets firms behave non cooperatively.

COURNOT COMPETITION

Cournot model of non-cooperative oligopoly.
We're going to return to the example of prisoner's dilemma, but instead of just facing two choices, talk or not talk, we're going to talk about firms facing a whole continuum of choices. *Firms choosing how much they produce in a non-cooperative equilibria situation.*
Let's imagine there's two airlines that fly between New York and Chicago, American and United. And let's imagine for simplicity those are the only two airlines, those are your only two options flying New York to Chicago. How do United and American decide how many flights to run and what price to charge?
If they're monopolies we'd know; if it was a perfect competition we'd know, but to figure this out if we are in oligopoly is by looking for the Nash equilibrium, which we also call Cournot equilibrium. *Basically the quantity chosen by each firm such that holding all other firms' quantities constant*, is a profit-maximizing quantity. *And if each firm can choose a quantity that makes the market function where this is met, then you're in Cournot equilibrium.*
You're in Cournot equilibrium when each firm has decided, I'm happy. It's the same as the Nash concept: I'm happy with what I'm producing given what everybody else is producing. If everybody feels that way then you're in a Nash equilibrium or a Cournot equilibrium.
The steps of how you'd solve a Cournot equilibrium:

1. Create, compute each firm's residual demand. Residual demand curves is the demand for my firm given the quantity absorbed by other firms in the market. In this case, it's quantities absorbed by the one other firm in the market, but in general you do this with multiple players.
2. Develop a marginal revenue function. You calculate your marginal revenue which will be a function of other firm's quantities.
3. Do one and two for all firms. So for each firm you end up with a marginal revenue function and a function of all the firm's quantities.
4. You have n equations and n unknowns and you solve. So you develop a series of equation where each firm's marginal revenue function is a function of each other firms quantities. That leaves you n equations and n unknowns you solve. If you can solve it then you reach equilibrium, if you don't have a solution then there is no stable Nash equilibrium, but if you can solve that there is a Cournot equilibrium and you solve for it.

Figure 16-1: Profit-maximizing output under monopoly



So we'll start by doing this graphically, so let's start by considering the case of American Airlines in figure 16-1.

And let's say that the demand curve in this market, is: $P = 339 - q$.

So there's 339,000 flights that are demanded each month in the whole market. And let's also assume the $MC = 147\$$.

If American Airlines was a monopolist, it would set marginal revenues which are $MR = 339 - 2q$, (just multiply the demand curve by q and then differentiate it).

$$MR = MC \rightarrow 339 - 2q = 147 \rightarrow q = 96$$

So, if it was a monopolist it would choose a quantity of 96 and it would choose a price of $243, which we would just get out the demand curve.

The marginal revenue curve intersects the marginal cost curve at a quantity of 96,000, you then go up to the demand curve to read off the price.

American recognizes that United is in the market and United is going to deliver some amount of flights q_u. They don't quite know yet what it is, but they know there's going to be some amount of flights q_u.

So the residual demand for American is $q_a = Q - q_u$.

So, for example, let's say that American just guesses that United will fly 64,000 passengers, so what you want to do is then you just re-solve the problem but using residual demand. So then you say, well if United is going to fly 64,000 passengers then my residual demand is $P = 339 - q_a - q_u$,which I think is $P = 339 - q_a - 64$. So my new residual demand is $P = 275 - q_a$, that's what's left.

So if I use this as my new demand function and re-solve, my marginal revenues are $MR = 275 - 2q_a$; and my marginal cost is the same which is $MC = 147\$$..

If I do that I'm going to get a $q_a^* = 64{,}000$ flights.

If I was a monopolist I would have deliver 96,000 flights, but given that United is delivering 64,000, that's it's going to be optimal for me to also deliver 64,000.

At 64,000 flights my price is: $P = 275 - q_a$=275-64 =211

If I think United is delivering 64,000 flights then I'm going to deliver 64,000 flights at a price of $211.