UNIVERSITY OF CHICAGO
GRADUATE SCHOOL OF BUSINESS

Business 33101
Advanced Microeconomics
Class Notes
Kevin M. Murphy

I. The Basics of Consumer Choice – Stable Preferences

A. Broadly speaking, economists are students of human behavior. Economists study the behavior of individuals, firms and markets. The basic model of consumer choice is central to this analysis.

B. The basic model of consumer choice used by economists is the rational consumer. In the rational consumer model, individual consumers make choices based on the opportunities available (what economists call the individual's opportunity set) and their preferences.

C. Economists actually have very little to say about preferences. Broadly speaking, economists take preferences as given. In some sense, consumers like what they like. What is key from the economic perspective is that these preferences are stable (i.e., what we learn about preferences from consumer behavior in one context can be used to predict behavior in other situations).

II. Positive and Normative Elements of Consumer Choice

A. Economists are typically interested in two types of predictions, *positive* predictions that describe what will happen and *normative* predictions that make statements about which outcomes are better or worse. For example, we could make positive predictions about how a change in tax policy will affect the economy or make normative predictions about whether the change in tax policy will make consumers better or worse off.

B. The fact that we take preferences as given in our analysis will have important implications for normative analysis. In essence, economists will say consumers are better off with outcome A than with outcome B if and only if they would choose outcome A over outcome B. In the standard economic paradigm, what makes consumers better off is synonymous with what they themselves would choose if they had the power to do so.

III. Example #1: Fish vs. Leisure

A. To illustrate the economic model of consumer choice, we consider an individual by himself on an island who can either relax (i.e., take leisure) or catch fish. For every hour he fishes, he catches two fish.

1. Note: Here scarcity (one of the basic elements of economics) is present since more fish can be caught only at the expense of less rest. The cost of a fish is the half hour of rest that must be sacrificed to catch that fish. The cost of a good is always the amount of the other good or goods that must be given up. Hence the *cost of fish* is expressed in terms of *hours of leisure* in this case!

2. Similarly, the cost of an hour of leisure is the two fish that could have been caught with the hour. Hence the cost of leisure is expressed in terms of fish as 2 fish. The fact that the cost of leisure is the 2 fish that could have been caught with that hour illustrates that costs are really *opportunity costs* (i.e. what alternative we give up – in this case fish - to obtain the good in question – in this case leisure). *The cost of a good is always what you give up in order to obtain that good.*

B. What choices does the stranded island dweller have?

1. This can be expressed as:

$$F \leq 48 - 2L$$

where

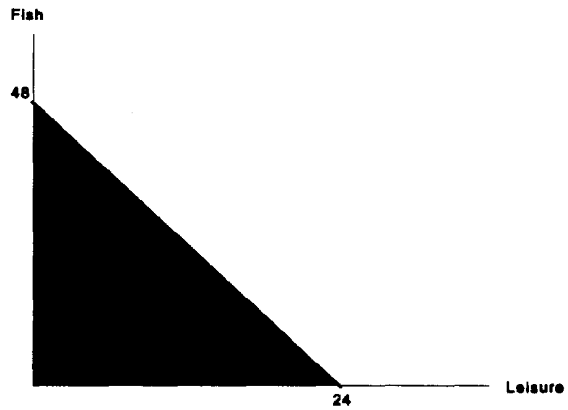
a. F = the number of fish per day
b. L = the number of hours of leisure per day
c. Our individual has 24 hours per day

C. The "less than or equal to" sign follows from the fact that he can always discard some of his catch (though he typically would not want to do so).

D. We compute this as follows. If he fishes for the full 24 hours he will catch $2 \times 24 = 48$ fish. For each hour of leisure he gets 2 fewer fish (the cost of leisure). For L hours of leisure he loses $2 \times L$ fish. Hence $F = 48 - 2L$ if no fish are discarded. Allowing him to discard fish yields the inequality $F \leq 48 - 2L$.

E. We can compute an equivalent expression by working in terms of hours rather than fish. We begin with 24 hours of leisure and no fish. For each fish he must give up .5 hours. Hence for F fish he must give up $.5 \times F$ hours. Hence $L = 24 - .5F$ if he doesn't discard any fish. Though this equation doesn't look exactly as that obtained above, some simple algebra will show it is the same as that obtained in part (D).

F. A convenient way to look at this problem (and one frequently used by economists) is to graph it. To do this we put hours of leisure on the horizontal axis and the number of fish on the vertical axis:



G. The shaded area is the area available to the individual. The heavy line represents the combinations of fish and rest available if no fish are discarded. The shaded area is called the individual's *opportunity set* since he can choose any point (i.e., some combination of F fish and L hours of leisure) that lies within this region. We define the opportunity set as the choices available to the consumer.

H. How will this choice be made? To answer this question we must make a brief digression and study what economists call preferences.

Preferences (a digression)

1. An individual's preferences tell us how that individual will choose among the choices available to him. In our example they tell us how many fish and how many hours of rest the individual will choose from the opportunities available to him (i.e., from the points in the shaded area above). Remember that consumers make choices between bundles of goods (in this case 20 hours of leisure and 8 fish or 16 hours of leisure and 16 fish).

2. Assumptions on preferences. We assume the following three properties hold for consumers. Here A, B, C, D represent points in the graph (e.g., A could be 6 fish and 21 hours of leisure, B could be 20 fish and 10 hours of leisure, . . .).

   a. The individual can rank any two possible choices A and B according to: A is preferred to B, B is preferred to A, or A and B are equally preferred (or indifferent).

b. If A is preferred to B and B is preferred to C, then A is preferred to C. The same holds if we replace "preferred" by "indifferent."

c. More is preferred to less: If A has more of every good than D, then A is preferred to D.

3. Note that by a point A we mean a point in the graph (i.e. an amount of fish and leisure). Hence preferences are defined over *consumption bundles*. A consumption bundle is a list of the quantities of each good associated with a particular choice. In our example (5 hours of leisure and 20 fish) is one consumption bundle as would be (3 hours of leisure and 30 fish). As stated above, the set of consumption bundles that is available to the consumer is known as the opportunity set.

4. We now define an indifference curve. We begin with any consumption bundle (for example point A) and look at all points that are indifferent to A according to the individual's preferences. The locus of points that are indifferent to a given point (and hence to each other by property (b) on preferences) is known as an *indifference curve*. The following properties of indifference curves follow from the assumptions about preferences.

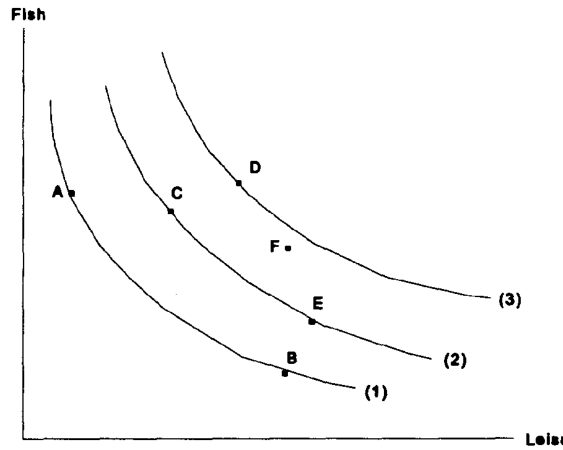
   a. Indifference curves slope down. This follows from assumptions (a) and(c) on preferences.

   b. Indifference curves cannot intersect. Assumptions (b) and (c) on preferences imply this.

   c. An indifference curve passes through every point. (This can be regarded as an assumption. It can be proved by more difficult means.)

5. We make one additional assumption about indifference curves that can be verified by observing consumer choice. This assumption is that the indifference curves are convex to the origin. That is, the slope of an indifference curve decreases in absolute value as we move down the indifference curve. What this means is that as an individual substitutes good x for good y (i.e., gets more x and less y) his value of y in terms of x increases. Hence, this is an assumption that as a good becomes more scarce to a particular individual the value of that good in terms of the other good increases.

6. These properties are illustrated by the indifference curves in the following diagram:



Here all the points on curve (1) are valued equally to each other. Similarly all the points on curves (2) and (3) are indifferent to all points on the same curve. By assumption (c) on preferences all points on (3) are preferred to all points on (2) which are preferred to all points on (1).

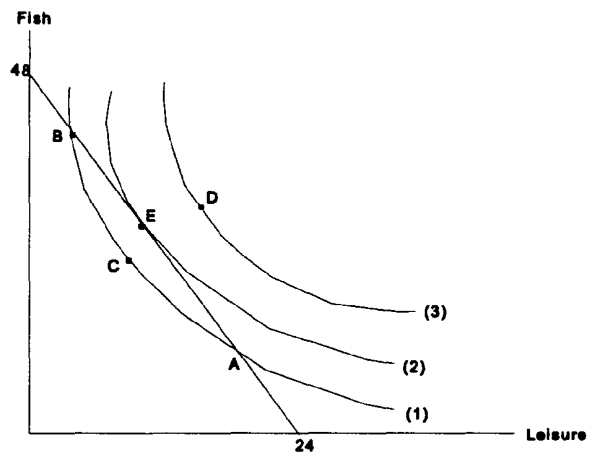
Note: We could have known that D was preferred to C without knowledge of the individual's indifference curves since D has more of both goods than C. However, to know if F is preferred to another point such as C we must know the shape of the individual's indifference curves. While C has more fish than F, F has more leisure; hence, we must know the individual's relative values of fish and leisure (i.e., his indifference curve) to know which point he will choose.

7. From this we can see that a key aspect of indifference curves is that they tell us how an individual values one good in terms of another. Just as costs are always expressed in terms of another good, we always talk of values in terms of another good as well. Economists define the *marginal value* of good X in terms of good Y as the amount of good Y an individual is willing to give up for an additional unit of X. The slope of the individual's indifference curve tells us how many fish he is willing to give up for an additional hour of leisure. This is known as the individual's *value of leisure in terms of fish.* A statement about consumer preferences is essentially a statement about the relative values a consumer places on the different goods in that bundle. Hence, if the slope of the indifference curve at point C in the graph above is –1.5 then this would say that a consumer currently consuming at point C would value 1 hour of leisure at the equivalent of 1.5 fish. As can be seen from the picture, the values that individuals place on goods are determined by both their preferences and the amount of each good they have to consume (i.e., the slopes of the indifference curves will not in general be the same at all points).
[This ends the digression on preferences.]

I. Using these ideas we can determine how our island dweller will choose among his opportunities. This is stated as the fundamental axiom of consumer choice:

An individual will choose the point from his opportunity set that yields the highest level of satisfaction or utility. In terms of indifference curves, the individual will choose the point in this opportunity set that puts him on the highest indifference curve.

J. The point the individual chooses must be like that in the graph below at point E.



Here his available points just allow him to reach the level of satisfaction associated with indifference curve (2). A point such as C will not be chosen since E has more of both goods. Since more is preferred to less we can rule out all points not on the edge of the opportunity set.

K. The key issue is that at point E the slope of the opportunity set is equal to the slope of the individual's indifference curve. <u>The slope of the indifference curve tells how many fish the individual is *willing* to sacrifice to obtain an extra hour of leisure</u>. <u>The slope of the opportunity set tells us how many fish the individual *must* sacrifice in order to get an additional hour of leisure</u> (2 in this case).

L. The preferred point E has the property that the individual's willingness to sacrifice fish for leisure (i.e., his marginal value of leisure in terms of fish) is equal to the marginal cost of leisure or how many fish must be sacrificed.

M. At a point like B the cost of an hour of rest is 2 fish but the slope of the individual's indifference curve is greater in absolute value, indicating that additional leisure is worth *more* than the two fish per hour that must be sacrificed. Since the value of leisure (in terms of fish) exceeds the cost of leisure (again in terms of fish), it will pay for the individual to move toward E and take more leisure and consume fewer fish!

The exact opposite holds at a point such as A.

N. The equilibrium at point E illustrates an important point. An individual is led to choose a point where the marginal rate of substitution in consumption (i.e., the marginal value of leisure in terms of fish) is equal to the marginal cost of the good (i.e., how many fish must be given up to get an additional hour of leisure).

IV. Example #2: Exchange

A. This example will illustrate how trade or exchange leads a free market economy to equate the marginal values of all individuals in an economy.

B. We consider an example with two individuals, Bill and Tim, and two goods, bananas and oranges. Bill starts out with 20 bananas and 5 oranges. Tim begins with 20 oranges and 5 bananas.

C. The two individuals' preferences are as follows. Bill is indifferent between the following bundles:

BILL	A	B	C	D	E	F
Bananas	25	20	16	13	11	10
Oranges	4	5	6	7	8	9

Tim is indifferent between the following bundles:

TIM	A	B	C	D	E	F
Bananas	4.5	5	6	8	11	15
Oranges	21	20	19	18	17	16

D. Both traders begin at position B. At this point Bill is willing to give up 4 bananas to get an additional orange whereas Tim will give up one orange to get an additional banana. In this example, at the initial endowments Bill values oranges (in terms of bananas) relatively more than Tim. Similarly, Tim values bananas (in terms of oranges) more highly than Bill. Hence we have potential gains to trade. Since Bill values oranges relatively more than Tim (remember, in terms of bananas) the trade will give oranges to Bill in exchange for bananas. Here Bill values an additional orange at 4 bananas while Tim is willing to give up an orange for 1 banana. This difference in marginal rates of substitution signals potential gains from trade.

E. After trading one unit, Bill values the next orange at 16 - 13 = 3 bananas, while Tim is willing to give up the orange for only 8 - 6 = 2 bananas; hence it will pay to trade another unit as well since the gain to Bill (in terms of bananas) is 3 while the cost to Tim is only 2. After trading 2 oranges Bill values an additional orange at only 2 bananas while Tim values it at 3; therefore, additional units will not be traded. From this analysis we can infer that both Bill and Tim would be better off if they trade 2 oranges for (say) 5 bananas. Actually they could trade the 2 oranges for any amount of bananas between 3 and 7 and still make both Bill and Tim better off than without trade.

F. With a trade of 2 oranges for 5 bananas Bill gets a gain from trade equal to this value of the two oranges, i.e., 4 + 3 = 7 minus what he paid (5), for a gain of 7 - 5 = 2 bananas. This example illustrates several concepts used in economics. Here 7 = 4 + 3 is called Bill's *total value* of the 2 oranges; the difference between this and the 5 bananas he pays is called Bill's *consumer surplus*.

G. We can do the same analysis for Tim. Tim gets 5 bananas for the 2 oranges he valued as 1 + 2 = 3 bananas for a surplus of 2 bananas

H. Note that in this example all values for oranges are expressed in terms of bananas in keeping with what we said earlier. An equivalent analysis can in principle be carried out using bananas as the good to be traded and expressing the values in terms of oranges.

V. Demand with Money Prices and Money Income

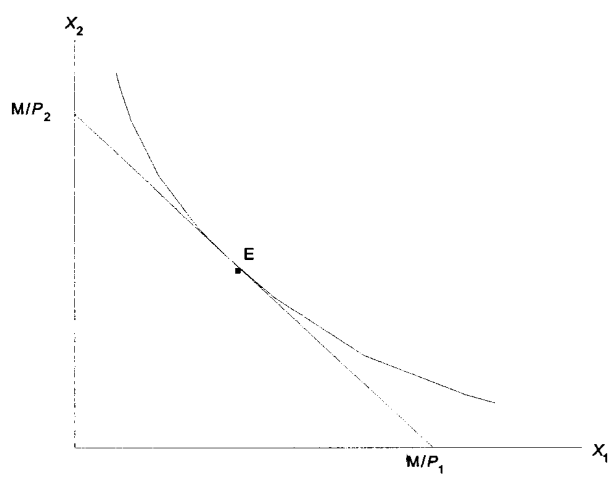
A. The preceding examples served to emphasize that the cost and value of a good are always in terms of another good or goods. In an economy with money, this fact is sometimes hidden. When we say the cost of a good is \$2.00 what we really mean is that the \$2.00 represents the value (in terms of money) of the other goods that must be forsaken.

B. Consider an individual with a money income of \$$M$ in a world with two goods, X_1 and X_2. He can buy X_1 at price P_1 and X_2 at price P_2.

C. The first step in our analysis is to determine his opportunity set. Letting X_1 be the amount of good 1 he purchases, we can determine opportunity set. In this case we sometimes call his opportunity set his *budget set* since it gives all the bundles that are within his budget. Since the price of good 1 is P_1, if he buys X_1 units his expenditure on good 1 is P_1X_1; similarly, his expenditure on good 2 is P_2X_2. Since he cannot spend more than his income, his budget set is clearly

$$P_1X_1 + P_2X_2 \leq M.$$

D. If we plot this set we obtain the triangular-shaped region shown below. Just as in the



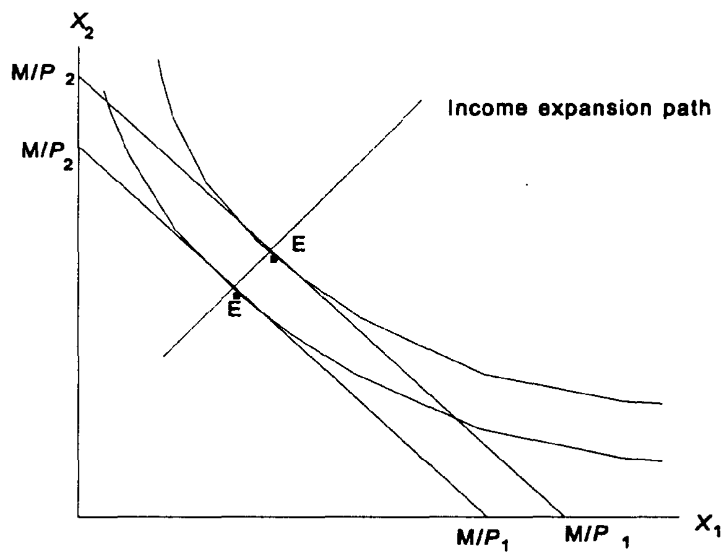
fish example, the point the consumer chooses must be like that at point E where the individual's indifference curve is tangent to the budget set. Clearly, if we double M, P_1, and P_2, the individual's budget set will not change and hence he will choose the same point since he faces the same alternatives. This implies that it is only relative prices that matter in determining the individual's choice. We can divide the budget constraint through by P_1 to obtain $X_1 + (P_2/P_1)X_2 < M/P_1$. Here P_2/P_1 is the price of good 2 in terms of good 1 (i.e., the cost to the consumer of good 2 in terms of good 1) and M/P_1 is the value of the individual's income in terms of good 1. Now this problem is exactly equivalent to our earlier fish vs. leisure example. Let X_1 be leisure, X_2 be fish, $P_2/P_1 = .5$ and $M/P_1 = 24$. This is the way in which adding money really does not change the basic consumer's problem.

E. If we do not know this person's preferences (i.e., the shape of his indifference curves), we cannot say where the tangency point will occur. In fact, since economists do not have a theory of what makes preferences the way they are, we cannot hope to explain why he would choose one point over another. For example, we can say people value cars and gold but dislike garbage based on what we observe, but we cannot say why.

F. Instead, economists prefer to try to explain how *changes* in opportunities (i.e., changes in *prices* and *income*) will change a person's consumption choices.

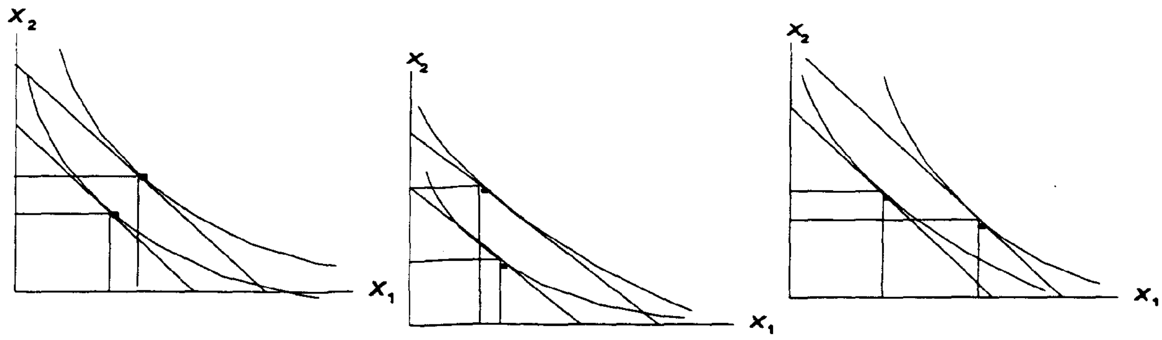
G. Since we do not have a theory of how tastes are determined, economists *do not try to explain changes in consumption choices by changes in tastes.*

H. With this in mind, we turn to a discussion of how changes in income affect consumption. As we saw in the fish vs. leisure example, since more is preferred to less, an individual will consume all of his income. When we increase the individual's income from M to M′ the budget constraint shifts out in a parallel fashion as in the graph below.



The slope of the budget line is $-P_1/P_2$ and since prices are not changing this slope will remain the same. If we continue to vary income we will trace out a curve which we call the individual's *income expansion path*. This curve tells us how the amount of each good purchased will change as we change income, holding prices fixed. Since the consumer will spend all his additional income and prices have not changed, consumption of at least one good must increase.

I. However, consumption of some goods may fall. For example, in the two-good case we have the following possibilities:



Case A Case B Case C

In case A consumption of both goods increases; this is the usual case. In case B consumption of good X_2 rises while consumption of good X_1 falls. In case C the opposite occurs: the consumption of X_1 rises while consumption of X_2 falls.

J. Goods for which consumption falls when income increases are called *inferior goods*. Goods for which consumption moves in the same direction as income are called *normal goods*. Normal goods are the rule and inferior goods tend to be the exception.

K. In order to describe in detail how consumption responds to changes in income we define the *income elasticity* of a good. The income elasticity of a good is the percentage change in the quantity of the good consumed divided by the percentage change in income (note we are holding all prices constant in this analysis). Denote consumption by X, income by M, the change in income by *dM*, and the change in consumption by *dX*. Using this notation we have the income elasticity, *N*, as:

$$N = \frac{dx/X}{dM/M} = \left(\frac{dx}{dM}\right)\left(\frac{M}{X}\right) = \text{Income Elasticity.}$$

L. For normal goods *dX* and *dM* have the same sign and hence the income elasticity is positive. For inferior goods *dX* has the opposite sign of *dM* (i.e., for an increase in M, $dM > 0$ and the change in consumption is negative: $dX < 0$), hence the income elasticity is negative for inferior goods.

M. We can use this definition to derive an identity for income elasticities. Consider two income levels, M and M′, with prices held fixed. At M the consumer chooses X_1 and X_2 and at M′ the consumer chooses X_1' and X_2'. We then have

$$X_1P_1 + X_2P_2 = M$$

and

$$X_1'P_1 + X_2'P_2 = M'.$$

Subtracting these two equations and collecting terms yields

$$P_1(X_1' - X_1) + P_2(X_2' - X_2) = M' - M.$$

But by definition $(X_1' - X_1) = dX_1$, $(X_2' - X_2) = dX_2$, and $M' - M = dM$. Hence, we have

$$P_1dX_1 + P_2dX_2 = dM.$$

Dividing by dM yields

$$P_1dX_1/dM + P_2dX_2/dM = 1.$$

This can be rewritten as

$$(\frac{P_1X_1}{M})(\frac{dX_1}{dM})(\frac{M}{X_1}) + (\frac{P_2X_2}{M})(\frac{dX2}{dM})(\frac{M}{X_2}) = 1.$$

But $P_1X_1/M = K_1$ is the fraction of income spend on good 1; $(dX_1/dM)(M/X_1) = N_1$ is the income elasticity for good 1; and the same is true for the corresponding terms of X_2. Using these definitions we have

$$K_1N_1 + K_2N_2 = 1.$$

This says that a weighted average of income elasticities is always equal to 1. The weights are the fraction of income spent on each good, which is commonly referred to as the budget share of the good.

N. Above we defined the income elasticity of demand for a good as the percentage change in the quantity of the good purchased divided by the percentage change in income. If the income elasticity of demand for good X is less than 1, then the percentage change in the quantity purchased must be less than the percentage change in income. Since the price of the good is unchanged, this would imply that the *fraction* of income spent on good X will fall as income rises. Conversely, if the income elasticity of demand is greater than 1 then consumption of (and hence the amount spent on) X increases by a greater percentage than income, leading to an increase in the fraction of income spent on good X.

O. Goods with income elasticities less than 1 (i.e., goods whose share of the budget falls as income rises) are called *necessities*. Goods for which the share of the budget rises as income increases are called *luxuries*. Hence, all inferior goods are by definition necessities. To review, we have

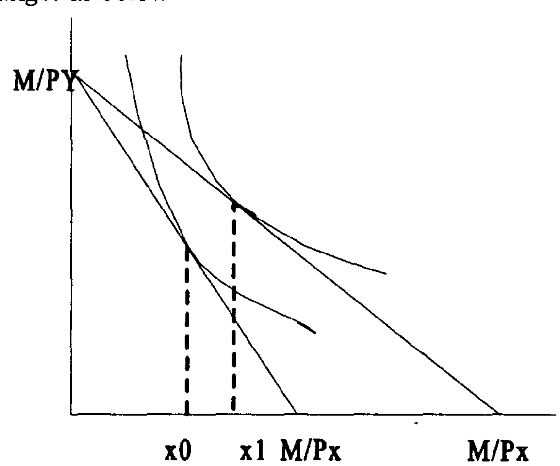
Income Elasticity (N)	Normal/Inferior	Necessity/Luxury
$N < 0$	Inferior	Necessity
$0 < N < 1$	Normal	Necessity
$N > 1$	Normal	Luxury

P. One should keep in mind that most broadly defined classes of goods such as food, clothing, or housing — taken as a whole — are normal goods. Inferior goods generally occur when we define goods more finely. For example, while "cars" as a whole are a normal good, people in certain income groups may reduce purchases of certain types of cars (e.g., Volkswagens) as their income rises.

Q. One should also keep in mind that the income elasticity can (and often does) vary with the level of income. Typically, narrowly defined goods are normal over a range of income and then become inferior at higher income levels as individuals move to higher quality products with further increases in income. No good can be inferior over all income levels (why?).

VI. Changes in Prices

A. Having considered how consumption of a good varies as we change income, the next logical experiment is to see how consumption varies when we change the price of a good. In our two-good example, with two goods x and y, if we lower the price of x the budget set changes as below:



Note that when prices fall from P_x to P'_x the budget line shifts out but *not* in a parallel fashion. The intercept on the y axis stays fixed. The change in the slope of the line represents the change in relative prices. That is, when the price of x fell from P_x to P'_x

one can obtain an extra unit of x without giving up as much y as before the price change. This induces individuals to substitute the now cheaper x for y. In addition, since the new budget line lies everywhere above and to the right of the old budget line the individual can now purchase more of *both* goods. Hence he has realized a growth in real income via the fall in price of good x.

B. We are interested in predicting how the consumption of x and y will respond to this change in price. Economists refer to the observed change in the consumption of good x when its prices changes as the **total effect.** This would simply be the difference X_1 - X_0 in the graph above.

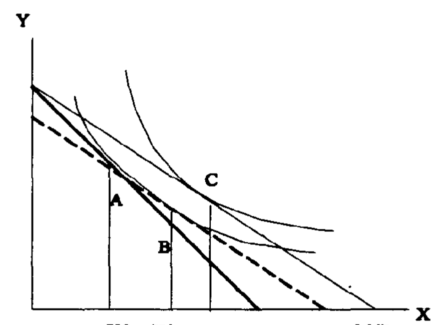
C. As we stated above, the budget set changes in two ways. First, the slope changes, which represents a change in relative prices and an incentive for the individual to substitute the now cheaper x for y. Second, the price change allows the individual an increase in opportunities to purchase both goods, hence a gain in real income.

D. For analytical purposes, it is often useful to separate out the effect of changing the relative price of good x (i.e. changing the amount of good y that must be given up to obtain an additional unit of good x) and the change in real income generated by the change in the price of good x. Economists typically separate out these two effects into what are called the **income** and **substitution** effects of a price change.

E. The change in the consumption of x due to the change in relative prices, holding the individual's "real income" or utility constant, is called the **substitution effect.** It tells us how the individual will substitute x for y with his "real income" held constant.

F. The *total effect* minus the *substitution effect* (or equivalently, the change resulting from moving to a new indifference curve) is called the **income effect.**

G. The reason for this terminology and the way to see these two effects in a graph can be seen below.



The shift in price of x from P_x to P'_x, leaving the individual on the same indifference curve, moves the individual from point A to point B. (We determine his point by drawing a new budget line parallel to the new price line but tangent to the original indifference curve.) His consumption of good *x* changes from X_0 to X_1. The distance X_1 - X_0 is called the substitution effect. The substitution effect always increases x and reduces y when we reduce the price of x (or increase the price of y).

H. The change from point B to point C is the response to a parallel shift in the budget constraint (e.g. if we changed the individual's nominal income and held prices fixed at their new level). This is exactly what happens when income changes. Hence the distance X_2 - X_1 is called the income effect of the price change

I. **The substitution effect always leads to an increase in the consumption of x when its price falls** and a decrease in the consumption of good *x* when the price of *x* rises.

J. The *income effect* can lead to either an increase or decrease in the consumption of *x* when P_x falls depending on whether *x* is a normal or an inferior good. In the picture above I assumed that *x* was a normal good and hence x_2 was larger than x_1 and the income effect reinforced the substitution effect. **For all normal goods the income effect of a change in the price of that good always reinforces the substitution effect** and a fall (rise) in the price of good *x* must lead to a rise (fall) in the quantity of *x* purchased.

K. If x is an inferior good then the income effect goes in the opposite direction of the substitution effect, making the sign of the total effect unknown in theory for an inferior good.

L. Note the size of the income effect is proportional to the share of income spent on the good. When the price of good x falls by k dollars the "gain" in income to the consumer is roughly k times the number of units of x purchased. This is the amount he no longer has to spend to finance his initial purchases, hence this is the increased amount he has to spend on additional x or other goods. If this change of $k \times X_0$ dollars is a small share of the individual's budget then the income effect will be small.

M. The only possibility for the income effect to dominate the substitution effect is if the good is inferior and it is also a very large fraction of the individual's budget. But as we stated above, inferior goods are narrowly defined goods and hence typically have small budget shares (and good substitutes). **No examples of the income effect outweighing the substitution effect for an inferior good have ever been observed.**

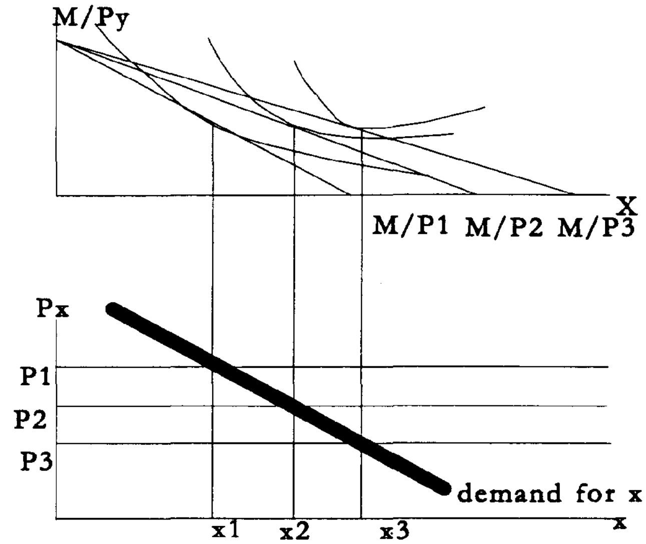
N. Given these results we can state the **Law of Demand**:

For all normal goods and all observed cases of inferior goods a decrease in the price of a good (money income and the prices of other goods held constant) leads to an increase in the quantity of the good consumed.

VII. The Demand Curve

A. The relationship of the quantity of good x purchased and its price P_x, holding money income and the prices of all other goods constant, is called the demand curve for good x.

B. We can derive this relationship as follows. We begin with two graphs. The first is the simple budget set diagram that we have been using so far. On the other graph we will put the quantity of x purchased on the horizontal axis and the price of x, P_x, on the vertical axis. In the second diagram we will plot the quantities of x the consumer chooses to purchase at each price P_x.



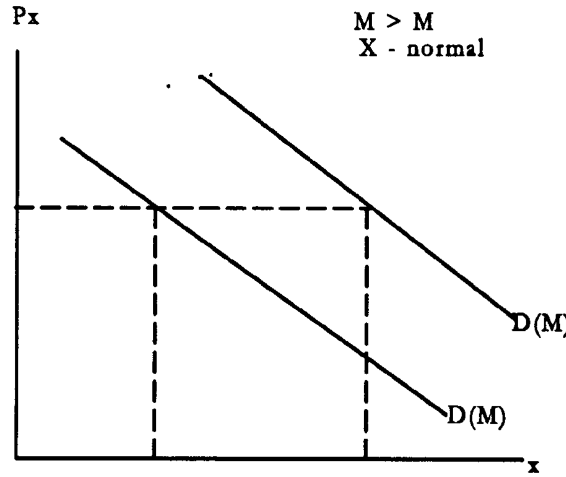
One should always remember that this relationship is derived *holding money income and the prices of other goods constant.* At a price of P_1 we see that the consumer chooses to purchase X_1 units of *x*. At a power price of P_2 his consumption rises (remember the Law of Demand) to X_2. Both of these points can then be placed in the second graph and similarly for P_3 and X_3. Due to the Law of Demand the demand curve must slope down. The demand curve tells us *how much an individual will consume at each price.* By proceeding as above we can then fill in quantities that the individual will consume at each price. Connecting these points we obtain the individual's demand curve. The demand curve is simply a schedule that summarizes an individual's choices (and hence his underlying preferences). It tells us how much of good *x* an individual will purchase at each price given a level of money income and fixed prices of the other goods.

C. How to use demand curves: To find out how much an individual will purchase at a given price we simply extend a horizontal line at the price we are interested in until it reaches the demand curve. The point where this line hits the demand curve then gives the quantity the individual will purchase. We will see more how to use demand curves in the upcoming lectures.

D. As we emphasized above, a demand curve is derived holding income and the prices of other goods constant. Changes in these other variables will lead to a shift in the demand curve. We call such a change a **change in demand** (to be distinguished from a change in the quantity demanded, defined below).

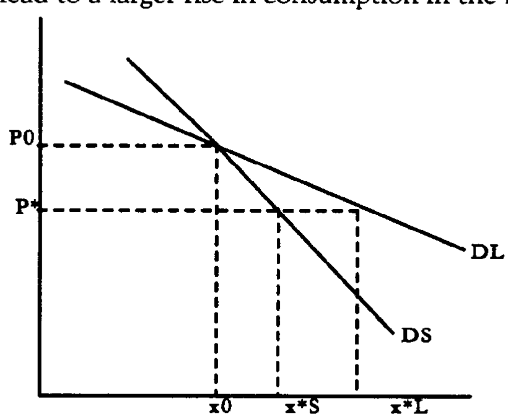
E. For example, if *x* is a normal good then a rise in income will lead to an increase in the quantity of *x* purchased at each price. This can be seen as a shift out in the demand curve for good x as in the graph. Similarly a decrease in money income will lead to an inward shift in the demand curve for a normal good. For an inferior good the effects would go in exactly the opposite direction.

F. Holding the price of x constant, changes in income change the quantity of x purchased by shifting the demand curve in accordance with the income elasticity for the good.



G. Changes in the quantity of a good purchased due to a change in its own price occur by *moving along a demand curve* to the new price. Such changes are called *changes in the quantity demanded* to emphasize that the demand curve does not shift.

H. It is also important to distinguish between the response to a price change in the short versus long run. For example, the response to a permanent rise in the price of petroleum will be different a year after the price rise than in the first week. Basically the long run allows for increased substitutability over the short run. While in the short run one can reduce the amount of driving and lower the temperature of one's house, in the long run one can purchase a smaller car and reduce heating costs further through increased efficiency of furnaces and insulation. Given this we have the relationship illustrated below between the short-run demand curve D_S and the long-run demand curve D_L. Hence for rises in price the reduction in quantity will be greater in the long run than in the short run. Similarly, a fall in price will lead to a larger rise in consumption in the long run.



VIII. Elasticity of Demand

A. Similar to our definition of income elasticity of demand for a good, we define the own **price elasticity** of demand for good *x* as the percentage change in the consumption of good *x* divided by the percentage change in the price of good *x*. If we denote the change in the price of good *x* as dp and the change in the quantity of *x* consumed by dx, we have the price elasticity E_x as:

$$E_x = (\% \text{ change in } x)/(\% \text{ change in } P_x) = (dx/x)(dp/p).$$

This can then be rewritten as

$$E_x = dx/dp\ (p/x).$$

B. The price elasticity is always negative. The Law of Demand implies that the change in *x* is always of the opposite sign as the change in P_x. If dp is positive, that is, prices rise, then dx will be negative, the quantity of *x* consumed must fall.

C. If the demand for one good has a more negative price elasticity (i.e., it is larger in absolute value) than for a second good, we say that the demand curve for the first good is *more elastic* than the demand for good 2. The greater the elasticity, the greater the percentage change in consumption for a given percentage change in price. Hence, more elastic demand curves are those with a greater quantity response to a price change.

D. We make one major distinction on the elasticity of demand. If $E_x < -1.0$ we say that demand is *price elastic*. If $0 > E_x > -1.0$, then we say that demand is *price inelastic*. If $E_x = -1.0$ then we say that demand is unitary elastic. The major significance of this distinction will be clear when we discuss the effect of elasticity on total expenditure below.

E. We define total expenditure as the amount spent on purchases of good *x*. In terms of X and P_x we have TOTAL EXPENDITURE = $X \times P_x$. Total expenditure is simply the market value of his purchases of good *x* in terms of dollars.

F. If the demand curve is inelastic (i.e., $0 > E_x > -1.0$) then a drop in the price of good *x* will lead to a percentage increase in consumption smaller than the percentage decrease in price. This implies that the total amount spent on good *x* will decline. Hence, when the demand for a good is inelastic expenditures on that good fall as we reduce its price, holding income and the prices of other goods constant.

G. If demand is *elastic* then the percentage increase in consumption due to a fall in price is greater than the percentage reduction in price, hence expenditure on that good (or total expenditure) will rise as we reduce price.

I. Earlier we introduced the concepts of consumer surplus and total value. In terms of the demand curve we can consider these same ideas. To fix ideas, it is useful to consider a good that must be consumed in integer amounts (i.e., the consumer must choose to consume 1 unit, 2 units, 3 units etc. but cannot choose 2.5 units. Consider the demand curve represented by the table below.

Price	Quantity Demanded	Total Expenditure (at highest price)	Total Value	Consumer Surplus (at highest price)
$P > 4$	0	0	0	0
$4 \geq P > 3$	1	4	4	0
$3 \geq P > 2$	2	6	7	1
$2 \geq P > 1$	3	6	9	3
$1 \geq P > 0$	4	4	10	6

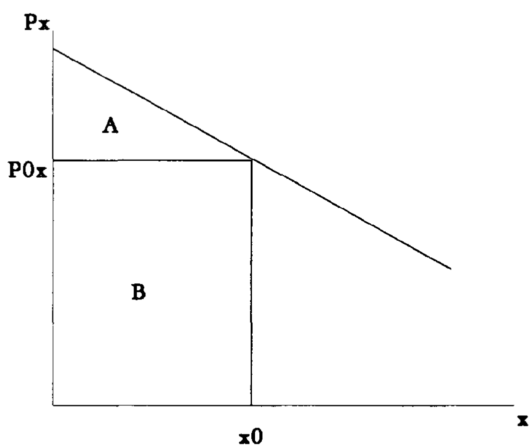
J. The first two columns describe the demand curve by giving us the quantity the individual will purchase at each price. At prices above $4 per unit the individual demands zero units. At prices between $4 and $3 the individual demands only one unit. When prices fall to $3 the individual demands a second unit.

K. **Total value** is defined as the total value to the individual of the quantity purchased. Since he is willing to buy the first unit for $4 but not at any price above $4, THE FIRST UNIT OF THE GOOD MUST BE WORTH $4 TO THIS INDIVIDUAL. If it were worth more he would buy it at a higher price. If it were worth less he would not buy it at $4. Similarly, the second unit must be worth $3 and the third worth $2. The *total value* to the individual of a given number of units is the sum of his marginal values, which is just the sum of the prices he is willing to pay for each unit.

L. **Consumer surplus** is defined as the individual's gains from trade or simply *total value* minus *total expenditure* (i.e., the amount he is willing to pay minus the quantity he actually had to pay). For example, if the price in the market was $2.50, the individual would buy 2 units. The individual's value of these two units is $7 while the amount he pays to obtain them is only $5 (2 times $2.50). The individual's consumer surplus is then $7 - $5 = $2. If the price fell to $1.50, the individual would now buy a third unit. The total value of 3 units is the $4 he was willing to pay for the first unit, plus the $3 he was willing to pay for the second unit plus the $2 he was willing to pay for the third unit for a total of $9. Total expenditures are 3 times $1.50 or $4.50 for a surplus of $9.00-$4.50 = $4.50. The fall in price from $2.50 to $1.50 increased the individual's surplus by $2.50. In essence, the individual pays $1.00 less per unit for each of the two units he was already buying for a gain of $2.00 plus gets the additional gains from trade of $.50 from buying the third unit (for which he was willing to pay $2.00 but only had to pay $1.50).

M. One should note that while total revenue can fall with a reduction in price (and a corresponding increase in the quantity demanded), total value and consumer surplus *must rise*. This simply says that at a lower price the consumer must get more gains from trade and hence is better off.

N. We can also do this graphically for a good for which the individual can purchase any amount, not just an integer number of units as in the above example. In this case the sum of the individual's values is the area under the person's demand curve. And the consumer surplus, when the price is P_x, is simply the area A in the diagram. Total value is A + B and total revenue is just B. By lowering price it is clear that the area B can either rise or fall (depending on the demand elasticity) but the total value A + B *must rise* and similarly consumer surplus, A, *must rise* as well.



O.The true measure of the total value of a good to a consumer or society as a whole is *not* the amount spent on the good but the amount spent *plus* consumer surplus. Hence while water is a small fraction of our budget it may in fact have an extremely large total value to society. Remember, prices measure marginal value (i.e. the price of the good is equal to the marginal value of the good). Hence, total expenditure is equal to marginal value times the quantity purchased. It is not equal to total value. The difference between total value and expenditures represents the gains from trade.

IX.Complements and Substitutes

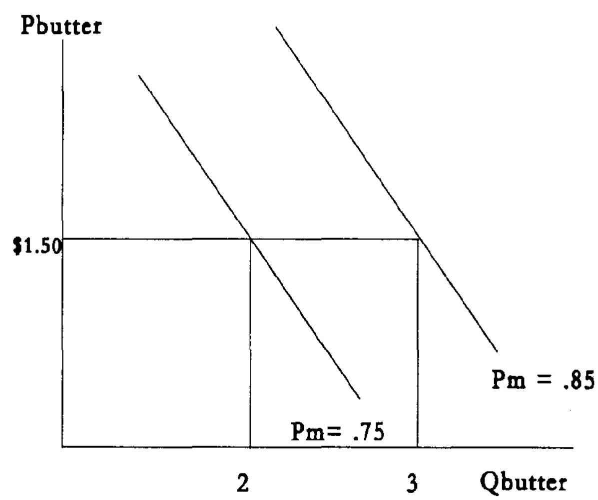
A. As we stated last time, the demand curve for a good is derived under the assumption that income and the prices of other goods are constant. We have already considered the effect of changes in income on the demand curve for a good. In this section we consider how the demand curve for a good can change when the prices of other goods change.

B. If an increase in the price of good *y* increases the demand for good *x* (note that since the price of good *x* is not changing this represents a change in demand), then we refer to *x* and *y* as **substitutes.**

C. Similarly, if an increase in the price of good *y* leads to a reduction in the demand for *x* we say that *x* and *y* are **complements.**

D. These definitions agree with the common usage of these words. For example, if the price of butter rises then we would expect more people to substitute margarine for butter and thus increase the demand for margarine at any given price. Our definition would call butter and margarine substitutes, which would seem reasonable. Similarly, a rise in the price of tennis racquets would decrease the demand for tennis balls, which would by our definition lead these to be called complements. This again agrees with intuition.

E. In the graph below we plot the demand for butter at margarine prices of 75¢ and 85¢. Note that since these are substitutes, the demand curve corresponding to the price of 85¢ lies to the right of the demand for butter when the price of margarine is 75¢. At the price of butter of $1.50, the quantity of butter purchased rises from 2 to 3 units, when the price of margarine rises from 75¢ to 85¢.



F. For complements, the shifts in the demand curve and the quantity purchased at a given price would go in the opposite direction.

X. When To Use Indifference Curves vs. Demand Curves

A. So far we have developed two methods of analysis. The first method was that using the budget set and indifference curves. The second is that using demand curves.

B. Generally demand curves are easier to deal with than the budget set and indifference curve analysis. If it is possible to analyze a problem in terms of demand curves it is generally preferable to do so.

C. Demand curves are easily used to find the effects of changes in a good's price, changes in income, or changes in the prices of complements or substitutes. For example, a rise in income leads to an outward shift in the demand curve for a normal good and hence an increase in the amount purchased at a fixed price. Similarly a rise in the price of a substitute shifts the demand curve out and leads to an increase in consumption at a fixed price for the good.

D. However, to determine the effect of more complicated changes, such as changes in wage rates, income taxes, or fixed costs of work or consumption, one must use indifference curves. Generally look to see if it is simply a price change of a purchased good or a change in income. If it is, then the demand curve analysis should work; if not, then the use of indifference curves is necessary.

XI. The Cost of Living (Price & Quantity Indexes)

A. We frequently hear about price indexes such as the CPI or wholesale price index. In addition, people frequently use these measures to say something about the cost of living.

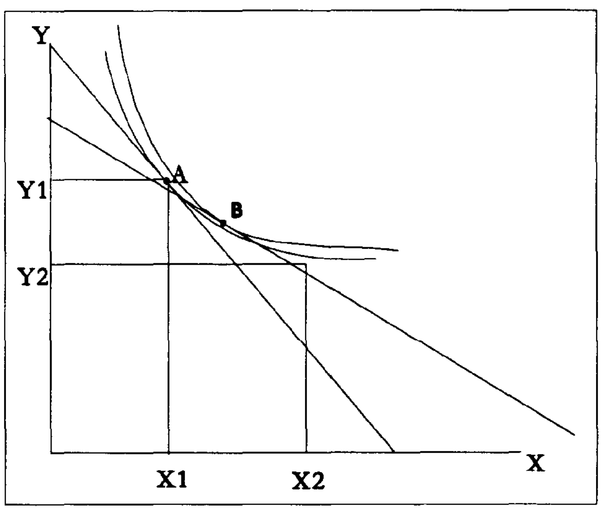
B. We will construct a price index for a single individual similar to CPI and see what we can say about the true cost of living for this individual as prices change.

C. Consider an individual who lives two periods, period 1 and period 2. There are two goods, X and Y, with prices PX_1 and PY_1 in period 1 and prices PX_2 and PY_2 in periods 2. The individual has equal levels of money income M in the two periods. Denote the bundle that he purchased in period 1 as X_1, Y_1 and the amounts of the two goods he purchased in period 2 by X_2, Y_2.

D. Since he spends all his income (which we assume), we must have $X_1 \cdot PX_1 + Y_1 \cdot PY_1 = M$ and $X_2 \cdot PX_2 + Y_2 \cdot PY_2 = M$. The a CPI type index can be defined as

$$\text{CPI} = \frac{X_1\ PX_2 + Y_1\ PY_2}{X_1\ PX_1 + Y_1\ PY_1}$$

E. If the CPI is less than 1 then the cost of his first-period bundle in period 2 is less than M. Hence the individual can still afford to purchase what he purchased in period 1. In this case the individual can be no worse off in the second period (with the same money income) and we can safely say that prices have fallen (for this individual).



F. When the CPI is greater than 1 then he can no longer afford to purchase the same bundle in period 2. In this case people usually say prices have risen. Unfortunately this inference cannot be made. While his old bundle may not be affordable, the change in relative prices may allow him to purchase a different bundle which yields the same or a greater level of satisfaction. Such a case is illustrated in the graph at left. Here, his first-period choice A lies outside of his second-period budget set. This implies $X_1 \cdot PX_2 + Y_1 \cdot PY_2 > M$. Hence the CPI > 1 but the individual is better off in period 2 with the same money income and it would be incorrect to say that the cost of living has increased.

G. Hence, when the CPI < 1 then we know the cost of living fell. But when the CPI > 1 we cannot say for sure if the cost of living rose. Can we ever say that the cost of living rose? The answer is yes. We can look at another index which uses the period 2 bundle to measure the growth in prices. Such a bundle would be like the implicit price deflators used in the GNP and product accounts. Such an index would be defined as:

$$DEF = \frac{P_{y2}Y_2 + P_{x2}X_2}{P_{y1}Y_2 + P_{x1}X_2}$$

H. If DEF>1 then the individual could afford to purchase his second-period choice in period 1 and have some income left over. Hence he must be at least as well off in period 1 and the cost of living must have risen in period 2. However, DEF suffers from the same problem as the CPI. DEF < 1 does not imply that the cost of living has fallen from period 1 to period 2. In the graph above we have CPI > 1 so that we can't say what happened to the cost of living; at the same time DEF < 1 so again we cannot say.

I. In general there are four possibilities:
   1. CPI < 1, DEF < 1. Here the CPI tells us that prices must have fallen and DEF is ambiguous. Hence we know that the cost of living must have declined.
   2. CPI > 1, DEF > 1. Here we can infer the cost of living rose since DEF > 1 and the CPI is ambiguous.
   3. CPI > 1, DEF < 1. Here both indexes are ambiguous and we cannot say what happened to prices.
   4. CPI < 1, DEF > 1. This is crazy. Here the CPI says that prices must have fallen and DEF says prices must have risen. No one set of indifference curves can generate these results and hence the person's tastes must have changed.

J. While the above results are correct for the CPI as defined for an individual as we did, the CPI is defined using the average consumption of all individuals for X_1 and Y_1. In this case the results on the CPI can be even more misleading, to the extent that an individual consumes commodities in very different proportions than used in the CPI weights! The basic result is that the CPI and all price indexes should be interpreted with some caution.

K. The methodology described above can be used to look at both changes in prices and changes in quantities. To see this we can note that changes in nominal expenditures between two periods (period 1 and period 2 in our example above) can be decomposed as follows:

$$(1+g_e)=\frac{X_2\,PX_2+Y_2\,PY_2}{X_1\,PX_1+Y_1\,PY_1}=\frac{X_2\,PX_1+Y_2\,PY_1}{X_1\,PX_1+Y_1\,PY_1}\bullet\frac{X_2\,PX_2+Y_2\,PY_2}{X_2\,PX_1+Y_2\,PY_1}=(1+g_q)\bullet(1+g_p)$$

where g_e is the growth in nominal expenditures, g_q measures the growth in consumption at fixed period 1 prices and g_p measures the growth in prices using the DEF measure introduced above. A symmetric decomposition can be performed using the CPI measure of the growth in prices as in

$$(1+g_e)=\frac{X_2\,PX_2+Y_2\,PY_2}{X_1\,PX_1+Y_1\,PY_1}=\frac{X_2\,PX_2+Y_2\,PY_2}{X_1\,PX_2+Y_1\,PY_2}\bullet\frac{X_1\,PX_2+Y_1\,PY_2}{X_1\,PX_1+Y_1\,PY_1}=(1+g_{q*})\bullet(1+g_{p*})$$

where g_{q*} now measures the growth in consumption at period 2 prices and g_{p*} uses the period 1 consumption bundle to measure the growth in prices (our CPI measure from above).

L. Several things should be noted about these measures. First we always measure the growth in prices by the increase in the cost of a fixed bundle between the two periods. The two measures differ in terms of the bundle used but the basic method remains the same. Second, we measure changes in consumption by calculating the change in the market value at fixed prices. Again the measures differ as to which set of prices are used but not in the basic methods. Finally, when we use period 1 prices to calculate the growth in quantities, consistency requires us to use period 2 quantities when measuring price growth and vise-versa. However, since changes in bundles and prices tend to be relatively small such distinctions are not crucial for short run comparisons.

M. When we wish to measure changes in consumption and prices over longer periods of time choices of consumption bundles and the base periods for prices are much more significant. One common method is to change the bundle used to compute prices (or the prices used to compute changes in consumption) over time. Producing a chained index does exactly this, a chained price index between period 1 and period 4 would be

$$\frac{Price\,Index_4}{Price\,Index_1}=\frac{X_1\,PX_2+Y_1\,PY_2}{X_1\,PX_1+Y_1\,PY_1}\bullet\frac{X_2\,PX_3+Y_2\,PY_3}{X_2\,PX_2+Y_2\,PY_2}\bullet\frac{X_3\,PX_4+Y_3\,PY_4}{X_3\,PX_3+Y_3\,PY_3}.$$

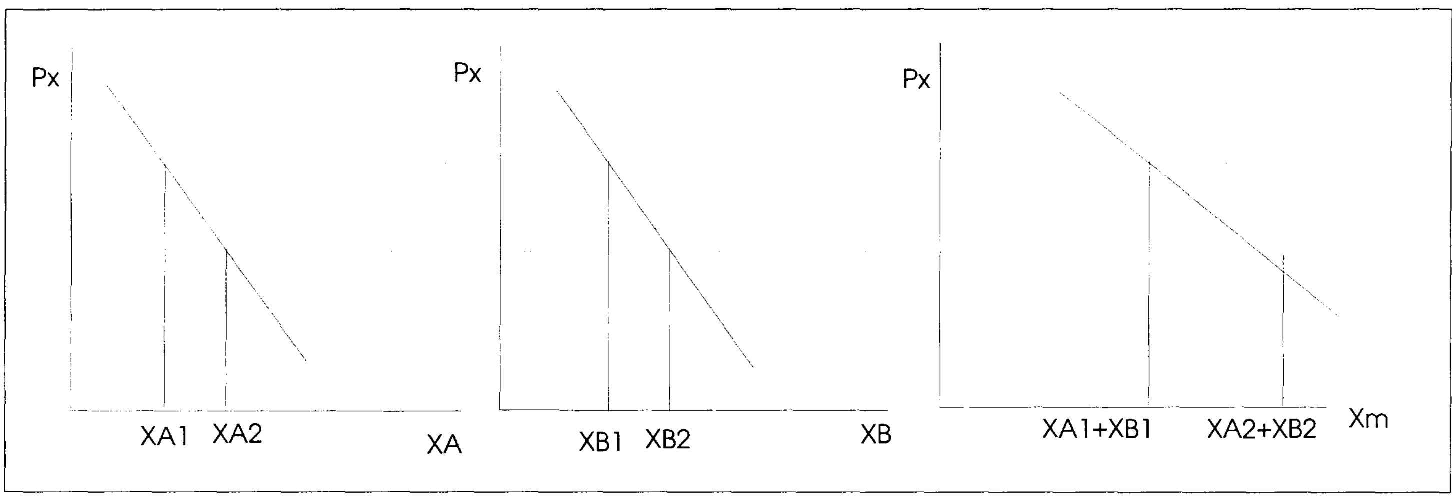
Here we measure the growth in prices from period 1 to period 4 as the accumulation of the change in prices from period 1 to period 2 (measured using the period 1 consumption bundle), the growth in prices from period 2 to period 3 (measured using the period 2 consumption) bundle and the growth in prices from period 3 to period 4 (measured using the period 3 consumption bundle). The corresponding consumption index would then be

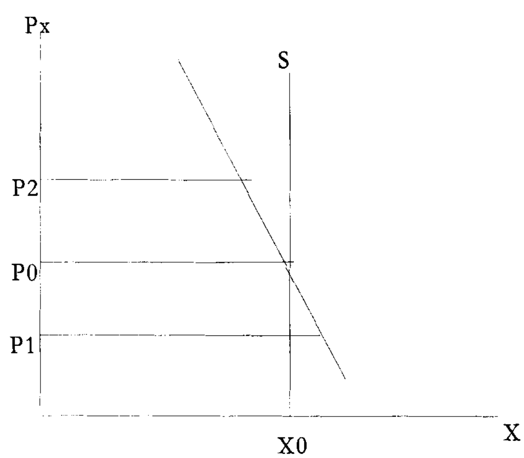
$$\frac{Quantity\,Index_4}{Quantity\,Index_1}=\frac{X_2\,PX_2+Y_2\,PY_2}{X_1\,PX_2+Y_1\,PY_2}\bullet\frac{X_3\,PX_3+Y_3\,PY_3}{X_2\,PX_3+Y_2\,PY_3}\bullet\frac{X_4\,PX_4+Y_4\,PY_4}{X_3\,PX_4+Y_3\,PY_4}.$$

XII. Market Demand

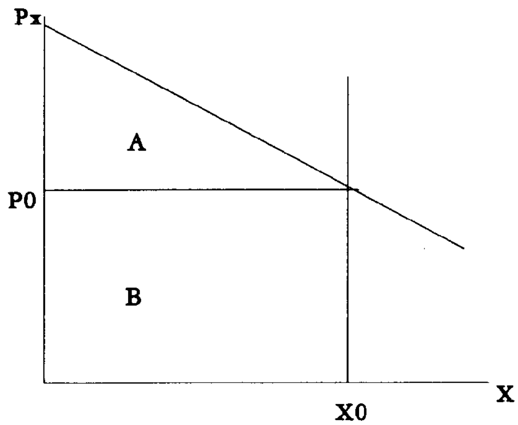
A. Up to now we have derived and employed demand curves for an individual. However, in order to employ these demand curves to determine market prices we must first determine what is known as a **market demand curve**. Just as an individual's demand curve for *x* gives the amount of good *x* he or she chooses to purchase as a function of the price P_x, the market demand curve tells us the total amount of good *x* that *all* consumers in the market wish to purchase at price P_x.

B. The market demand is then the horizontal sum of the individual demand curves. With two individuals, Mr. A and Mr. B, we determine the market demand curve as follows. At price P_1 individual A wishes to purchase QA_1 and individual B wishes to purchase QB_1. The total demand at price P_1 is then given by $AQ_1 + QB_1$. Similarly, at price P_2 the total demand is $QA_2 + QB_2$ as illustrated in the graph below.





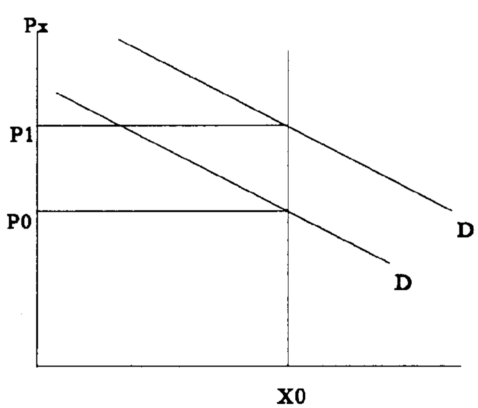
C. If the total quantity of good *x* is available X_0 then we can determine the market price as follows: first, we draw a vertical line at X_0 until it hits the demand curve. The price where this intersection occurs, P_0 in the graph at left, must be the market price. Here is why. At a price below P_0 (say, P_1) the amount people wish to demand is greater than the amount available, X_0. This excess demand will lead to competition among buyers and a rise in the price of the good until the price reaches P_0, where people are willing to purchase exactly the amount available. At prices above P_0 (say,



P_2) the amount people are willing to buy is less than the available supply, which will lead to competition among sellers and a fall in price back to P_0, where supply and demand are equated.

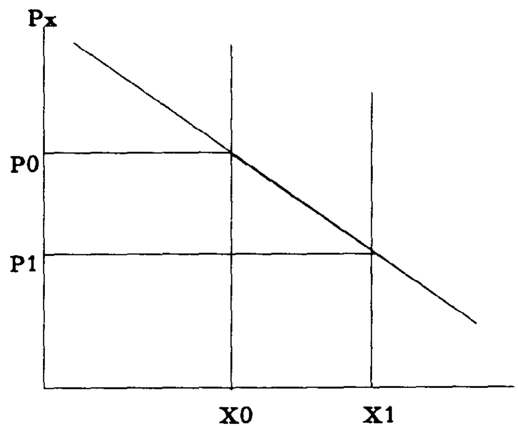
D. Just as in the single consumer case, we can define the total revenue, total value, and consumer surplus. In this case total revenue is simply $X_0 \cdot P_0$, which is the total amount spent on the good, or simply the sum of each individual's total revenue. In terms of the graph above, total revenue is equal to the area B. Just as in the single consumer case, total value is the areas A + B and consumer surplus is equal to the area A. Just as for total revenue, these areas are simply the sum of the corresponding areas for each consumer. Hence, the total value of a good is given by the area under the demand curve and the total amount of consumer surplus is given by the area between the demand curve and the price line.

E. Changes in income or in the prices of complements and substitutes will shift the individual's demand curves and hence shift the market demand curve. These changes will then lead to changes in the equilibrium.



F. For example, if we increase money income to all individuals and the demand for x is a normal good to these consumers, then each individual's demand curve will shift out, causing an outward shift in the market demand curve, as in the graph at left. This leads to a rise in the equilibrium price. Similarly, a rise in the price of a substitute good or a fall in the price of a complementary good will lead to an outward shift in the market demand curve and a rise in price. A fall in income or, equivalently, a fall in the price of a substitute good, will lead to an inward shift of the market demand curve and a fall in price.

G. In this way the market demand curve allows us to determine the effects of changes in demand on the market price.



H. Another change that is important to analyze is the effect of a change in supply on the quantity demanded and the product price. Such a change is illustrated at left. Here the supply of good x (i.e., the amount available) increases from X_0 to X_1. This change results in excess supply at the price P_0. Competition among sellers then drives prices down to P_1 where once again we have supply equal to demand. Here an increase in supply has induced an increase in the *quantity demanded* (*not* an increase in demand) through the effect of the increased supply lowering prices.

XII. Markets and Allocation

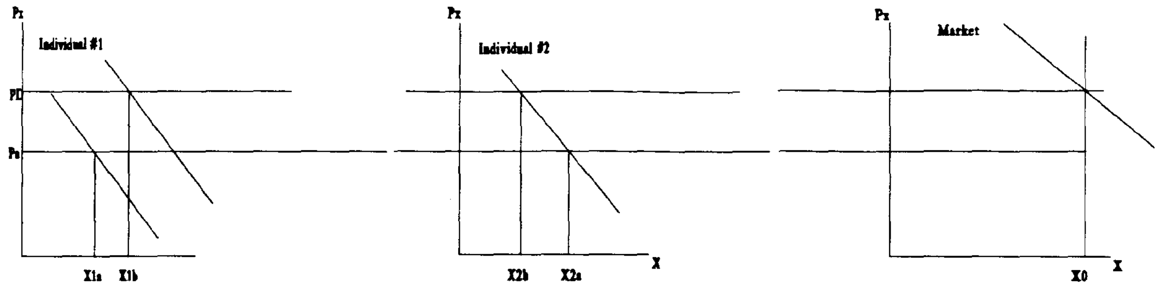
A. In equilibrium each person sets his marginal rate of substitution between two goods equal to the price ratio (as illustrated by the tangency of his indifference curve with his budget set). Since the price ratio is common to all individuals, this implies that the marginal rates of substitution are equal across individuals.

B. Since the marginal rates of substitution are equal across individuals each person's value of good *x* in terms of good *y* is equal to that of all other consumers. (Remember that from lecture 1 differences in MRSs across individuals signal potential gains from trade.) In fact, one can easily show that the results of a competitive equilibrium, such as we have described, where prices are determined by supply and demand, are efficient in the sense that it is impossible to reallocate goods so as to make everyone better off than they currently are in a competitive market!

C. A major role of market prices is that they serve to allocate goods to consumers. That is, goods flow to their highest-valued uses in a competitive market. If one prevents the competitive market from working, then this efficiency of allocation is no longer guaranteed.

D. For example, consider a reduction in supply from X_0 to X_1 in the graph below. The pricing mechanism leads to a rise in price from P_0 to P_1. Since the supply of this good is now X_1 regardless of the price it is often argued (even by some so-called economists) that the rise in price from P_0 to P_1 serves no role; it simply makes some people better off at the expense of others. However, the price rise does serve an important function: it leads to the available supply's being allocated to its highest-value users, represented by the units demanded at the higher price of P_1. If we held the price at P_0 then some people who value the good at only P_0 may still obtain the good (via a rationing scheme, for example); while those with the value of P_0 may gain from this, their gain is more than offset by the loss to the higher-valued user.

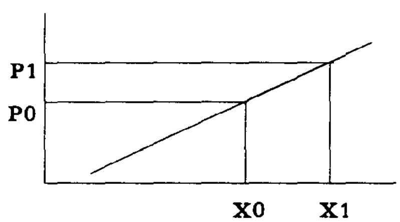
E. A similar situation occurs when some individuals in the economy have an increase in demand (due to a rise in income, for example). This shifts their demand curves out and hence causes an outward shift in the market demand. For example, with two individuals, a rise in demand by individual 1 (illustrated in the leftmost graph below) leads to a rise in price. This rise in price equates the fixed supply with the now higher level of market demand. The higher price causes the consumption by the individual whose demand increased.

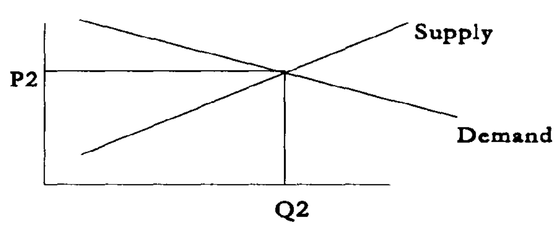


F. In this case the goods flowed to the user whose value increased until the market price (and the value to the other user) rose to equate demand to the available supply.

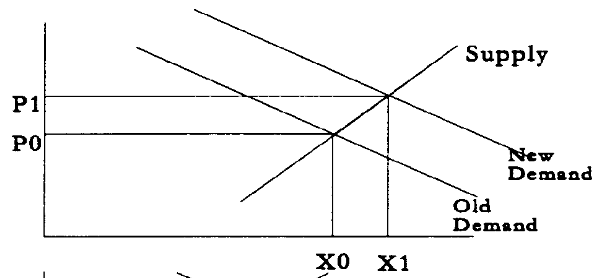
XIII. Price with Upward-Sloping Supply

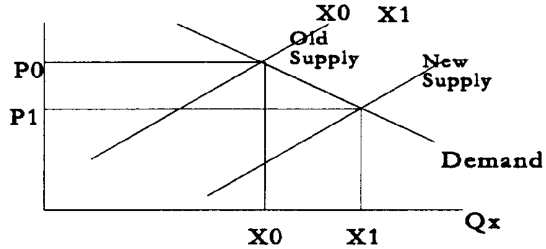
A. As we will see in the second half of the course, as prices rise sellers are usually willing to provide greater amounts of the good. In such cases we cannot treat the supply as fixed as we did in the previous example. Rather we can represent supply via an upward-sloping curve known as a supply curve. The supply curve is used exactly as one would a demand curve. For example, in the graph at left at a price of P_0 sellers are willing to produce X_0 units for sale. At a price of P_1 they are willing to sell X_1 units of output. In this case the equilibrium price is given by the intersection of the demand and supply sources as illustrated in the second graph. Here, the equilibrium price is P_2 and the equilibrium quantity is Q_2. At prices above P_2 sellers are willing to sell more than the consumers wish to purchase. This would lead to excess supply and a fall in prices. Similarly, at a price below P_2 demand exceeds supply, which leads to a rise in price.





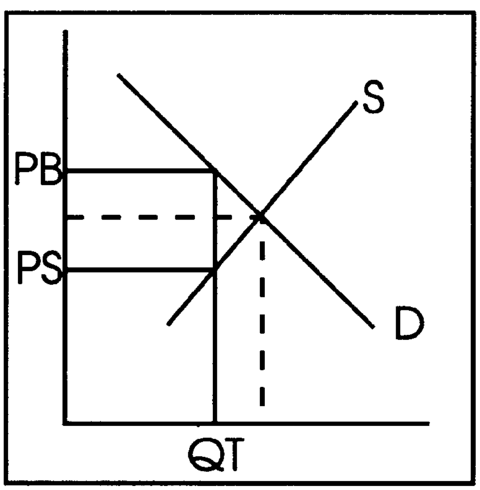
B. With upward-sloping supply, an increase in demand (as illustrated at left) will lead to a new, higher equilibrium price (P_1 vs. P_0) and an increase in the quantity sold from X_0 to X_1. Hence, changes in demand lead to prices and quantities moving in the same direction. Both increase for an increase in demand and decrease with a decline in demand.





C. An increase in supply is when sellers are willing to sell more at a given price. This corresponds to a rightward shift in the supply curve as shown in the diagram. This leads to an increase in the equilibrium price and a rise in the equilibrium quantity transacted. Changes in supply cause price and quantity to move in opposite directions. Prices fall and quantities rise for an increase in supply; prices rise and quantities fall for a decrease in supply.

D. The effects of taxes: consider the effect of a $1 per unit tax on an item. The effect of a tax is to drive a wedge between the price paid by the buyer and the price received by the seller. If we denote the price paid by the buyer by P_B and the price received by the seller by P_S, then we must have $P_S = P_B - 1$. The equilibrium for the market is then found where the quantity supplied by the sellers (which depends on PS) is equal to the quantity demanded by the buyers (which depends on PB).



E. The equilibrium is illustrated in the graph at the left where we see the equilibrium prices paid by the buyer, PB, and received by the seller, PS, and the equilibrium quantity transacted, QT. The dashed lines mark what the competitive price and output would be absent the tax. As can be seen in the figure, the equilibrium price paid by the buyer with the tax exceeds the competitive market price (PC) that would prevail absent the tax while the equilibrium price received by the seller is lower than the competitive price that would prevail absent the tax. The quantity sold on the market is lower with the tax than without it.

F. The results shown above illustrate that the price paid by the buyer will rise when we impose a tax while the price received by the seller will fall. Thus the burden of the tax is shared between the two parties. In fact how this burden is shared does not depend at all on which party pays the tax to the government. How much is paid by each party depends on the slope of the supply and demand curves. If the supply curve is vertical then the suppliers pay the entire tax (i.e., the net price to consumers must stay the same to make them willing to purchase the same fixed quantity available). If the supply curve is perfectly flat then the consumers pay the whole tax. In this case the before-tax price cannot fall (or all suppliers drop out of the market) and consumers must pay the entire amount. Changes in the slope of the demand curve have the opposite effect.

XIV. Example: Labor Supply

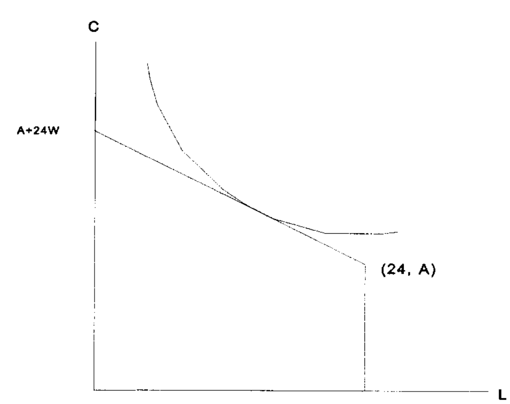
A. Here we consider a simple model of how much an individual should work at a fixed wage of *W* per hour.

B. The basic model considers a single period (for example, one day). The individual has some nonwage income of \$A and can work as many hours as he wishes at a fixed wage of *W*.

C. Most jobs specify a fixed number of hours rather than offer the individual the ability to choose his hours freely. However, hours of work do differ across jobs. One can think of this model as explaining the individual's choice of which job to work at (e.g., a job with 5 hours per day or one with 8 hours per day).

D. We think of this individual's preferences as being defined over two goods: consumption, *C*, and hours of leisure or time spent not working, *L*. We also assume that both of these goods are normal goods and the indifference curves have the usual convex shape.

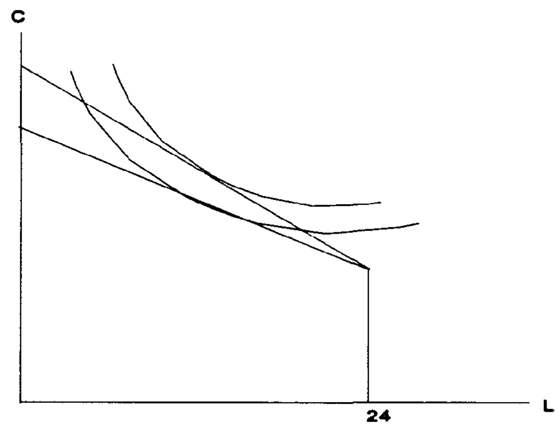
E. The individual's budget set is that given in the graph at the right. Here leisure time is given on the horizontal axis and the consumption is on the vertical axis. If he takes 24 hours per day of leisure then his consumption can be at most equal to his non-wage income, A. This gives the vertical line at L = 24. After this, in order to increase his consumption the individual must reduce his leisure. For each hour of increase leisure (i.e., an increase of one hour of work) the individual's income and hence consumption can increase by *W*.



Hence, the budget line is simply a line with slope -*W* from the point (24, A) to where it hits the vertical axis at (0, A+24*W*).

This gives the vertical line at L = 24. After this, in order to increase his consumption the individual must reduce his leisure. For each hour of increased leisure (i.e., an increase of one hour of work) the individual's income and hence consumption can increase by W. Hence, the budget line is simply a line with slope $-W$ from the point (24, A) to where it hits the vertical axis at (0, A+24W).

F. An increase in the wage leaves the point (24, A) unchanged and shifts the budget set out as shown by the dotted line in the graph at right. We can analyze this effect of a wage change in terms of income and substitution effects. The substitution effect leads the individual to work more since he can now obtain more consumption for a given decrease in leisure. Hence, an increase in the wage represents a decrease in the relative price of consumption which leads toward a substitution toward consumption away from leisure. Equivalently, a rise in the wage represents an increase in the relative price of leisure and therefore a substitution effect away from leisure into consumption.



G. The income effect of a rise in the wage is positive. For a person working h hours per day a rise in the wage of \$1 is roughly an increase of \$$h$ per day in income. The individual now earn \$$h$ more per day even without a change in hours of work.

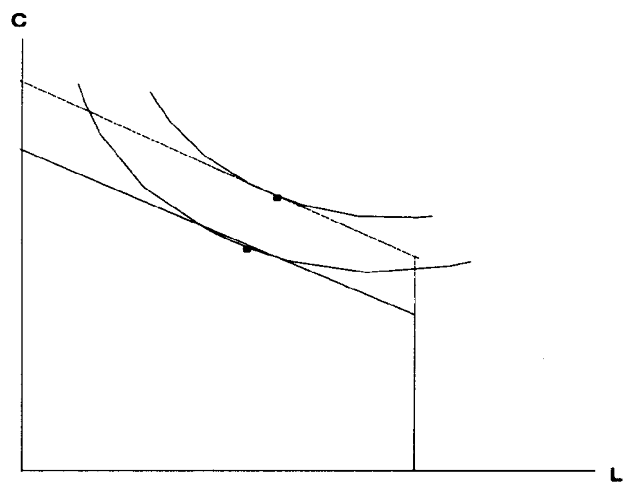
H. Since a rise in the wage has a positive income effect, the income effect will be to increase both consumption and leisure if both are normal goods. Further, this income effect is likely to be quite large for individuals who derive a large fraction of their income from hourly work. For example, if A = 0 then a 10% rise in the wage will raise income by 10% if there is no change in hours of work.

I. We can then summarize the total effect as follows:

ITEM	Substitution Effect	Income Effect	Total Effect
Leisure	-	+	?
Hours worked	+	-	?
Consumption	+	+	+
Wage income	+	+	+

J. The effect on hours or work is uncertain because the higher wage makes working more profitable (the substitution effect) but at the same time makes the individual "wealthier" which will lead him to take increased leisure time to consume his additional income (the income effect). This causes the change in hours when the wage changes to be unknown. By observing how people actually make these decisions it has been found that the income effect dominates the substitution effect and hours of work fall with an increase in the wage.

K. We now consider the effect of an increase in A. In this case the shift in the budget set is as shown at the right. The new budget line is parallel to the old one but farther to the northeast. Since the wage hasn't changed, the relative prices of leisure and consumption are unchanged. This is a pure income effect. Since an increase in A represents an increase in income, the income effect will be to increase both consumption and leisure (hence decrease hours worked).



L. Therefore, the total effect of a rise in non-wage income is the same as the income effect and we have the following results:

ITEM	Substitution Effect	Income Effect	Total Effect
Leisure		+	+
Hours worked		-	-
Consumption		+	+
Wage income		-	-

Note that while an increase in A leads to a rise in total income (wage-income plus non-wage-income), wage income falls. This follows directly from the fact that the wage doesn't change and the income effect is to reduce hours of work and increase leisure.

XV. Allocation Over Time

A. Just as we can think of individuals choosing the amounts of each good to consume at a point in time, we can consider how individuals and the economy as a whole allocate goods over time. This is known as capital theory. Here, rather than looking at different goods at the same point in time, we consider the same good (usually just called consumption) over time.

B. The first key point is that a dollar today is not the same as a dollar tomorrow. Since, even in the absence of inflation, the interest rate is positive, one can turn a dollar today into more than $1 tomorrow. By putting that dollar in the bank at an interest rate of 10% one can have $1.10 tomorrow. The price of income tomorrow in terms of income today is $1/(1+r)$ where r is the interest rate. For a 10% interest rate this would be $1/(1 + .10)$.

C. For example, if the interest rate is 10%, a tree that produces a net profit of $10 today and in each of the next two years would be worth

$$V = \$10 + \$10/(1 + .10) + \$10/(1 + .10)^2$$

in terms of present dollars. The second year is discounted by $1/(1+r)$, the relative price of second-period income, which is the amount of period-1 income needed to get $1 in period 2. Similarly, the third period's income is discounted by $1/(1 + .10)(1 + .10)$, which is the value of third-period income in terms of period 1 (i.e., the amount of period-1 income needed to get $1 in period 3).

D. The calculation of the present value of the tree, V, is known as computing the *capital value* or *present value* of the tree. V simply gives the value of the tree's output in terms of present dollars.

XVI. Consumption Over Time

A. A simple mechanism to analyze the allocation of consumption over time uses the indifference curve and budget set analysis that we have been using so far.

B. For simplicity we consider a world with two periods and one aggregate good called consumption valued in dollars. Here there are two goods: C_1 (consumption in period 1) and C_2 (consumption in period 2). We assume that the individual has the usual sort of preferences over these goods with convex indifference curves.

C. The individual has income I_1 in period 1 and income I_2 in period 2. In addition, the individual can freely borrow or lend at a constant interest rate of r. Since the individual can always not save or borrow, his budget set must contain the point with $C_1 = I_1$, $C_2 = I_2$. In addition, he can increase consumption in period 2 by $(1+r)$ for each dollar he saves from period 1. Hence, if the individual saves S dollars (here borrowing is simply negative savings) we have:

1. $C_1 = I_1 - S$

2. $C_2 = I_2 + (1+r)S$

   Equation (1) implies that $S = I_1 - C_1$ (i.e., the difference between period 1 income and consumption). Substituting this for S in equation (2) yields

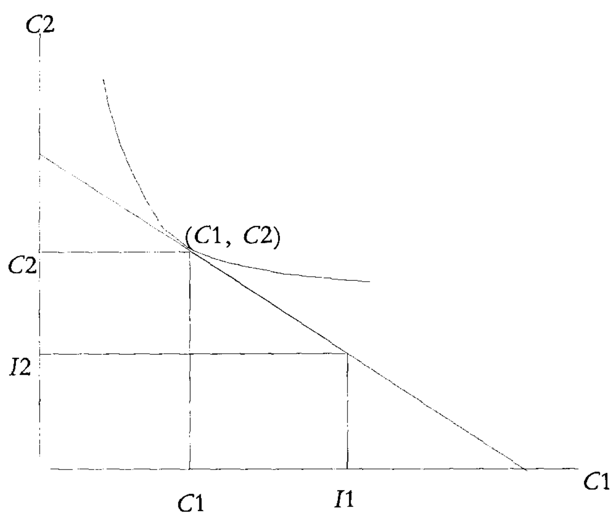
3. $C_2 = I_2 + (1+r)(I_1 - C_1)$

   Solving out for C_2 and C_2 on the same side yields:

4. $C_2 + (1+r)C_1 = I_1 + (1+r)I_2$.

   Equation (4) simply says that the value of his consumption (in period-2 dollars) must equal the value of his income (again in period-2 dollars). Dividing through by $(1+r)$ yields equation (5), which says that the present value of income must equal the present value of consumption:

5. PV(income) $= I_1 + 1/(1+r)I_2 = C_1 + 1/(1+r)C_2 =$ PV(Cons)

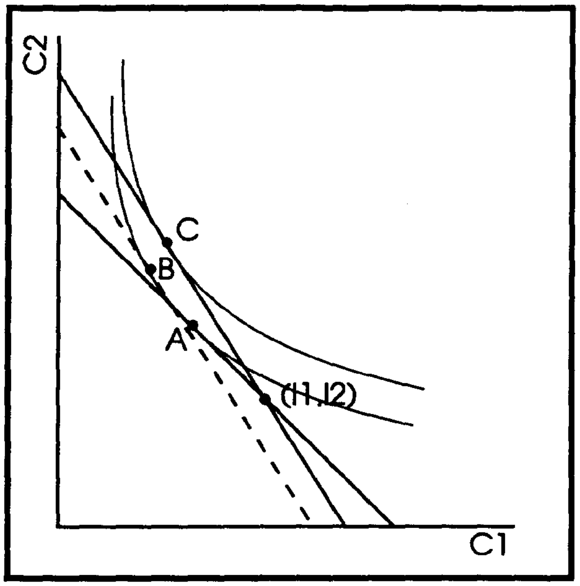


D. In terms of the budget set, we have the graph at left. The line passes through the endowment point I_1, I_2 and has slope equal to $-1*(1+r)$. The maximum amount of period-1 consumption possible is $I_1 + I_2/(1+r)$ and the maximum available for consumption in period 2 (i.e., with no consumption in period 1) is $I_2 + I_1(1+r)$. The maximum he can consume in period 1 is simply the present value of his income, which would imply no consumption in period 2. Similarly, by consuming 0 in period 1 and saving his total income, I_1, he can consume I_2, his income in period 2, plus $I_1(1+r)$, his income in period 1 with accumulated interest.

E. His equilibrium will be like that shown in the graph above. In this case he saves $(I_1 - C_1)$, which he consumes along with the interest and his period-2 income in the second period.

F. The first effect we wish to study is that of a rise in interest rates. Here this corresponds to a pivoting of the individual's budget line through the point (I_1, I_2). As interest rates rise, the line becomes steeper. As usual we look for the income and substitution effects of the price change.

G. The substitution effect is that one can now obtain more period-2 consumption for each dollar of savings (i.e., for each dollar of period-1 consumption given up). This effect will decrease period-1 consumption and make period-2 consumption rise, holding utility constant. This is illustrated in the graph at left. The rise in interest rates shifts the budget line from the flatter solid line to the steeper solid line in the figure. The substitution effect moves the individual from point A to point B, while the income effect moves the individual from point B to point C. The total or observed effect is to move the individual from point A to point C in the figure.



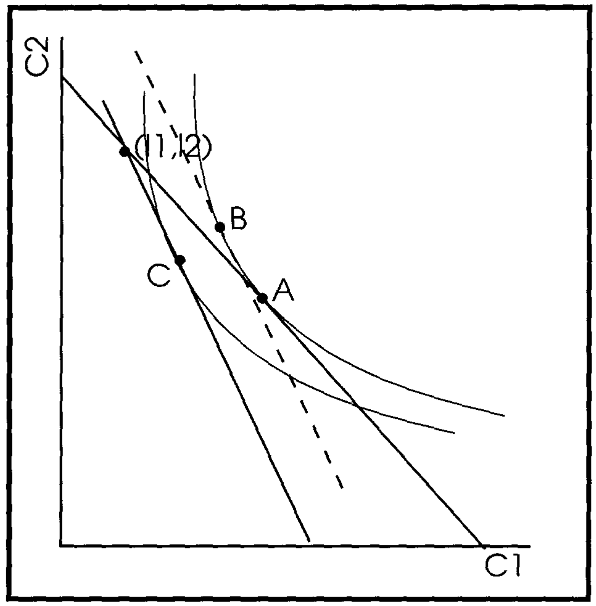
H. Since this person is a saver, the rise in interest rates represents a positive income effect (his old consumption point is now inside his budget set). With the rise in interest rates he can now have more period-2 consumption with the same level of consumption in period 1. This gives us a positive income effect. This effect will tend to raise C_1 and C_2, as seen in the graph.

I. To summarize the effects of a rise in the interest rate on a saver, we have:

	C_1	C_2
Income Effect	+	+
Substitution Effect	-	+
Total	?	+

Period-2 consumption must rise, since both effects go in the same direction. Provided that savings are a small fraction of the individual's wealth, the substitution effect should dominate the income effect for C_1 and first-period consumption should fall, though in principle the rise in interest rates could now allow an individual with a large amount of savings to consume more today, since he can get the same consumption tomorrow with less savings. However, with the usual amount of savings (as a fraction of income) a rise in the interest rate should (and does) lead to increased savings.

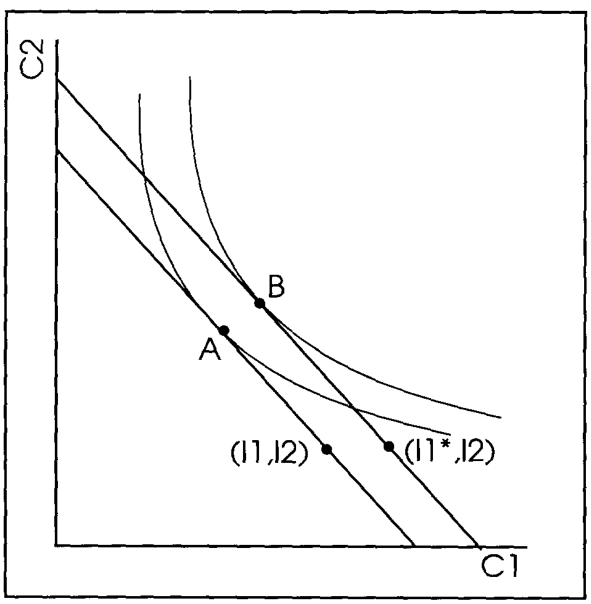
J. For a borrower the income effect goes in the opposite direction. As shown in the graph at the right, the same rise in interest rates shifts the budget line through this individual's endowment point as well. However, the borrower can no longer afford his old consumption points (C_{1a}, C_{2a}). Here the substitution effect is the same as for a saver (i.e., C_2 should rise and C_1 should fall), but the rise in interest rates represents a fall in real income.



K. For the borrower we have:

	C_1	C_2
Income Effect	-	-
Substitution Effect	-	+
Total	-	?

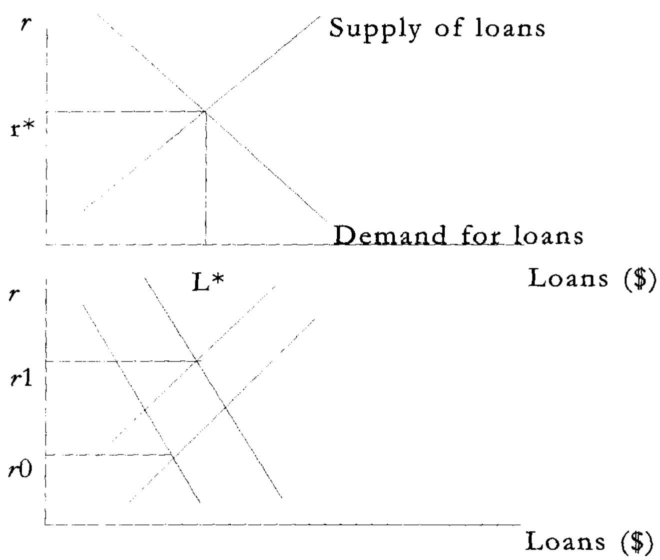
L. Since I_1 is unchanged and borrowings are equal to C_1 - I_1, the rise in interest rates must reduce borrowings and borrowers will borrow less.



M. Similarly, we can study the effects of a rise in I_1 or I_2, i.e., a rise in the person's income in one of the two periods. This change is illustrated for an increase in I_1 in the graph at left. Since the interest rate has not changed, the budget line shifts out parallel, indicating a pure income effect. This implies that both C_1 and C_2 will increase. The interest rate is unchanged, therefore a rise in C_2 necessitates a rise in savings (since I_2 is also fixed). A rise in savings for savers and a decline in borrowings for borrowers.

N. The effect of an increase in I_2 is also a parallel shift (the graph is left).

Again the income effect says C_1 and C_2 both rise. Since I_1 is unchanged, the individual will save less (so C_1 rises) or borrow more. Hence a rise in future income will lead to less saving today and more borrowing by an individual.



XVII. Interest Rate Determination

A. Deriving the equilibrium market interest rate, and the associated level of savings involves setting the supply of loans (savings) equal to the demand for loans. This is illustrated by the intersection of the supply curve and demand curve in the graph.

B. If we increase I_2, raise second-period income, then the supply of loans decreases (remember a rise in I_2 leads to a decline in savings) and an outward shift in the demand for loans (since an increase in I_2 increases borrowings at a given interest rate). The result of these changes is shown in the second graph. Clearly an increase in I_2 for everyone must lead to a rise in interest rates.

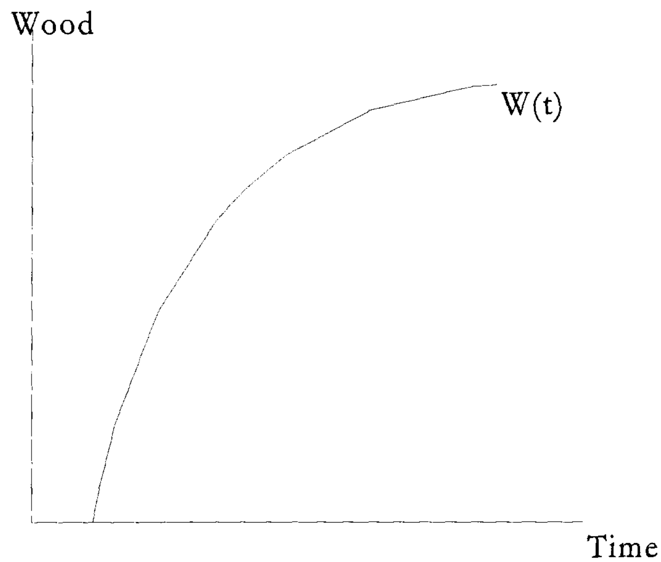
C. Similarly, a rise in I_1 will lead to a decrease in the demand for loans and a rise in savings, i.e., an increase in the supply of loans. These together will lead to a fall in interest rates. Hence, interest rates are higher; the higher future income is relative to present income.

XVIII. Applications of Capital Values

A. As an example of how to use capital values, consider the problem of when to cut a tree (or when to sell a bottle of wine).

B. In the tree-cutting problem we will assume for simplicity that the interest rate is constant and that the price of wood is also constant. The amount of usable wood in the tree is assumed to grow over time and is given by W(t).

C. The graph below plots the amount of wood in the tree as a function of time. There is no cost to growing or to cutting it down. The problem is when to cut the tree. The answer is to let the tree grow until the amount of usable wood is increasing at a rate equal to the interest rate. This can be seen as follows. We wish to maximize the present value of our



income from cutting the tree. This is simply W(t) * Pw/(1+r). Here Pw is the price of wood. We will cut the tree at t if the value of cutting at t exceeds the value from cutting at t+1. We can state this as

$$\frac{W(t)\ Pw}{(1+r)^t} > \frac{W(t+1)\,Pw}{(1+r)^{t+1}}.$$

Dividing both sides by the left-hand side and multiplying by 1+r yields

$$1+r > W(t+1)/W(t).$$

This is equivalent to the condition

$$\frac{W(t+1)-W(t)}{W(t)} < r,$$

or that the tree will grow at a rate less than the interest rate between t and t +1. Also the value of cutting at t must exceed that from the cutting at t-1 or after doing the same as above.

$$\frac{W(t)-W(t+1)}{W(t-1)} > r.$$

Hence, if it grows slower than the interest rate from t-1 to t it will pay to cut sooner.

D. This problem is easily solved in continuous time using calculus. In continuous time the discount factor to future values in terms of present values is $\exp(-r^*t)$. In this case the maximization problem is

$$\max_{t} \exp(-r^*t)W(t)Pw.$$

Differentiating with respect to t yields

$$-r \exp(-r^*t)W(t)Pw + \exp(-r^*t)W'(t)Pw = 0.$$

Simplication then yields:

$$W'(t)/W(t) = r.$$

Hence, the optimal cutting time involves setting the rate of growth of the tree equal to the interest rate.

E. If the tree is growing faster than the interest rate, then it pays to keep the tree, since its return exceeds the return in the market (the opportunity cost of your investment). If the tree is growing slower then it pays to cut the tree and invest the money elsewhere where you get a higher rate of return.

F. A similar problem is when to sell an asset that is increasing in price over time (e.g., a bottle of wine). Let $P(t)$ be the price of the good as a function of time. In this case the maximization problem is

$$\max_{t} \exp(-r^*t)\, P(t).$$

Differentiating with respect to t yields

$$-r \exp(-r^*t)P(t) + \exp(-r^*t)P'(t) = 0.$$

Solving this equation implies

$$P'(t)/P(t) = r.$$

This says one should sell the asset when its value rises slower than the interest rate.

A. In the previous examples we considered a case of allocation over time where the decision was whether to consume in one period ***or*** another and whether to cut down a tree this year ***or*** next year. In both of these examples, the trade-off was the choice of when to take an economic action. The case of a durable good represents the opposite case where the choice involves an asset that provides services in more than one period (a sort of joint production problem).

B. Durable goods provide a stream of services over time. How long such goods last can be measured by their depreciation rate (i.e. what fraction of the good is lost after one period of use). The simplest model is for a good that depreciates at a constant rate so that a fixed fraction of what remains, δ, is lost each period. In this case if we have K units of the good today we will have K $(1-\delta)$ units left next period.

C. For durable goods we must make distinctions on both the quantity and price sides. First in terms of the quantities we must distinguish between the stock of the durable good (i.e. how much is available at a given point in time) and new production of the good (which is commonly called gross investment). If we denote the stock of the good in year t by S(t) and new investment by I(t) then we have the following stock-flow equation:

(1) $S(t) = (1-\delta)\, S(t-1) + I(t).$

In addition, we must distinguish between two prices for the good, the capital price, P(t), equal to what it costs to purchase the good and the rental price, R(t), equal to what it costs to use the good for one period. It is crucial to remember that both such prices are defined and relevant whether there exists a purchase market only, a rental market only, or both a purchase and rental market for the good in question. These prices are economic concepts which can be defined independent of the market structure and independent of the particular contractual arrangement used for a transaction (i.e. whether one buys or rents).
Just as stocks and investments are related via the equation above so too are rental prices and capital prices. In particular, the rental cost of using an asset is simply the cost of buying the good and re-selling it after one period. Thus we must have

(2) $R(t) = P(t) - P(t+1)\,(1-\delta)/(1+r),$

Where r is the nominal rate of interest and P(t+1) is next year's price for the good in question.

D. Equation 3 can be rearranged to yield

(4) $R(t) = [\, rP(t) + \delta\, P(t+1) + (P(t) - P(t+1))\,] \,/\, (1+r).$

Which implies that the rental cost of an asset consists of three components, the interest cost, rP(t), depreciation, δ P(t+1), and market re-evaluation, P(t) – P(t+1). All are discounted by (1+r) since all are measured in period t+1 dollars. The rental cost will be higher, the higher is r, the greater is the rate of physical depreciation and the faster the price of the asset is declining. The formula in equation (4) is often used to calculate the implied rental rate for assets where no formal rental market exists.

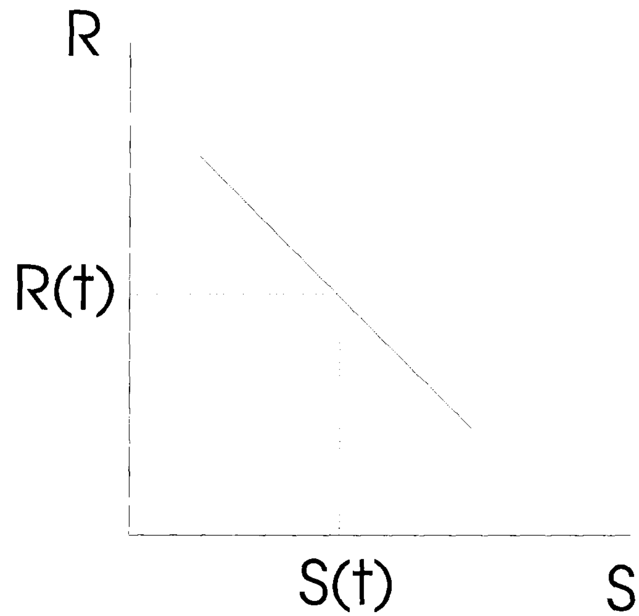
E. When we have rental rates and desire to go in the reverse direction to calculate the implied capital prices we must use the present value equation

(5) $P(t) = R(t) + R(t+1)\,(1-\delta)/(1+r) + R(t+2)\,(1-\delta)^2 / (1+r)^2 \ldots\ldots$

This equation can be obtained by recursively substituting for future prices in equation (3) above.

F. When we look at a market equilibrium for the housing market at any one point in time we must realize that today's market is influenced by both the past and the future. The effect of the past comes through the effect of past production decisions on the current stock of housing. The effect of the future comes from the effect of future expected rental rates on the current price (i.e. via equation 5). At each point in time there are four equilibrium conditions:

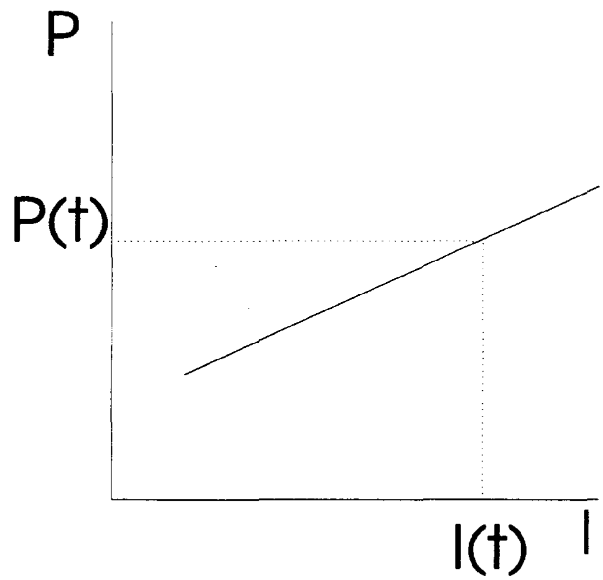
First the rental market must be in equilibrium. The demand for housing services today depends on the flow cost of housing services (i.e. the rental rate, R(t)). Housing services are provided by the stock of housing S(t). As a result, the demand side of the market links the current rental price, R(t), and the current stock as in



Or in mathematical terms S(t) = D(R(t)).

Second, prices must be equal to discounted rental rates so that
(6) $P(t) = R(t) + R(t+1)\,(1-\delta)/(1+r) + R(t+2)\,(1-\delta)^2 / (1+r)^2 \ldots\ldots$

The supply of new construction, investment depends on the current price. This can be represented by a supply relationship of the form



Or in mathematical terms, I(t) = I(P(t)).

Finally, we have stock adjustment equation above which implies that

(7) $S(t) = (1- \delta)\ S(t\text{-}1) + I(t)$.

We can then put all of these equations together as

(8.1)	$S(t) = D(R(t))$	Rental Market Equilibrium
(8.2)	$P(t) = R(t) + R(t+1)\ (1\text{-}\delta)/(1+r) \ldots\ldots$	Asset pricing equilibrium
(8.3)	$I(t) = I(P(t))$	Investment market equilibrium
(8.4)	$S(t) = (1\text{-}\delta)\ S(t\text{-}1) + I(t)$	Stock adjustment

G. Since the current equilibrium is linked to the past through the lagged stock of housing, S(t-1) and is linked to the future through the expectations of future rental rates (working through the current price) today's equilibrium will in general depend on supply and demand conditions in both the past and the future. The easiest way to analyze a model such as this is to consider the case where demand and supply conditions are stable (i.e. constant) over time so that rental rates, prices, the stock and investment are all constant. In this case R(t), P(t), S(t) and I(t) are all constants so that the four equations above simplify to

(9.1) $S = D(R)$

(9.2) $P = R/[1-(1-\delta)/(1+r)]$

(9.3) $I = I(P)$

(9.4) $I = \delta S$.

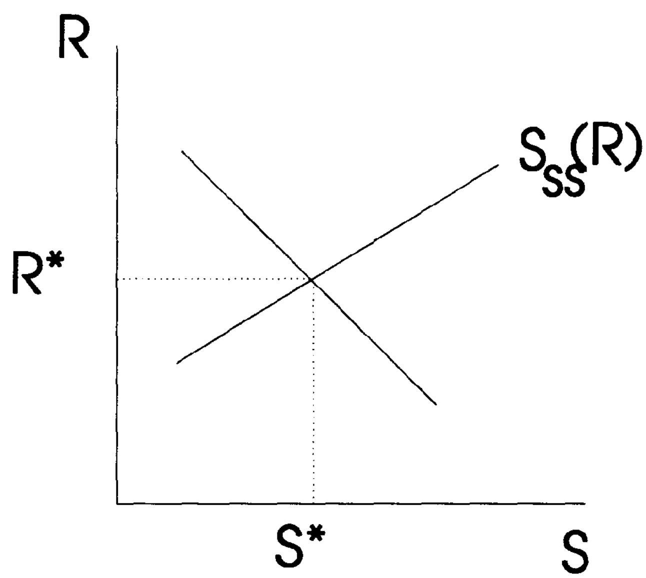
The first and third equations are just equilibrium in the rental and investment markets respectively. Equation (9.2) recognizes that in the steady state the rental price must be interest and depreciation on the capital prices so that the capital price is proportional to the rental price with a factor of proportionality inversely related to the interest and depreciation rates. Finally, equation 9.4 recognizes that in a steady state new investment must exactly offset depreciation on the steady state stock.

H. To solve this system we simply substitute 9.2 into 9.3 and the result into 9.4 to yield a steady state supply equation which gives the steady state stock as a function of the steady state rental price. The basic idea is that we translate any given rental price to a capital value via the present value equation (9.2). Once we know the price we can then determine steady state investment via 9.3 and then translate this into an equivalent stock using 9.4. Mathematically this comes down to

(10) $S_{ss}(R) = (1/\delta)\ I(R/[1-(1-\delta)/(1+r)])$

so that the steady state equilibrium is given by

STEADY STATE EQUILIBRIUM



I. The steady state behavior of the equilibrium for a durable good is the same as for any other good. For example, the price and quantity responses to a tax depend on the elasticities of supply and demand just as in the regular case. The difference with durable goods comes in the ***transition phase, i.e. the process by which we get to the steady state equilibrium.***

J. Analysis. When analyzing any problem it is best to think about the effect of any change (like a tax, change in demand, change in production costs, etc.) on the steady-state graph and hence on the steady-state stock, rents, price and construction. To analyze the dynamics it is best to stick with the four equations we had above and use them in that order as

(1) Demand and stock S(t) => R(t) (figure 1)

(2) Rents R(t) discounted back to today => P(t)

(3) P(t) and the costs of new construction => I(t) (figure 2)

(4) I(t) and depreciation determine the change in the stock

The stock then feeds back into equation (1) to determine future rents and back through the chain. If you keep things in this order and start at the point of the change you should be able to follow the dynamics to get to the steady state.

K. Example. What if we increase the cost of new construction? The effect of this on the steady state is shown in Figure 4, where we have shifted the steady-state supply curve to reflect the higher costs of construction. It leads to higher rents and a lower stock, which implies a higher steady-state price and a lower steady-state rate of construction. This tells us where we are going relative to where we start. The dynamics go as follows:

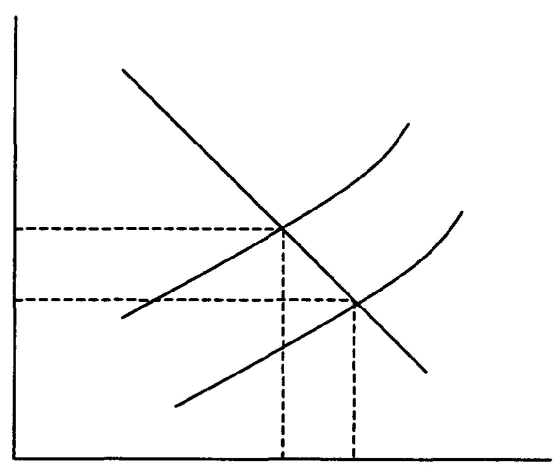


Figure 4

We start with equation (3) since we are shifting supply and that is where it enters the model. The shift in supply lowers new construction today.

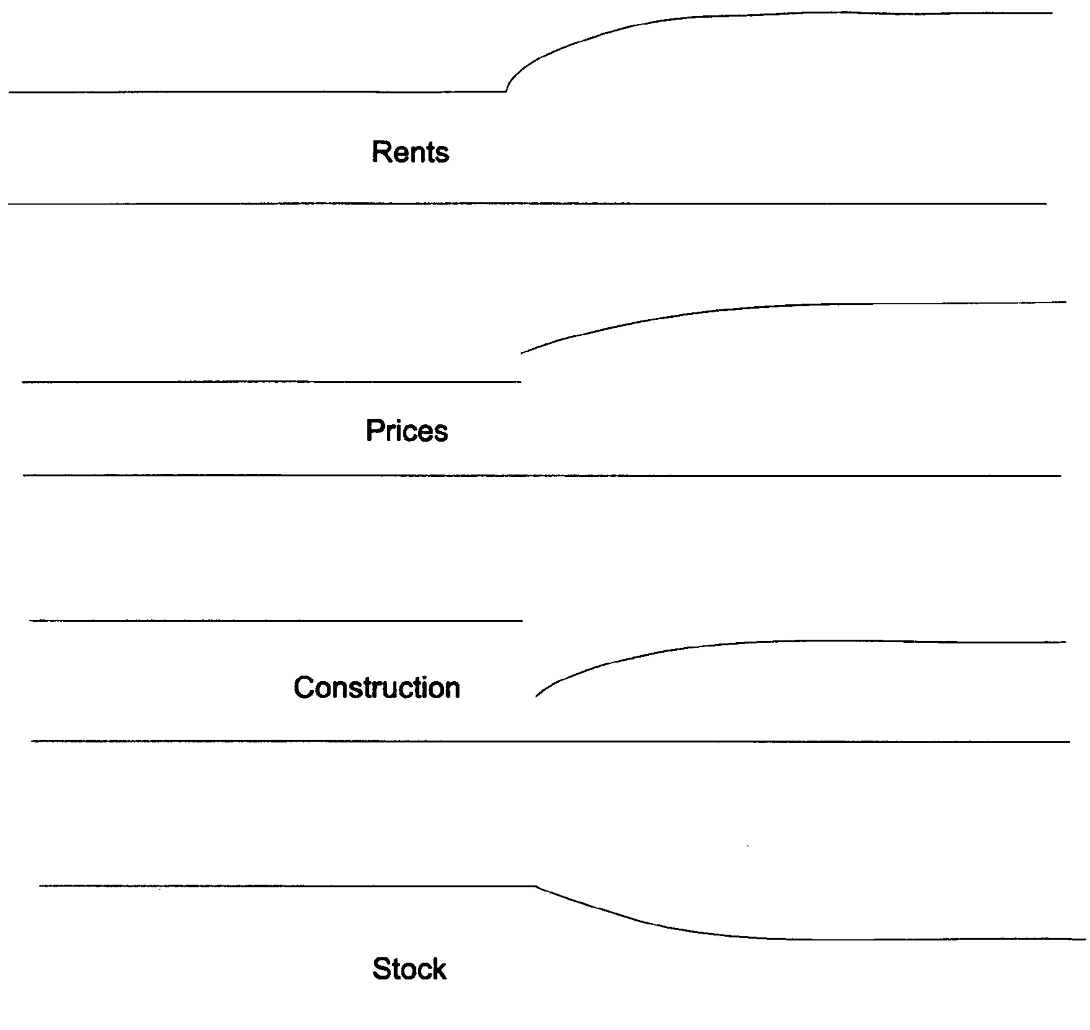
Less construction causes the stock to begin to shift.

The declining stock causes rents to begin to rise.

Rising rents and the prospects of higher future rents (remember steady-state rents are higher) cause the current price to rise.

As prices rise over time, construction rebounds.

This analysis tells us that rents, prices, and construction will be rising as we approach the steady state and the stock will be falling. In terms of time lines this looks like


Rents
Prices
Construction
Stock

XIX. What Is Cost?

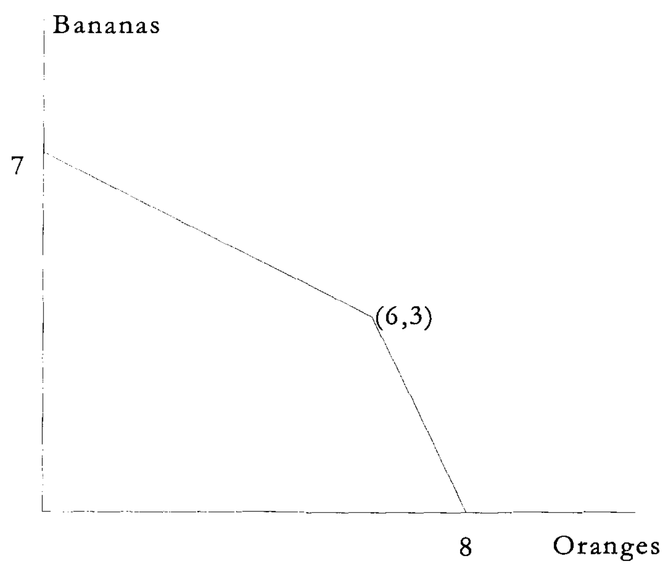
A. Example: Comparative Advantage and Cost

1. We begin our discussion of cost and production with a simple example that will illustrate the true nature of cost and the problem society faces when deciding how much of each good to produce.

2. In order to keep things simple, we begin with two goods, bananas and oranges, and a single individual by himself. He has two plots of land of equal size (say, 1 acre each). Plot 1 can yield 6 oranges per acre or 4 bananas per acre. Similarly, plot 2 can yield 2 oranges per acre or 3 bananas per acre. The problem this individual faces is how much of each crop to produce and how to divide this production between the two plots.

3. In order to see how this individual will make this choice, we must determine his opportunity set from which we can determine the *costs* of production and in particular we will be able to determine *marginal cost* which will be relevant for his decision.

4. If we plot oranges on the horizontal axis and bananas on the vertical axis as in the graph at left, we can plot his opportunity set as follows. If he produces nothing but oranges he will get 6 from plot 1 and 2 additional from plot 2 for a total of 8.



5. The next question is where he should produce his first banana. Since plot 1 can produce 4 bananas per acre, whereas plot 2 can produce only 3, it might seem that plot 1 is the low-cost producer of bananas. However, this is *not* the case. On plot 1 an increase of 1 banana costs $6/4 = 1.5$ oranges (remember costs are always defined as opportunity costs) whereas on plot 2 the cost of a banana is simply $2/3 = 6.7$ oranges. These costs follow from the fact that on plot 1 we have a choice of 6 oranges or 4 bananas for a tradeoff of 6/4 oranges per banana. For plot 2 this same analysis yields the tradeoff as $2/3 = .67$ oranges per banana.

6. From this it is clear that even though plot 2 is less productive in absolute terms than plot 1 (i.e., plot 1 can produce more of either fruit), plot 2 is still the *low-cost producer* of bananas. Similarly, the cost of oranges from plot 1 (in terms of

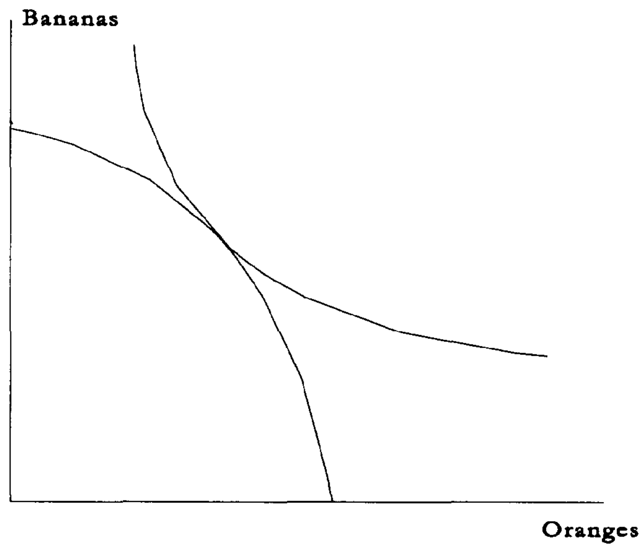
bananas) is simply 4/6 = .67 and the cost of oranges from plot 2 is the 3/2 bananas that must be given up. In this case plot 1 is the low-cost producer of oranges.

7. Hence it is clear that bananas should be produced first on plot 2 until that plot is entirely used for banana production. At that point additional bananas should be produced on plot 1 if more are desired. Using this, we obtain the opportunity set shown in the figure above.

8. If additional plots were available, we could simply determine the cost of bananas (in terms of oranges) for each plot. Production of bananas would then use the lowest-cost plots first, followed by the next-lowest and so on until the highest-cost plot would be used last. With three plots, a figure such as the first one below would result from this optimal assignment of production. In the limit with a large number of different plots the graph would approach the smooth, outward-bending curve shown in the second graph below.

9. The individual's choice can then be made by finding a tangency between his indifference curve and his opportunity set as in the graph.

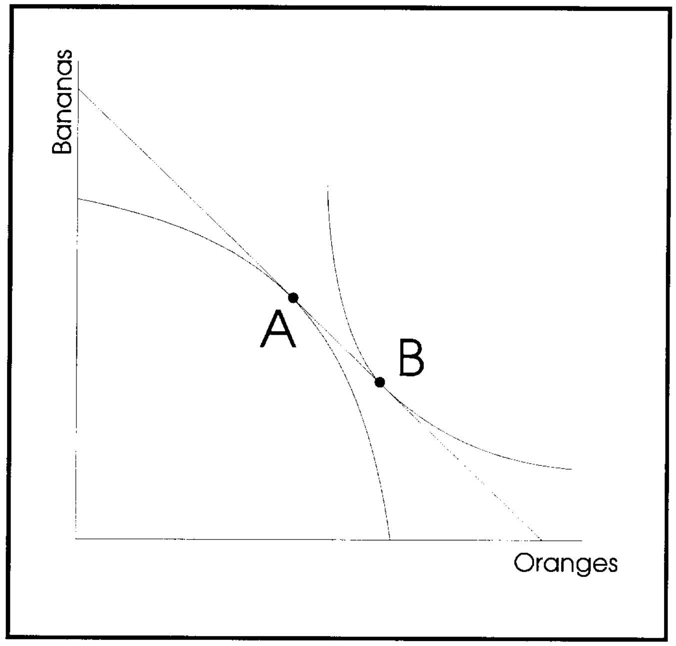
10. At this point the slope of the individual's indifference (equal to the rate at which he is willing to substitute bananas for oranges) is equal to the slope of the opportunity set or the cost of an additional orange (again, in terms of bananas).



11. In this case the individual's opportunity set is referred to as his *production possibility set* (since it gives all his alternative production choices). The slope of the opportunity set (in absolute value) gives the cost of an additional orange. This is known as the **marginal cost** if an orange. The individual's optimal choice sets the ***marginal cost*** of an orange ***equal to*** his ***marginal value*** of an additional orange.

12. If we now allow this individual to buy or sell oranges in the market at a fixed exchange rate (i.e., a fixed relative price of bananas and oranges), we can determine his optimal production and consumption choices. In the previous case he was forced to consume whatever he produced. Now he can produce whatever gives him the largest income (in terms of bananas, oranges, or dollars; they will be equivalent) and then simply purchase his desired consumption bundle.

13. The equilibrium choice for the individual will be like that in the right-hand graph. He will choose a production point where the marginal cost of oranges (in terms of bananas) is equal to the price of oranges in terms of bananas such as at the point A. This is simply a tangency between the production possibility set and the price line. His consumption point will be like that at point B where the same price line is tangent to his indifference curve. By choosing the production point that is tangent to a price line, he achieves the highest possible level of income. This income can then be spent on the desired proportions of the two goods.



14. This equilibrium illustrates two important points. First, an individual's production decision is independent of the individual's consumption preferences. In addition, the individual's optimal production point is one where marginal cost equals the market price of the good (in this case in terms of bananas). These results lead to two proportions that will carry through to production in any competitive market.

    a. Individuals will produce the level of output that yields the highest market value. That is, they will maximize profit, or the market value of output minus the market value of the inputs used in production.

    b. Profit is maximized by setting the marginal cost of producing the good equal to the market price.

XX. Types and Measures of Cost

A. In the above example we talked about the cost of oranges in terms of bananas and thereby emphasized the opportunity nature of cost. However, such explicit accounting of costs is too awkward for most discussions on the cost of producing a good. For example, one could describe the cost of producing a car by listing the amount of energy, steel, man hours, etc., that go into the production of a car, but this would be hopelessly complicated. Instead we transform these forgone resources into common units (namely dollars) by multiplying by the appropriate market price of the good. Since in equilibrium the market price of a good equals its marginal value to all users (in terms of money), this will yield an appropriate measure of the opportunity cost of the resources used to produce the car. We say the cost of the car is equal to the price of steel times the amount of steel used plus the wage rate times the number of man hours plus ... In this way we obtain a simple dollar value representation of the opportunity cost of producing the car. While this cost is expressed in dollar terms, one should remember that it is the alternative use value of these resources that represents the true cost. The money cost provides a simple measure of the value of these alternative uses.

B. In order to discuss cost and the firm's associated production decisions we must first define some terminology and definitions for cost.

C. First we distinguish among different types of costs as they relate to the level of output produced. These are:

1. **Fixed Cost** (FC) = Costs that are independent of the output level produced (given some production). This could be the cost of renting a plot or warehouse.

2. **Variable Cost** ($VC(x)$) = The part of costs that depend on the chosen level of production (such as the number of workers hired or the amount of inputs). The notation $VC(x)$ illustrates the dependence of variable cost on the chosen level of output x.

3. **Total Cost** ($TC(x)$) = Fixed Cost +Variable Cost. This gives the total amount spent to produce x units of output. Since VC depends on x, TC will depend on the level of output as well, hence $TC(x)$.

4. **Sunk Cost** = The part of the fixed cost which must be paid whether production occurs or not. These are generally present after some initial investment or commitment has been made.

5. **Salvageable Cost** = The part of fixed cost that can be recovered if one decides not to produce after making the initial fixed-cost investment or commitment.

D. Next we can define several measures of cost. These are:

1. **Average Cost**. Average total cost is simply defined as $ACT(x) = TC(x)/x$ or the cost

per unit of producing x units of output. Another useful average cost measure is average salvageable cost, which is the opportunity cost of production once the fixed-cost investment has been made. We commonly call this $ASC(x)$ = (Salvageable Cost + Variable Cost)/x.

2. As we saw above, a very important notion of cost was marginal cost, which in this case is defined as $dTC(x)/dx$ or the ratio of the change in total cost divided by the change in output.

F. As we saw above, the firm will maximize profits by producing where price equals marginal cost. However, we must also check to see whether the firm makes positive profits at this level of output to see whether production is profitable at all. We state this as proposition 1.

Proposition 1: The firm will produce at the output level where price equals marginal cost so long as at this output level the firm make positive profits. If the firm cannot make a profit at the point where price equals marginal cost then the firm's best choice is not to produce at all.

G. The condition that the firm make positive profits is simply that $P \times x + TC(x) > 0$. Dividing through by x yields the simple condition $P > ATC(x)$. Therefore, the firm can make a profit and hence will produce so long as price is greater than average cost (i.e., the per-unit revenues exceed the per-unit costs).

H. Next we consider how the firm's decision problem changes if the fixed-cost investment has already been made. In this case the firm will still produce where price equals marginal cost (if it decides to produce at all). However, now if the firm decides not to produce it is out the *sunk cost* investment. The firm will produce if producing at the point where price equals marginal cost yields a level of profits greater than -sunk cost. This condition is simply $Px - TC(x) > -SuC$. This can then be rewritten as equation (2):

$$Px > TC(x) - SuC. \tag{2}$$

Equation (2) simply states that the firm will produce as long as it can cover its salvageable cost. Dividing through by x yields the equivalent condition that the firm will produce where price equals marginal cost so long as at this level price exceeds average salvageable cost.

I. These results can then be summarized as proposition 2.

Proposition 2: The firm will produce where price equals marginal cost after making the fixed investment so long as at this output level the firm can cover its cost or simply price exceeds average salvageable cost. If the price is less than average salvageable cost, then the firm will shut down and not produce.

J. Since the firm will make the fixed costs investment and produce as long as price exceeds the minimum value of average total cost (i.e., the firm can make a profit), the minimum value of

ATC is commonly called the firm's **entry price**. Since the firm will continue to produce (after making the fixed-cost investment) as long as the output price exceeds the minimum value of average salvageable cost, the minimum value of salvageable cost is called the firm's **exit price**.

XXII. The Firm's Decision

A. The previous propositions lay out two important principles about how firms will behave in a competitive market environment. These are:

1. Firms will decide to make the fixed-cost investment and produce where price equals marginal cost, so long as this price exceeds minimum average total cost.

2. After making the fixed-cost investment, firms will produce where price equals marginal cost if price exceeds minimum average *salvageable cost.* In particular, sunk cost investments are not relevant for decisions once they have been made. Remember, ***sunk costs are sunk!***

B. A convenient way to approach this problem is by graphing the firm's average and marginal cost functions as a function of the output level. To do this we first derive a useful relationship between marginal and average cost. By definition we have the following:

$$TC(x+1) = TC(x) + MC(x+1). \tag{1}$$

But $TC(x) = AC(x) \times x$; using this we obtain

$$AC(x+1)(x+1) = AC(x)x + MC(x+1). \tag{2}$$

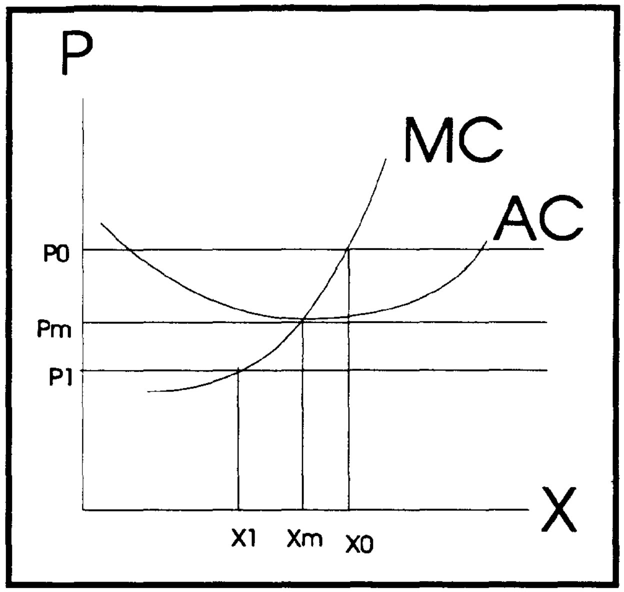
Dividing through by $x+1$ and simplifying yields

$$AC(x+1) - AC(x) = \frac{MC(x+1) - AC(x)}{(x+1)}. \tag{3}$$

If marginal cost exceeds average cost then the right-hand side of (2) will be positive. This implies that $AC(x+1) > AC(x)$, or that average cost is rising. If average cost exceeds marginal cost then equation (2) implies that average cost is falling. Given these results we have the following relationship between marginal and average cost:

Marginal Cost > Average Cost ⇒ Average Cost Is RISING
Marginal Cost = Average Cost ⇒ Average Cost Is FLAT
Marginal Cost < Average Cost ⇒ Average Cost Is FALLING

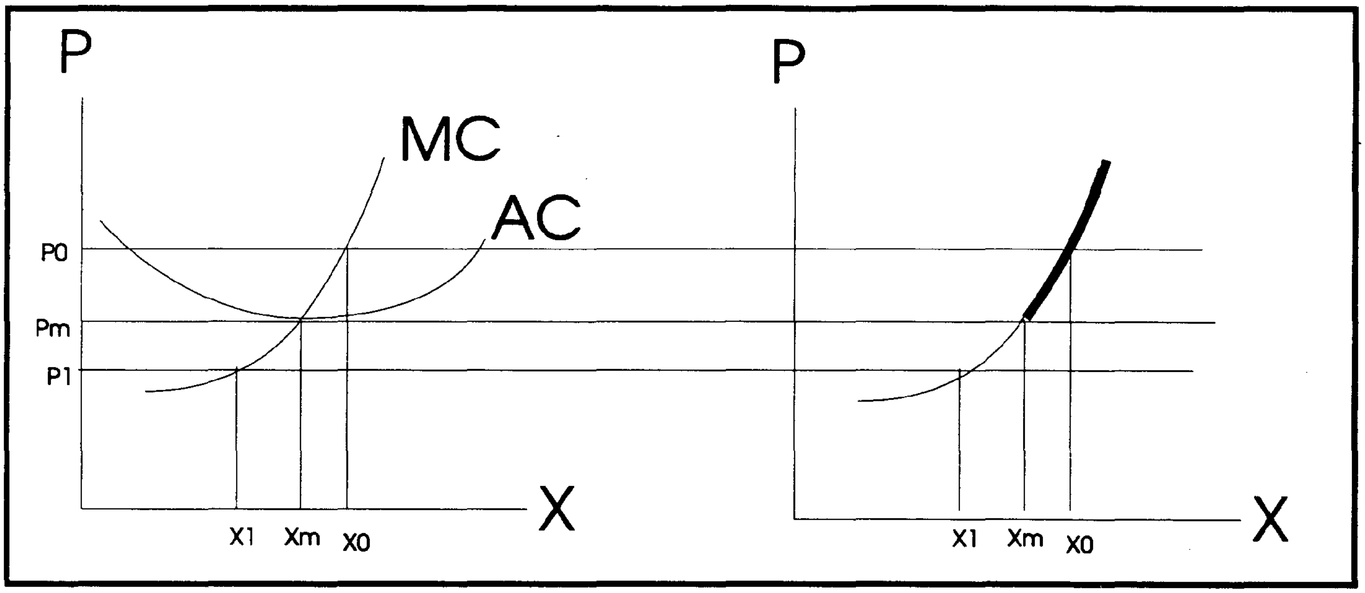
C. Often a firm's average cost will first declined as output increases. This initial decline is due to the fact that the initial fixed cost can be amortized over a greater number of units and perhaps to increased efficiency of larger scale production. However, as output continues to rise, the firm begins to experience diminishing returns and average cost begins to rise. These increased costs may be due to decreased efficiency of large-scale production, reaching the capacity of the fixed inputs such as a given plant size, or the fact that fewer and fewer productive resources must be used as the low-cost production opportunities are used up. (An example of this was our example where we had two plots; after using up the capacity of the low-cost plot, the marginal cost of bananas increased.)



D. An average cost curve with this U shape is plotted in the graph at the right above. The output level X_m represents the lowest per-unit cost of production. Since average cost is falling to the left of X_m, marginal cost must exceed average cost in this range. Using this we can draw an appropriate marginal cost curve as shown in the graph at right above. If the product price is P_0 then the firm's production rule would lead the firm to produce X_0 units of output, where price equals marginal cost. At this level of output, price exceeds average cost and the firm can earn a profit. Therefore, it will pay the firm to produce at X_0 when the price is P_0. With the price at P_1, setting marginal cost equal to the product price gives X_1 as the optimal level of output given that the firm decides to produce. However, at this level of output, price exceeds average cost and it will not pay the firm to produce. In fact, the firm will not produce at output level less than X_m. (Why?)

E. The above argument can now be used to find the firm's *supply curve*. Just as an individual's demand curve tells us how many units of a good the individual will purchase at each price, an individual firm's supply curve tells us how many units of output the firm will produce at each price. To do this we can make two graphs, one with the firm's marginal and average cost on the vertical axis and output on the horizontal axis, and the other with price on the vertical axis and the quantity supplied by the firm on the horizontal. These are shown in the two figures below:

The supply curve shown in the figure above right can be derived as follows. At price P_0 the firm will supply X_0 where price equals marginal cost, since at this level price exceeds average cost. Similarly, at all prices above P_m, the firm will produce where price equals marginal cost since price will exceed average cost. At prices like P_1, average cost will exceed price and the firm will produce zero output. Hence, the firm's supply curve is zero for prices below minimum average cost (i.e., prices below P_m) and the supply curve is equal to the firm's marginal cost curve for prices above this level.



F. The above example was done under the assumption that the firm earned zero profits if it decided not to produce. (Hence the firm would produce only if it could make positive profits.) This would be appropriate if the firm had not yet made any sunk cost investments. However, after making a sunk cost investment the firm will make negative profits equal to the cost of the sunk investment. (Since by definition these costs must be paid even if the firm decides not to produce.) If the firm's sunk costs are SuC then the firm will produce as long as profits from producing exceed -SuC. The firm's profits from producing x units of output are

$$\text{PROFIT} = P \times x - (FC + VC(x)). \qquad (1)$$

Hence the firm will produce as long as the following inequality holds:

$$P \times x - (FC + VC(x)) > -SuC \quad (2)$$

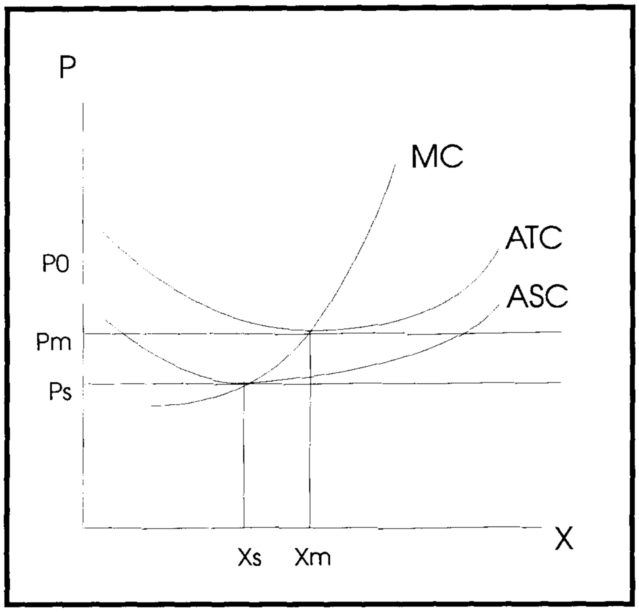
or, equivalently, as long as

$$P \times x > FC - SuC + VC(x). \qquad (3)$$

But FC - SuC = SaC or the salvageable part of fixed cost. Dividing through by x yields:

$$P > \frac{(SaC + VC(x))}{x} = ASC(x) \tag{4}$$

as an equivalent condition. Equation (4) states what we said before: after making the sunk cost investment the firm will produce as long as price exceeds average salvageable cost. This result simply says the firm considers the cost of the fixed investment at what the firm can get for it elsewhere, namely the opportunity cost of the investment *once it has been made.* The sunk costs are not considered after they have been made. Once again, sunk costs are sunk.

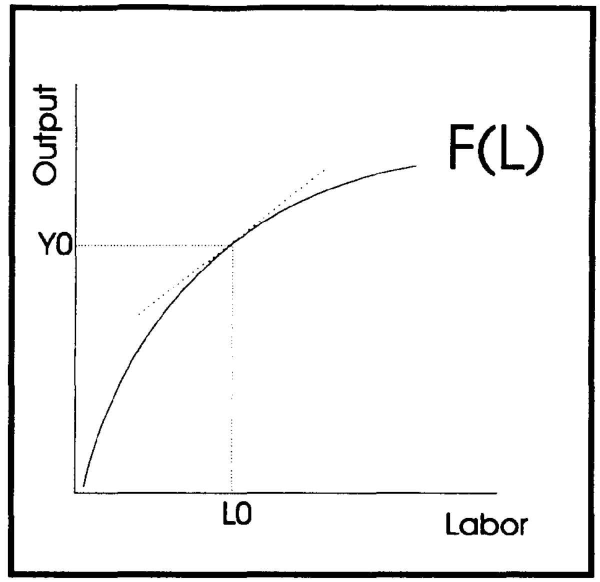


G. If some costs are sunk then total cost will exceed salvageable fixed cost plus variable cost by the amount of the sunk cost investment. In this case average total cost will always be greater than average salvageable cost. This is illustrated in the example at left. In this case, after making the sunk cost investment the firm will produce where price equals marginal cost at any price above P_S. At prices below P_S the firm cannot cover even its salvageable cost and so the firm will no longer produce. Hence P_S is often referred to as the firm's *exit price.* At prices below this level the firm will exit the industry and not produce. We saw above that the firm will make the investment and produce at prices above P_m. Hence P_m is referred to as the firm's *entry price.* The difference between these two prices implies that existing firms will often continue to produce even when they cannot cover their *total cost* (including the sunk cost) so long as they can cover the opportunity or salvageable cost. Existing firms will continue to produce even though it is not profitable for others to enter the industry and produce.

XXIII. <u>Factor Demand</u>

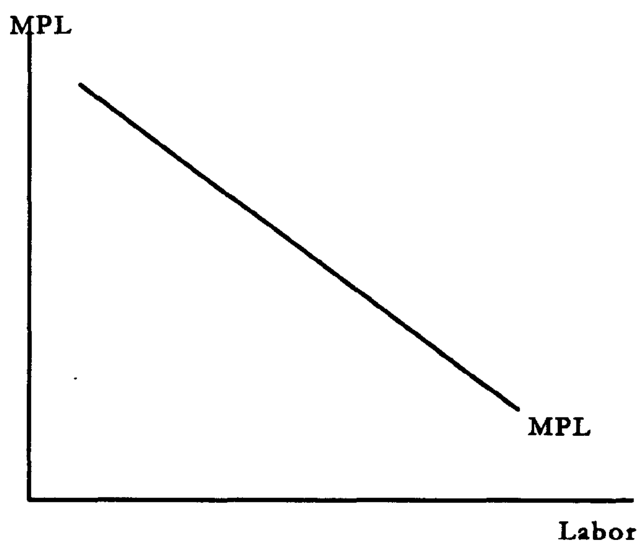
A. Our analysis to this point has focused mostly on the choice of output given the cost conditions of the firm. However, the cost curves themselves are the result of an underlying optimization process in which the firm chooses the best mix of inputs to produce its output. Economists refer to the inputs used in the production process as factors of production and the demand for these inputs as factor demand. Economists typically focus on two types of factors, labor, the services of people and capital, the services of physical assets. The demand for labor and capital is a derivative of the demand for the goods and services produced by the firms that hire these inputs. It is the ability of labor and capital to produce goods and services that will determine the demand for labor and the demand for capital.

B. In order to understand the nature of this derived demand, economists introduce the notion of a production function. Consider the case of a firm that produces good, Y, using only a single input, labor. In this case the production function, $F(L)$, gives the amount of the good produced as a function of the quality of labor employed. This function is illustrated in the graph at right. As more labor is employed by the firm, the amount of the good produced increases.



C. However, since other factors such as capital (i.e., the size of the plant) are being held fixed, output increases at a decreasing rate as the quantity of labor is increased. This represents decreasing returns to labor or decreasing marginal productivity. From the production function $F(L)$ in the graph above, we can define the marginal product of labor curve, MPL, which gives the rate of change of output with respect to the quantity of labor employed. The slope of the dashed line in the figure above would measure the marginal product of labor at L0 units of labor. Since the additional output from another unit of labor declines with the level of labor employed (i.e., the slope of $F(L)$ decreases with the quantity of labor in the top figure), the marginal product of labor schedule will slope downward as shown in the graph at the right.

D. Since firms desire to maximize profits, the firm will employ labor to the point where the marginal gain to adding additional labor is exactly equal to the cost of an additional unit of labor. The return to additional labor is the amount of output that labor can produce times the price of output, or $P_y \times MP_L$. The cost of labor is the wage, W. Hence the optimal choice of labor by the firm will be where



(1) $P_y \times MP_L = W$.

The term $P_y \times MP_L$ is the price of output times the marginal product of labor or simply the **value marginal product** of labor, or VMP.

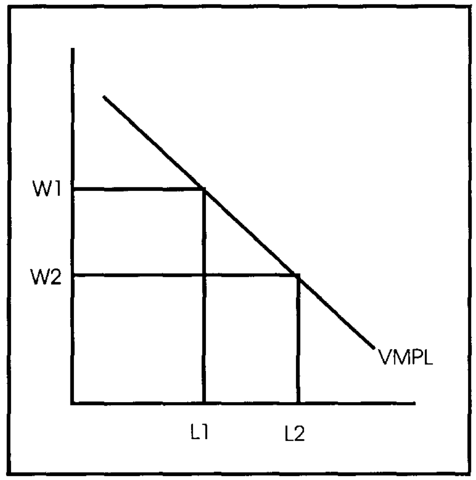
E. This condition can be derived simply using calculus. The profits of a firm that employs L units of labor are simply the value of the output produced less labor and non-labor costs. If we denote the non-labor costs as C then the firm's profits are given by

(2) $P_y \times F(L) - C - W \times L$,

or simply the value of output produced, $P_x \times F(L)$, less costs of C less the wage $W*L$ for each unit of labor employed. Differentiating with respect to L gives the first-order condition that

(3) $P_y \times dF/dL - W = 0$ or $P_y \times \mathrm{MP_L} = W$.

F. As this discussion illustrates, firms will hire labor to the point where the value of marginal product equals the wage rate. This can be illustrated graphically as in the graph at right. At a wage of W_1 the firm will operate where the VMP equals the wage and hire L_1 units of labor. At a lower wage of W_2 it will pay the firm to hire more labor, in this case L_2 units of labor. Hence, the VMP schedule represents the firm's demand for labor curve since it gives the amount of labor demanded at each wage.



G. A decrease in the wage causes the firm to move along the VMP schedule and demand more labor. An increase in the price of output (i.e., P_x) shifts the VMP curve upward and increases the quantity of labor demanded at a fixed wage. With a higher price for output the value of the goods or services produced by an additional unit of labor rises proportionately. With a higher value of the output produced it pays the firm to use more labor.

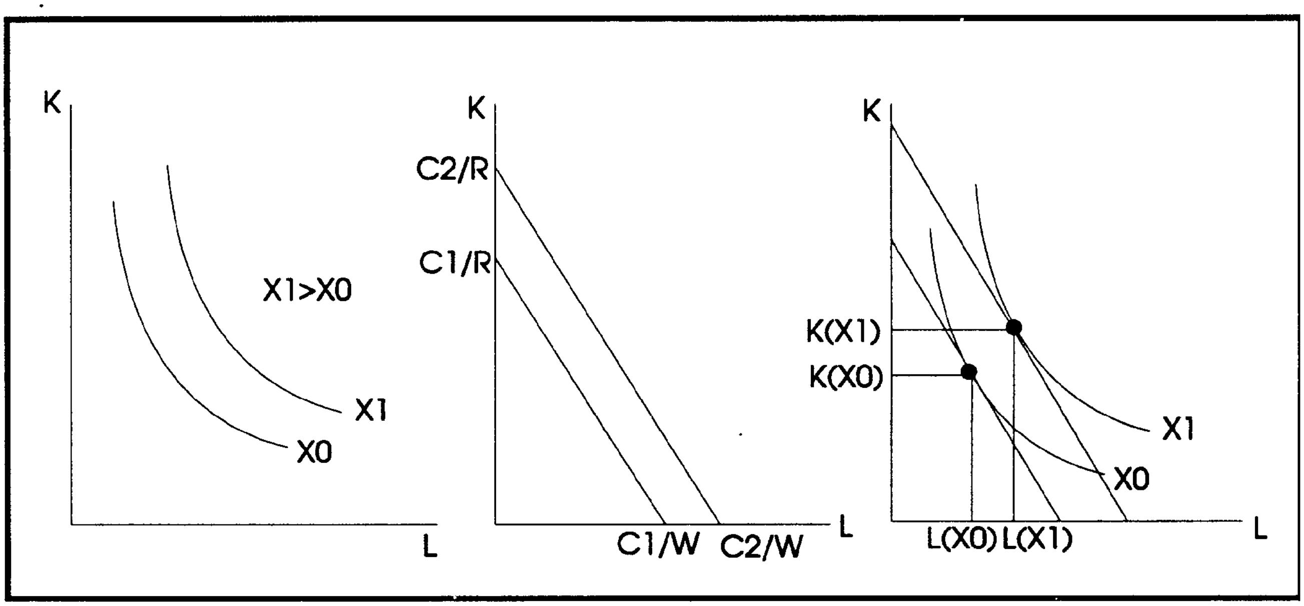
XXIV. <u>Long-Run Factor Demand</u> (Two Inputs)

A. In the previous case we regarded the choice of labor as the only decision variable open to the firm. In general firms use multiple inputs to produce output. For example, a manufacturing firm may use labor together with plant and equipment to produce its output. In this context, our discussion above is often regarded as the "short run." The basic idea is that in the short run the size of the plant together with the amount of physical equipment can be regarded as fixed. In this case the VMP schedule used above can be regarded as the marginal product of labor holding the level of these other inputs (what we will collectively call capital) fixed at the current level.

B. In general, the current level of capital will substantially influence the demand for labor (i.e., the size of the current plant together with the amount of machinery and other equipment available will determine the amount of output produced by a given quantity of labor as well as the additional amount of output generated by adding more labor (the marginal product of labor)). In the long run the firm will decide on the optimal level of both labor and capital jointly in order to maximize profits.

C. We can think of this choice of capital and labor as occurring in two stages. In the first stage the firm decides the mix of labor and capital that can produce a given level of output at the least cost. Once we have determined the optimal amounts of labor and capital to use to produce each level of output we can determine the cost function for the firm (i.e., the cost of producing at each level of output). In the second stage the firm determines the optimal level of output by producing where price equals the marginal cost of production. The optimal levels of labor and capital are then the cost-minimizing levels for this chosen level of output.

D. A convenient method for analyzing the problem of cost minimization for a given level of output is to consider all the combinations of labor and capital that can be used to produce the given level of output. If we make a graph with the quantity of labor employed on one axis and the amount of capital employed on the other axis, then the collection of points that can produce a given level of output would look like an indifference curve. In this case these points are called an isoquant (for equal quantity). Isoquants slope downward due to the fact that in order to keep the level of output constant a decrease in the amount of capital employed must be offset by an increase in the quantity of labor employed. In addition, as we decrease the amount of capital, capital becomes more "scarce"" and the amount of labor required to substitute for a unit of capital will increase. The amount of labor needed to substitute for a unit of capital (i.e., the amount labor must increase to maintain output at the current level of production) is known as the *marginal rate of substitution* of labor for capital. Similarly, the marginal rate of substitution of capital for labor is the amount of capital required to compensate for a one-unit reduction of labor and is simply the slope of the isoquant as shown in the first graph below. Isoquants corresponding to higher levels of output require either more labor, more capital, or more of both inputs. Hence, isoquants for higher levels of output lie to the northeast of the isoquants corresponding to lower levels of output. The isoquant X_1 lies above that for X_0 as long as $X_1 > X_0$ as shown in the figure.

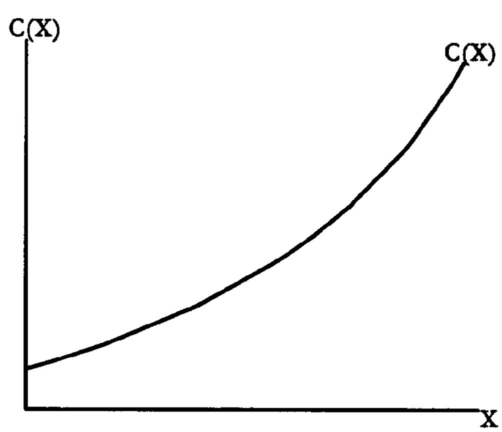


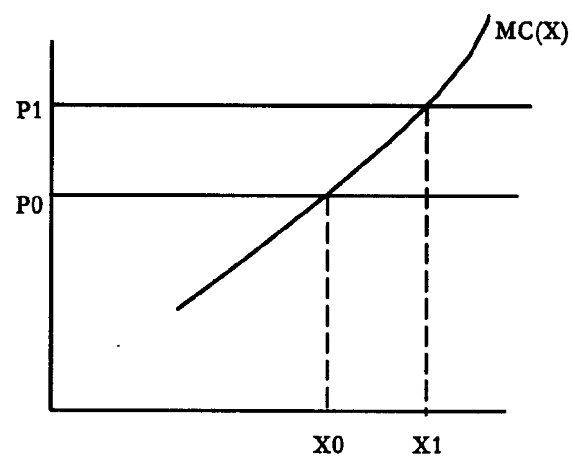
E. The isoquants determine the ability of the firm to substitute capital and labor in production. The rate at which capital and labor can be substituted in the marketplace is determined by the prices of capital and labor. With a price of capital of R and a price of labor of W, one unit of labor can be substituted for W/R units of capital. This ability to substitute is illustrated by the ISOCOST curves shown in the second figure above. Just as the isoquant curves give the combinations of labor and capital that can produce a given level of output, the ISOCOST curves give the combinations of labor and capital with equal costs (these are just like budget lines in the case of consumers). In this case the ISOCOST curves closer to the origin represent lower cost and hence the firm will attempt to produce a given level of output (say X_0) at the least possible cost corresponding to the lowest possible isocost curve. Hence, the cost-minimizing solution will be for the firm to find a tangency between the isoquant for X_0 and an isocost curve. This tangency is illustrated in the third figure and yields $L(X_0)$ and $K(X_0)$ as the optimal levels of labor and capital to produce output X_0. Similarly, the minimum cost inputs to produce the higher level of output X_1 are $L(X_1)$ and $K(X_1)$ as shown in the figure.

F. If we were to find the optimal levels of labor and capital to use for each level of output, then we could determine the cost of producing each of these output levels directly. For example, at output X_0 the optimal choices of labor and capital are $L(X_0)$ and $K(X_0)$, respectively, and hence the cost of producing X_0, $C(X_0)$, is simply

(1) $C(X_0) = L(X_0) \times W + K(X_0) \times R$.

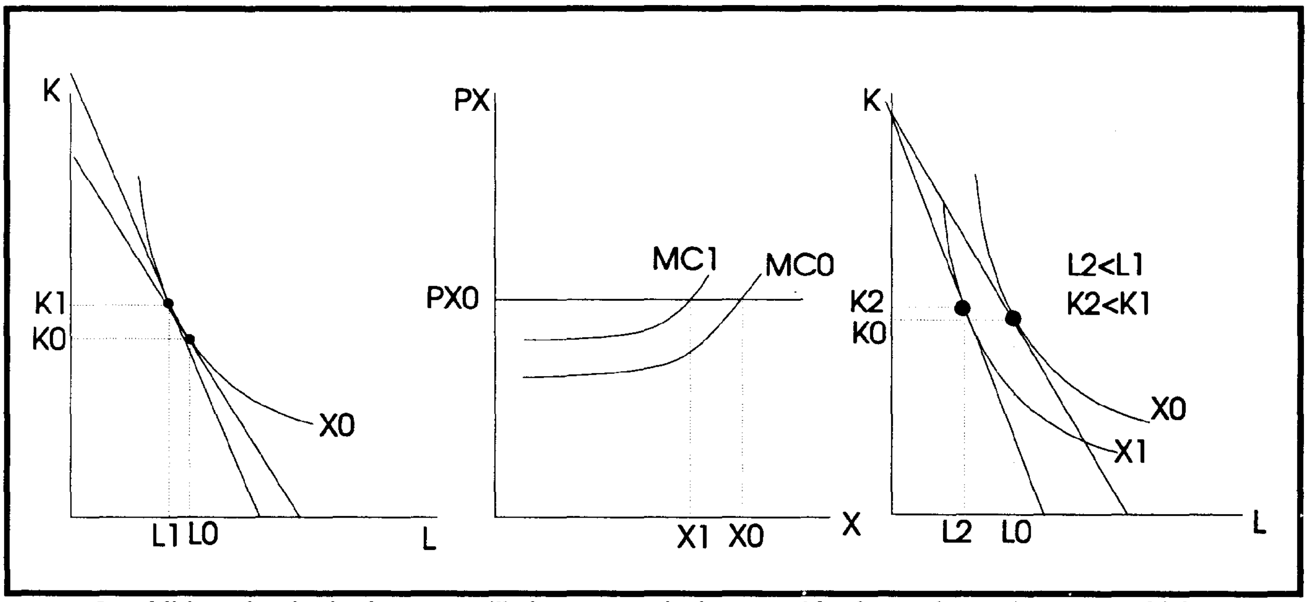
Similarly, the cost of producing the output level X_1 is $C(X_1) = L(X_1) \times W + K(X_1) \times R$. By computing the cost in this manner for each level of output, we can determine the firm's total cost curve $C(X)$ as shown in the first graph below. The marginal cost curve is simply the derivative of this total cost curve and gives the cost of an additional increment to output at each level of production as shown in the second graph below. As in the usual case the firm's optimal level of output will be where marginal cost equals the product price. Hence, as the second graph illustrates, the firm will choose output level X_0 when the price of output is P_0 and the higher level of output, X_1, when the price of output is P_1.





G. As the second graph illustrates, an increase in the market price for output will lead to an increase in the level of output of the firm and will therefore affect the amount of labor employed. If this increase in output increases the amount of labor employed (at fixed prices of labor and capital, i.e., fixed W and r), then we refer to labor as a *normal factor of production.* Similarly, a factor of production (such as labor) that decreases in utilization as the level of output increases is called an *inferior factor.* Most factors of production are normal factors (including labor). Hence, an increase in product price will increase the amount of labor employed if labor is a normal factor of production and will decrease the amount of labor employed if labor is an inferior factor.

H. This same methodology can be utilized to determine the long-run effects of a change in the wage rate (or a change in the rental price of capital). An increase in the wage causes the isocost curve to become steeper and leads to a substitution of capital for labor, holding the level of output fixed as shown in the first graph below. The reduction in labor from L_0 to L_1 is called the substitution effect and always leads to less labor employed as the wage increases.



In addition, the rise in the wage will change marginal cost and lead to a change in output which we call the *scale effect.* Marginal cost at a level of output X is simply the change in total cost for an increase in output to X+1 (approximately). This is simply

(2) $MC(X) = L(X+1) \times W + K(X+1)\ R\ - (L(X) \times W + K(X) \times R)$

$= W \times (L(X+1) - L(X)) + R \times (K(X+1) - K(X)).$

Hence, an increase in the wage will raise marginal cost as long as $L(X+1) > L(X)$, i.e., as long as labor is a normal factor of production. If labor is an inferior factor of production then an increase in the wage will lower marginal cost, since $(L(X+1) - L(X))$ is negative in this case.

I. If labor is a normal factor of production, a rise in the wage will lead to a reduction of labor due to the substitution effect and a rise in marginal cost (via equation (2)). This rise in marginal cost will reduce output and lead to a further reduction in the utilization of labor (from X0 to X1<X0 in the second figure above). The total effect is then to shift labor from L0 to L2 in the third figure above (and change capital from K0 to K2). Putting the substitution and scale effects together implies that a rise in the wage reduces the amount of labor employed and therefore that the long-run demand for labor is downward-sloping when labor is a normal factor of production. The effect on capital is ambiguous since the substitution and scale effects of a rise in the wage go in the opposite direction for capital

J. If labor is an inferior factor then a rise in the wage reduces labor due to the substitution effect (as above) and lowers marginal cost. With lower marginal cost, output expands but the amount of labor *decreases* since labor is inferior in this case by assumption.

K. As we stated above, this same analysis implies that the effect of a rise in the wage on the quantity of capital employed is ambiguous. The higher wage causes capital to be substituted for labor at any given level of output (thus increasing the usage of capital). But the higher wage also raises the marginal cost of output (assuming that labor is a normal factor of production) which will lead to a reduction in the usage of capital (assuming capital is also a normal factor). The net effect will depend on which of the two effects (the substitution effect or the scale effect) dominates.

L. The analysis presented here illustrates that the cost curves of a firm are themselves a result of an optimization process, where the firm determines the least costly way of producing any given level of output. Hence when we assume that the firm is choosing the profit-maximizing level of output given its cost function, we are assuming that the firm is already minimizing costs for any given level of output. Typically when we focus on questions regarding output decisions we ignore the cost minimization stage and focus directly on the output choice for a given set of cost curves. When we wish to address issues of input usage we typically start with the more primitive production function and cost minimization approach.

XXV. <u>Productivity and Technological Change</u>

A. Productivity refers to the amount of output that can be produced from a given set of inputs. Technological change refers to changes in the production process that increase (or decrease) the amount of output that can be produced from a given quantity of inputs and/or alter the optimal mix of inputs used to produce a given level of output.

B. The study of productivity and technological change is largely empirical in nature. Measuring productivity requires measuring the inputs and output of the production process and using these measures to determine both the level and rate of change of the output obtained from given inputs.

C. The simplest measure of productivity is labor productivity. Labor productivity is most often measured as output per man-hour. In order to measure labor productivity, we require measures of output (in physical or "real" units) and the amount of labor input (in man-hours). For example, labor productivity in the steel industry might be measured by tons of steel produced per hour of labor used.

D. If we denote the physical quantity of output produced by Y and the number of man-hours of labor used by L then labor productivity, or the average product of labor, AP_L, is simply Y/L. In general, the level of labor productivity will depend on both technology and the amount of other inputs (e.g. capital) used in the production process.

E. Most often we are interested in measuring how productivity has changed over time. For this purpose it is convenient to look at the change in labor productivity. If we denote the percentage change in a variable x by Δ_x then by definition we have

(1) $\Delta_{APL} = \Delta_Y - \Delta_L.$

Hence, if output is growing at 4% per year while labor input is growing at only 3% per year then labor productivity must be growing at 1% = 4% - 3% per year.

F. We can also examine productivity growth for a competitive industry using the VMP theory of labor demand. In a competitive market the wage rate will be equal to the value marginal product of labor or

(2) $W = P \bullet MP_L,$

Where P is the price of output and MP_L is the marginal product of labor. Using this equation and looking at rates of change we then have

(3a) $\Delta_W = \Delta_P + \Delta_{MPL}$ or

(3b) $\Delta_{MPL} = \Delta_W - \Delta_P.$

Since $\Delta_W - \Delta_P$ is equal to growth in the real wage rate (where real is defined relative to the price of the output being produced), this equation tells us that we can measure growth in <u>marginal productivity</u> by growth in the real product wage. For example, if the price of output for a competitive industry is growing at 5% per year while the wage rate is growing at 6% per year then marginal productivity must be growing at 1% per year (i.e. 1% = 6% - 5%).

G. It is important to note that the two notions of the productivity of labor we have been discussing are distinct, $\Delta_Y - \Delta_L$ measures growth in the average product of labor while $\Delta_W - \Delta_P$ measures growth in the marginal product of labor. How are these related? By definition, the average product of labor is Y/L, while the marginal product of labor is W/P. Hence, the ratio of the marginal product to the average product of labor is then

Where S_L is labor's share. Equation (4) implies therefore that

(5a) $\Delta_{MPL} - \Delta_{APL} = \Delta_{SL}$, or

(5b) $\Delta_{MPL} = \Delta_{APL} + \Delta_{SL}$.

Hence, marginal productivity will grow faster (slower) than average productivity when labor's share is rising (falling). Average and marginal productivity grow at the same rate when labor's share is constant. Empirically, for the economy as a whole labor's share is relatively stable over time so that growth in labor productivity measured by the average and marginal products of labor tend to be relatively similar over the medium to long run. The similarity of the two series is illustrated in the figure below, which plots the time series of real compensation per hour and real output per hour for the U.S. non-farm business sector from 1959 to 1999.

As can be seen from the figure, the major deviation between the two series when average productivity (i.e. output per hour) declines more than marginal productivity (i.e. real compensation per hour) in an economic downturn (e.g. 1975, 1982-83).

H. In practice, the most difficult part of measuring labor productivity is measuring the physical quantity of output. For a firm or industry with a single output or many very similar outputs we can measure output in terms of a common physical unit like tons of steel or bushels of corn. For many industries, such uniform measures do not exist and we are forced to measure output (more accurately output growth) using an index number approach. To see how this works consider an industry that produces several outputs that sell at different prices. If we denote the quantity of output j in year t produced by X_{jt} and its corresponding price by P_{jt}, then total industry revenues in year are given by

(6) $REV_t = \Sigma X_{jt} P_{jt} = X_{1t} P_{1t} + X_{2t} P_{2t} + \ldots X_{nt} P_{nt}$,

where n is the number of products produced. If we compare revenues at two dates, t and t+1 we have

(7a) $REV_{t+1}/REV_t = \Sigma X_{jt+1} P_{jt+1} / \Sigma X_{jt} P_{jt}$ $\Rightarrow$

(7b) $REV_{t+1}/REV_t = (\Sigma X_{jt+1} P_{jt} / \Sigma X_{jt} P_{jt}) \bullet (\Sigma X_{jt+1} P_{jt+1})/ (\Sigma X_{jt} P_{jt})$

(7c) $= (Y_{t+1}/Y_t) \bullet (P_{t+1}/P_t)$,

where, Y_{t+1}/Y_t measures the growth in real output and P_{t+1}/P_t measures the growth in output prices from t to t+1. The output growth term measures both growth in the overall quantity of outputs and changes in the composition of output from lower to higher valued products. The price growth term is a standard price index, where the weights used in the price index are the quantities of outputs at t+1.

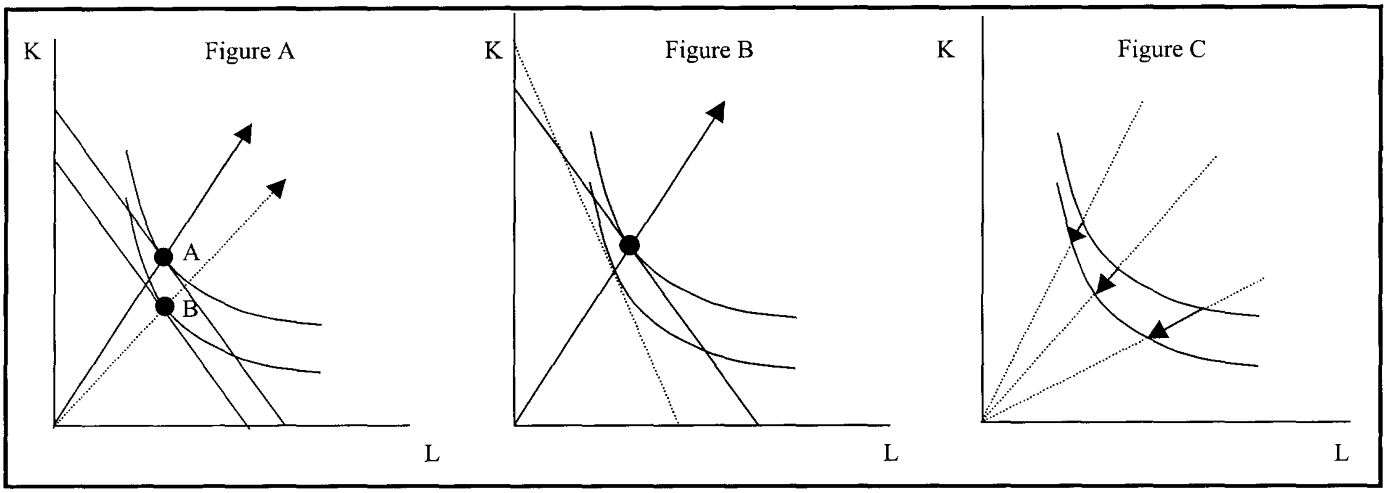
I. While labor productivity is the most common measure of productivity reported, often we are interested in measuring the growth in productivity taking account of the growth in all inputs (not just labor). Such a measure is total factor productivity. Measuring growth in total factor productivity is analogous to measuring growth in labor productivity with the only distinction being that we measure the growth in output relative to the growth in all inputs (not just labor). With two inputs, labor and capital, TFP growth is measured as

(8) $\Delta_{TFP} = \Delta_Y - (S_L \Delta_L + S_K \Delta_K)$.

The term $(S_L \Delta_L + S_K \Delta_K)$ is often referred to as the growth in total inputs since it combines the growth in labor and capital to get a measure of the inputs overall (note: the weights are simply the cost shares of the two inputs). We can also measure TFP growth using prices (just as we measured labor productivity using the real wage) as

(9) $\Delta TFP = (SL\ \Delta W + SK\ \Delta R) - \Delta P.$

J. Often we are interested in measuring how technological progress changes the relative demands for capital and labor (or skilled and unskilled labor). When technological change alters the relative demand for labor and capital we refer to such change as technological bias. The figures below show how technological bias can increase the demand for labor relative to capital.

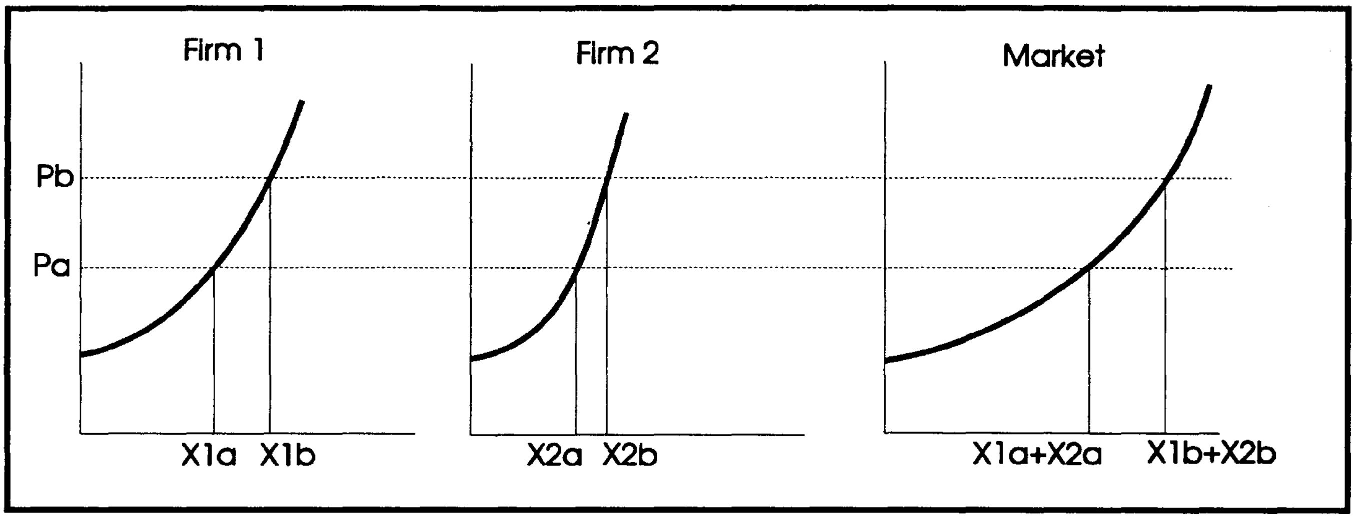


<u>The impact of biased technological change.</u>

The above figures illustrate the impact of biased technological change. In this case the bias is in favor of labor. All three Figures show the same shift in the unit isoquant. Figure A illustrates that at the same relative price of capital and labor (i.e. the same W/R ratio) the firm would switch to using relatively more labor (i.e. K/L would fall). Figure B illustrates that at the same factor ratio (i.e. along the same ray from the origin) the relative price of labor would have to go up (i.e. W/R would have to rise). This is shown by the steeper dashed isocost line tangent to the new isoquant. Figure C illustrates that with technological change biased in favor of labor, the rate of technological improvement is greatest at higher L/K (i.e. lower K/L) ratios.

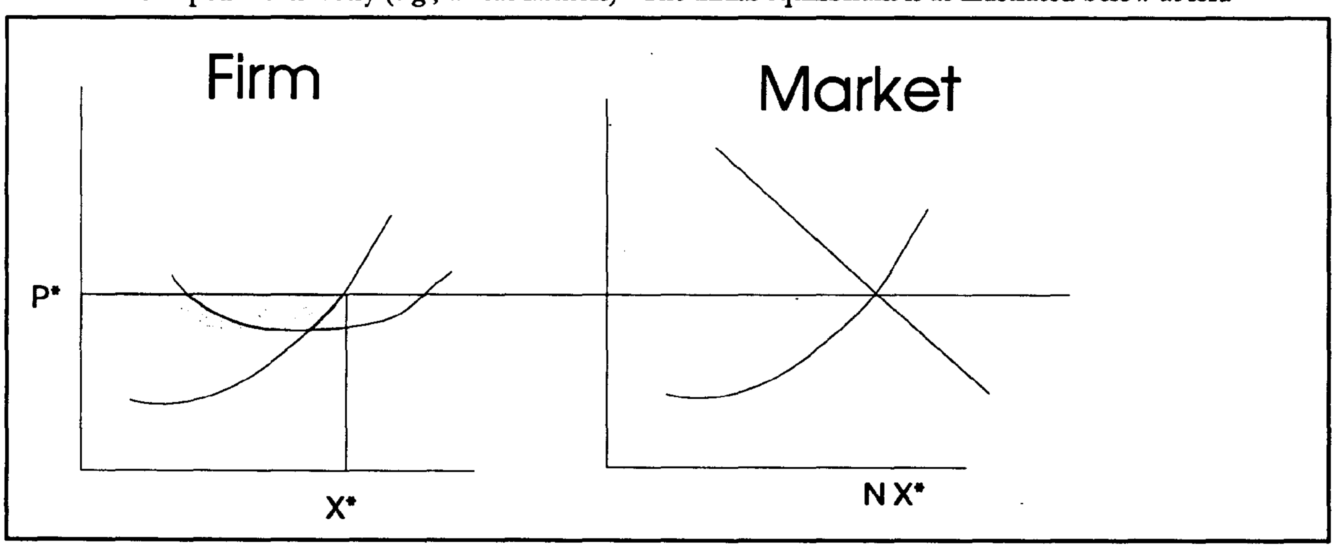
XXVI. Market Equilibrium (Identical Firms)

A. Above, we derived the supply curve of an individual firm. To derive the market supply curve for a fixed number of firms we simply sum the individual firm's supply curves horizontally to obtain the ***market supply curve (see below for an important exception to this).*** The example of two firms is illustrated in the figures below: at the price P_a firm 1 produces X1a and firm 2 produces X2a for a market supply of Xa = X1a + X2a. At the higher price, P_b, firm 1 ***increases its output*** to X1b and firm 2 produces the higher quantity X2b; the total output on the market therefore increases to X_b = X1b + X2b.



B. Once we have derived the market supply curve of the firms in the industry we can determine the equilibrium price by the intersection of the supply and demand curves for the product. With the market price we can then determine the output and profit levels of the individual firms. We begin by assuming that all the firms in the industry have the same cost curves. This yields a useful model of production and competitive supply.

C. We begin by assuming that there are currently *N* identical firms producing output in a competitive industry (e.g., wheat farmers). The initial equilibrium is as illustrated below at left.

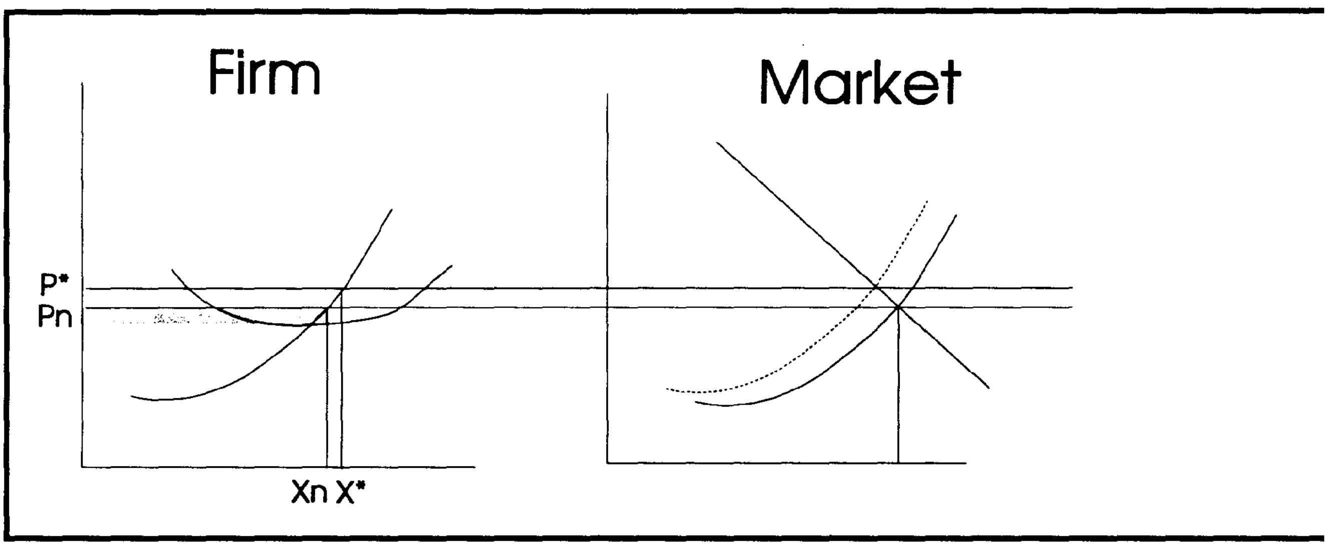


The sum of the individual firms' market cost curves (i.e., their supply curves) determines the market supply. The intersection of the supply curve with the market demand curve gives the equilibrium price P* as in the figure at right above. At this price each firm produces X* units of output and earns profits equal to the shaded area in the figure at left. With a fixed number of firms in the industry, this represents the **short-run equilibrium.**

D. This equilibrium was defined holding the number of firms in the industry constant (at *N*). In fact this is the reason we referred to this as the short-run equilibrium. It seems reasonable to assume that other firms cannot enter the industry immediately and hence the above equilibrium will prevail for a while. However, in a longer time frame there is no reason why additional firms will not enter the industry or existing firms exit the industry in response to the industry's level of profits.

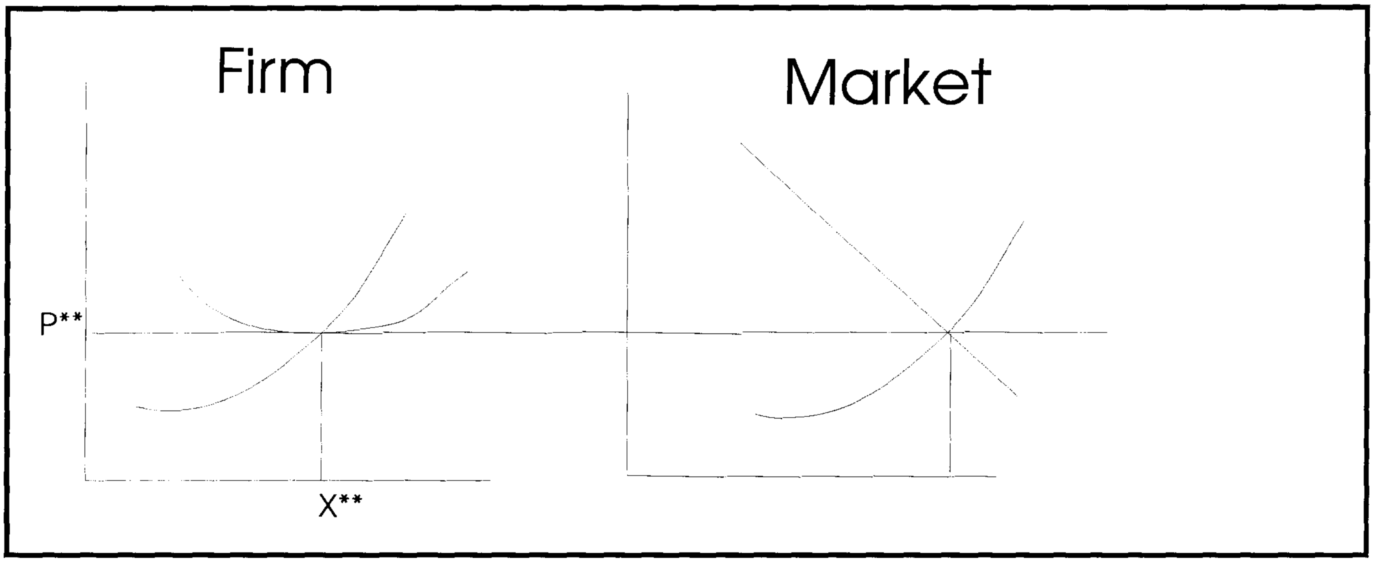
E. In the above equilibrium, existing firms are earning profits equal to the shaded area in the figure on the left. This implies that these firms are more than covering their costs and thereby making a greater than normal rate of return on their investment. Since all firms have the same costs of production, additional firms not currently in the industry see these profits and will find it attractive to enter the industry. Hence, additional firms will enter the industry and the short-run equilibrium where the firms earn profits cannot be maintained in the long run when we allow free entry.

F. When additional firms enter the industry, the supply curve shifts to the right. This increases the equilibrium quantity sold on the market and decreases the market price, as shown in the figures below.



G. The new industry supply curve is represented by the solid line and the old supply curve by the dashed line. The new market equilibrium price is given by P_N. This reduction in price causes each firm in the industry to cut back sales (which are picked up by the new firm) and reduces the profits of each firm in the industry.

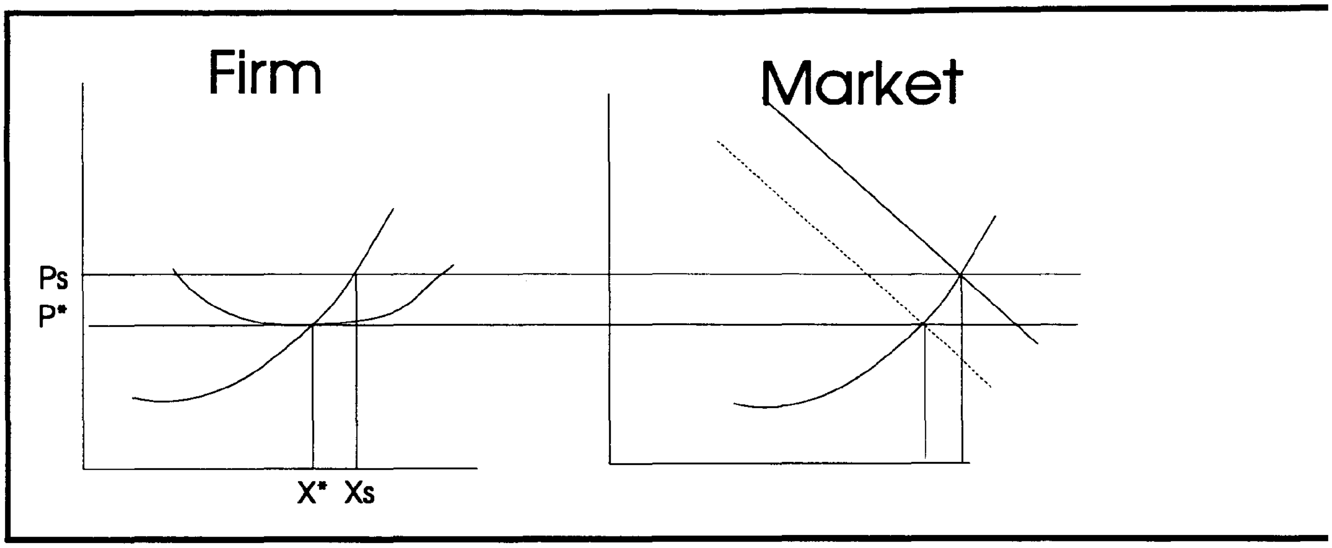
H. Even after these firms have entered, existing firms are still making a profit since price exceeds average total cost as illustrated in the figure at left. Hence, entry will continue until the market price is driven down and profits are eliminated when price equals the level of minimum average cost of the identical firms. This **long-run equilibrium** is illustrated in the figures below.



I. In the long run, all firms earn zero profits and produce at the minimum average cost level of output. At this point consumers receive the product at the lowest possible production cost, namely the level of minimum average cost. Since firms earn zero profits, it will not pay for additional firms to enter the industry and this equilibrium will be maintained.

XXVII. Example: Changes in Demand

A. In order to consider the effects of an increase in demand on the market equilibrium we begin with the long-run equilibrium position outlined above. In the figure below on the right, the dashed line represents the old level of demand and P* is the corresponding long-run equilibrium price equal to minimum average cost. At this price all firms produce at the minimum average cost level of output as illustrated in the diagram at left and earn zero profits. If demand increases to the solid line in the diagram at right, the price will rise to P_S along the short-run supply curve (equal to the sum of the marginal cost curves of the existing firms). This rise in price induces each firm to expand output to Xs where price equals marginal cost, as illustrated in the figure on the left.



B. At this new short-run equilibrium the price has risen, the total quantity supplied has risen, and the output of each firm has increased as have each firm's profits. The rise in profits will induce other firms to enter the industry which will lead to a new long-run equilibrium. As entry occurs the market supply curve will shift to the right, which will further increase the quantity sold on the market (thus meeting more of the increased demand) and will begin to lower the price back toward its original level.

C. Firms will continue to enter as long as they can earn a profit (i.e., as long as price is above minimum average cost). The free entry of these firms will therefore drive the market price back to the level of minimum average cost and lead to the long-run equilibrium. Note that in the long run we have more firms, a lower price, greater total output and smaller output per firm than in the short-run equilibrium following the increase in demand.

D. These results are listed in the chart below. An increase in product demand has the following effect on the market equilibrium. The first two columns compare the short- and long-run equilibria *after* the demand increase to the long-run equilibrium that prevailed prior to the increase in demand. The final column compares the long-run equilibrium after the demand increase to the initial short-run equilibrium following the increase in demand.

Effects of a Rise in Demand on a Competitive Market with Identical Firms

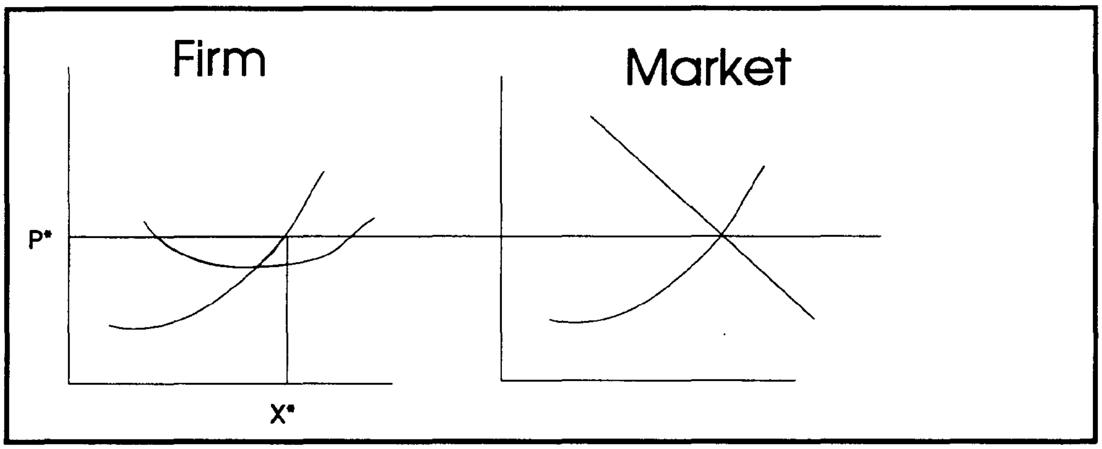
	Short Run	Long Run	Long vs. Short Run
1. Price	Increase	Unchanged	Decrease
2. Total Output	Increase	Increase	Increase
3. Output per Firm	Increase	Unchanged	Decrease
4. Profit	Increase	Unchanged	Decrease
5. Number of Firms	Unchanged	Increase	Increase

XXVIII. Market Equilibrium: Non-Identical Firms

A. As we will see later in the course, the above model is useful for analyzing the market effects of changes in demand or other conditions. However, the assumption that all firms have the same cost curves seems unrealistic. More importantly, this assumption ignores the important role of comparative advantage that we saw in our example of the two plots of land.

B. If we allow the firms to have different cost curves, then we can determine which firms will enter the industry and which will remain out of the industry.

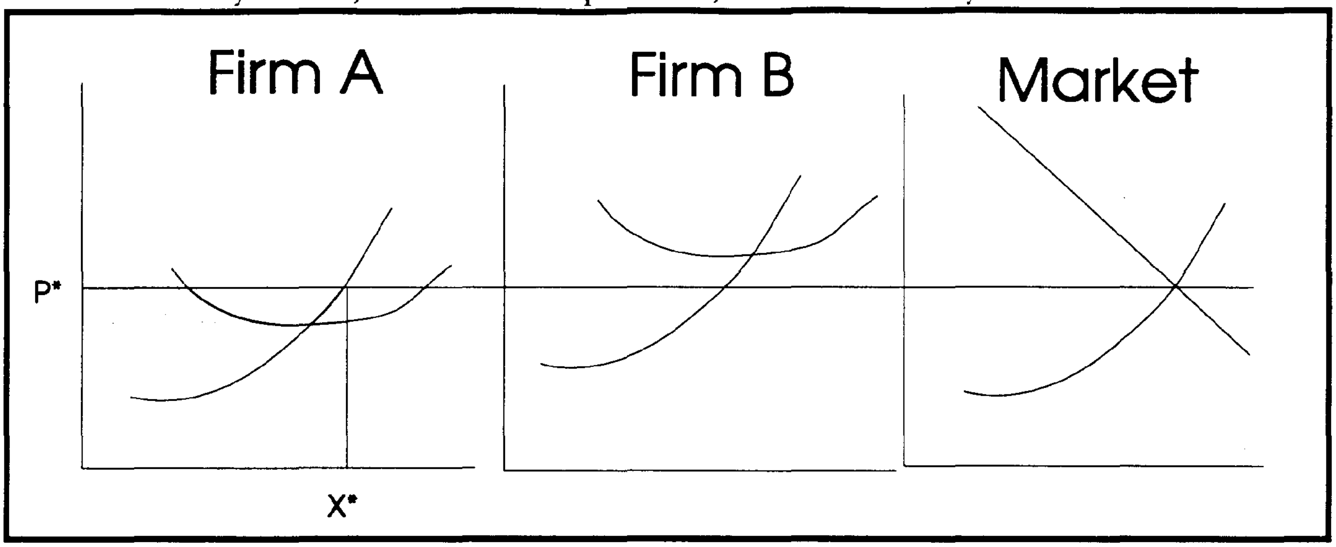
C. We begin with a fixed number of firms in the industry as we did in the case of identical firms. The market equilibrium for an individual firm will be like that illustrated below, which is the same as the short-run equilibrium with identical firms. The market equilibrium and the position of an individual firm will be like that illustrated in the figures below:



This particular firm is therefore earning profits equal to the shaded area in the figure at left.

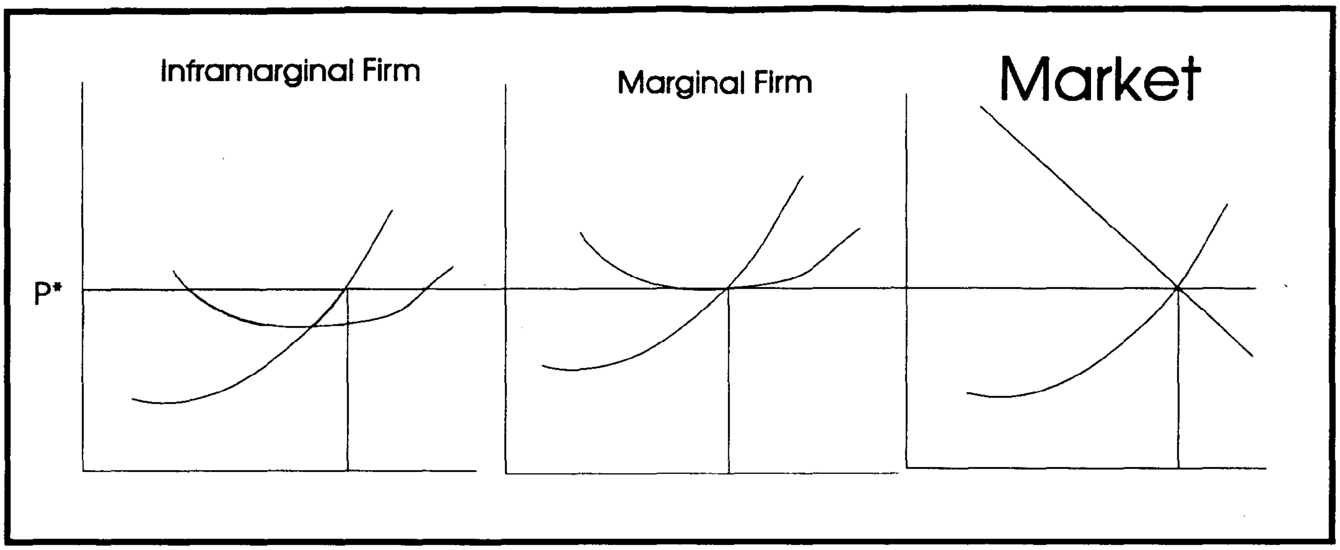
D. The difference between this and the identical firm case comes in the long run. In particular, what will distinguish firms is the level of minimum average cost. A firm will enter the industry as long as price exceeds minimum average total cost (i.e., as long as price exceeds the firm's entry price). In the long-run equilibrium, all firms outside the industry will have minimum average costs greater than the current equilibrium price. (Hence it will not pay for them to enter.) All firms in the industry will have minimum average cost at or below the current price. (Hence it will not pay them to leave the market.)

E. Such an equilibrium is illustrated in the three figures below. In this case firm B will not be in the industry. Firm A, with lower cost of production, will be in the industry.



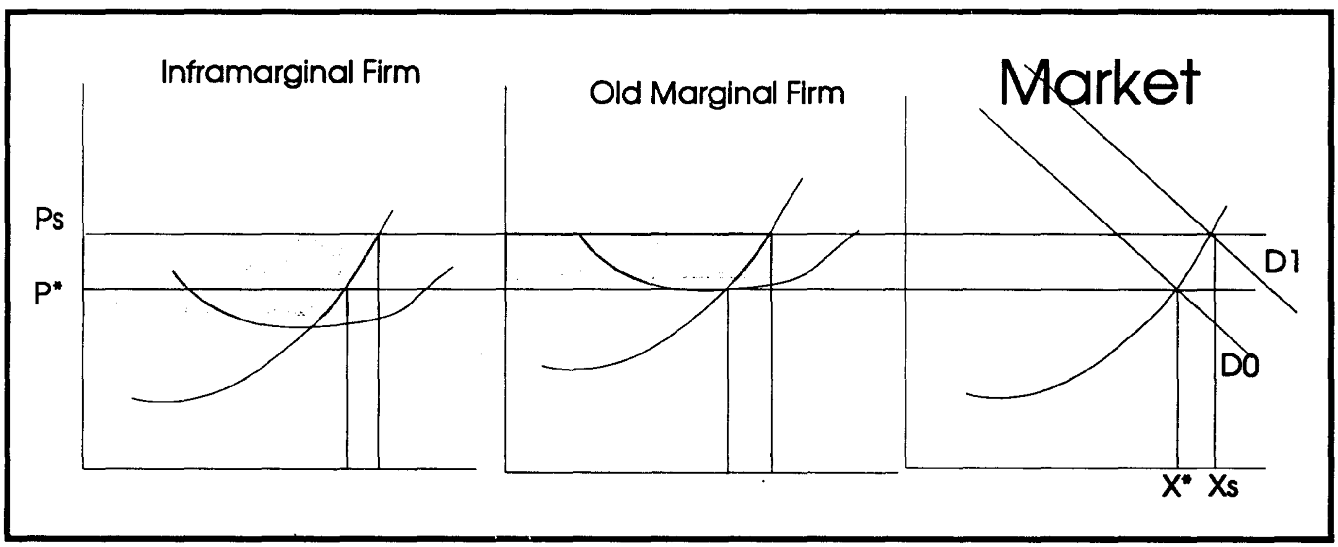
F. A useful idea in this case is the notion of the **marginal firm**. If there are many firms with different cost curves then the highest-cost firm currently in the industry will earn approximately zero profits (if not it would pay a slightly less productive firm to enter the industry, since it too could then make profits). This firm is the current marginal firm. That is, it will be the first firm to leave the industry if price falls. In the long-run equilibrium, the marginal firm will earn zero profits (approximately) and it will not pay additional firms to enter. In this case those firms with lower cost will earn positive profits.

G. The long-run equilibrium is illustrated in the three figures below. The one on the left shows the long-run position of a relatively low-cost producer or **inframarginal firm**, which earns positive profits in the long run. At this price the firm in the middle figure it he marginal firm and hence earns zero profits. The figure on the right shows the equilibrium price as the intersection of the supply curve of the existing firms and the market curve.



H. Above, we illustrated the long-run equilibrium when we have non-identical firms. In order to illustrate the properties of this equilibrium and contrast it with the identical-firms case, we will now consider the short- and long-run effects of an increase in demand.

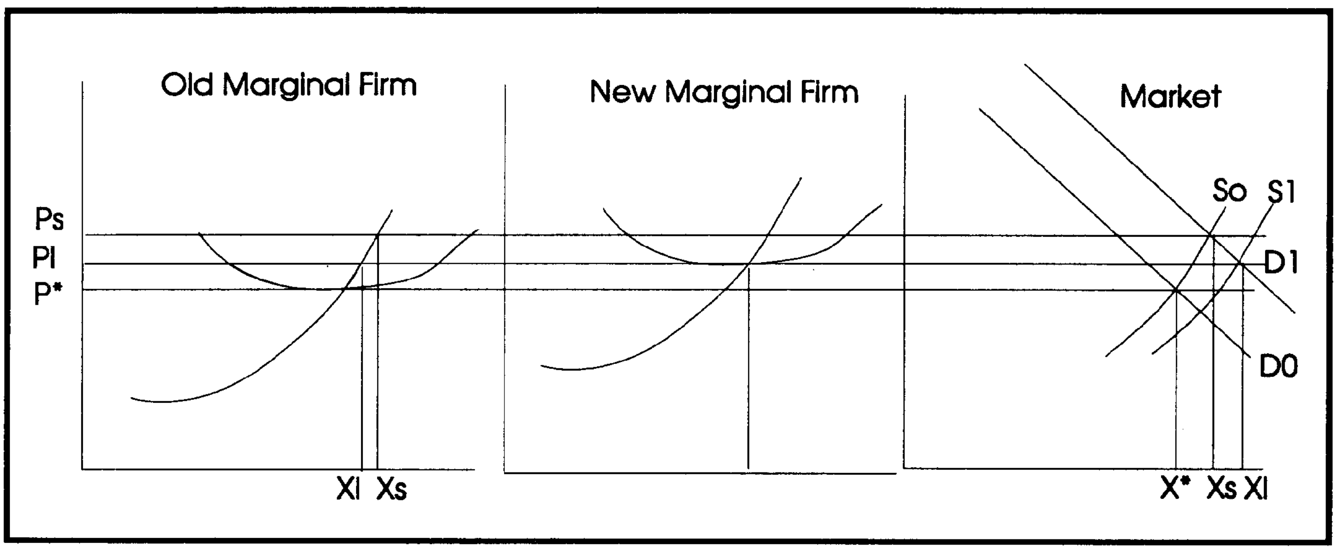
I. The short-run effect of the increase in demand is the same as in the identical-firms case. The rightmost figure below shows an increase in demand from the original demand curve D_0 to D_1. Since the short-run supply curve is simply the sum of the marginal cost curves of the existing firms, this leads to a rise in price and an increase in the quantity supplied. The effect of this on an inframarginal firm is shown in the leftmost figure below. The profits of this firm increase along with sales.



J. The effect on the marginal firm is similar and is illustrated in the middle figure above. This firm formerly earned zero profits in the long-run equilibrium and now is earning positive profits on an increased level of sales in the short run.

K. The fact that the marginal firm is earning positive profits will induce additional firms to enter the market. As they enter, the supply curve will shift to the right — further increasing the total quantity sold and decreasing price. Entry will stop when the market price has fallen below the minimum average cost level of firms still outside the market. The last firm to enter (the new marginal firm) will now earn zero profits and hence produce at the minimum average cost point. Since these entering firms have higher costs than the existing firms the new market price will still be higher than the original market prices (or else the new firms would not enter the market).

L. The new long-run equilibrium is illustrated in the three figures below. The new supply curve is shown as the solid line S_1 and the old short-run supply curve by the dashed line S_0 in the rightmost figure. The figure on the left shows the new marginal firm earning zero profits. The figure on the left illustrates the new position of the *old marginal firm*. This firm, which was earning zero profits before, now has increased sales and earns positive profits.



M. In the case of identical firms, the long-run supply curve was horizontal at the level of minimum average cost. The previous example of an increase in demand illustrates that in the case of non-identical firms, the long-run supply curve is upward-sloping. That is, the price must rise in order to induce additional firms to enter the industry and expand industry output.

N. The effects of a change in demand can be summarized in the following table:

Effects of a Rise in Demand on a Competitive Market with Non-Identical Firms			
	Short Run	Long Run	Long vs. Short Run
1. Price	increases	increases	decreases
2. Total output	increases	increases	increases
3. Output per firm	increases	increases	decreases
4. Profit	increases	increases	decreases
5. Number of firms	unchanged	increases	increases

XXVIII. A Note on Profits

A. The previous example of non-identical firms illustrated the fact that firms can earn positive economic profits even in the long run. That is, firms can have receipts in excess of the opportunity costs of their inputs even in the long run. Does this imply that these firms will earn accounting profits in the long run?

B. The answer is that they in general will not appear to be earning a greater than normal rate of return in the long run. A simple example will illustrate the idea. Assume that we purchase a wheat farm which provides us with exactly a normal rate of return on our investment. For example, we pay $100,000 for the land and earn a net return of $10,000 per year on our investment and the interest rate is 10%. In this case the $10,000 is exactly a normal rate of return on our investment of $100,000. In this case we now make zero accounting profits.

C. Now assume that there is an increase in the demand for wheat, which leads to an increase in the market price of wheat, even in the long run. We now earn a return of $15,000 per year on our investment of $100,000 and hence it would appear that we would now make $5,000 a year in profit (i.e., the $15,000 return minus the $10,000 opportunity cost or normal rate of return on our investment). However, this neglects the fact that the value of the land we now own will rise!

D. At a price of $100,000 investors would flock to buy the wheat land which currently provides a return of $15,000 per year for a 15% rate of return (i.e., the market currently yields 10% by assumption the interest rate equals 10%). However, this will bid up the price of the land until the $15,000 per year return represents a normal rate of return on the new higher price of the land. In this case the price of the land rises to $150,000, which makes the $15,000 per year a normal rate of return at the 10% market rate.

E. At the newer, higher price for the land we once again make zero accounting profits. This illustrates several important points. One of these is the idea of an **economic rent**. Since we were willing to use the land to grow wheat even for a return of $10,000 per year, the return on the land in its next best alternative (example, corn farming) must be less than $10,000, say $8,000 in this case. In this case the opportunity cost of the land (in another use) is $8,000 per year. At the same time the opportunity cost of the investment in the land, $10,000 per year, would be equal to the rental rate of the land.

F. Hence, the competitive rental price is $10,000 and the opportunity cost in its next best use is $8,000. The difference between the competitive price and the next best alternative (outside the current market) is referred to as *economic rent* or *quasi-rent.* Simply stated, the quasi-rent on an asset such as land is the rise in price above its next best alternative which is necessary to eliminate profits.

G. After the rise in demand and the price of the land rose to $150,000, then rental rates rose to $15,000 per year. Since the next best alternative is still $8,000, this rental rate is made up of $8,000 of opportunity cost and $7,000 of economic or quasi-rents. In this case accounting profits are eliminated by the fact that the economic profit (i.e., $15,000 - $8,000) is allocated to the accounting cost of the land.

H. This example illustrates that in equilibrium economic rents are allocated to specific assets. (I.e., the assets that allow us to earn the greater than market rate of return on the opportunity cost in another use. The specific asset in this case is the land, since it is what yields the $15,000 return in the production of wheat.)

I. From an economic perspective these rents are very different than costs. Costs affect the production decision of firms and economic rents *do not.* Economic rents are determined by the market price of the product and do not affect supply. This is easily illustrated using two different farms. Assume that one farm is the same as in our previous example, i.e., it yields a return of $15,000 per year in the production of wheat and has a next best alternative of $8,000 per year in corn production. On this plot we make zero accounting profits but earn $7,000 in economic profits. Consider also a second plot of land that also yields $15,000 per year in the production of wheat but has an opportunity cost of $15,000 in the production of corn. While both firms earn zero accounting profits, only the latter earns zero economic profits and hence only the latter is a marginal firm.

J. For example, if the price of wheat falls so that the return on both farms falls to $14,000 per year, only the second farm will no longer supply wheat. The first farm will continue to produce wheat and simply suffer a $1,000 loss in quasi-rents.

K. Land is not the only example of a specific asset that can earn quasi- or economic rents. Any asset that yields a rate of return greater than its opportunity cost in another use is a specific asset (i.e., its return is specific to this use) — e.g., patents, productive facilities, rights to production, highly productive managers, government licenses, etc.

XXIX. Monopolistic Markets

A. Up until now we have been discussing competitive markets, where firms can sell as much output as they wish to produce a given product price. At the other extreme of the spectrum is the monopolist who faces the market demand curve for his product. In this case the amount the monopolist chooses to sell has a direct effect on the price he can charge for his product. Instead of being able to produce as much as he wishes at a fixed market price, the monopolist is constrained to choose some price-output combination that lies on the market demand curve for his product.

B. Some typical reasons for monopoly are:

1. Government-granted monopoly, such as a utility.
2. Having a patent on a particular product.
3. Ownership of a very specific productive asset (e.g., a unique natural resource).

C. The basic objective of the monopolist is the same as that of a competitive firm, namely, to maximize profit. The monopolist will choose his output level, *x*, to maximize

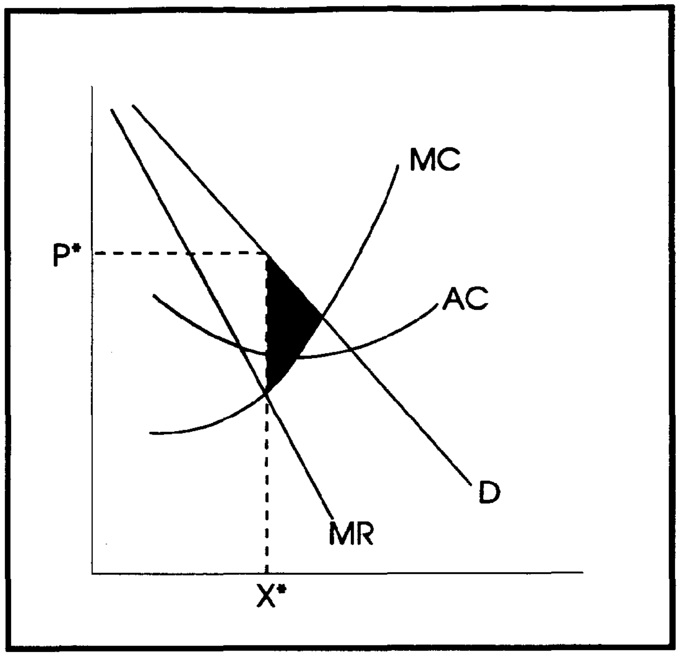
$$P(x)\,x - TC(x). \tag{1}$$

The notation P(x) is to remind us that the price and quantity the monopolist charges must lie on the demand curve. By definition P(x) x - TC(x) = Total Revenue in the market when *x* units are sold.

D. Differentiating with respect to *x* yields the equality that *marginal revenue equals marginal cost* as the optimal choice for the monopolist. This result is intuitive. The monopolist produces where his increase in revenues from an additional unit of output (i.e., marginal revenue) equals the increased cost of production (i.e., marginal cost). If marginal revenue exceeds marginal cost, then it would pay the monopolist to continue to expand output until marginal revenue equals marginal cost. Similarly, if marginal cost exceeded marginal revenue it would pay the monopolist to cut back output.

E. In a previous lecture we derived a relationship between marginal revenue and the elasticity of demand. This equality is simply MR = P(1 + 1/E), where E is the elasticity of demand. Since E is negative (i.e., a rise in price always leads to a fall in the quantity demanded) marginal revenue is *always* less than price. In addition, if demand is inelastic (i.e., 0 > E > -1), then marginal revenue is negative. Hence it will never pay the monopolist to produce in the inelastic portion of the demand curve.

F. The figure below illustrates the equilibrium position for a monopolist. Setting marginal revenue equal to marginal cost yields X_e as the monopolist's optimal level of output. At this quantity the demand curve gives the market price as P_e. At this price and quantity the monopolist's profits are equal to the lightly shaded area in the figure. Since there is no entry into the market, this represents the long-run equilibrium for the monopolist.

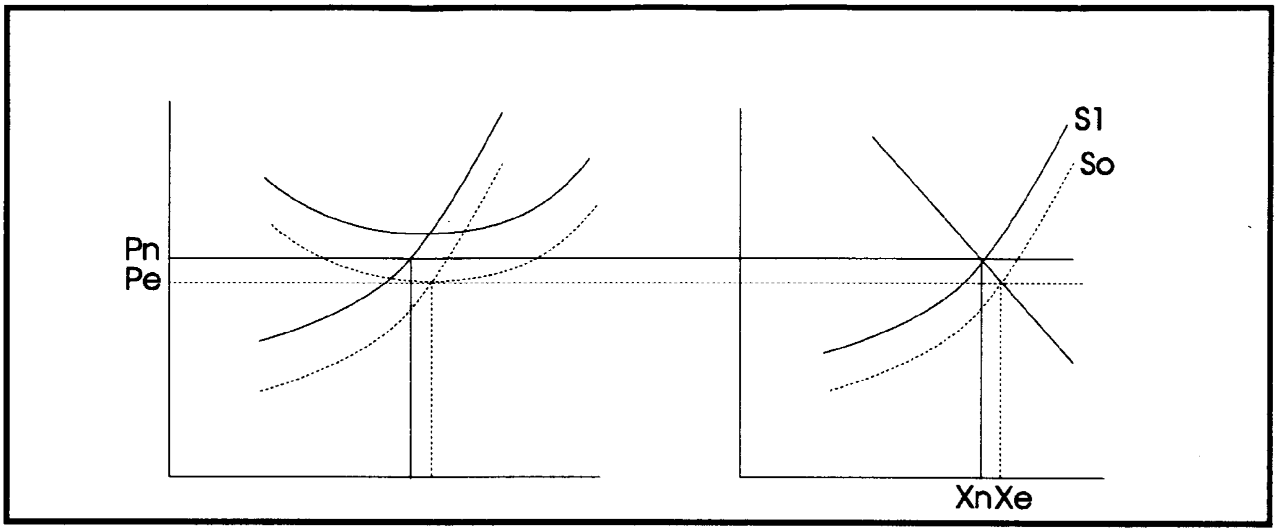


G. Economists frequently talk about the social costs of monopoly. These are seen as follows. Since the monopolist produces where marginal revenue equals marginal cost, and marginal revenue is always less than marginal cost, we must have price *greater* than marginal cost for the monopolist. This is illustrated in the bottom figure at left by the difference $P_c = MC_c$. Since the height of the demand curve equals the marginal value to society of an additional unit of output and the marginal cost of the monopolist represents the opportunity cost of production, the monopolist always produces where the social value of an additional unit of output exceeds the social cost. This represents a loss to society. In fact, the difference between the demand curve and the marginal cost curve for all units of output between X_c (the monopolist's output) and SC (the competitive output) represents a cost to society. For these units of output, the value to society exceeds the cost to society. Since the monopolist does not produce these units, the lost opportunities represent a cost to society. This social loss is given by the darkly shaded area in the figure.

XXX. <u>The Effects of a Tax</u>

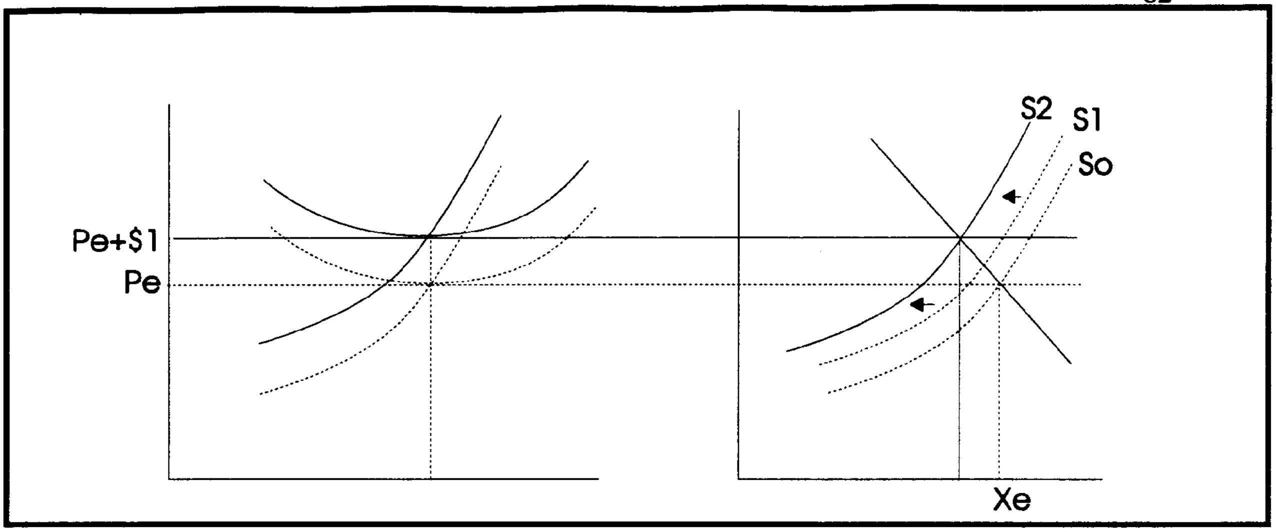
A. With the above models of competitive and monopolistic firms we can now consider the effects of a per-unit tax on the equilibrium.

B. To consider the effects of a tax on a competitive industry we simply consider its effects on the supply curve of firms. For simplicity we consider the case where the firm pays a tax on \$1 on every unit it sells. This \$1 per unit tax increases marginal and average cost of the firm by \$1 at every level of output. Since marginal cost increases by \$1, the supply curve of each firm (remember this is equal to its marginal cost curve) will shift up by \$1. The effect of this on the market price and on firm output and profits is illustrated in the two figures below for the case of identical firms.



C. The original supply curve is given by the dashed line S_0 in the figure on the right and the original cost curves by the dashed line in the figure on the left. The original price is P_e and the new equilibrium price is P_n. Note that since the demand curve is downward-sloping and the supply curve is upward-sloping, the price rises by ***less than*** $1. At this new price, current firms are making losses (i.e., they are not covering total costs). If there are no sunk costs, firms will immediately exit the industry and we will go directly to the long run, illustrated below. More likely there will be some sunk costs and the firms may not exist immediately since they may still be covering salvageable costs. (Remember sunk costs are sunk!) In this case the figures above will give the short-run equilibrium. The tax leads to an increase in price to consumers from P_e to P_n The quantity sold on the market declines from x_e to x_n; output per firm falls, as do firm profits from the imposition of the tax.

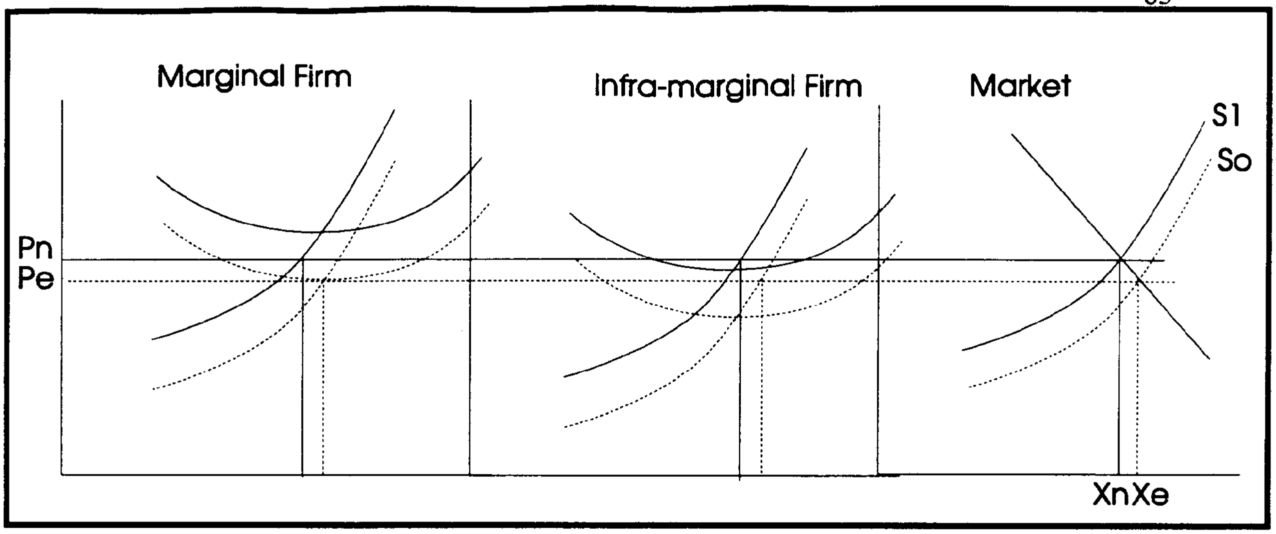
D. In the long run, firms will not reinvest in the sunk assets and firms will exit the industry until the price rises to cover total costs (including the $1 tax) and firms will once again make zero profits. Since minimum average costs are $1 higher than before the tax, the new long-run equilibrium price will also be higher than before by exactly the amount of the tax or $1. This new equilibrium is illustrated in the figures below. The new cost and supply curves are once again given by the solid lines with the pre-tax curves illustrated by the dashed lines. The shift in the market supply curve from So to S1 is the original effect of the tax, while the shift from S1 to S2 is the result of firms exiting the industry.



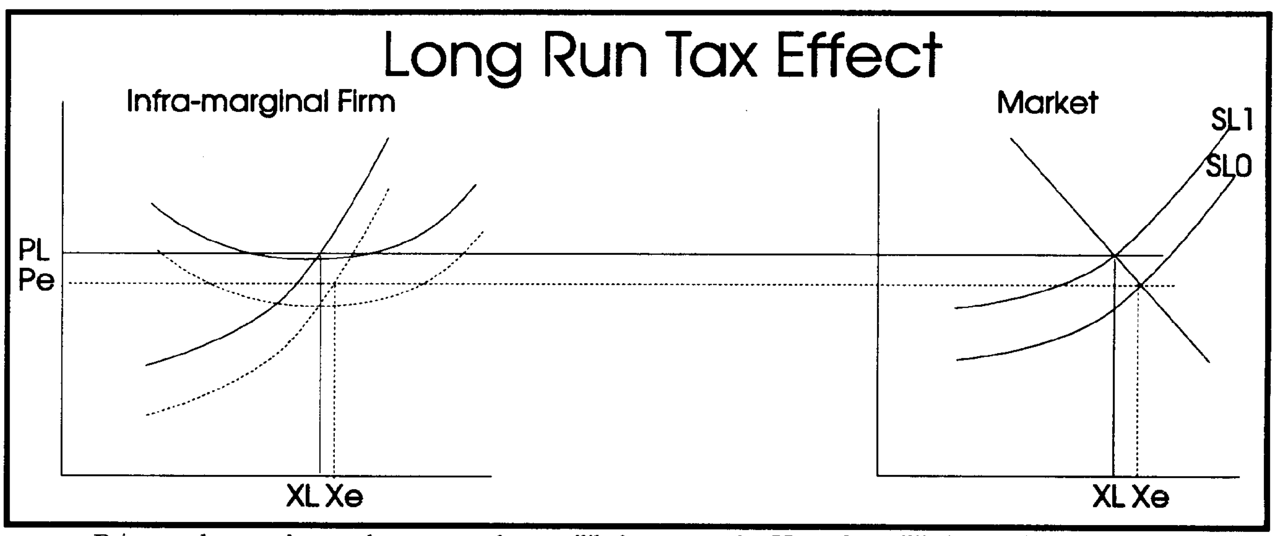
E. The result — that consumers will pay all the tax in the long run — can be easily understood. Since the long-run supply curve is horizontal with identical firms, the tax leads to a shift upward in the supply curve by $1. Since the supply curve is horizontal, the price will rise by exactly one dollar, as illustrated in the figure below.

F. In the case of non-identical firms, the short run is the same as with identical firms. However, since the long-run supply curve is upward sloping, in this case the price will not rise by the full $1 and firms will pay some of the tax even in the long run. Output and profits of the infra-marginal firm will be reduced by the imposition of the tax in the long run.

G. This is illustrated as follows. The tax shifts the marginal and average cost curves of the firms up by $1, just as in the identical firms case. This shifts the short-run market supply curve from S_0 to S_1 in the figure on the far right below. Again the curves that existed before the tax are in dashed lines with the new curves in solid lines. The market price rises from P_e to P_n and the quantity sold falls from X_e to X_n in the short run. The short-run effect on an inframarginal firm is illustrated in the middle figure. Profits and sales are reduced as compared with the pre-tax state. The old marginal firm now is no longer covering total cost as illustrated in the leftmost figure, and will leave the industry in the long run.

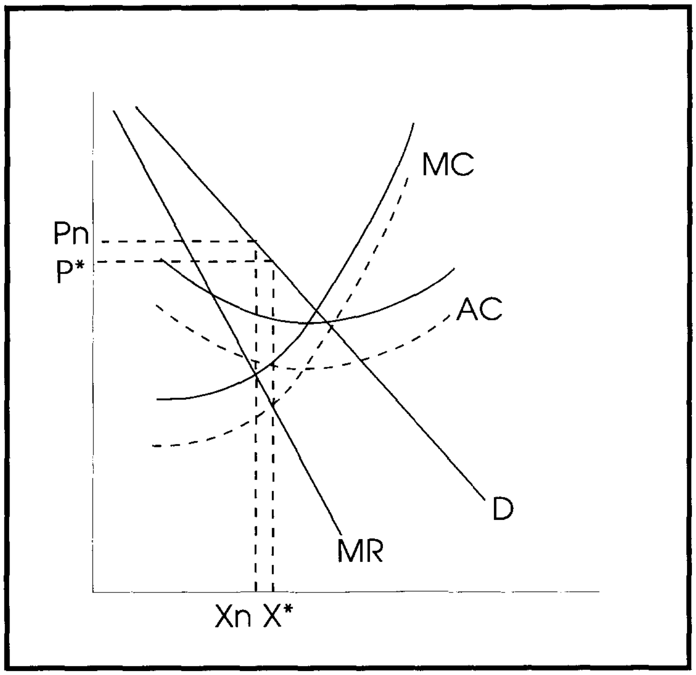


H. In the long run the firms that are marking losses will leave the industry and the price will rise until all firms left in the industry are no longer making losses. This can be seen using the long-run supply curve for the industry as illustrated in the figure at right.



Prior to the tax the market was at the equilibrium quantity X_e and equilibrium price P_e where the original long-run supply curve, SL_0, intersects the demand curve. The new long-run supply curve, SL_1, is \$1 above the old long-run supply curve due to the imposition of the tax. The price will rise to P_L in the long run. (Note P_L will be above P_n since the long-run supply curve is flatter than the short-run supply curve.) This will affect an inframarginal firm, as illustrated in the figure on the left above. The firm's output falls from the pre-tax level of X_e to the new long-run equilibrium level X_l and profits are reduced since the price increases by less than the $1 tax. Once again, the dashed lines give the cost curves before the imposition of the tax for reference.

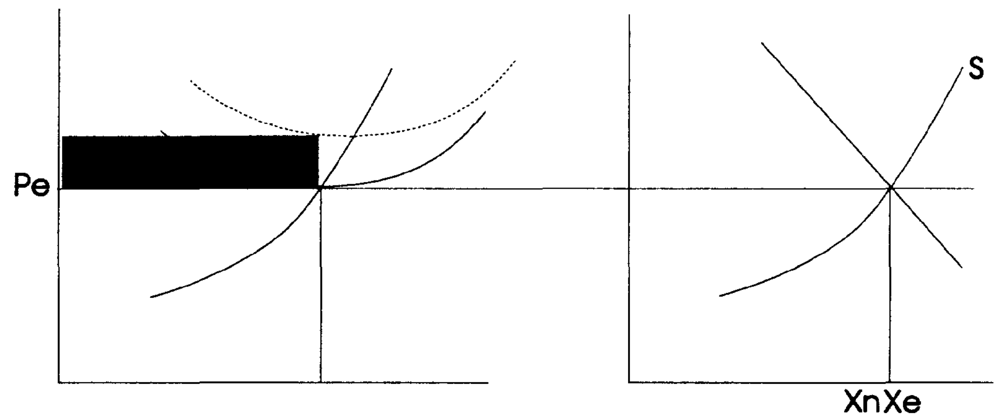
I. The effect of a per-unit tax on a monopolist is easily analyzed as in the figure at right. The tax increases the monopolist's marginal cost by $1. This causes him to reduce output from X^* to X_n where marginal revenue equals the new higher marginal cost. The reduction in output implies that the market price will rise, though one cannot say by how much. Hence, a per-unit tax leads to an increase in price and a reduction in the output of a monopolist. Since costs have risen, the profits of the monopolist will also be reduced.



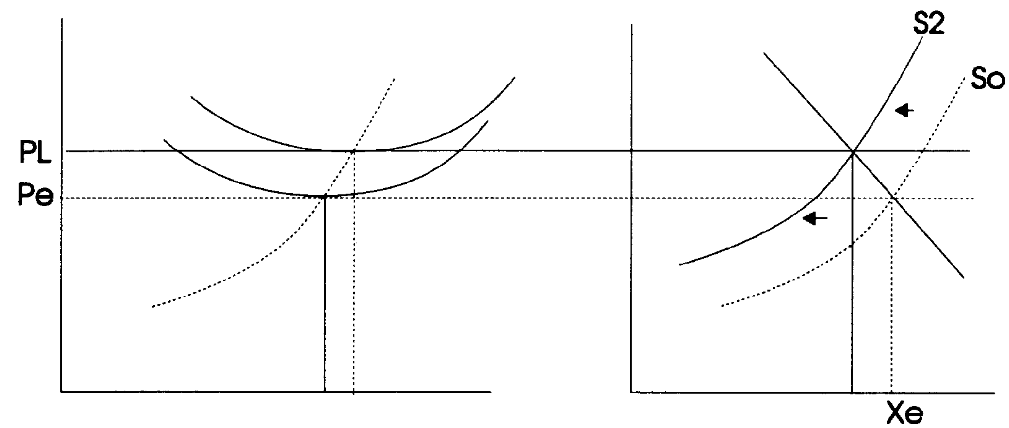
XXXI. Lump Sum Taxes

A. Earlier we considered the effects of a per-unit tax on a competitive industry and a monopolist. Another form of taxation is to tax each firm in an industry a fixed amount independent of output.

B. We consider first the case of identical competitive firms. In this case a fixed tax of $T per firm will not change marginal cost, since the tax is simply an additional fixed cost. Instead it will raise average cost at every level of output and leave marginal cost unchanged, as illustrated in the figure at right. Since marginal cost has not changed, the short-run supply curve will not shift. If the firms have no sunk cost, then firms will exit the industry and we will go to the long run discussed below. More typically, firms will have some sunk costs and the short-run equilibrium will result in no change in output or price. (Since marginal costs and hence supply have not changed.) This is illustrated in the figures below. The firm's new average cost curve is given by the dashed line in the figure below left. The market price is unchanged at P_c and the firm makes losses equal to the shaded area, which is equal to the tax, $T.



C. In the long run, firms will leave the industry until the price rises to the new higher level of minimum average cost, which includes the lump sum tax. This is illustrated in the figure below. The pre-tax long-run tax was P_e and the new long-run price will be P_L and firms earn zero profits. In this case consumers once again pay the entire tax. In addition, the figure also shows that the output of each firm will be larger after the lump sum tax.



D. The case of non-identical firms is the same in the short run as the identical firm case (i.e., there is no short-run effect). The tax causes the marginal firms to earn negative profits, which leads them to leave the industry (how fast depends on the amount of sunk costs) which leads to a rise in price until all remaining firms no longer make losses and the new marginal firm makes zero profits. *The tax lowers profits of all inframarginal firms* in the long and short run (though less in the long run).

E. The effect of a lump sum tax on a competitive industry can then be summarized in the following table:

Effects of a Lump Sum Tax on a Competitive Market with Identical Firms

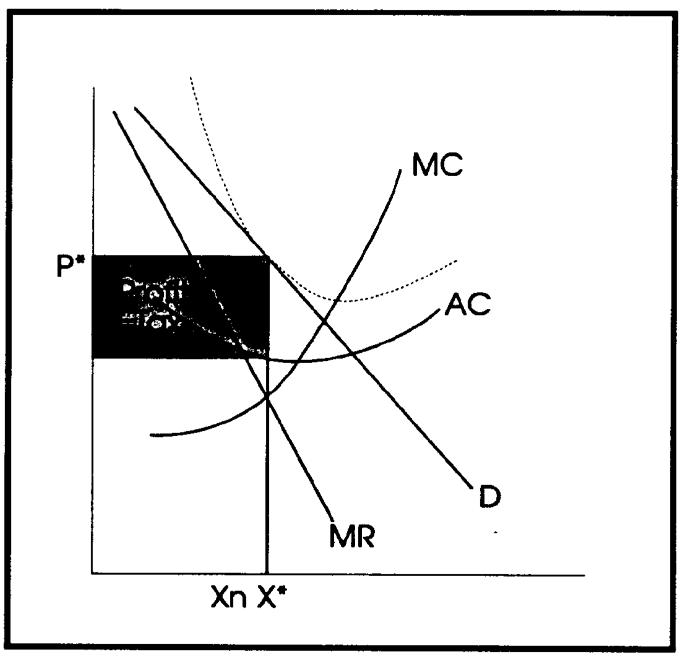
	Short Run	Long Run	Long vs. Short Run
1. Price	no change	increase	increase
2. Total output	no change	decrease	decrease
3. Output per firm	no change	increase	increase
4. Profit	decrease	no change	increase
5. Number of firms	no change	decrease	decrease

Effects of a Lump Sum Tax on a Competitive Market with Non-Identical Firms

	Short Run	Long Run	Long vs. Short Run
1. Price	no change	increase	increase
2. Total output	no change	decrease	decrease
3. Output per firm	no change	increase	increase
4. Profit	decrease	decrease	increase
5. Number of firms	no change	decrease	decrease

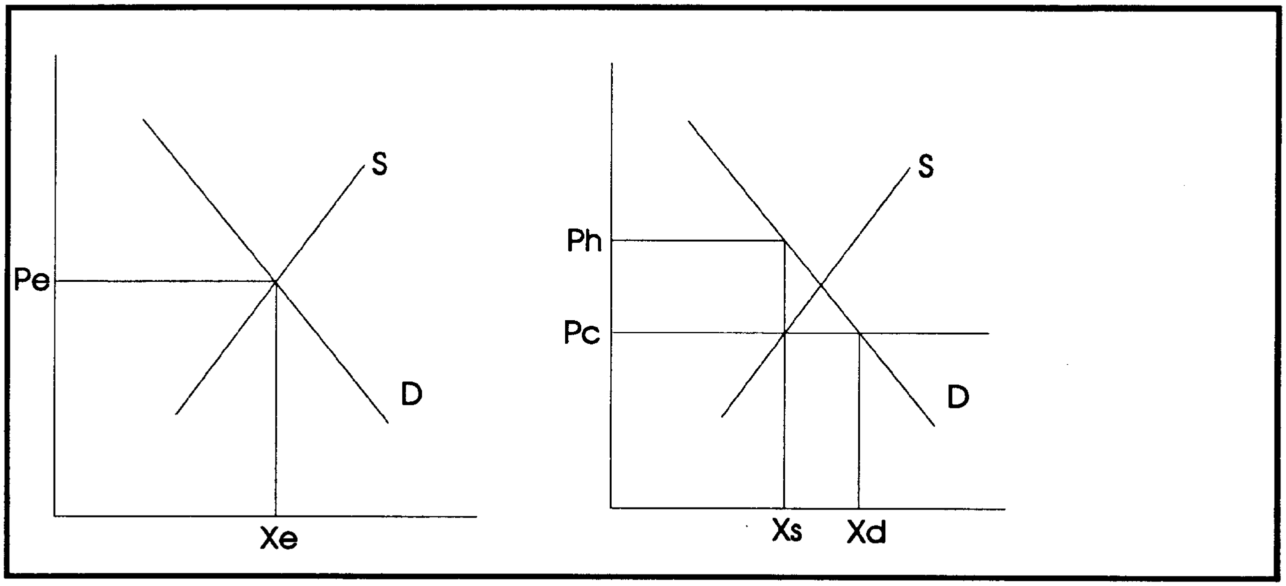
F. We can also analyze the effects of a lump sum tax on a monopolist in a similar manner. Since it does not affect marginal cost it will not affect the monopolist's choice of output. The tax will simply lower the monopolist's profits by the amount of the tax. If we attempt to tax the monopolist an amount greater than the monopolist's initial profit he will go out of business and will not produce (eventually, if there are sunk costs).

G. By taxing the monopolist an amount equal to his pre-tax profits, we can lower his profit to the competitive level of zero profits. Basically we simply increase his fixed cost until all profits are eliminated. This is illustrated in the figure below. Note that while we have eliminated the monopolist's profit, we have *not* eliminated the social cost of monopoly. In fact, since the lump sum tax does not affect output, the monopolist's choice of output and the social cost of monopoly are unaffected by the lump sum tax!



XXXII. The Effects of Price Controls

A. Using the tools of supply and demand we can also analyze the effects of a price control on a competitive market. At the intial equilibrium, supply and demand are equated at the long-run equilibrium price P_e in the figure below at left. Then a price control is imposed at a price of P_c below the equilibrium price, as illustrated below right. At this price the quantity demanded, X_D, exceeds the quantity supplied, X_S. Normally this shortage would put upward pressure on price, causing the price to rise back to the equilibrium level. In this case the price cannot rise and another analysis is called for.

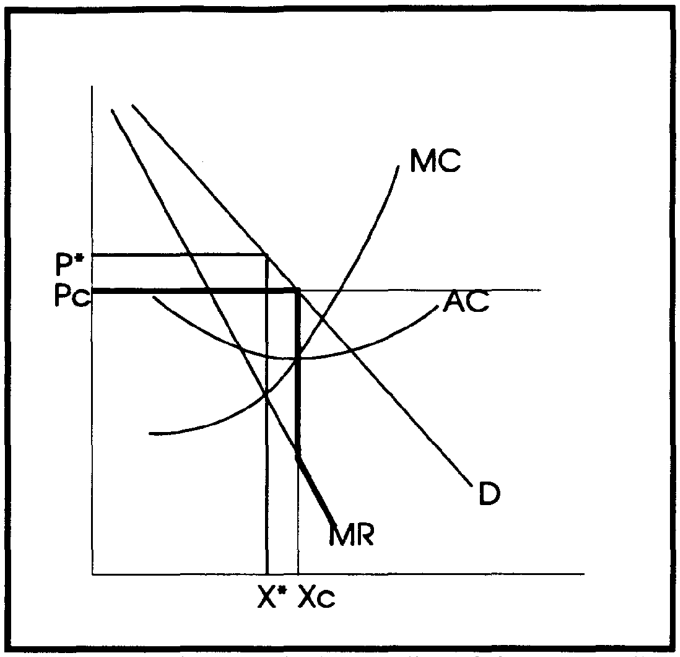


B. Given the price P_C, the supply curve tells us that firms will supply the quantity X_S. At this quantity, consumers are willing to pay P_h for an additional unit of output. Since the available supply will be only X_S, consumers will compete for the available supply by standing in line or other methods. In fact, consumers will continue to wait in lines until the full cost of the product, including time spent in line, equals the marginal value of an additional unit or P_h.

C. The net effect of the price control is to reduce the quantity available to X_S, the amount sellers are willing to supply at the controlled price. At this price the consumers compete away the difference P_h - P_C in time spent in line. The net effect of this price control will be to reduce the price to firms to the controlled price P_C and raise the price to consumers (including time spent in line) to P_h. In this case a price control leads to a *higher* full price to consumers and makes both consumers and firms worse off!

D. The story for the monopolist is somewhat different. In this case imposing a maximum price below the monopolist's optimal price is illustrated in the figure below. The monopolist's marginal revenue is now P_C, the controlled price, until output reaches X_C, the quantity demanded at the controlled price. Outputs beyond this imply a market price below the controlled price and hence the old marginal revenue schedule is applicable to the right of X_C.

In this case the monopolist's effective marginal revenue schedule is given by the heavy line in the figure below.



E. Setting marginal revenue equal to marginal cost will lead the monopolist to supply the quantity demanded at the controlled price in this case. This can be seen intuitively. The maximum price makes it impossible for the monopolist to charge a higher price even if he reduces the quantity he sells. Therefore, it will pay the monopolist to sell as much as he can at this price so long as the price exceeds marginal cost.

F. In fact, if we place a maximum price at the point where price equals marginal cost we can induce the monopolist to produce the competitive level of output. The maximum price makes marginal revenue equal to the competitive price, which causes the monopolist to act as a competitive firm would.

G. However, this is not the end of the story in the case of competitive firms or a monopolist. In the case of competition, firms have customers lined up to buy the product. Hence, they can sell even a somewhat lower quality product at the same price. This would simply reduce the length of the line with no reduction in the firm's revenues. The cost saving from producing the lower quality would then yield a higher level of profits. Hence, the price control will induce firms to ***reduce product quality*** in order to get around the price control.

H. Similarly, the monopolist would like to raise price above the controlled level. One way to raise the effective price is to reduce product quality. In so doing, the monopolist can increase his profits. Hence, the monopolist also has an incentive to cut quality in the face of a price control.

MONOPOLY PRICING: Part II

A. Another interesting case is where a monopolist sells output in two different markets. Denoting these as markets 1 and 2 we have the monopolist's maximization problem as the usual: maximize revenues minus costs, or

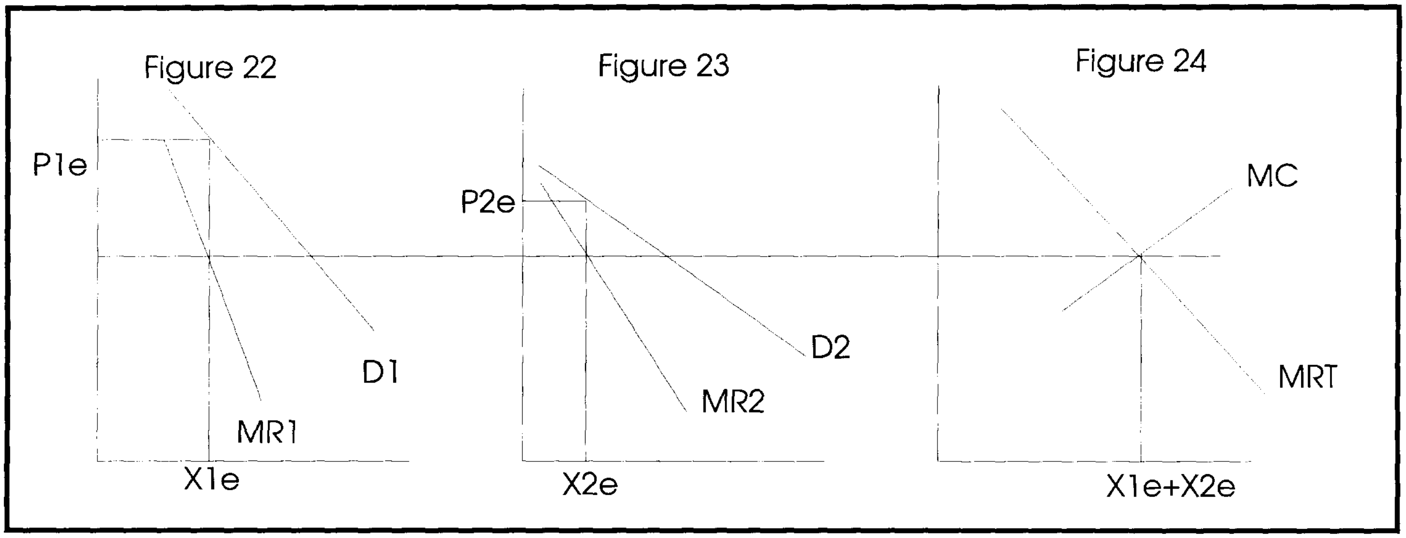
(2) $$\max TR_1(X_1) + TR_2(X_2) - TC(X_1 + X_2).$$

Here, $TR(X_1)$ gives total revenue in market 1 as a function of the quantity sold in market 1, X_1. $TR_2(X_2)$ gives total revenue in market 2 as a function of market 2 sales. Differentiating with respect to X_1 and X_2 yields (3a) and (3b):

(3a) $MR_1 = MC$
(3b) $MR_2 = MC$

The monopolist will produce where marginal revenue in each market equals the common marginal cost of production.

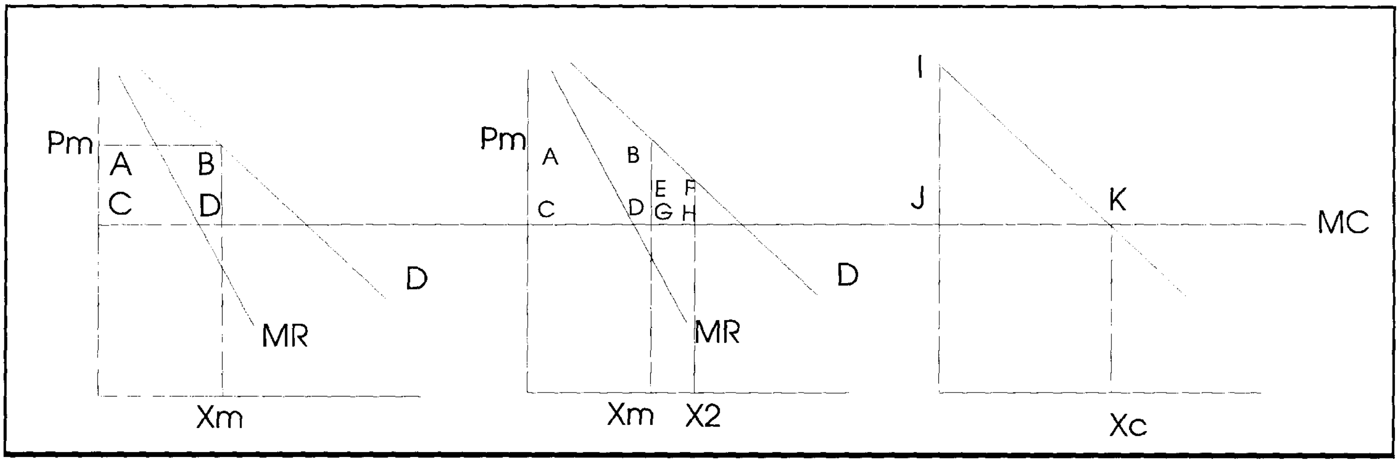
B. Will prices be the same in the two markets? The answer is NO in general. Earlier we used the equation that $MR = P(1 + 1/E)$, where E is the elasticity of demand. Since marginal revenue is equal across the two markets, prices will be equal if and only if the two market demand curves have the same elasticity. Otherwise it will pay the monopolist to sell at different prices in the two markets, charging a *higher* price to the market with the *less elastic* demand curve. Since increases in price imply smaller reductions in sales in the market with the less elastic demand, it will pay the monopolist to sell at a higher price in the market with the less elastic demand. The figures below illustrate this equilibrium. Here the monopolist sums the marginal revenue curves horizontally to obtain the curve MRT in the third figure.



C. The intersection of MRT with the marginal cost curve gives the equilibrium level of total output X_e and the equilibrium marginal revenue MR_e. This level of marginal revnue then yields the equilibrium quantities $X1_e$ and $X2_e$ together with the corresponding equilibrium prices $P1_e$ and $P2_e$. Note that the market with the less elastic demand is charged the higher price.

D. This is known as price discrimination. The monopolist is discriminating by charging one market a higher price than the other. The monopolist's profits are higher than if he was forced to charge one price in the two markets. Hence, it will always pay for the monopolist to try to divide the market into different groups which can be charged different prices.

E. There are other ways to price discriminate as well. One of these is by giving quantity discounts to individual customers. To see how this works we consider a simple example with one customer and a monopolist with constant marginal cost MC. If the firm acts as a simple monopolist then we will charge the price Pm in the first figure below, sell Xm units and make profits equal to the area ABCD.



In addition, the individual would be willing to purchase the additional units X2 – Xm in the second figure at price P2. If the monopolist makes this offer of additional sales he will now make the area EFGH in addition to the area ABCD in profits. This can be achieved by giving the consumer a quantity discount. The monopolist offers to sell up to Xm units at the monopoly price Pm and any additional units purchased at a price of P2. In this case the consumer will purchase a total of X2 units and the monopolist will earn profits equal to ABCD + EFGH.

F. However, the monopolist can do even better. He can do the best he possibly can by using a *two-part pricing scheme.* To achieve the maximum level of profits the monopolist can charge the consumer the area IJK in the third figure as a lump sum and then charge the consumer MC per unit. Here the monopolist charges the consumer marginal cost and thereby achieves the competitive level of output. The consumer is charged his entire consumer surplus (less some small amount to make the whole thing worth the consumer's effort) as a fixed fee. In this case the monopolist earns the entire area IJK as profits and captures all the consumer surplus.

G. An example where such pricing can be used is an amusement park. If the demand curve represents the demand for rides at the amusement park, then the firm can charge the area IJK as the fee to get into the park and charge MC per ride.

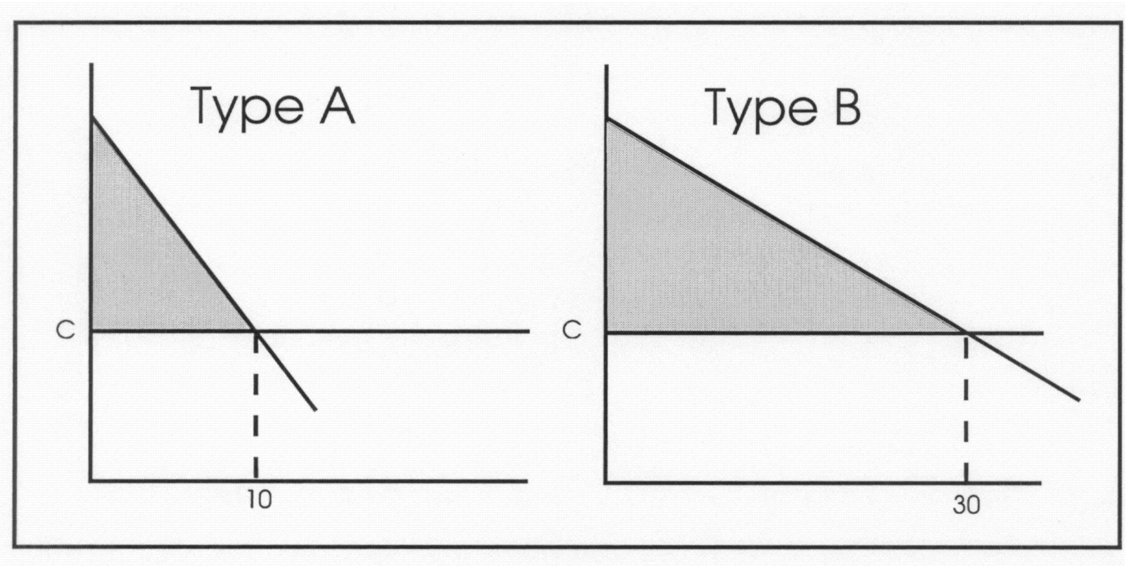
H. However, a major limitation to all price discrimination schemes is that people will attempt to get around them. For example, in the case of a two-market monopolist, he may not be able to charge very different prices in the two markets. If the difference between the two prices exceeds the cost of buying in one market and reselling in – or simply traveling to – the other market, then consumers will all buy in the low-price market and prevent the monopolist from charging very different prices.

I. Similarly, in the case of quantity discounts or two-part pricing, if resale is possible then it will pay one consumer to buy a lot of units and then resell them to other consumers and thereby save on the fixed fee or high cost of the initial units. This makes two-part pricing and quantity discounts possible for services and other items that cannot be resold but severely limits their applicability to items which are easily resold between consumers. In the amusement park case rides cannot be resold to people who do not pay the entry fee and the two-part scheme will be viable.

Monopoly Pricing II (continued)

J) The two part pricing paradigm is useful for illustrating a more general point: firms have an incentive to design their pricing structure in a way that raises prices to those with less elastic demand without raising prices to those with more elastic demand.

K) Consider a firm that sells to two types of customers (type A and type B). The firm sets two prices, a fixed fee, F, and a per unit charge, P. For simplicity, we will assume that the firm has constant marginal costs of C per unit. If the firm could price separately to the two types they would set the per unit price, P, equal to marginal cost, C, and charge two fixed fees (call them FA and FB) that would extract all of the consumer surplus from each of the two types. This strategy maximizes profits of the firm since the firm sells the efficient output to each type and captures all of the gains from trade. The picture is shown below:



As can be seen in the figure, type B customers buy more units (30 versus 10) and would be charged a higher fixed fee since the shaded area is larger for them. We denote the shaded area for type A by FA and the shaded area in figure B by FB. The firm's pricing structure would be a per-unit charge of C and fees of FA and FB to the two types.

L) What would happen if the firm could not charge different prices to the two types? If we stick with a unit price of P=C, then the firm has two real choices: charge a fixed fee of F_A and sell to both types or set a fixed fee of F_B and sell only to the higher demand type B's. If there are N_A customers of type A and N_B customers of type B then it would pay to sell to both types if $(N_A+N_B)\ F_A > N_B\ F_B$. We will assume that it pays to sell to both types in order to make the problem interesting. Under this pricing scheme the firm's revenues are

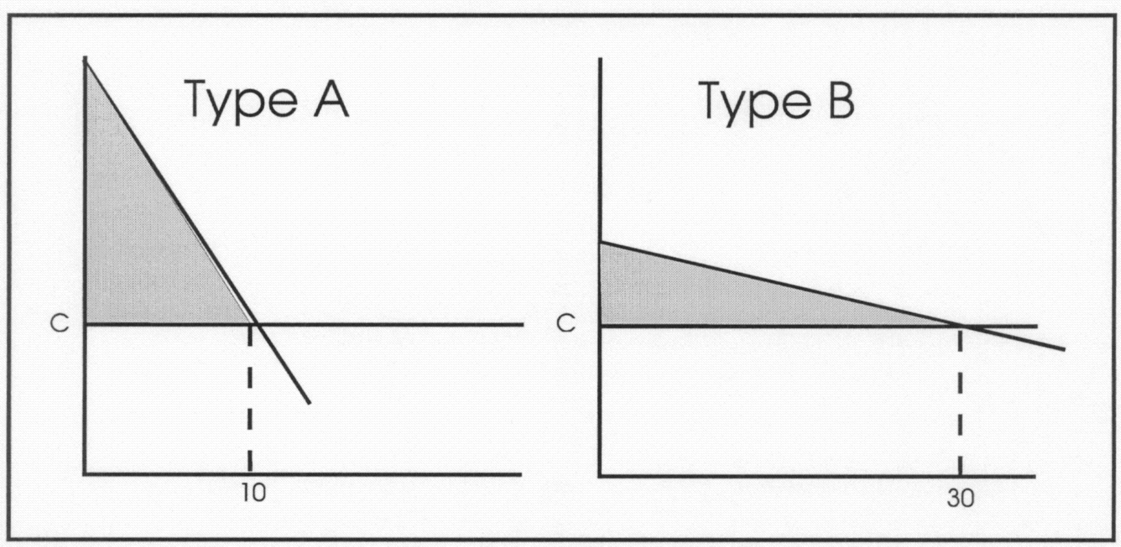
$N_A\ (F_A + 10\ C) + N_B\ (F_A + 30\ C)$ and its profits are

$(N_A+N_B)\ F_A$.

M) The question we wish to ask is whether it still pays to set the per unit charge equal to C. If the firm raises the per unit charge by some small amount Δ then it must reduce the fixed fee by 10Δ in order to retain the type A customers, since they are currently earning zero consumer surplus on the deal. This will leave the firm's profits from the type A customers unchanged (actually they will fall very slightly since the consumer will buy slightly less than 10 units). However, profits from the type B customers will rise since the 10Δ reduction in the fixed fee is more than offset by the 30Δ increase in revenues from the higher per unit charge. Hence, it will pay the firm to charge more than marginal cost on the unit charge.

N) The rationale for this result is fairly simple: the type B customers are willing to pay more overall and the firm, if it could, would charge them a higher fixed fee. When the firm cannot charge different fixed fees to the two types, it pays to raise the per unit charge since the type B customers buy more units; hence, charging a higher per unit price allows the firm to collect more from the type B customers than from the type A's. This is a general result. Firms will try to find ways to charge more to those who are willing to pay more. This involves up-charges on the goods that the high value consumers consume relatively more of and (as we will see below) down-charges on the goods they buy less of.

O) In some cases, it pays the firm to reduce the per unit charge below cost. Such a case is illustrated below. Once again, the type B customers buy more units. But now the type A customers are willing to pay a higher fixed fee. In this case the fixed fee will be limited to F_B (assuming it pays to sell to both types).



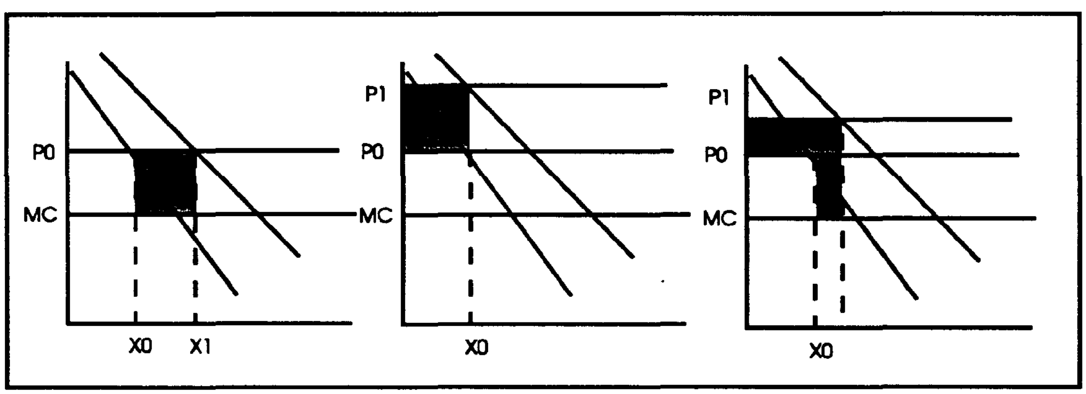
P) Now, if the firm reduces the per unit charge by Δ, they can raise the fixed fee by 30Δ since that will allow them to retain the type B customers. Once again, things are a wash for this group. But the firm now collects more from the type A customers (the high value group now) since the 30Δ increase in the fixed fee more than compensates for the 10Δ reduction in the per unit charges. Once again, profits are increased. In this case the firm profits by lowering the per unit price below cost since the high value customers are less intensive users of the good.

Q) There are numerous examples that fit this general paradigm. The most famous case is the IBM machine cards case, where IBM required users of its computers to buy computer cards from IBM at a price significantly above the competitive market price. The basic logic was that the number of cards purchased served as a meter of the customer's willingness to pay for the computer as a whole. Those who valued the computer more generally used more cards. A high per card charge allowed IBM to extract more from those willing to pay more without losing the low valued users as they would if they raised the rental price of the machine itself. One example of the down-charge is the case of amusement parks that often charge only a fixed fee and no charge per rides. The most logical interpretation is that the low valued (high demand elasticity) users are teenagers who actually go on more rides than the average customer.

Advertising

A) The demand models we have considered so far describe the demand for a product as a function of prices and income. However, other variables also affect the demand for a product. One of these variables is advertising. The simplest economic model of advertising simply adds advertising as an additional variable affecting consumer demand such as $X = D(P_X, A)$, where X is the quantity demanded, P_X is the price of good X and A is expenditures on advertising. For simplicity, we consider the case with constant marginal costs of production.

B) We can look at this problem graphically as shown below. An increase in advertising shifts the firm's demand curve outward. As shown in the figures below, the firm can earn additional profits by increasing sales at the same price (as shown in the left-hand figure) or by increasing price at a given quantity (as shown in the middle figure) or lead to some combination of an increase in price and quantity. The firm will keep the price constant when demand increases due to advertising if changes in advertising do not affect the elasticity of demand. The firm will increase price with advertising if advertising reduces the elasticity of demand and will actually reduce price if advertising makes demand more elastic.



C) Mathematically, we can examine the firm's optimal choice of advertising as follows: profits are given by

Profit = $P_X X(P_X,A) - TC(X(P_X,A)) - A$.

To find the profit-maximizing level of advertising we differentiate this expression with respect to A to obtain:

$(P_X - MC)\, dX/dA = 1$.

This can the be rewritten as

$(P_X - MC)/P_X\, dX/dA\, (A/X) = A/XP_X$.

The optimal pricing of a monopolist implies that $(P_X\text{-}MC)/P_X = -1/\varepsilon_X$. dX/dA (X/A) is the advertising elasticity of demand (the percentage increase in the quantity demanded generated by a 1% increase in advertising); which we denote by ε_A. A/XP_X is the advertising to sales ratio. This implies that the rule for optimal advertising is:

(Adv/Sales) = $-\varepsilon_A/\varepsilon_X$.

This formula shows that the firm will spend more on advertising when the advertising elasticity is larger and when the price elasticity of demand is lower. However, the incentive to advertise does not depend (as is sometimes alleged) on how advertising affects the elasticity of demand. In general, firms do not advertise in order to reduce the elasticity of demand for their product; they advertise in order to increase the demand for their product (whether it makes demand more elastic or less elastic).

We could formulate the same problem by looking at how advertising affects the willingness to pay at a fixed level of output. This would amount to looking at how the consumer's willingness to pay, P_X, depends on the quantity consumed, X, and the level of advertising. Such a function would be $P_X(X,A)$. In this case, the firm's problem becomes

Profit = $P_X(X,A)\, X - TC(X) - A$.

This optimal level of advertising then solves

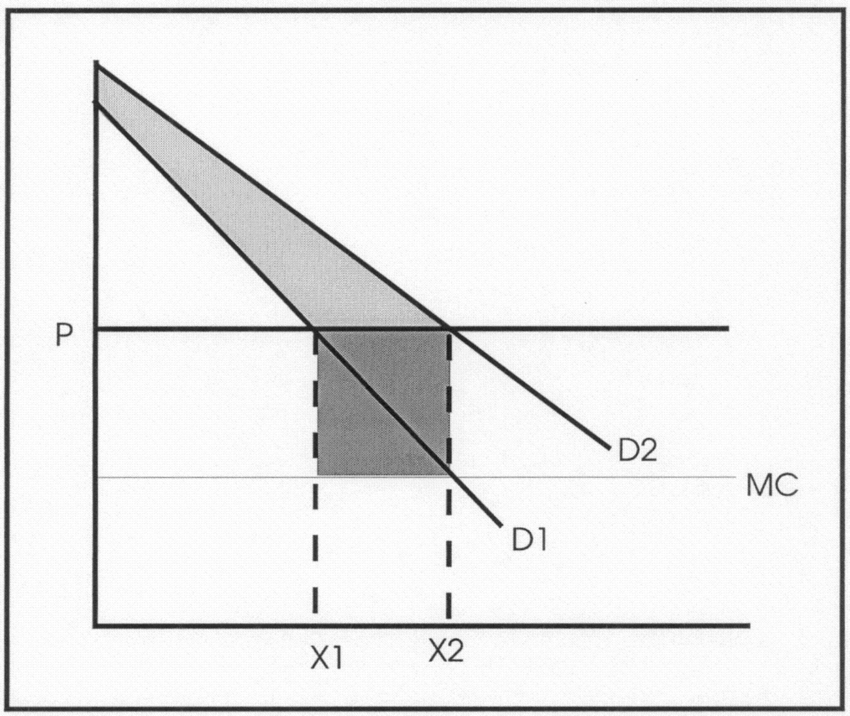
$dP_X/dA\; X = 1$, or

$dP_X/dA\; A/P_X = A/(XP_X)$ or

(Adv/sales) = ε_{PA},

where ε_{PA} is the elasticity of willingness to pay with respect to advertising. This formula is consistent with the one obtained before, since $\varepsilon_{PA} = -\varepsilon_A/\varepsilon_X$. Hence another way to express the optimal advertising condition is that firms will spend more on advertising when advertising has a greater impact on the consumer's willingness to pay. Note that the elasticity of demand does not enter this expression. As this formula makes clear, the gain from advertising depends on how much advertising shifts the firm's demand curve upward independent of the elasticity of demand. Hence, the elasticity of demand matters for the level of advertising only when we hold the effect of advertising on sales constant.

D) The effects of advertising on consumer welfare depend on how advertising affects price. Consider first the case where advertising affects the level of demand but does not affect its elasticity. In this case prices are unchanged and the picture looks like:



In the figure, advertising shifts the demand curve from D1 in the figure to D2. Since the elasticity of demand is unchanged and marginal costs are constant, the optimal monopoly price is unaffected. As a result, quantity increases from X1 to X2 and the monopolist's revenues increase by the darker shaded region. What about consumer welfare? Clearly, as shown in the figure, willingness to pay was increased by the advertising. Even with the increase in expenditures, consumer surplus grows by the lightly shaded region. Unless the advertising lowers utility substantially at zero consumption the greater consumer surplus would imply that the consumer has gained from the advertising. The advertising has increased the consumer's value of the good to the consumer and this generates increased surplus.

E) If advertising increases price significantly, the story can be different. In this case, the increase in demand increases consumer surplus (at a fixed price) but the increase in price reduces consumer surplus. If the price increase is large enough, consumer surplus and hence consumer welfare can decline. Advertising can also reduce price. Of course this makes it even more likely that consumers will gain.

Antitrust Economics

Background and Overview

A) The area of antitrust deals with a fundamental conflict in economics. On the one hand, competitive market forces drive firms to exploit the gains from trade and comparative advantage. Competition also motivates firms to create new and improved products and expand markets for existing products. Fundamentally, the competitive process is driven by greed. Greed is a powerful motivator and it generates a tremendous amount of good through the forces described above. However, greed also leads firms to try to fix prices, collude and/or monopolize an industry. It can also lead firms to use their market positions to exclude competitors or otherwise restrain competition. Antitrust policy attempts to deal with this more negative side of greed without interfering with its more positive aspects.

B) Broadly speaking, the area of antitrust economics examines the market forces and policies that affect the level of competition and thereby the efficiency of market outcomes. There is a wide range of opinion on the need for antitrust policy. Some see real world markets as dominated by monopoly, collusion and inefficient outcomes while others see the market as dominated by competitive forces that generate good outcomes for consumers and society. The truth of course is somewhere in between. The major policy question is not whether there should be antitrust enforcement but what is the proper scope for policy. I will say upfront that I am an advocate for a clearly articulated but limited antitrust policy.

C) The most important development in antitrust economics over time is the focus on <u>consumer welfare and economic efficiency as the goals for antitrust policy</u>. From an economic standpoint, the <u>*goal of antitrust enforcement is not to protect producers*</u> from each other but to enhance the efficiency of market outcomes. However, there is some tension between the consumer welfare and overall economic efficiency goals. In particular:

   Under the consumer welfare criterion, policy should be focused on what benefits consumers (i.e. what generates the most consumer surplus in our standard paradigm).

   Under the overall efficiency standard, policy should be based on what generates the most efficient outcome (i.e. what generates the most total surplus = consumer surplus + producer surplus).

D) Economic analysis typically argues for a focus on overall efficiency, but consumer welfare is the dominant standard in antitrust policy. In many cases the two criteria give the same answer but certainly not always. In some cases (see our discussion of price discrimination below) a focus on consumer welfare as the ultimate goal will lead us to focus on overall efficiency as an intermediate standard.

E) A direct application of the consumer welfare or the economic efficiency criteria would lead to a policy that always attempted to achieve the "best" outcome possible in each individual market or situation. However, real world policy must recognize that courts, economists and politicians are far from perfect. Our ability to analyze market outcomes and prescribe solutions that will improve those outcomes is limited. As a result, we must be cautious not to move in the direction of trying to tightly regulate market outcomes.

F) While the competitive process is not perfect, firms and consumers have a built-in incentive to move toward more efficient outcomes. Even the classic monopoly distortion drives firms to find more efficient pricing mechanisms that more greatly exploit the gains from trade. The profits generated by the monopoly and the forgone gains to trade create incentives to generate competing products and meet consumers' needs. In general, consumers and firms are likely to be better at finding solutions to many problems because they have something directly at stake (their individual welfare in the case of consumers or their firms' profits in the case of firms). Politicians, economists, and courts lack those same motivations.

G) If I had to provide one guideline for thinking about antitrust issues it would be:

> *Competition is a powerful force. Competition occurs at many levels. When competition is restricted at one level or in one dimension, firms and individuals will have an incentive to compete at other levels or along other dimensions. When profits exist or unexploited gains to trade remain, firms and consumers will find ways to compete for those profits and capture those gains to trade.*

H) Sometimes such competition will lead to inefficient outcomes, as when long lines ration consumption of a price-controlled good. But even in those cases, the market has an incentive to provide more efficient solutions through the development of a black market or substitution toward other goods that are not controlled or rationed.

I) Those that advocate an activist antitrust policy basically assume that markets will not be successful at solving many antitrust problems and that courts, economists and politicians will be reasonably successful at finding better solutions. Those that advocate a much more restrained approach (such as myself) place greater faith in market outcomes and considerably less faith in economists, politicians and courts.

J) The U.S. has an extensive body of antitrust law and several means of antitrust enforcement. The most important laws in the antitrust area are:

Law	Primary Focus
Section 1 of the Sherman Act	Price fixing and collusion
Section 2 of the Sherman Act	Monopolization
The Clayton Act	Anticompetitive mergers/business practices
The Robinson-Patman Act	Price discrimination

K) The primary avenues of antitrust enforcement are:

(1) The Antitrust Division of the Justice Department

(2) The Federal Trade Commission (and FCC for communications related matters)
(3) Private antitrust enforcement through both State and Federal Courts

L) Antitrust policy deals with a wide range of issues. We will discuss only a few of these and cover them only at a very general level. We will discuss collusion and price fixing, mergers, vertical restrictions, and price discrimination. In order to provide a background for the discussion we will first return to a comparison of competition and monopoly.

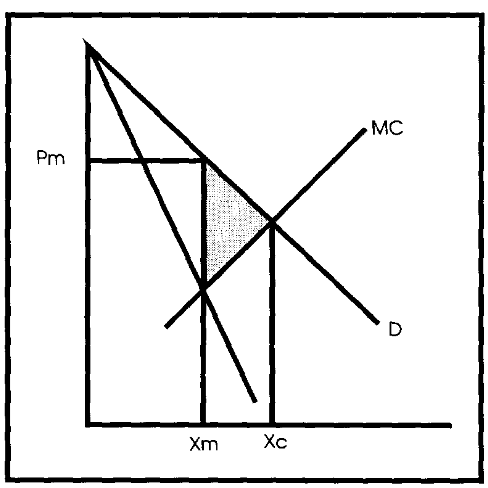
Competition and Monopoly

A) As we have seen throughout the course, competitive markets have many desirable properties. In general, **competitive markets are efficient** in that they equate the marginal costs of production with the marginal values of consumers. Competitive markets provide consumers with the products they desire at the lowest possible cost. By contrast monopolies tend to be inefficient (at least in a static sense); they tend to produce too little output and charge prices in excess of what we would see in a competitive market with the same cost structure.

B) While competitive markets are desirable, they are not always possible. Comparative advantage and economies of scale often limit some markets to just a few or even a single firm. In general, markets have fewer firms when there are greater cost differences between firms, the minimum efficient scale (the point of minimum average cost) is a large fraction of market demand and products are better substitutes.

C) **Limiting the ability of firms to exercise monopoly power is a major goal of antitrust law and antitrust enforcement**. In general, firms can attempt to obtain monopoly power in a wide variety of ways. Three important ways are through **(1) collusion, (2) horizontal mergers and (3) monopolization.**

D) How large is the monopoly distortion? Recall the picture:



The social loss from monopoly is equal to the shaded area in the figure. This distortion is roughly a triangle with one side of length P-MC and height (measured horizontally) equal to X_c-X_m. Hence the social loss in dollar terms is 1/2 (P_m-MC)(X_c-X_m) = 1/2 (P_m-MC)/P_m (X_c-X_m)/X_m * (P_m*X_m) = 1/2 (% of markup in price) (% reduction in output) (monopoly revenues). Hence for a monopolist with a markup of 20% who reduces output 20% below the competitive level we would have a monopoly distortion of roughly 1/2*.20%*.20% =2% of the monopoly revenues. In contrast, for a monopoly with a margin of 40% who reduced output by 40% we would have a distortion of 1/2*40%*40% = 8% of the monopoly revenues. ***In general, as these examples show, the social cost of monopoly is not large for even modest increases in price above marginal cost.***

E) In practice we seldom see true monopolists (i.e. producers that face no competition from other firms). Rather we tend to find firms that have "market power" which economists define by the ability of a firm to significantly affect prices in the marketplace. As you can tell from the definition, the existence of market power is often somewhat subjective. What is the marketplace and what constitutes a significant effect? These questions are the source of endless court battles and often form the core of the debate.

F) A model that is very useful for understanding market power issues is the **dominant firm model**. In the dominant firm model a firm with a significant market share faces competition from other suppliers. In particular we start with the market demand for the product D(P) and the supply curve of the other firms in the market $S^0(P)$. If we denote the quantity sold by the dominant firm by X we must have $D(P) = X + S^0(P)$. The demand facing the dominant firm is then $D^*(P) = D(P) - S^0(P)$. To measure the market power of the firm we need to determine the elasticity of this demand curve. To calculate this elasticity we need to calculate the effect of changes in the dominant firm's output, X, on the market price P. Looking at percentage changes we then have

$$\varepsilon_D \Delta_P = S_f \Delta_X + (1-S_f) \varepsilon_S \Delta_P,$$

where S_f is the market share of the dominant firm, ε_D and ε_S are the elasticities of demand and supply, and Δ_P and Δ_X are the percentage changes in the market price and the dominant firm's outputs. These equations imply that the elasticity of demand facing the dominant firm, $\varepsilon^* = \Delta_X/\Delta_P$, will be

$$\varepsilon^* = \varepsilon_D/S_f - (1-S_f)/S_f\, \varepsilon_S.$$

G) As this formula makes clear, the elasticity of demand facing the dominant firm will be smaller in absolute value (i.e. the firm will have more market power) when

(1) The market demand for the product is less elastic

(2) The firm has a larger share of the market

(3) The supply curve of other firms is less elastic

H) Example: $\varepsilon_D = -2.0$, $\varepsilon_S = 4.0$, $S_f = 40\%$. We then have

$\varepsilon^* = -2.0/.40 - .60/.40\ 4.0 = -11.0.$

I) Since both the elasticity of supply and the elasticity of demand are greater in the long run, the firm's ability to exercise market power (i.e. raise price significantly above competitive levels) may be greatly reduced in the long term. In fact, if the supply of competitors (ε_S in our model) is perfectly elastic, as it would be in the case of identical firms with free entry, the firm will have no market power in the long run.

J) Using our rule that $P = MC/(1+1/\varepsilon_D)$ we see that a monopolist in this market would set price at twice marginal cost (i.e. a markup of 100% over marginal cost) while for the same market a dominant firm with a share of 40% would price at only 10% over marginal cost. In contrast, if we were to increase the firm's share to 80% with these same elasticities we would have $\varepsilon^*=-3.5$ and get an equilibrium markup of 40% over marginal cost. If we decrease the firm's market share to 20% we get $\varepsilon^* = -26.0$ and a markup of only 4% over marginal cost. The table below lists the implied elasticity and the resulting equilibrium markups over marginal cost for various market shares of the dominant firm.

Demand Elasticity	-2.0
Supply Elasticity	4.0

Share	Elasticity	Markup
10%	-56.0	1.8%
20%	-26.0	4.0%
30%	-16.0	6.7%
40%	-11.0	10.0%
50%	-8.0	14.3%
60%	-6.0	20.0%
70%	-4.6	28.0%
80%	-3.5	40.0%
90%	-2.7	60.0%
100%	-2.0	100.0%

The key lesson from the table is that the relationship between market share and market power is very non-linear. In this example, even at a 60% market share, the markup is only about 20% as large as with a pure monopoly. At an 80% share it is only 40% of the monopoly markup.

K) We can use this analysis to study many issues in antitrust economics. In particular we can study

Collusion - Collusion can be modeled as an increase in the share of the dominant firm where the new share is equal to the sum of the shares of the competing firms. The effect of such collusion will be to increase price with no offsetting beneficial effects. In general economists take a very hard line on collusion.

Mergers - Mergers affect market power in much the same way as collusion. They generate an increased market share. In general we would expect this to lead to a somewhat greater monopoly distortion. However, we must net against this any increase in productive efficiency from the merger (so called "pro-competitive" benefits). These pro-competitive benefits may come about through synergies, more effective use of firm assets, lower costs of contracting etc. When the market shares of the merging firms are small the increases in market prices are also likely to be small. To know whether a merger is efficient we must compare the induced monopoly distortion to the improvement in efficiency. One key question is whether we should look at just the effect on consumers (i.e. consumer benefits) or the effect on overall efficiency (including both consumer and producer benefits).

L) While it is tempting to focus on the anticompetitive pricing effects of market power and monopoly there is another side. We often need to ask why a firm has market power. Monopolies do not occur by accident. Often, monopolies or an equilibrium with few firms is the outcome of competition between firms even if that competition does not lead to the classic competitive outcome with price equal to marginal cost. As such, these equilibriums often have at least some desirable properties. Sometimes, large market shares and prices above marginal cost result from a high degree of comparative advantage where other firms are unable to constrain prices charged by the dominant firm. In other cases, having a large dominant firm is efficient due to network effects (which generate economies from a large number of consumers using the same product or standard such as a software platform).

M) In other cases, market power results from large fixed costs and low marginal costs. Taken together, these aspects greatly limit the number of competitors that can survive in the market. Often such markets will have price roughly equal to average cost even though prices remain above marginal costs.

N) Often, market power and even monopoly positions result from patent and copyright laws that give exclusive rights to a particular firm. These policies trade off the losses from monopoly pricing against the gains from improved incentives for research and development. Protecting intellectual property sometimes involves tolerating some market power after the fact.

O) In each of the cases described above, the equilibrium with market power represents a trade-off. The economy is better off with one firm having low costs (i.e. comparative advantage) than all firms having a higher level of costs. In cases with large fixed costs and low marginal costs, the equilibrium with price equal average cost is not first best. However, improving on this equilibrium would require subsidies to firms or restrictions on entry. Few economists would advocate either policy.

Market power generated by patents and copyrights represent an explicit trade-off between innovation and market distortions. This is more general; firms will compete up front for the profits generated by monopoly through efforts to produce the products and innovations that generate monopoly power. Hence, monopoly profits are not all bad. Some economists, such as the famous economist Joseph Schumpeter, even argue that it is the pursuit of monopoly that motivates much of our progress.

P) Many of the most durable monopolies are granted or perpetuated by governments themselves. Governments grant monopolies to local providers of gas, electricity, even phone service and cable television and restrict entry into many arenas. Often these industries are declared natural monopolies (industries where competitive forces are commonly asserted not to work or to lead to inefficient outcomes). Unlike the monopolies described above which resist displacement by market forces (and hence typically have some inherent advantages) government granted monopolies often do not arise from any particular efficiency but from a desire to benefit individuals or particular groups.

Collusion

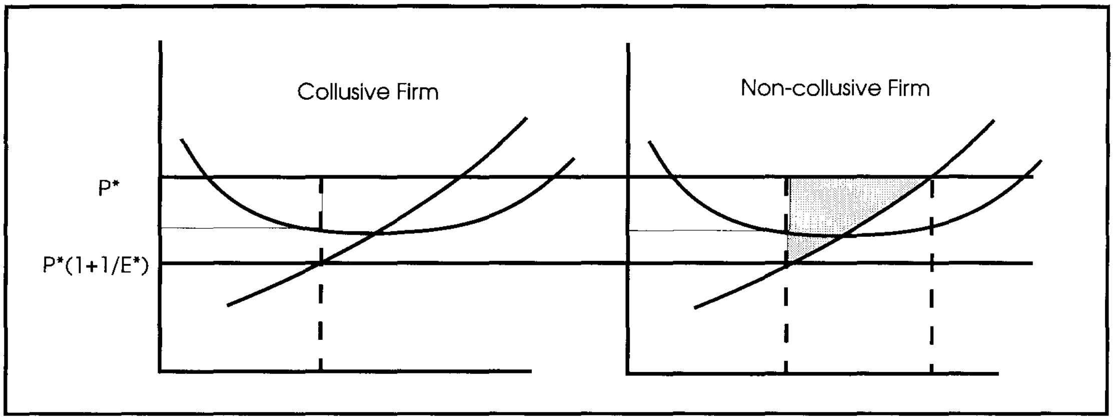
A) Adam Smith recognized that when sellers get together the conversation inevitably turns to how they can raise price. George Stigler, one of the greatest industrial organization economists of the 20th century, recognized that the incentive to raise price above competitive levels was universal but that the ability to do so was much more limited. In his view, we should focus our attention on those markets where collusion is likely.

B) In a collusive arrangement, two or more firms get together in an effort to raise market prices above competitive levels. Since market demand slopes down, increasing the market price requires these firms to reduce market output. As in the case of the dominant firm, the elasticity of demand facing the colluding firms (once we account for the increased production by their competitors) is

$$\varepsilon^* = \varepsilon_D/S - (1-S)/S\ \varepsilon_S,$$

where S is the equilibrium market share of the colluding firms and ε_S and ε_D are the elasticities of supply and demand. As we saw in the case of the dominant firm, the profitability of reducing output is greater the smaller is this elasticity. As Stigler recognized, this equation says that collusion is more profitable when demand is less elastic (at the competitive price), when the supply of other firms is less elastic, and when the colluding firms have a larger share of the market.

C) However, while producers always have the incentive to collude, collusive agreements contain the seeds of their own destruction. To illustrate the incentives involved in a collusive agreement we start with a competitive industry that has a fixed number of identical firms. A fraction of these firms, f, collude in an attempt to raise their profits. As they reduce their output, prices will rise and the firms outside of the collusive agreement will expand their output. As a result of the collusive firms decreasing their production and the firms outside of the agreement increasing their output, the equilibrium share of the colluding firms, S, will be less than f. The optimal collusive price for these firms will satisfy $MC = P(1+1/\varepsilon^*)$, where ε^* is as defined above. At the optimal collusive price, the situation for the colluding firms and the firms outside of the collusive agreement will look like:



As the figures show, the collusive firms now produce less output than under competition but now make profits (shown by the lightly shaded box in the left hand figure). The non-collusive firms also make profits. In fact, the non-collusive firms make more profits than the collusive firms since they sell at the same high price but are able to expand rather than contract their output. Hence, while firms gain from being part of the collusive group, they gain even more from being outside of the group. Indeed, this disparity increases the more the collusive firms reduce output since an even higher price further increases output for the non-collusive firms and further increases profits while those inside the agreement must further restrict output. The extra profits made by the non-collusive firms are measured by the darker shaded region in the right-hand figure (equal to the gap between price and marginal cost for the additional output the non-collusive firms produce).

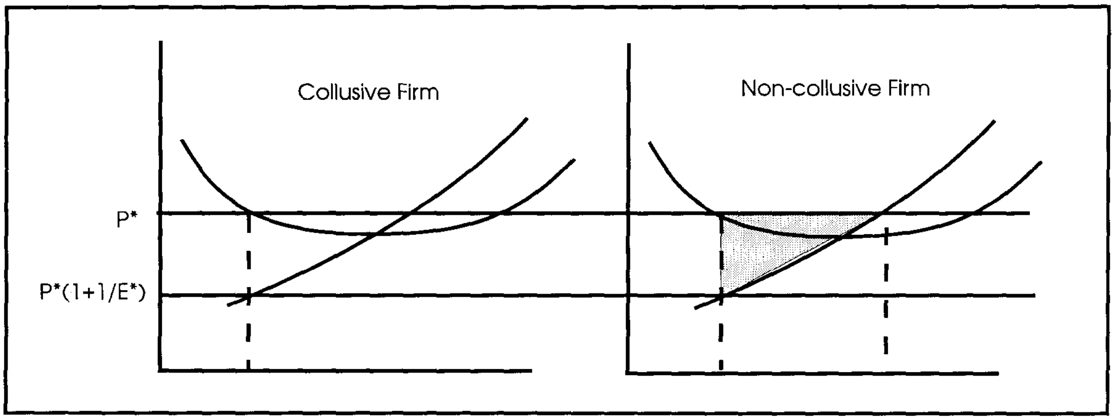
D) The greater profits earned by the non-collusive firms creates the classic problem for collusive agreements. It pays firms to cheat and the gain to cheating rises faster than the returns to maintaining the agreement. This makes price fixing agreements (particularly agreements that raise price significantly) difficult to sustain.

E) As Stigler recognized, any such agreement must have some means to induce individual firms to join the agreement and prevent firms that join the agreement from cheating. Getting individual firms to join the agreement is difficult since they can potentially earn more by operating outside of the agreement. To get them to join it is necessary to convince them that the agreement will not go forward without them.

F) Policing such an agreement requires that firms are able to detect and punish firms that cheat on the agreement. Collusion will be more likely when it is easier to detect cheaters (i.e. firm level prices and/or production are more easily monitored). Maintaining and enforcing an agreement is likely to be easier when there are fewer firms (i.e. each firm has a larger share of the market) and when prices and output are easily monitored (e.g. products are homogeneous and prices and quantities are publicly observable).

G) The problem of punishing cheaters is much more difficult since antitrust laws make court enforcement of collusive agreements illegal. The most common method of enforcement is the threat that the other members will no longer go along with the agreement if one member cheats or the threat of retaliatory price cuts. However, in many cases such threats will not be credible since it is not clear how the firms' incentives to collude or cheat in the future are affected by one firm cheating today.

H) Even if the colluding firms are successful at detecting and preventing cheaters, the profits generated by collusion are unlikely to persist in the long run. Entry by additional firms and expansion by those firms that are outside of the collusive agreement will force these firms to further reduce output and/or reduce the collusive price. In the extreme case where all potential firms have identical cost structures, entry will continue as long as profits are positive. Even if the entering firms are absorbed into the agreement, output per firm and prices will continue to fall until profits for those in the cartel are driven to zero (i.e. even the collusive price generates zero profits). Such a solution would look like:



At this point, firms would still desire to enter the non-collusive sector, making profits for the collusive firms negative. Obviously, the agreement would have long since broken down since there is no reason to stay in the agreement in the long term. The lack of long-term benefits would tend to make the agreement unravel even earlier.

I) Given the incentives to cheat on the agreement in the short run and the propensity for entry to erode profits in the long run, most economists are skeptical of the ability of most collusive arrangements to survive. Even groups like OPEC that have explicit meetings and are not subject to antitrust enforcement have a difficult time keeping their members in line. Cheating by members and the expansion of supply by countries outside of OPEC has caused OPEC's prices to erode over time. Collusion by firms in domestic markets is even more difficult since they must deal with antitrust enforcement as well as market forces.

J) Cases involving explicit collusive agreements are rare these days. Conspiring to fix prices is per se illegal. That is, it is illegal to make an agreement to fix prices whether the agreement is successful or not. Discovering an actual written agreement or direct evidence of an oral agreement is relatively rare but such cases do exist. The most egregious cases sometimes lead to criminal prosecution of those involved.

K) Many cases lack direct evidence of the agreement itself. Often, they have indirect evidence of the potential for an agreement (such as meetings between the parties allegedly involved in the agreement) and a pattern of behavior that suggests the existence of an agreement. Such analyses rely on finding ways in which individuals behave that would not be in their own self-interest absent an agreement. Such an inference obviously requires accurate knowledge of what is in the individual's self-interest. Such knowledge is often difficult to obtain. This is greatly complicated by the fact that the law allows firms to take account of the reaction of their competitors when setting prices or production levels. For example, not cutting price for fear that your competitor will follow and make the price cut unprofitable is NOT a violation of the law. Not cutting price because it violates an agreement not to do so would be illegal. Obviously telling these cases apart is difficult.

L) Since direct or even indirect evidence on agreements is rare, antitrust enforcement often focuses on eliminating situations and business practices that facilitate collusion. These are often referred to as facilitating devices. Examples of such devices are (1) private communication between competitors (particularly the communication of pricing and sales information); (2) communication of pricing and/or sales information through industry trade associations or other third parties, and (3) coordination of actions and alignment of incentives through joint ventures or other business relationships between horizontal competitors.

M) In general, economists take a hard line on collusion. As both Smith and Stigler emphasized, the incentive for firms to collude is ubiquitous and there are no significant benefits from collusion. Fortunately, the ability of firms to successfully do so is much less common. The incentive to remain outside the agreement, the incentive to cheat on the agreement by those who join, and the erosion of profits through entry are market forces that work to limit the success of collusive agreements. Successful collusion is most likely when:

(1) There are few firms
(2) Entry is difficult
(3) Demand is relatively inelastic
(4) Supply is relatively inelastic
(5) Prices and quantities are easily monitored

N) A prudent antitrust policy will focus attention on collusion in such cases and deal harshly with cases of explicit collusion. Inferring collusion from behavior is much more difficult. Perhaps the best policy is to limit facilitating devices that do not have a good pro-competitive justification when the other precursors for collusion are present.

Mergers

A) Mergers are common in U.S. markets. Typically we divide mergers into three categories: horizontal mergers (mergers between firms that compete in the same product market), vertical mergers (mergers between a firm and an upstream supplier or downstream customer), and conglomerate mergers (mergers between firms that have no close ties). Antitrust scrutiny is toughest for horizontal mergers, somewhat less for vertical mergers, and very light for conglomerate mergers.

B) The largest concern in most ***horizontal mergers*** is the enhancement of market power. In order to measure both the level and increase in market power we must first define a market. The merger guidelines define a market by the following hypothetical experiment: we start with the merging firms and then expand the market (expanding geographically or by adding additional products) until we get to the point where a hypothetical monopolist would profitably raise the market price 5% above its current level. Since this is a hypothetical experiment, there is frequently great disagreement over the answer. In general, those arguing for the merging firms argue for a wide market definition while those opposed to the merger argue for a narrower one.

C) Once the market is defined, the courts typically look at three things: (1) the shares of the merging firms, (2) the elasticities of supply and demand for the market, and (3) the ease of entry into the market. As we know from our analysis above, mergers are more likely to lead to an anticompetitive outcome when the merging firms have larger shares of the market, supply and demand are less elastic, and entry is more difficult.

D) Since horizontal mergers will typically increase market power (at least a little), you might wonder why such mergers are allowed at all. The basic answer is that mergers can have pro-competitive benefits by increasing the efficiency of the combined operation. When the merger involves two relatively small firms, the merger is typically not subject to much (if any) antitrust scrutiny because there is no effect on market power and hence the merger is assumed to be driven by efficiency gains. The amount of scrutiny increases when the firms have larger market shares or other conditions of the market make anti-competitive effects more of a concern. Mergers in this intermediate range must demonstrate a substantial ***pro-competitive benefit*** that results from the merger. Typically, this would involve more efficient use of assets or extending the use of a new technology or product to a wider market. The presumption is that such efficiency enhancements will expand rather than contract output and benefit rather than harm consumers.

E) One way in which pro-competitive and anti-competitive horizontal mergers differ is in their effects on competitors. Generally, anti-competitive mergers that enhance market power will benefit competitors as the merged firm restricts output and increases the market price. On the other hand, pro-competitive mergers will typically lead the merged firm to expand output or improve its products, which will harm the firm's competitors. Hence, when competitors complain about a horizontal merger it is likely that the merger is pro-competitive rather than anti-competitive.

F) Vertical mergers (i.e. mergers between a firm and its suppliers or its downstream customers) typically involve less of a direct concern about market power. Typically, the concern is an indirect one. How will the merger of a firm and its upstream supplier or downstream customer affect competition between the firm and its horizontal competitors. The most common concerns are that the firm will use these vertical relationships to exclude its rivals from downstream markets or deny its rivals access to key inputs.

G) In spite of the initial appeal of such foreclosure arguments, making a convincing case is often much more difficult. The basic problem is that if there are profits to be had from denying competitors access, why wouldn't the downstream or upstream firm exercise that power before? Indeed, the upstream or downstream firm can often extract even more than the combined firm by threatening to deny access to the firm and its competitors. Instead, any exclusionary gains typically come from the interaction of what the firms can do at the different levels.

H) There are also many efficiency reasons to vertically integrate (see below for a discussion of vertical contracting issues). Due to the relatively large number of potential efficiency gains and the reduced concern over anticompetitive gains most economists take a much more permissive stance on vertical mergers than on horizontal mergers.

Vertical Restrictions

A) While horizontal relationships (i.e. relationships between competing firms) raise significant issues of collusion, **the outputs of most vertical relationships are complements** rather than substitutes (i.e. a lower price for steel raises profits in the automobile business and lower costs of retailing benefit a manufacturer). As a result, most such agreements are designed to lower rather than raise the costs of these services. In general, economists look much more favorably on vertical contracts (including price setting contracts) than on contracts between horizontal competitors.

B) The most common vertical relationships are between a producer and the downstream retailer or distributor of its products. Some of the most common vertical restrictions studied be economists are:

Exclusive Territories: where a producer grants a downstream marketer an exclusive right to market in a geographic area or to a particular market segment.

Exclusive Dealing: where a producer requires a downstream marketer not to market other products

Vertical Ties: where a producer requires the downstream marketer to take one product as a condition for being able to sell another

Full Line Forcing: an extreme version of tie-ins where a producer requires a downstream marketer to carry the producer's full product line

Resale Price Maintenance: where the producer set either a minimum or maximum price at which the retailer can resell the product to consumers

C) One could teach an entire course on vertical restrictions. Obviously a complete treatment is far beyond what we can hope to accomplish here.

D) Perhaps the best place to start is the equilibrium without any vertical restrictions. In the simplest case, selling a unit of the producers product requires one unit of the product and one unit of what we will call "retailing services." If we assume that retailing is perfectly competitive with identical suppliers and that the minimum unit costs of retailing services are C_R then a upstream producer will face the problem of choosing the optimal level of output (or equivalently the optimal wholesale price). The basic setup would be:

Sales of the product = X
Retail Price = $P_R(X)$, we allow for the retail price to depend on output (i.e. downward sloping demand)
Wholesale Price = P_W
Retailing Cost per Unit = C_R.
Production Cost = C(X).

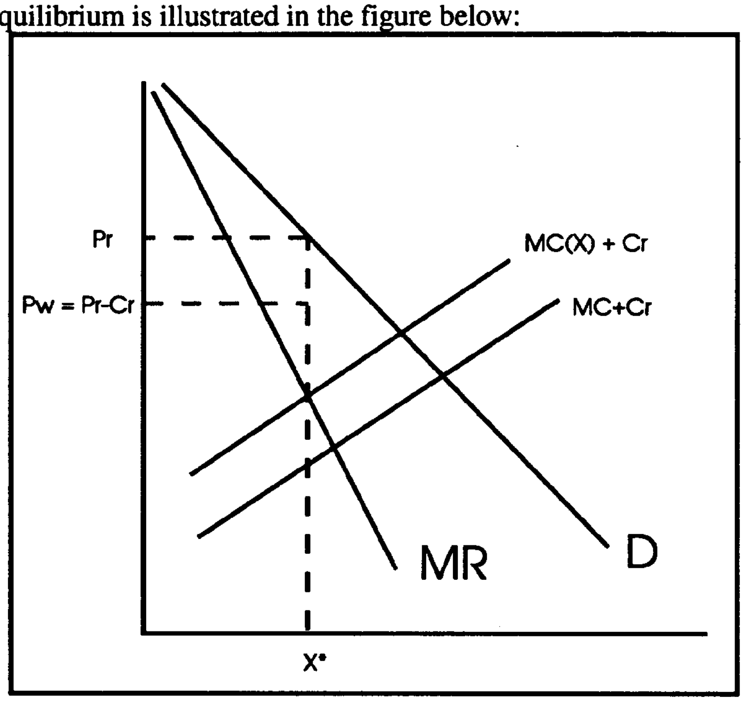
The producer wishes to maximize profits which are:

Profit = $X\,P_w - C(X) = X\,(P_R(X) - C_R) - C(X)$.

The firm's optimal production level will be where

$P_R + X\,dP_R/dX = MC(X) + CR$ or $MR_R = MC + C_R$.

Hence, the firm will produce where marginal revenue at retail is equal to full marginal costs (i.e. marginal production costs, MC(X), plus marginal retailing costs C_R). This equilibrium is illustrated in the figure below:



E) In order to implement this equilibrium, the producer simply sets the wholesale price, P_W, equal to the desired retail price less the competitive costs of retailing. Competition between retailers will insure that retailing services are provided at the minimum average cost and hence that the retail price will be set at the optimal level.

In this equilibrium, the producer harnesses the competition between retailers to provide retailing services at minimum cost. The producer sells his product to the retailer who does the rest.

F) Vertical restrictions represent a deviation from the framework described above. First, it should be clear that vertical restriction are not motivated by a desire to increase the retail price. The producer can increase the retail price as much as he desires simply by raising the wholesale price. Something else must be at work.

G) The basic reason for vertical restraints is that retailing is not one-dimensional. Manufacturer's desire retailers to perform many services, promotion, distribution, maintain inventories, regulate product quality etc. The level of these services provided by retailers directly affects the amount of the product the manufacturer sells. Achieving the lowest possible retail price (for a given wholesale price) is not the producer's ultimate goal.

H) When it comes to product quality, the manufacturer and retailer incentives are often (but not always well aligned). Greater product quality raises the demand for the manufacturer's product but also raises the demand for the retailer's product (assuming customers can identify quality differences across retailers). That is why we see many items sold as described above. Making the product as appealing as possible to consumers (in terms of the price quality combination) is in the interest of both the retailer and the producer.

I) On other dimensions the alignment is often less than perfect. Retailers tend to focus on promotional and other efforts that allow them to increase sales even if the bulk of these sales come from other retailers. The producer would like retailers to focus their efforts on things that increase total sales of the producer's product. The basic distortion is that retailers tend to focus heavily on what shifts sales across outlets while the producer cares only about what shifts sales in the aggregate.

J) To focus the retailer's efforts in specific directions, the producer will often adopt vertical restraints. For example, **exclusive territories** work by reducing competition between retailers. Similarly, **exclusive dealing** helps to focus the retailer's efforts on the producer's product rather than the products of other producers and can facilitate monitoring. Cooperative advertising and promotion subsidies provide direct subsidies to retailers for the provision of services that increase the demand for the producer's product generally.

K) **Resale price maintenance** can serve to switch the focus of competition from price to other dimensions (like promotion) and is particularly effective when competition between retailers focuses too much on price and not enough on other dimensions (i.e. when price sensitive customers are the marginal customers across firms but service of promotion sensitive customers are the marginal customers for the manufacturer).

L) In general, anticompetitive motivations for vertical restraints examine how vertical restraints imposed by one producer effect competition horizontally between producers. For example, **exclusive dealing** requirements can be used to foreclose rivals from gaining access to a downstream market if such contracts cover a significant part of the downstream market. Similarly, the use of a common exclusive selling channel by a number of producers is sometimes thought of as an effort to facilitate collusion among producers.

M) In the case of collusion, the potential for pro-competitive effects is small or non-existent. As a result, economists typically take a very hard line on collusion. In contrast, the potential for pro-competitive effects from vertical restraints is much greater. In addition, the avenues for anticompetitive effects in the downstream market are much more limited since the producer already has the ability to control the downstream price via changes in the wholesale price

N) In general, it will be difficult for vertical restrictions imposed by an individual producer to have anticompetitive effect unless the producer has a significant degree of market power to begin with. For example, exclusion requires that the producer tie up a significant amount of the downstream capacity and that the producer is able to profit from the exclusion. Even a relatively large producer will have a hard time tying up a sufficient amount of the downstream capacity to exclude its rivals and in order to profit it must be able to affect prices in the upstream market. Since this indirect method of exercising market power is likely to be less effective than more direct methods, the existence of market power at the producer level is often taken as a threshold standard for examining the anticompetitive effects. For many vertical contracts (not resale price maintenance), the absence of market power is taken as sufficient for the absence of anticompetitive effects while such contracts are subject to a rule of reason (where we weigh the pro-competitive and potential anticompetitive effects) for firms with market power.

Predatory Behavior

A) Predatory behavior refers to efforts by a firm to drive out its rivals and benefit from the resulting reduction in competition. In the classic predatory pricing paradigm, the firm cuts prices in an attempt to drive its rivals from the market. Once the rivals are gone, the firm raises price above the competitive level and recoups its losses.

B) Most economists are skeptical of the predatory pricing model on both theoretical and empirical grounds. On the theoretical side, it is unclear why the strategy would be successful. Both the predator and the prey take losses during the period of predation. Unless the predator has superior access to capital it is unclear why it will be able to outlast its prey. This problem is exacerbated by the same force that plagues collusion; by definition, the predator firm is not maximizing profits in the predatory phase while the prey firms can still maximize their profits (given the behavior of the predator).

C) In addition, even if the firm is successful in driving its rivals from the market, it still may not be able to raise price significantly above competitive levels since the firms may still be able to return.

D) Even if the policy is modestly successful its effect on consumer welfare is ambiguous. Consumers gain from the lower prices received during the predation phase and these gains can more than out way the costs imposed on consumers during the recoupment phase.

E) Perhaps, the most damaging theoretical criticism of predatory theories is the difficulty in identifying predatory behavior. Competitors lose when a firm cuts it prices whether these price cuts are predatory or part of the normal competitive process. The incentive for competitors to complain is independent of whether these price cuts will lead to later price increases to consumers. If anything, competitors lose more when the price cuts are permanent.

F) A competitor fighting for market share is an integral part of the competitive process. Even efforts to drive ones rivals from the market should be regarded as part of competition. Typically both the firm and consumers benefit from such aggressive behavior even if the firm does not subsequently raise price. Indeed, given the ubiquitous incentive for firms to collude ***we should be extremely hesitant to prosecute firms for price-cutting***.

G) Predatory theories offer suffer from a lack of empirical support. While cases of alleged predatory behavior are common, documented cases where firms successfully raised price as a result of predatory gains are few and far between. The bottom line (in my opinion) is that both theory and empirical evidence suggests that predatory tactics are unlikely to be successful. In any case, consumers gain in the attempt. Most importantly, efforts to curtail predatory behavior run the risk of blocking some of the far more common cases where price cuts are pro-competitive.

Price Discrimination

A) As we saw in our previous discussion, firms gain by price discrimination in two ways. They are able to charge more to high value users who would purchase the product even in the absence of price discrimination and they collect addition profits on sales to low value consumers that would not occur absent discriminatory pricing.

B) **The efficiency consequences of price discrimination are unclear**. Typically such behavior leads to lower prices for some consumers and higher prices for others. Output can be expanded as when price discrimination is very efficient but output can also decline when discrimination is more effective at extracting surplus from high valued users than expanding sales to low valued users. Even if output remains constant, price discrimination can reduce efficiency by misallocating output among consumers.

C) Another way to think about the effect of price discrimination on consumer welfare is in terms of two offsetting effects. The **surplus extraction effect** represents the fact that firms are better able to capture surplus from consumers at any given level of sales. The **output expansion effect** reflects the fact that better methods of discriminating allows consumers and firms to exploit more of the gains from trade. Both effects tend to increase the profits of firms.

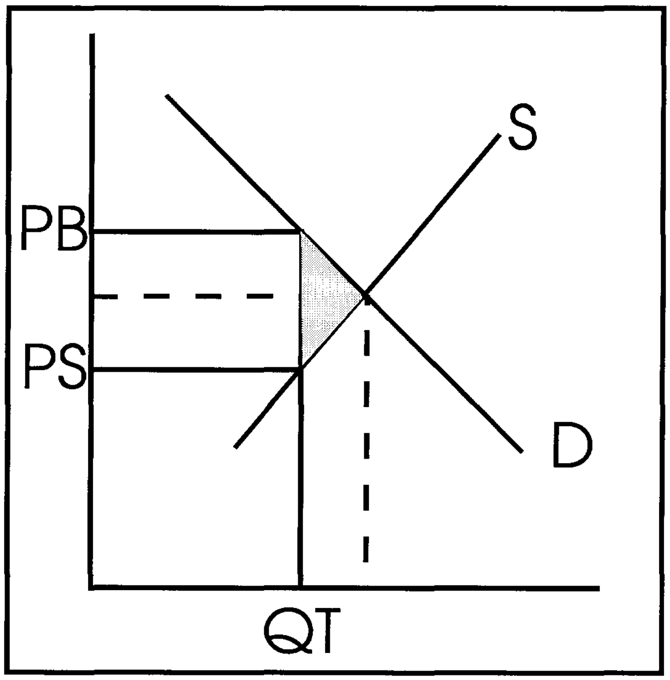
D) However, this is not the end of the story. In the long run, we expect that profits will be competed away, typically through lower prices to consumers or through enhanced incentives for product development and innovation. This will offset the surplus extraction effect and significantly increases the probability that consumers will benefit.

Microeconomic Analysis of Policy Issues

Taxes

A) The study of taxation is one of the classic topics in microeconomics. One can teach an entire course in tax theory. Obviously a full treatment is beyond the scope of what we can accomplish in one week. We will focus on three central questions, what goods should be taxed, what level of taxes should be set, and what is the optimal tax structure? We will address these issues in both the areas of taxing specific commodities and the general taxation of income.

B) The best place to start is with the taxation of a single commodity. The basic picture looks like.



The tax reduces consumption and of course production of the commodity while raising the price (inclusive of the tax) paid by buyers and lowering the net price received by the sellers. The tax is equal to the gap between the buyer's price, P_B, and the seller's price, P_S and government revenues are equal to $(P_B-P_S)*Q_T$. To quantify things we measure everything in percentage terms. This yields

$$\Delta Q = \varepsilon_D \, \Delta P_B = \varepsilon_S \, \Delta P_S \text{ and } \Delta P_B = \Delta P_S + \tau,$$

where, ΔQ is the percentage change in quantity generated by the tax and ΔP_B and ΔP_S are the percentage changes in the buyers' and sellers' prices respectively. The tax, τ, is measured in percentage terms as well. Together, these equations imply that $\varepsilon_D \, \Delta P_B = \varepsilon_S \, (\Delta P_B - \tau)$ and hence that

$\Delta P_B = \tau\ \varepsilon_S/(\varepsilon_S-\varepsilon_D)$
$\Delta P_S = \tau\ \varepsilon_D/(\varepsilon_S-\varepsilon_D)$
$\Delta Q = \tau/(1/\varepsilon_D - 1/\varepsilon_S)$.

C) In general, governments tax commodities for three basic reasons, to reduce consumption (such as with alcohol and cigarettes), pay for related services (such as federal transportation taxes on gasoline) and as a source of revenues (such as income taxes).

D) In general, as the equations above imply, the effects of the tax depend on the elasticities of supply and demand. In particular,

(1) The incidence of the tax depends on the relative elasticities of supply and demand. In particular, if the elasticity of demand is greater (in absolute value) than the elasticity of supply then sellers will pay the larger share of the tax. When the elasticity of supply is greater than the elasticity of demand (in absolute value) buyers pay the larger share of the tax. For example, if the demand elasticity is twice as large as the supply elasticity then sellers pay 2/3 of the tax and buyers pay only 1/3. If either supply or demand is perfectly inelastic (perfectly elastic) then that side pays all (none) of the tax.

(2) The reduction in quantity is greater the larger are the elasticity of demand and the elasticity of supply.

(3) The changes in prices and output are proportional to the tax (holding the elasticities fixed).

E) When taxes are levied solely as a source of revenue, they reduce economic efficiency. We can measure this inefficiency in the same way that we measured the social cost of monopoly. In particular, the tax generates a distortion in the market by driving a wedge between the marginal value of the good to buyers (measured by the buyers' price) and the marginal cost of the good (measured by the sellers' price). This is identical to the social loss from monopoly and is measured by the shaded triangle in the figure above. The base of the triangle (measured vertically) is τP while the height (measured horizontally) is $\tau Q\Delta Q$. Thus the social loss is $\frac{1}{2}\tau^2 PQ/(1/\varepsilon_D-1/\varepsilon_S)$. Since PQ equals total revenue, TR, we have

Social Cost = $\frac{1}{2}\tau^2\ TR/(1/\varepsilon_D-1/\varepsilon_S)$.

To calculate the added social cost from increasing the tax we can differentiate this expression to yield

d(Social Cost)/dτ = $\tau\ TR/(1/\varepsilon_D-1/\varepsilon_S)$.

These formulas imply that:

(1) The social cost of the tax grows with the square of the tax rate. This implies that doubling a tax roughly quadruples the social cost. Equivalently, the

social loss from increasing the tax is proportional to the starting level of the tax. Thus the loss from increasing a tax from 20% to 21% is roughly double the loss from increasing the tax from 10% to 11%.

(2) The social loss from taxation is smaller when demand or supply are less elastic.

(3) From the point of view of raising revenues with the least distortion, it is optimal to tax most heavily those goods that have inelastic demand and/or supply.

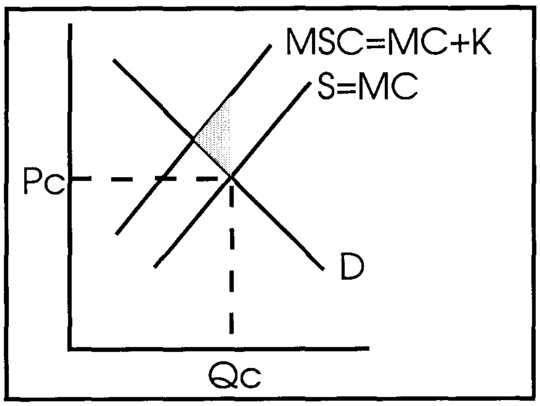
(4) Broad based taxes tend to yield lower social costs since the social cost rises with the square of the tax rate (i.e. two 5% taxes will generate roughly ½ of the social cost of one 10% tax).

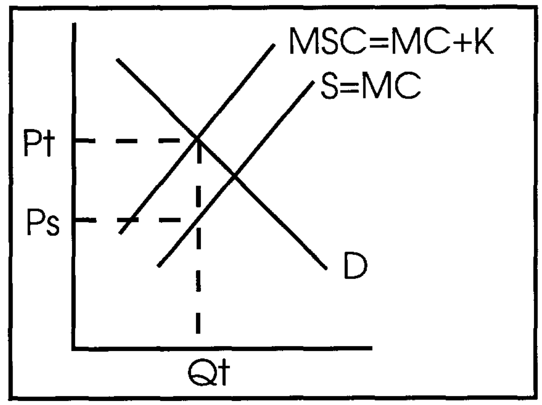
F) These principles on the incidence of taxes and the efficiency losses are illustrated in the following table, which gives the effects of various tax rates and elasticities.

Tax	Elasticity of Demand	Elasticity of Supply	Change in Buyer's Price	Change in Seller's Price	Change in Output	Social Cost (% of TR)
10%	0.0	0.5	10.0%	0.0%	0.0%	0.0%
10%	0.0	1.0	10.0%	0.0%	0.0%	0.0%
10%	0.0	2.0	10.0%	0.0%	0.0%	0.0%
10%	-0.5	0.0	0.0%	-10.0%	0.0%	0.0%
10%	-1.0	0.0	0.0%	-10.0%	0.0%	0.0%
10%	-2.0	0.0	0.0%	-10.0%	0.0%	0.0%
10%	-0.5	0.5	5.0%	-5.0%	-2.5%	-0.1%
10%	-0.5	1.0	6.7%	-3.3%	-3.3%	-0.2%
10%	-0.5	2.0	8.0%	-2.0%	-4.0%	-0.2%
10%	-1.0	0.5	3.3%	-6.7%	-3.3%	-0.2%
10%	-1.0	1.0	5.0%	-5.0%	-5.0%	-0.3%
10%	-1.0	2.0	6.7%	-3.3%	-6.7%	-0.3%
10%	-2.0	0.5	2.0%	-8.0%	-4.0%	-0.2%
10%	-2.0	1.0	3.3%	-6.7%	-6.7%	-0.3%
10%	-2.0	2.0	5.0%	-5.0%	-10.0%	-0.5%
20%	-1.0	1.0	10.0%	-10.0%	-10.0%	-1.0%
30%	-1.0	1.0	15.0%	-15.0%	-15.0%	-2.3%
40%	-1.0	1.0	20.0%	-20.0%	-20.0%	-4.0%
50%	-1.0	1.0	25.0%	-25.0%	-25.0%	-6.3%

G) Sometimes taxes are designed to enhance efficiency. Taxes can improve efficiency when there are externalities (as when the production of electricity generates air pollution) or when taxes are used to charge for complementary services that are not priced separately (such as gasoline taxes used to pay for roads).

marginal social cost of the good will exceed the marginal production cost of the good (reflected in the supply curve) since sellers fail to internalize the costs imposed on others by the production and/or consumption of the good. Since the supply curve measures the marginal cost of the good to the producers, the social marginal cost of the good will be $K higher at each level of output. This is illustrated in the left hand figure below where we have shifted the marginal cost curve up by $K at each level of output.





I) The left hand figure shows the equilibrium without a tax. The competitive market equilibrium will be at the intersection of the supply and demand curves. This yields Pc and Qc as the equilibrium price and quantity. At this level of output there is excessive consumption. The shaded triangle measures the social cost of this excess consumption since it measures the difference between social marginal cost and marginal value for the units between the optimal level of consumption (where demand crosses the marginal social cost curve (labeled MSC) and the competitive level of output.

J) The right hand figure illustrates the equilibrium with a per unit tax set equal to the social externality, $K. Output will now be at the optimal level, Qt, with consumers paying price, Pt, and sellers receiving price, Ps. In this case the equilibrium with the tax is socially efficient.

K) Taxes such as those illustrated in the right hand figure above improve efficiency as well as generate revenues. In this case, the reduction in output generated by the tax is desirable. Such taxes work best when the external cost (such as pollution) or the cost of providing related services (such as roads) are close to proportional to the level of production (or consumption) of the good. When the link is weak, the tax will be less effective.

L) In general it is better to tax the aspects of consumption that are most closely tied to the social cost. For example, it is better to tax emissions than electricity production or drunk driving rather than alcohol consumption. In fact, when the costs imposed on society differ greatly across users or producers (such as in the case of alcohol consumption) a broad based tax will do little to improve efficiency. A tax on alcohol will cut alcohol consumption across the board and will not necessarily reduce it most among the group producing the external cost (i.e. drunk drivers). A targeted tax would be much more efficient.

Income Taxes

A) While taxes on individual commodities are common, taxes on income and property are the most important in terms of revenues. Taxes on labor income include, federal and state income taxes, Social Security and Medicare payroll taxes, as well as broad based consumption taxes. Taxes on capital income include the corporate income tax as well as the personal income tax on capital income (dividends, interest, capital gains, etc.)

B) In fact we really could include a much wider range of taxes in a discussion of income taxation. In some sense, almost all taxes represent income taxes. People earn income in order to spend it. Taxes on consumption goods increase the effective price of consumption and hence reduce the return to work (though not necessarily the return to investing) just as direct tax on income would.

C) Taxes are an important part of modern economies. To illustrate the importance of taxation and government spending, Table 1 gives revenues and outlays for different levels of government in the United States for 1999.

	Federal Revenues	Federal Outlays	State & Local Revenues	State & Local Outlays
Billions of Dollars	$1,874.6	$1,750.2	$1,142.7	$1,092.7
% of Gross Domestic Product(GDP)	20.2%	18.8%	12.3%	11.8%
% of Net National Product (NNP)	23.1%	21.5%	14.1%	13.4%

D) Taken together, Federal, state and local government revenues amounted to about 32.5% of national output (GDP) and 37.2% of net national product (the value of output after depreciation). Government outlays were only slightly smaller (equal to 30.6% of GDP). Taxes on personal income accounted for 47% of Federal revenues, while corporate taxes accounted for 10% and payroll taxes accounted for an additional 33%.

E) In terms of income taxation, there are three fundamental questions. What to tax, what tax structure to use, and at what rates? Broadly speaking, the question of what to tax is a question of dividing the burden of taxation between labor and capital income. The question of what tax structure to use is a question of how progressive the tax structure should be. The question of what rates has two dimensions, the overall level of taxes and the distribution of the tax burden over time. We will address each of these questions in turn.

F) The choice between taxes on capital and taxes on labor can be addressed using the analysis of commodity taxation described above and the models of factor markets we studied earlier in the quarter. In our discussion of the relationship between productivity, output prices and input prices we derived the equation

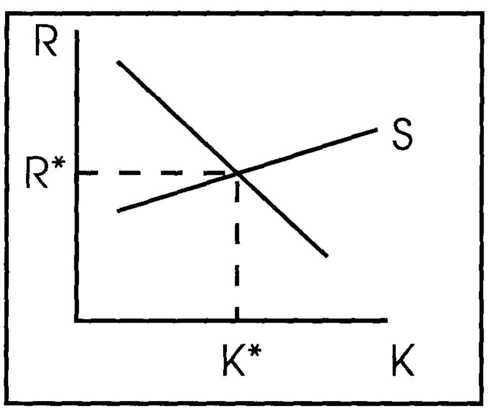
$$\Delta_P = S_L\, \Delta_W + S_K\, \Delta_R - \Delta_{TFP}.$$

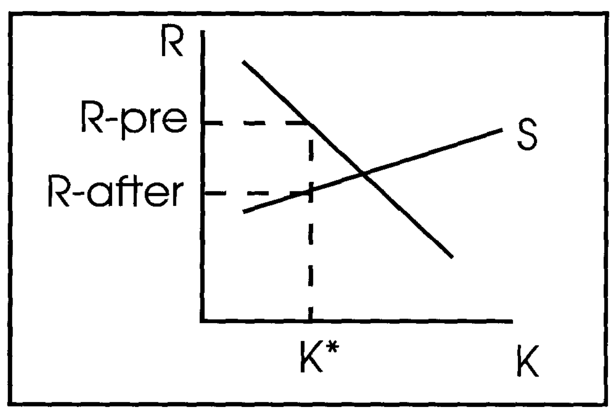
This equation reflects the fact that in a competitive market in long run equilibrium, price will be equal to unit costs. The equation reflects the fact that increases in input prices increase output prices while increases in productivity lower output prices (holding input prices fixed) or raise input prices holding output prices fixed. If we rewrite this equation in real terms we get

$$\Delta_{TFP} = S_L(\Delta_W - \Delta_P) + S_K\,(\Delta_R - \Delta_P) = S_L\, \Delta_{W/P} + S_K\, \Delta_{R/P}.$$

Here, $\Delta_{W/P}$ is the change in real wages and $\Delta_{R/P}$ is the change in the real return to capital. Both of these prices are measured pre-tax (i.e. they represent the input prices faced by the firm). To examine the effects of changes in the tax structure, we hold productivity fixed so that $\Delta_{TFP}=0$ and hence $\Delta_{W/P} = -S_K/S_L\, \Delta_{R/P}$. Thus, if taxation raises the pre-tax return to capital then it must by necessity reduce the pre-tax return to labor and vice-versa.

G) The real price of capital will be determined in the capital rental market as in the left hand figure below. A tax on capital will increase the pre-tax return to capital and decrease the after-tax return as shown in the right hand figure.



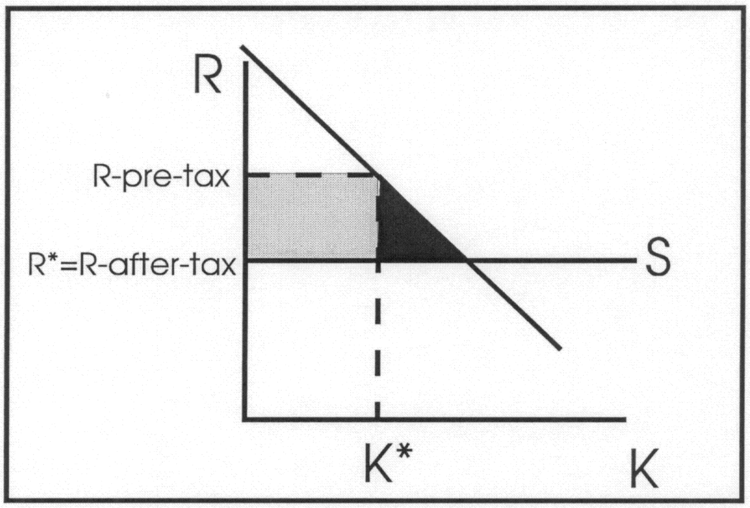


H) With a higher pre-tax return to capital the pre-tax return to labor must fall. In the U.S., the share of capital in GDP is about 38% while the share of labor in GDP is about 62% so that a 1% increase in the pre-tax return to capital will decrease the pre-tax return to labor by about .38/.62 = .61%. The key question is how the capital tax is split between a decline in the after-tax return and a rise in the pre-tax return.

I) The answer to this question is very different in the short run and in the long run. In the very short run, the supply of capital is likely to be quite inelastic so that a tax on capital will mostly reduce the after-tax return with little increase in the pre-tax return.

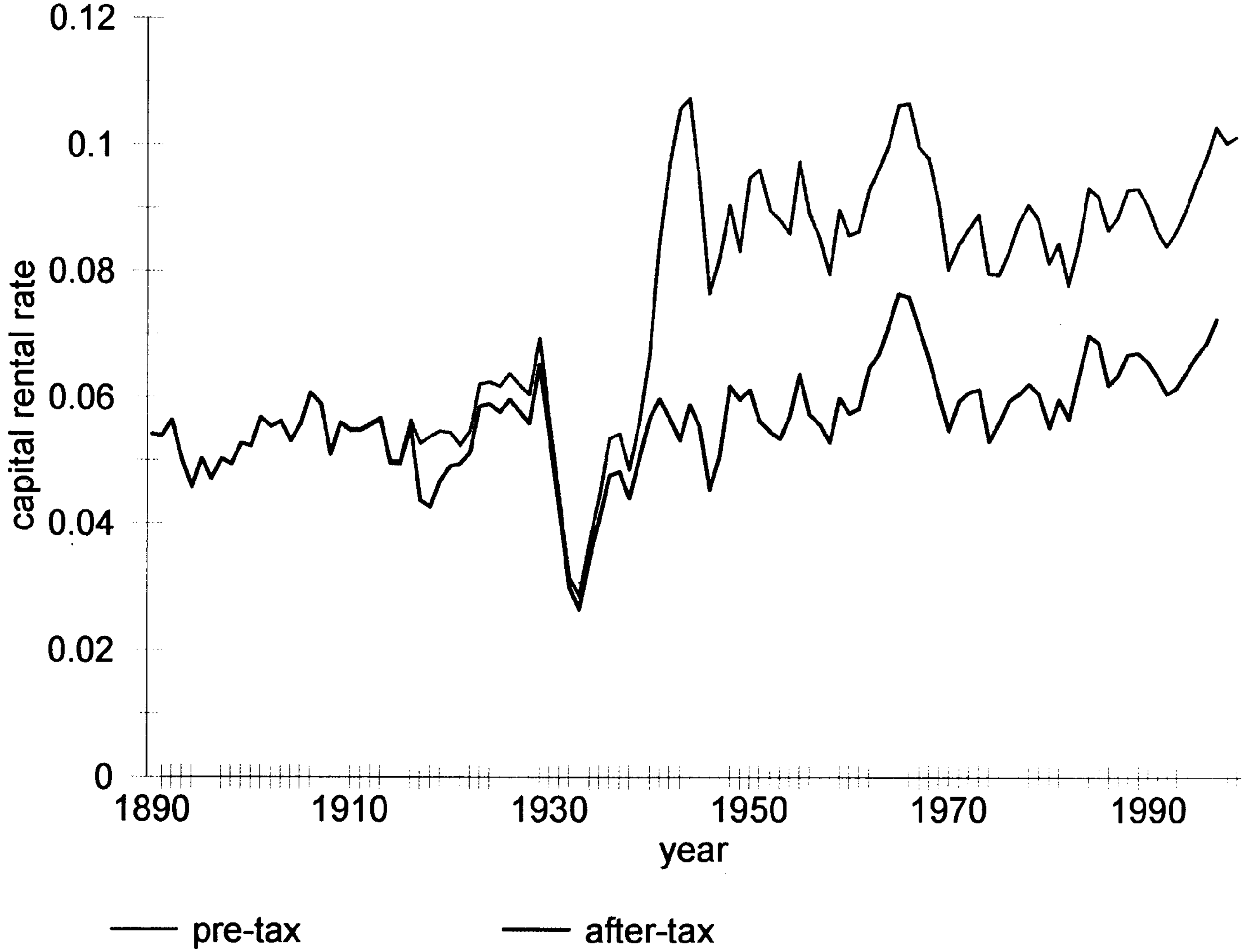
In contrast, in the long run the supply of capital is very elastic, perhaps perfectly elastic. Empirical evidence suggests in fact that the after-tax return will tend to be about 6-7% independent of the level of capital taxation. This can be seen in the graph on the following page which tracks the pre-tax and after-tax return to capital in the U.S. over the past century. The large increase in capital taxation in the post WWII era led to a substantial rise in the pre-tax return with almost no change in the after-tax return.

J) If the supply of capital is taken to be perfectly elastic at R*then the equilibrium in the capital market will be as illustrated below.



With constant returns to scale (so that profits are zero) pre-tax labor income will be equal to the triangle above R-pre-tax (consumer surplus in the usual supply and demand graph). The contribution of capital taxes to government revenues is given by the lightly shaded rectangle. Reducing the capital to zero (i.e. moving R-pre-tax down to R*) would increase pre-tax labor income by the shaded rectangle PLUS the darkly shaded triangle. Hence, the increase in pre-tax labor income will exceed the lost government revenues from the capital tax. As a result, we can increase the labor tax enough to compensate for the loss in capital tax revenues and have the triangle left over for additional revenues or to increase the after-tax wage.

K) The basic idea behind this point is actually quite simple. As we showed above, increases in the pre-tax return to capital must reduce the real wage. When the supply of capital is perfectly elastic all of the capital tax is passed on and results in higher output prices and/or lower wages. Either effect reduces the real wage. As a result all of the tax on capital is passed through to workers. In addition the distortion in the capital market reduces labor income more than it increases tax revenues. In general, with constant returns to scale and a perfectly elastic supply of capital, the optimal long-run tax on capital is zero even if we focus attention only on the incomes of labor.

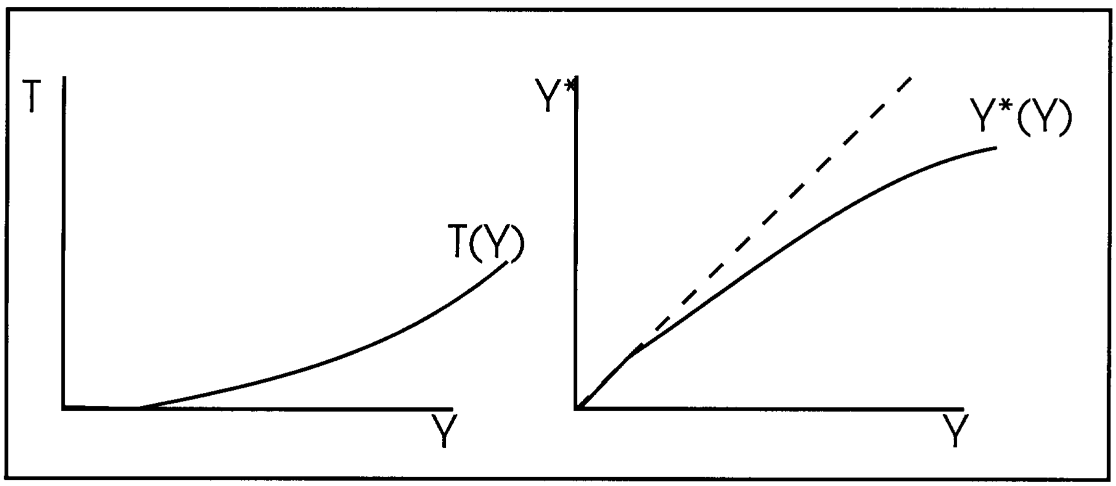

0.12
0.1
0.08
0.06
0.04
0.02
0
capital rental rate
1890
1910
1930
1950
1970
1990
year
pre-tax
after-tax

The Tax Structure

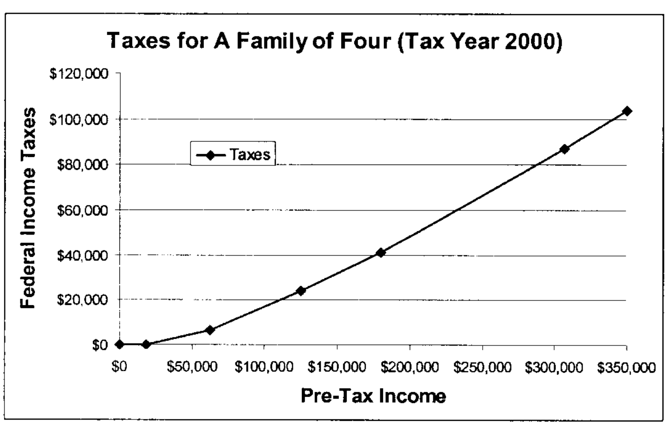
Other Tax Issues:

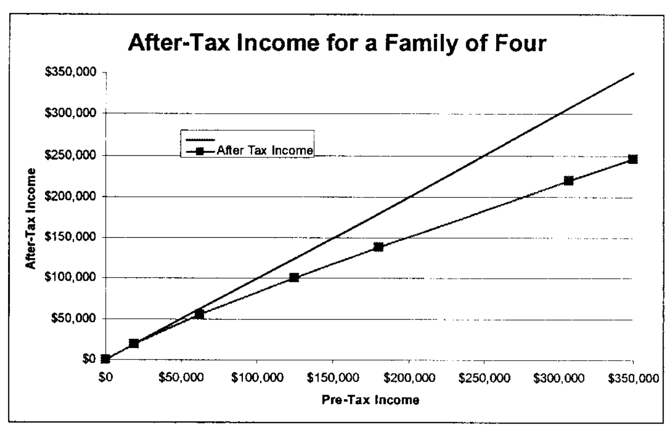
A) One key issue besides the choice of what income to tax (i.e. capital income versus labor income) is the choice of the tax structure. At its most basic level, the tax structure choice comes down to a choice of exemptions, deductions and the structure of tax rates. Here, we will examine the issue of progressivity (i.e. the extent to which tax rates rise with income).

B) To keep things simple we will deal with labor income taxes only and consider an individual who chooses work effort (i.e. hours). We start with a progressive tax structure with rising marginal tax rates. We can model this by a tax schedule T(Y) that gives taxes, T, as a function of pre-tax income, Y. A hypothetical schedule is shown in the left hand figure below. The corresponding after-tax income schedule is shown in the right hand figure, where we give after-tax income, Y*, as a function of pre-tax income, Y.



C) The figure below gives the actual Federal Income Tax structure for a married couple with two children claiming the standard deduction (see problem set #1) where we have ignored the EITC to keep things simple.

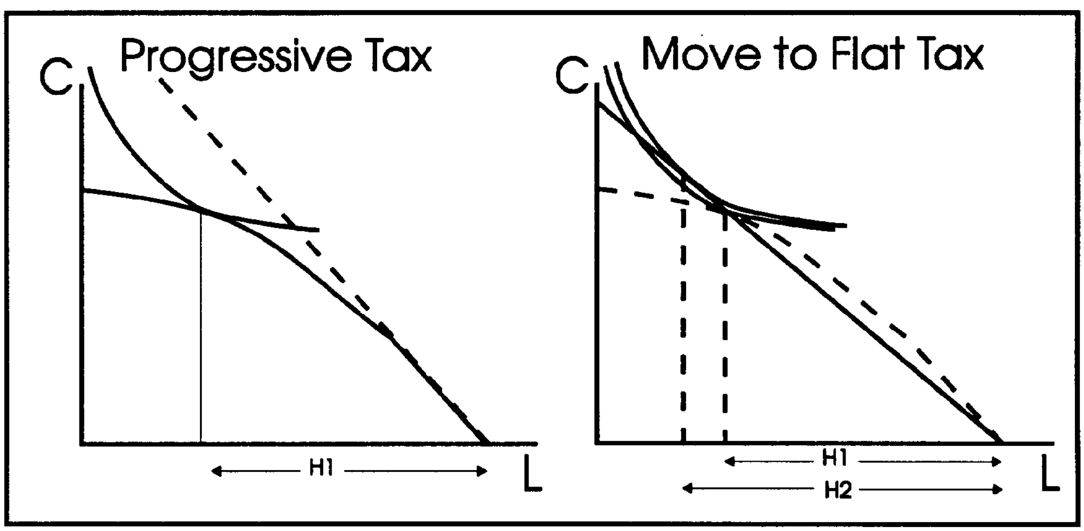




D) The table below describes the above tax structure in terms of the resulting marginal and average tax rates at different levels of family income. As can be seen from the table, marginal tax rates are greater than average tax rates for all households (typically by about 10 percentage points).

Income	Taxes	After-Tax Income	Marginal Tax Rate	Average Tax Rate
$15,000	$0	$15,000	0.00%	0.0%
$30,000	$1,718	$28,283	15.00%	5.7%
$60,000	$6,218	$53,783	15.00%	10.4%
$100,000	$17,106	$82,895	28.00%	17.1%
$150,000	$31,871	$118,130	31.00%	21.2%
$200,000	$48,371	$151,630	36.00%	24.2%
$250,000	$66,371	$183,630	36.00%	26.5%
$350,000	$103,922	$246,078	39.60%	29.7%

E) As we saw in data exercise #1, the marginal return to work is equal to the marginal after-tax wage rate (i.e., $W(1-\tau)$, where W is the pre-tax wage and τ is the marginal tax rate). We can analyze this problem in the usual consumption-leisure choice framework. With a progressive tax structure, the individual's budget constraint and choice of hours of work will be as illustrated in the left hand figure below. In the figure, the dashed line gives the individual's pre-tax earnings while the curved line gives the individual's after-tax earnings. As always, the individual will find a tangency between his indifference curve and the opportunity set. In this figure, the individual chooses H1 hours of work.



G) The right-hand figure illustrates the effect of eliminating the progressive tax structure and replacing it with a flat tax with the same average tax rate (i.e., the person would pay the same amount of taxes at the previous level of income). As can be seen in the figure, the individual would now choose to work more (since the marginal return to work has increased) and will end up better off (due to the decreased deadweight loss). The government will collect more revenues as well since the person now earns more with the same average tax rate.

H) Indeed, if all individuals were the same, a progressive tax structure like that shown above would make no sense. It would generate less revenue and provide more disincentives (higher social cost) than a flat tax with the same average rate. However, not all individuals are alike -- that is what makes things tough. While flat tax structures result in smaller disincentives in general, they also redistribute the burden of the tax away from richer households toward poorer ones. Equivalently, if the goal of the tax structure is to redistribute wealth from high-income households to low-income households then disincentives are unavoidable. This can be seen quite simply. By its very nature redistribution is at best zero-sum (i.e., what we give to some we must take away from others). We say "at best" because the process of redistribution will generate a deadweight loss through taxation.

Tax Cuts

A) Just like households, the government faces a budget constraint. The present value of receipts (taxes) must be equal to the present value of expenditures. If we denote expenditures in year t by E_t, taxes in year each year by T_t, and the current value of the outstanding debt by D_t, then we know that:

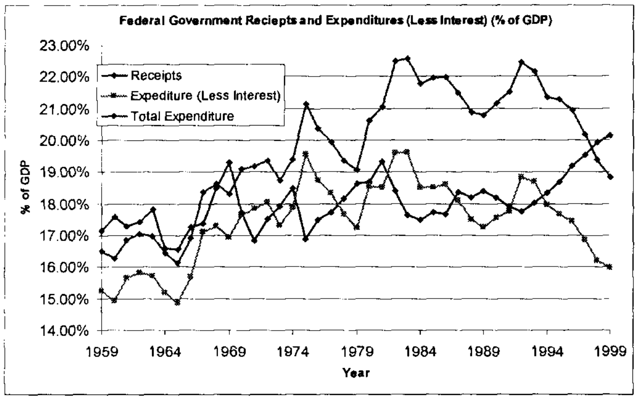
$$T_t + T_{t+1}/(1+r) + T_{t+2}/(1+r)^2 + \ldots = D_t + E_t + E_{t+1}/(1+r) + E_{t+2}/(1+r)^2 + \ldots$$

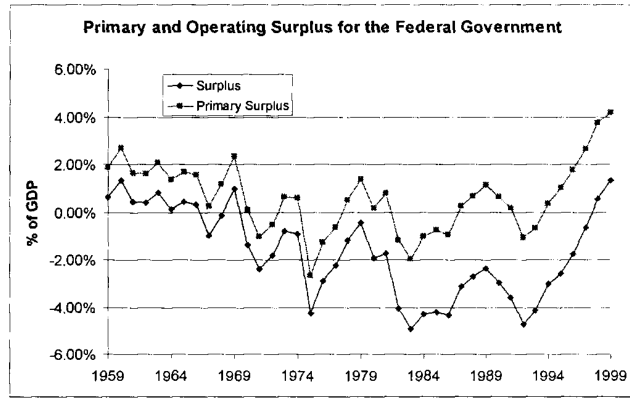
B) It is important to remember that as defined in the equation above, expenditures, E, do not include interest payments. The reduction to present value already accounts for interest. This equation can be rewritten in a useful way by collecting all of the terms except the current debt on one side as in:

$$D_t = (T_t - E_t) + (T_{t+1} - E_{t+1})/(1+r) + (T_{t+2} - E_{t+2})/(1+r)^2 + \ldots$$

This tells us that the government must run fiscal surpluses in each year that are equal in present value to the current outstanding debt. However, this equation does not imply that the government needs to run a surplus (as it is commonly defined = T_t-E_t-Interest Payments).

C) The exclusion of interest from the expenditure numbers is important empirically. This can be seen by looking at the two figures below. The first figure tracks government outlays (with and without interest) and government revenues over time. The second figure tracks the primary surplus (taxes – expenditures) and the surplus as it is conventionally measured (taxes – expenditures – interest). All are measured as a share of GDP to make them easier to understand. As can be seen from the second figure the government has run a primary surplus for all years after 1986 (except 2) and has hence been contributing to offset the debt even though it has run a surplus as conventionally defined in only the final two years. The first figure also shows that government expenditures have grown much more slowly since the mid 1970s once we exclude inertest payments.



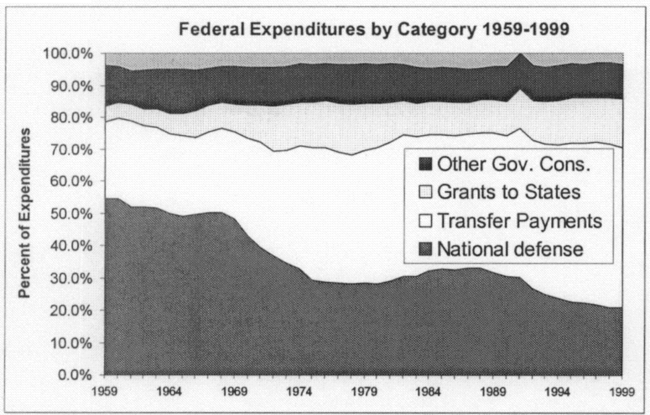


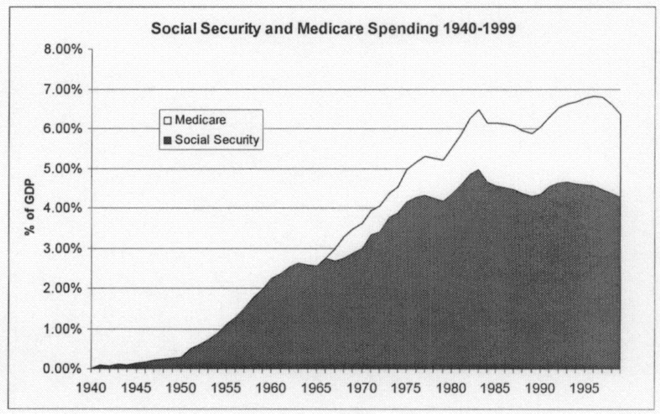
D) The equations above tell us that in the long-run cuts (increases) in taxes must be matched by cuts (increases) in spending that are equal in size in terms of present value. To spend more, one must tax more. While taxes and spending must match in the long run, they often diverge in the short term as the government runs either a surplus or a deficit. One way in which they diverge is through tax cuts that are not matched by cuts in spending. This raises the debt that is carried forward into the next period and as we showed above must be matched by some combination of future spending cuts and future tax increases. In practice, we often see a combination of both. More debt curtails the growth in spending but also puts upward pressure on taxes (this is what happened in the 1990s in my opinion). However, as long as all of the debt is not compensated by increases in taxes (i.e. some is compensated for by future cuts in spending) then a cut in current taxes is a way to reduce future spending. This seems to fit empirically since it is hard to believe that we would have run a fiscal surplus so many years after 1986 if it were not for the large debt accumulated in the earlier years.

E) This analysis also sheds considerable light on the most recent debate, whether the currents surplus should be used to reduce the debt or used to give tax cuts. While reducing the debt would seem to be the more fiscally conservative thing to do it is opposed by most republicans (and favored by most democrats). However, once we realize that reducing the debt now will lead to more spending in the future while reducing taxes today will reduce future spending these positions become much easier to understand. Those in favor of reducing the size of government would advocate tax cuts while those that favor more government spending would favor reducing the debt. Tax cuts represent a commitment to spend less while debt reduction frees up the government to spend more in the future.

Social Security and Medicare

A) The largest growth in government spending has been in the area of entitlements. This is particularly true of entitlements to the elderly (Social Security and Medicare). The left hand figure below gives a breakdown of government spending by type over time. The second figure illustrates the growth in Social Security and Medicare over time (measured as a percent of GDP). The growth in entitlements is shown in the left hand panel by the rapid expansion in transfer payments and to some extent by the growth in grants to states. The explosive growth in payments to the elderly is illustrated in the right hand figure. While things have looked somewhat better since the mid 1980s in terms of growth in programs for the elderly, the picture will worsen rapidly in the coming decades as the baby boom cohorts retire and per-capita medical spending continues to increase.





B) As we saw in data exercise #3 the Social Security system has two major problems: (1) current tax rates are insufficient to fund the system in the long run and (2) the system provides young workers with a very poor return on their retirement savings. Of course, raising tax rates to balance the system (i.e. solve problem #1) will make problem #2 even worse. Both of these problems have motivated people to look for alternatives.

C) The hottest topic on the agenda today is privatization. Privatization, it is often claimed can solve #2 while a temporary tax increase (often called a transition tax) can be used to solve problem #1. Unfortunately, things are not so simple.

D) First, if we were not giving young workers such a poor deal on their investment then problem #1 would be much worse. The poor return to young workers is needed to pay for the large un-funded debt that we owe to those at or near retirement. The current trust fund balance is an order of magnitude smaller then the debt owed to those age 50 and above. This difference must be made up by giving young workers a poor return on their retirement investments. This is true whether we have a public or private system. In the current public system it shows up as a low return on their tax dollars. In a privatized system it would show up as an additional tax that young workers must pay to fund the benefits of older workers. Privatization per se does nothing to solve this problem.

E) Some advocates of privatization point to the large gap in the returns paid on private investments (e.g., stocks and corporate bonds) relative to the implicit return paid by the current Social Security system. However, as stated above, this comparison makes no sense. The return in the current system includes the tax that is needed to pay the benefits promised to older workers and current retirees. This tax would need to be levied in any system. If any comparison should be made it would be between the return on private-sector assets and the assets held in the Social Security trust fund (government bonds). However, this difference represents a premium for risk (and possibly liquidity) and cannot be taken as a measure of what is being given up.

F) In lieu of privatization, some call for trust fund assets to be invested in the stock market rather than government bonds. Indeed, this would provide the trust fund assets with a higher expected return but would also subject those assets to greater risk. However, such a plan simply involves the government issuing more bonds to the public and using the proceeds to buy stocks. If this provides a net gain, then the government should do it, even if there were no Social Security or indeed even if Social Security did not exist. Few would argue for this as a general policy and hence I would hesitate to recommend it for the Social Security surplus.

G) Medicare has all of the issues of Social Security and one key addition, the growth in spending. Medicare spending, like spending on medical care in general, has been growing rapidly over time. However -- contrary to what people often say -- most of this growth is growth in the *quantity* of care rather than growth in the *cost* of care. Over time we have moved to more aggressive and widespread treatments and these cost more than the old methods of treatment (see the handout on heart attack treatments for a simple example).