№ 22

Reading Materials for:

Economics 101A

Microeconomic Theory

Inst: David Card

University of California, Berkeley

Fall 2008

Copy Central

northside@copycentral.com

$8.69 2483 Hearst Avenue • Berkeley, CA 94709 • (510) 849-0700

ECONOMICS 101A

Microeconomic Theory

Instructor: David Card

Course Description and Reading List

This is a course in intermediate microeconomics, emphasizing the applications of calculus and linear algebra to the problems of consumer decision making, firm behavior, and market interactions. The text is Walter Nicholson *Microeconomic Theory: Basic Principles and Extensions* (Ninth Edition), which should be available at the campus book store. A more theoretical treatment of the same material is in Hal Varian's textbook, *Microeconomic Analysis* (3rd edition is most recent), published by Norton. Students who are thinking of pursuing graduate work in economics or other social sciences might find this a good investment.

The lectures will not follow the textbook very closely, although most of the same material will be covered. Lecture notes are available at Copy Central on the North Side of campus (2483 Hearst). Additional material and lecture updates will be passed out in class.

The class will consist of two lectures (Tuesday and Thursday, 12:30-2 in 213 Wheeler Hall) and one discussion section per week. The section times are:

Section 101	Monday/Wed	4-5pm #2 Evans Hall (basement floor)
Section 102	Monday/Wed	5-6pm #2 Evans Hall (basement floor)

The discussion sections will present some additional material **(for which all students will be responsible)** and will also cover the answers from problem sets, old exams, problems from the lectures, etc.

Weekly problem sets will be handed out, starting in the second week of the course. Problem sets are due at the end of the Tuesday lecture. **We will not accept late problem sets.** Instead, we will drop your two worst scores in calculating the problem set average score. Thus, you can miss up to 2 problem sets without any penalty.

Course grades will be determined by a combination of grades on weekly problem sets (20 percent), two midterm exam (15 percent each), and the final exam (40 percent). The midterm exams will be held in class, Thursday October 2 and Thursday November 13.

Economics 101A Fall 2008 Approximate Course Outline A1+A41

Lecture	Content	Lecture Notes Reference	Text Reference	Suggested Problems in Text
1	Optimization Methods	1	Chapter 2	2.9, 2.10
2	Consumer Preferences and Consumer Choice	2	Chapter 3 Ch 4, pp. 94-105	3.2, 3.3, 3.8
3	Applications of Indifference Curve Analysis Expenditure Functions	3	Chapter 4 Ch. 5 Ex E.1	4.7, 4.8
4-5	Comparative Statics and the Slutsky Eq.	4	Chapter 5	5.4, 5.10
6	Market Level Demand and Supply	5	Chapter 10 pp. 292-295	10.4
7	Labor Supply	6	Chapter 16	16.2, 16.5
8	Intertemporal Consumption and Saving	7	Chapter 17	17.8, 17.10 (advanced)
9-11	Production and Cost Sheppard's Lemma	8, 9	Chapters 7-8 Example 8.4	8.9,8.10
12-13	Supply Determination	10-11	Chapters 9, 10	
14	Monopoly	12	Chapter 13	13.1, 13.2, 13.7
15	Consumer/Producer Surplus and Applications	13	Chapter 11 Ch. 10, pp. 306-308 Ch. 5, pp. 145-150	11.1-11.8
16-17	Duopoly	14,15	Chapter 14	14.2
18-20	Game Theory	16-17	Chapter 15	15.1, 15.4
21-24	Uncertainty	18-21	Chapters 18-19	18.5, 18.6, 18.7
25-26	Auctions	22-23	Ch. 15 Ex 15.9	15.12
27-28	Finance: CAPM and Efficient Markets	24-25		
27-28	Public Goods and Externalities	26-27		
28	Empirical Methods	28		

Lecture 1

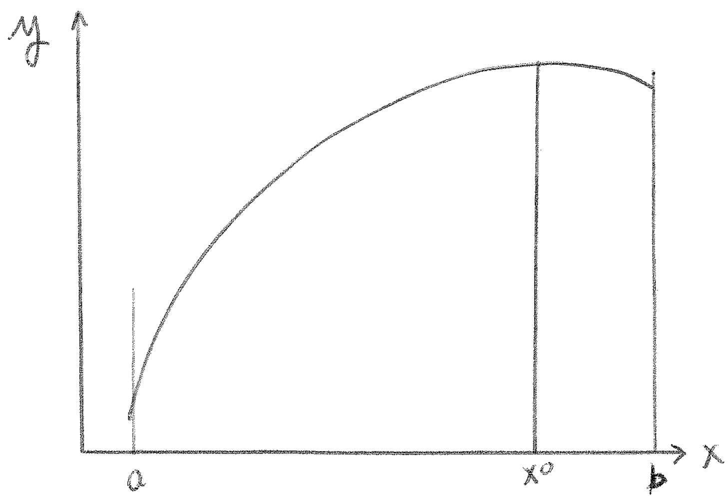
Optimization

I. Unconstrained Optimization

Suppose we had a function

$$y = f(x) \qquad a \le x \le b.$$

How could we go about finding a point x^0 such that $y^0 = f(x^0)$ is as big as (bigger) than f(x) for any other x?



In this picture $f(x^0) = \max_{a \le x \le b} f(x)$. Read this as "$f(x^0)$ is the maximum of f(x) when x is selected from the interval $a \le x \le b$ ".

What can we say generally? Obviously, if x^0 is a potential candidate for a maximizer, then it better be true that we can't move around x^0 and get to a higher function value. But this says that $f'(x^0) = 0$. Why?

If $f'(x^0) > 0$ then $f(x^0 + \epsilon) > f(x^0)$ for a small positive ϵ, using the def'n of $f'(x^0)$

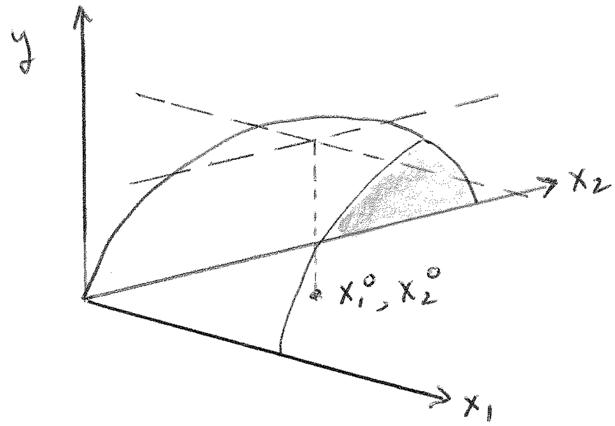
If $f'(x^0) < 0$ then $f(x^0 - \epsilon) > f(x^0)$ for a small positive ϵ.

This leads to:

RULE 1 if $f(x^0) = \max_{a \le x \le b} f(x)$ and $a < x < b$, then $f'(x^0) = 0$.

This is called the first order necessary condition (FONC) for an interior max.

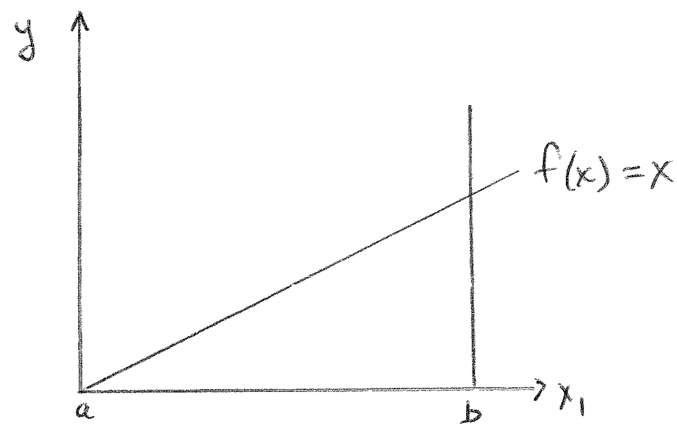
Notice that the same rule holds for a function of several variables:



Does $f'(x^0) = 0$ always mean that x^0 is a maximizer? Are there maximizers with $f'(x^0) \neq 0$?

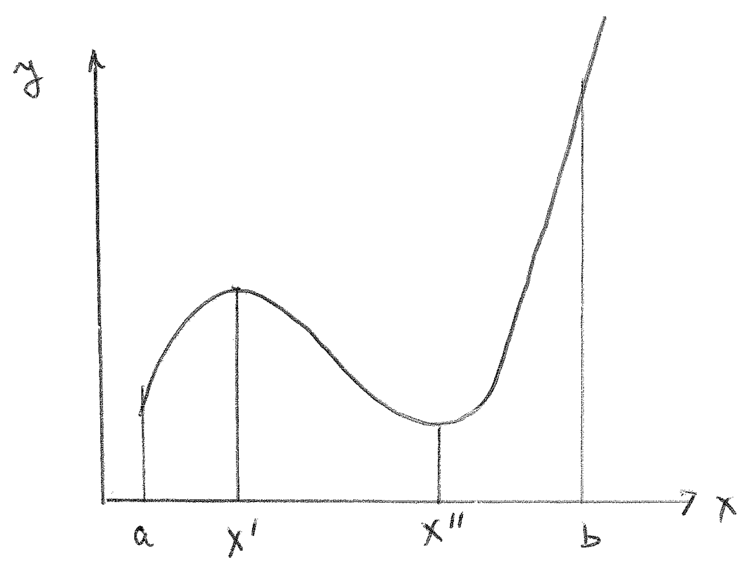
Consider the following examples:

(a)



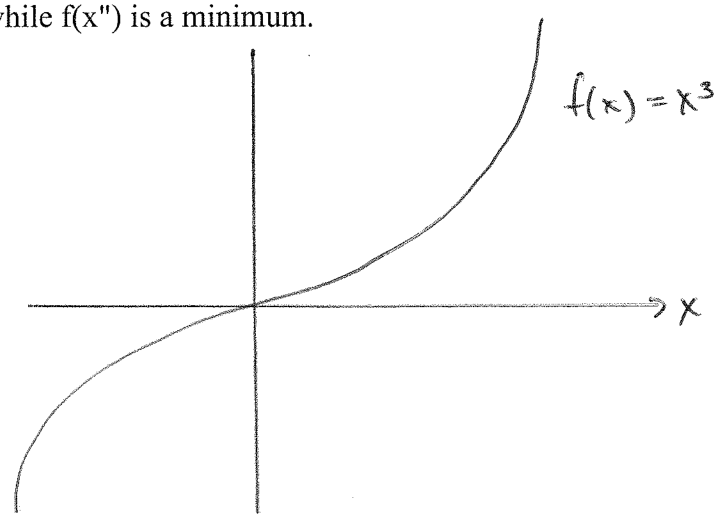
In this example $f(x) = x$. Thus $f(b) = \max_{a \le x \le b} f(x)$. But $f'(b) = 1$. Max f(x) occurs on the boundary.

(b)



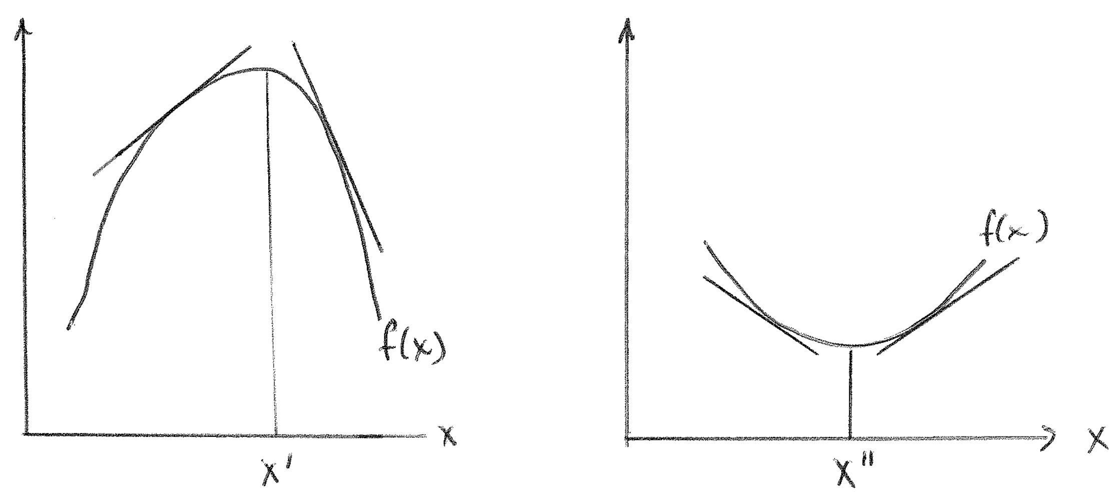
In this example $f'(x^0) = 0$ has two solutions: x' and x". But neither is the maximum. f(x') is a local maximum while f(x") is a minimum.

(c)

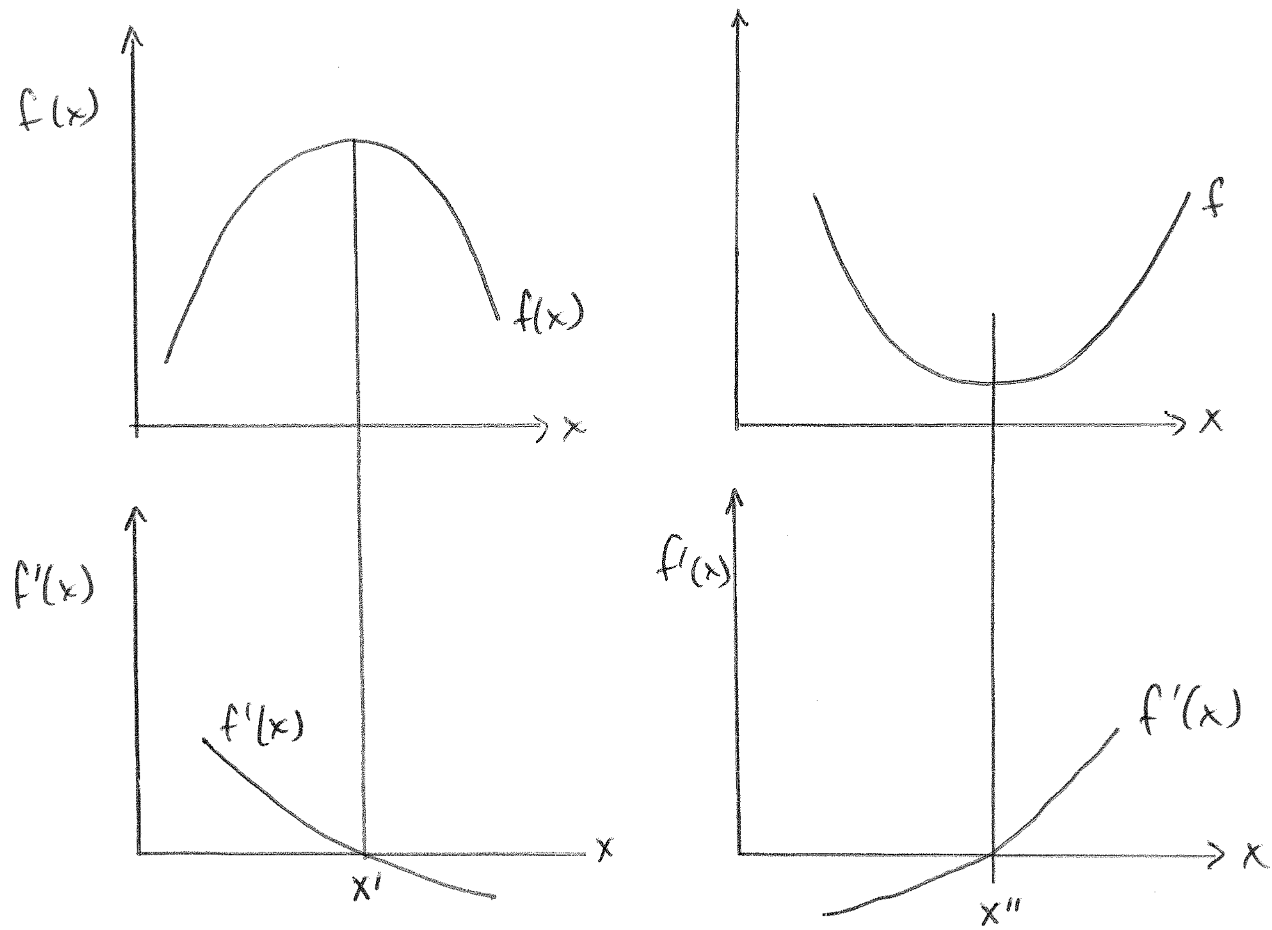


In this example $f(x) = x^3$. Solving $f'(x) = 0$ for x gives x = 0, which is a point of inflection.

How can we be sure that we have a maximum (not a minimum)? Consider the properties of $f'(x)$, **which is itself a function of x**:



At x' note that the function $f'(x)$ approaches from the left as a positive function, and approaches from the right as a negative function. At x", on the other hand, the function comes in from the left positive and comes in from the right negative. At a local maximum the function $f'(x)$ is negatively sloped or in other words $f''(x)$ is negative. At a local minimum the function $f'(x)$ is positively sloped or $f''(x)$ is positive. What happens at a point of inflection?



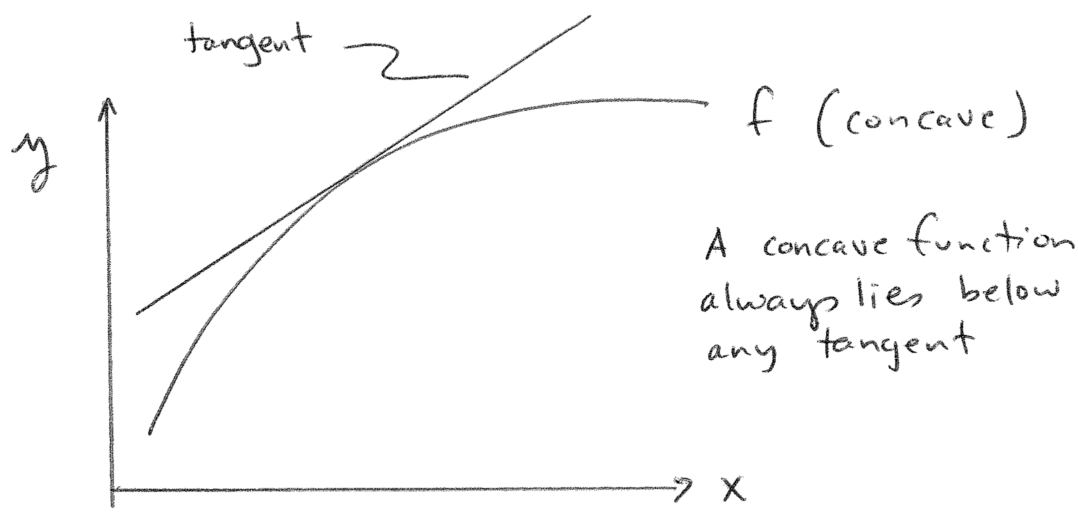
These considerations lead us to:

<u>RULE 2</u> If $f'(x^0) = 0$ and $f''(x^0) < 0$ then x^0 is a local maximum.

If $f'(x^0) = 0$ and $f''(x^0) > 0$ then x^0 is a local minimum.

This rule generalizes to 2 or more dimensions.

How can we be sure that a local maximum is a global maximum? If $f''(x) < 0$ for all x and $f'(x^0) = 0$ then x^0 is a global maximum. A function with $f'' < 0$ for all x is called a concave function.



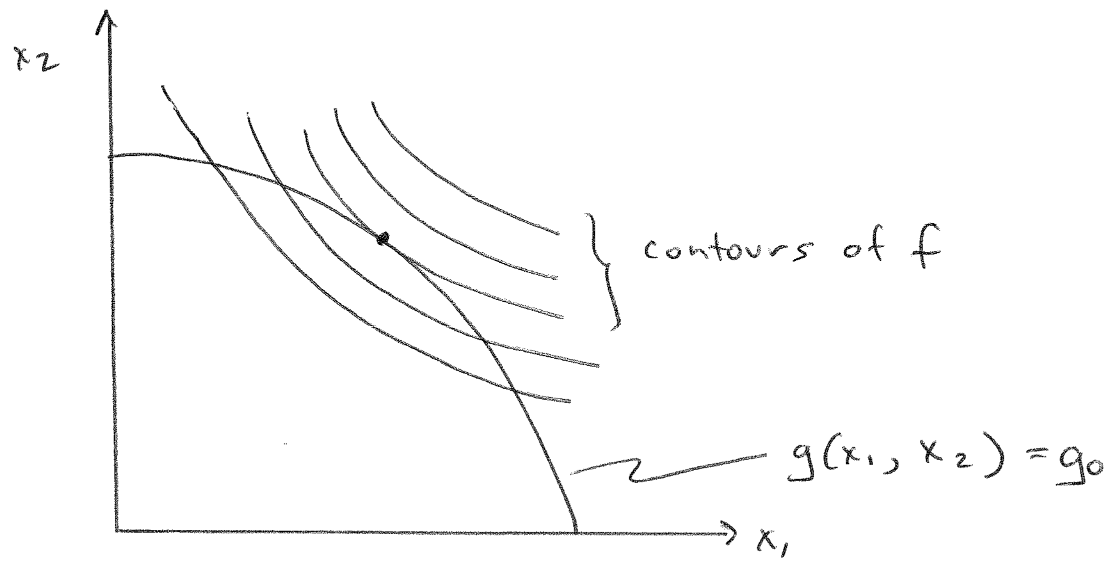
II. Constrained Optimization

Now we consider maximizing a function $f(x_1,x_2)$ subject to some constraint on x_1 and x_2 which we will denote $g(x_1, x_2) = g_0$. The two important examples are:

a) in the study of consumer behavior: maximize utility $U(x_1, x_2)$ subject to the budget constraint $p_1x_1 + p_2x_2 = I$.

b) in the study of firm behavior: maximize profit py - wx s.t. the production function $y= f(x)$.

How would you go about graphically analyzing the problem max $f(x_1, x_2)$ s.t. $g(x_1,x_2) = g_0$?



A two-step approach:

(1) plot the contours of the function $g(x_1,x_2)$.

E.g. $g(x_1,x_2) = x_1^2 + x_2^2$; $g(x_1,x_2) = k$ gives a circle with radius $k^{1/2}$ around the origin.

(2) plot the contours of the function $f(x_1,x_2)$.

E.g. $f(x_1,x_2) = x_1x_2$; $f(x_1,x_2) = m$ gives a hyperbola.

The constrained maximum of the function f occurs where a contour line of f( ) is tangent to the contour line of the function $g(x_1,x_2)$ corresponding to g_0.

Why? Suppose I consider adding a "bit" dx_1 to x_1 in such a way as to keep $g(x_1, x_2)$ constant. If this is true then I must have a corresponding <u>reduction</u> in x_2 such that the total differential of g is zero: i.e.

$$dg = g_1(x_1, x_2)dx_1 + g_2(x_1, x_2)\, dx_2 = 0$$

implying that $\dfrac{dx_2}{dx_1} = - \dfrac{g_1\,(x_1,x_2)}{g_2\,(x_1,x_2)}$.

If I increase x_1 by 1 unit, I must change x_2 by $-g_1(\,)\,/\,g_2(\,)$ in order to keep the value of g constant. The net effect of such a change in x_1 on the value of the function f is:

$$df = f_1(x_1,x_2)dx_1 + f_2(x_1,x_2)dx_2 =$$

$$f_1(x_1,x_2)\; dx_1 + f_2(x_1,x_2)\left|\frac{dx_2}{dx_1}\right|dx_1 \quad =$$

$$[\;\; f_1(x_1,x_2) - f_2(x_1,x_2) \cdot \frac{g_1(x_1,x_2)}{g_2(x_1,x_2)} \;\;]\;\; dx_1.$$

now in order for (x_1^0,x_2^0) to be a constrained optimum, it must be true that I can't increase f by either adding or subtracting a bit of x_1 <u>and</u> keeping the value of the constraint function g

$$\frac{f_1(x_1^0,x_2^0)}{f_2(x_1^0,x_2^0)} = \frac{g_1(x_1^0,x_2^0)}{g_2(x_1^0,x_2^0)} \quad .$$

constant. But this means that the above expression must be zero for all dx_1, or in other words, But this expression says that at the constrained optimum, the contours of $f(x_1,x_2)$ and $g(x_1\,,x_2)$ are tangent (they have the same slope). Note that this argument only works if the proposed optimum is not on the boundary, so that both x_1 and x_2 can be increased or decreased.

How could I convert a constrained maximization problem into an unconstrained one? A mathematician named Lagrange noted that you get the right answer by setting up an "artificial" (unconstrained) optimization problem with a new variable λ:

$$L(x_1,x_2,\lambda) = f(x_1,x_2) - \lambda (g(x_1,x_2) - g_0),$$

and then finding the first order necessary conditions for maximizing the function $L(x_1,x_2,\lambda)$ with respect to x_1,x_2, and the new variable λ. The first order conditions are the following:

$$L_1 (x_1,x_2,\lambda) = f_1(x_1,x_2) - \lambda g_1(x_1,x_2) = 0,$$

$$L_2 (x_1,x_2,\lambda) = f_2(x_1,x_2) - \lambda g_2(x_1,x_2) = 0,$$

$$L_\lambda (x_1,x_2,\lambda) = -g(x_1,x_2) + g_0 = 0.$$

Dividing the first of these by the second gives the equation:

$$\frac{f_1(x_1,x_2)}{f_2(x_1,x_2)} = \frac{g_1(x_1,x_2)}{g_2(x_1,x_2)} \quad ,$$

while the third first order condition simply restates the constraint! Thus by writing down the "Lagrangian"and setting its first derivatives to zero, we get the necessary conditions for a constrained optimum.

We also get a new variable, λ, called the Lagrange multiplier. What is its interpretation? It turns out that the value of λ tells you how much the maximized value of $f(x_1,x_2)$ would increase if you relaxed the constraint a small amount. Suppose that you were to maximize f ( ) subject to the constraint $g(x_1,x_2) = g_0$. Call the solution (x_1^0, x_2^0). Now suppose you find that you actually have $g_0 + dg_0$ units of constraint to use up (in other words, the constraint is really $g(x_1,x_2) = g_0 +dg_0$). How would you change your choices of x_1 and x_2? Suppose you decide to use more x_1: enough to "use up" the added constraint. Since the total differential of $g(x_1, x_2)$ is

$$dg = g_1(x_1,x_2)\, dx_1 + g_2(x_1,x_2)dx_2,$$

if you change <u>only</u> x_1, the amount you can change x_1 to satisfy the new constraint is:

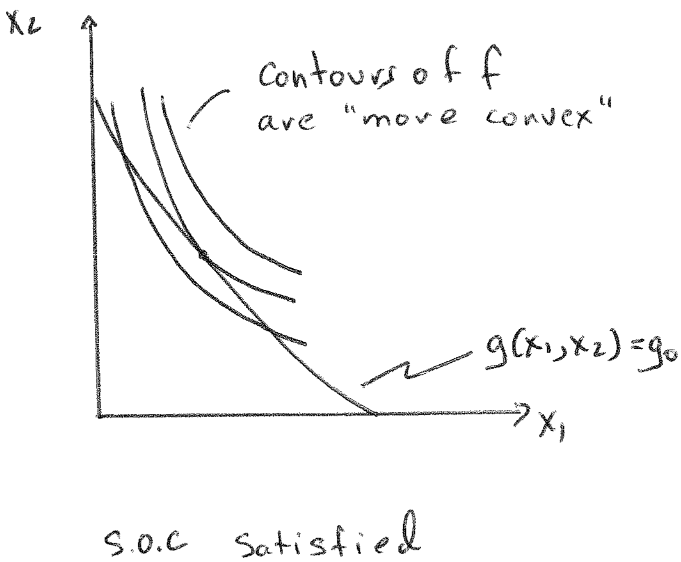
$$dx_1 = \frac{1}{g_1(x_1^0, x_2^0)} \, dg_0.$$

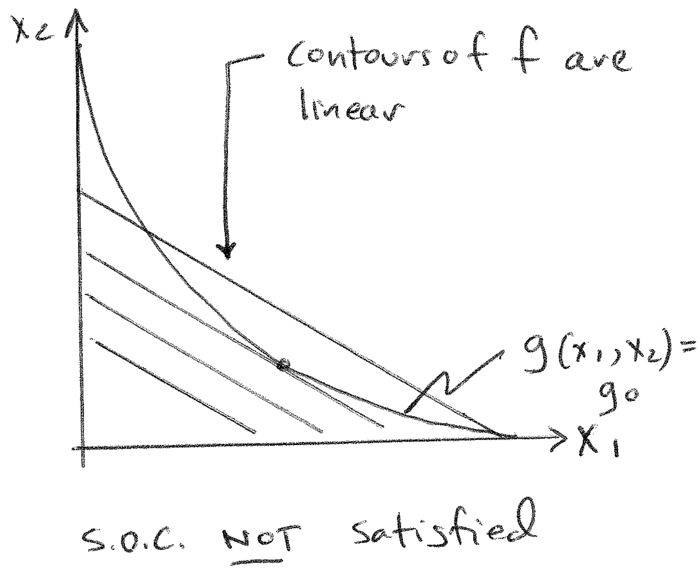
With this increase in x_1, the increase in the value of our objective function $f(x_1,x_2)$ is:

$$df = f_1(x_1^0, x_2^0) \, dx_1 = \frac{f_1(x_1^0, x_2^0)}{g_1(x_1^0, x_2^0)} = \lambda.$$

You should check for yourself that if you were to use up the extra constraint with x_2, the increase in the value of f would also be λ. This gives another interpretation of the "tangency" conditions: if we had a bit more constraint we would be indifferent as to using it up in terms of x_1 or x_2 at an optimum.

As in unconstrained optimization, there are also second order conditions. These can be expressed algebraically. They amount to the condition that the function to be maximized has contours that are “more convex” than the constraint function.





Lecture 2

Consumer's Optimum

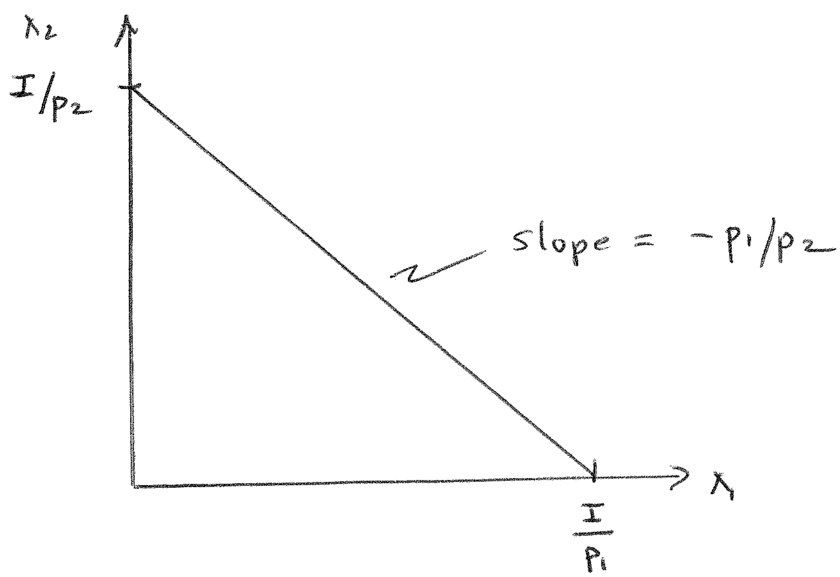
In this lecture, we apply the optimization techniques of the first lecture to analyze consumer choice subject to a budget constraint. There are three elements of the problem:

1) describe the constraint
2) describe the consumer's objective
3) set up and analyze the constrained optimization.

(1) Budget constraints

We assume that a consumer must choose among bundles of commodities (x_1, x_2, ... x_n) that satisfy her/his budget constraint. In the two good cases, prices of the first and second goods are p_1 and p_2, respectively. Assuming that money income is I, bundles (x_1, x_2) are affordable if and only if $p_1x_1 + p_2x_2 < I$

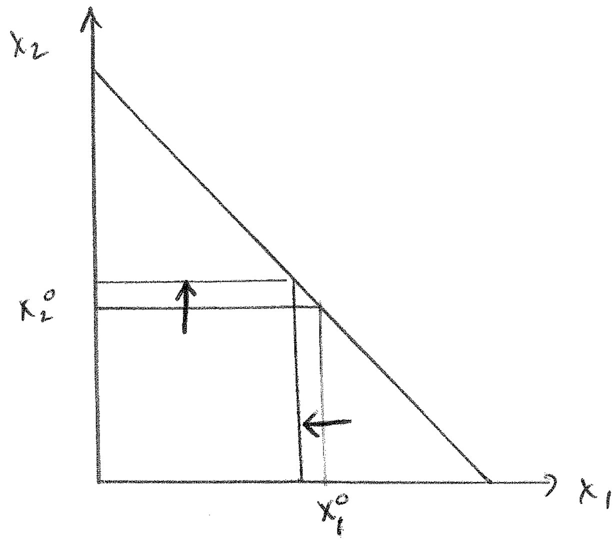
Graphically, the set of affordable bundles (also called the budget set) looks like a wedge:



Notes: (a) if all income is spent on x_1, the total amount of x_1 available is I/p_1.

(b) we are implicitly assuming that you can't buy negative amounts of x_1 or x_2.

(c) the slope of the "budget line" (the outer boundary of the budget set) is $-p_1/p_2$.

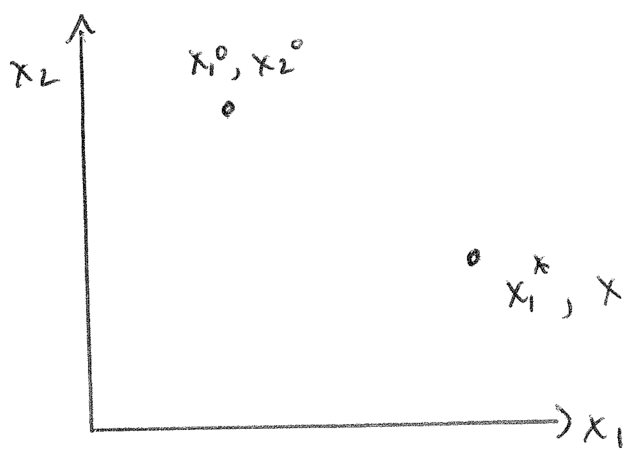


Why? If I give up 1 unit of x_1, I save $\$p_1$. If I then spent this on x_2, I can buy p_1/p_2 units of x_2. The market trades x_1 for x_2 at the rate p_1/p_2. This ratio represents the <u>relative price</u> of x_1 and x_2.

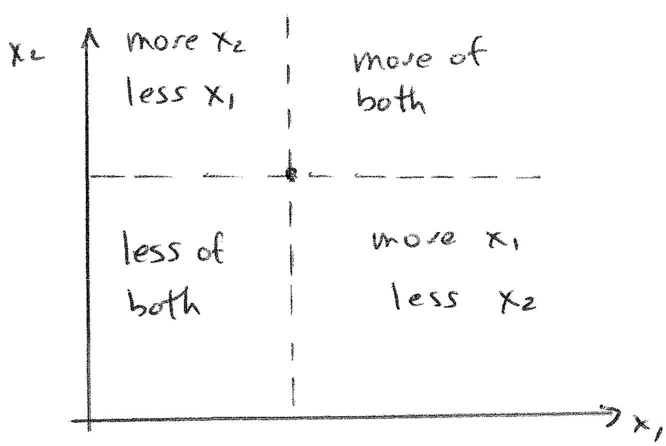
(2) Consumer Objectives

We want a simple way to summarize how the consumer evaluates alternative bundles:

say (x_1^0, x_2^0) vs. (x_1^*, x_2^*).



Graphically, the device we use is the indifference curve: a curve linking up bundles that are "as good." Consider an indifference curve for bundles "as good" as (x_1^0, x_2^0):



If both x_1 and x_2 are "desirable", then points with more of x_1 and x_2 must be better than (x_1^0, x_2^0). By the same token, points with less of x_1 and of x_2 must be worse. The only possibility is that indifference curves are negatively sloped.

In more advanced treatments of economic theory, indifference curves are "derived" from a set of assumptions on how people evaluate alternative bundles. Some types of preferences cannot be represented by indifference curves. The classic example is "lexicographic preferences": I evaluate a bundle (x_1, x_2) first by the amount of x_1, and then by the amount of x_2. Any bundle with more x_1 is strictly preferred. However, if two bundles have the same amount of x_1, I then look at the amount of x_2. (This is how 'alphabetical order' works). As an exercise – try to graph the indifference curves for a consumer with lexicographic preferences.

Analytically, we represent consumer preferences by a utility function $U(x_1, x_2)$ defined over the set of possible consumption bundles. Naturally, higher values of U are preferred (we choose U to make it true).

Examples: $U(x_1, x_2) = x_1 \cdot x_2$

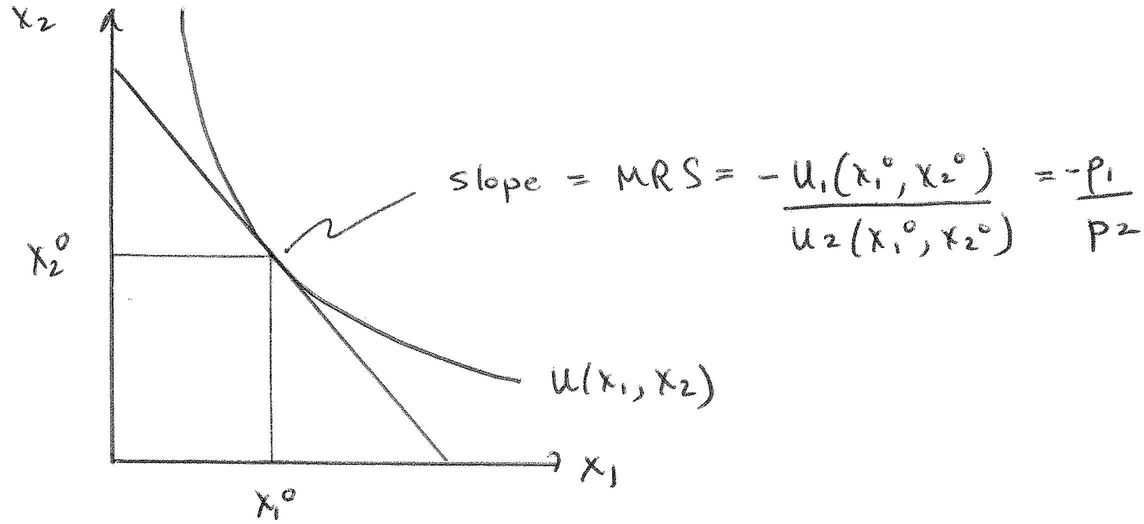
$U(x_1, x_2) = x_1 + x_2$

$U(x_1, x_2) = \min(x_1, x_2)$ (where min(a,b) = smaller of the two).

Facts: (a) The contours of the function U are the indifference curves. If (x_1^0, x_2^0) and (x_1^*, x_2^*) and are on the same indifference curve then $U(x_1^0, x_2^0) = U(x_1^*, x_2^*)$.

(b) If more of x_1 is always preferred, then $U(x_1 + dx_1, x_2) > U(x_1, x_2)$ for $dx_1 > 0$ or $U_1(x_1, x_2) > 0$. Similarly $U_2(x_1, x_2) > 0$.

(c) The slope of an indifference curve at a point (x_1^0, x_2^0) is $-\frac{U_1(x_1^0, x_2^0)}{U_2(x_1^0, x_2^0)}$. We call this the marginal rate of substitution. Why? Because it tells me how much x_2 I' have to get back per unit of x_1 given up in order to keep utility constant.



Examples: i) $U(x_1,x_2) = x_1^{\alpha} x_2^{\beta}$ (Cobb-Douglas)

$U_1(x_1,x_2) = \alpha\, x_1^{\alpha-1} x_2^{\beta}$

$U_2(x_1,x_2) = \beta\, x_1^{\alpha} x_2^{\beta-1}$

$$MRS(x_1,x_2) = \frac{U_1(x_1,x_2)}{U_2(x_1,x_2)} = \frac{\alpha\, x_2}{\beta\, x_1}$$

ii) $U(x_1,x_2) = x_1 + x_2$

$$MRS = \frac{U_1}{U_2} = 1, \textit{ constant for all } x_1,x_2.$$

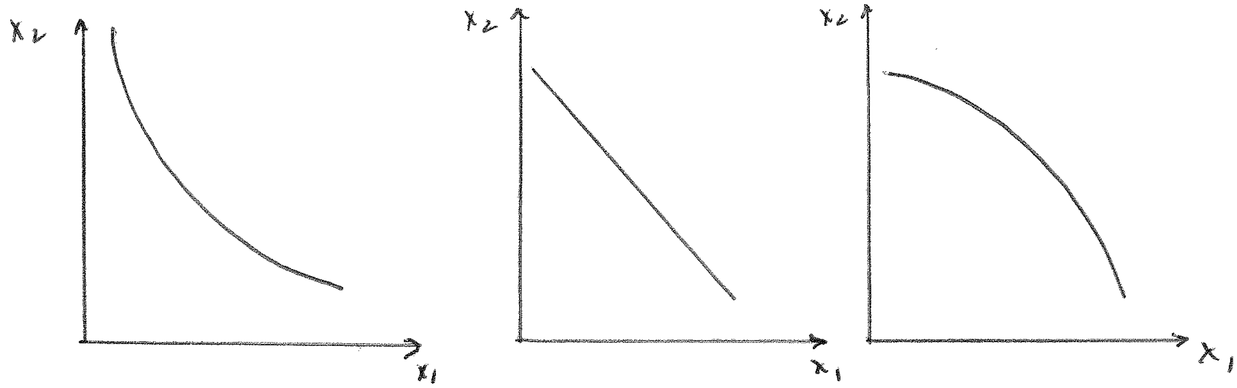
iii) $U(x_1,x_2) = 2 \log x_1 + x_2$

$$MRS = \frac{U_1}{U_2} = \frac{2/x_1}{1} = \frac{2}{x_2} \textit{ independent of } x_2.$$

Exercise: Graph the indifference curves for these three examples.

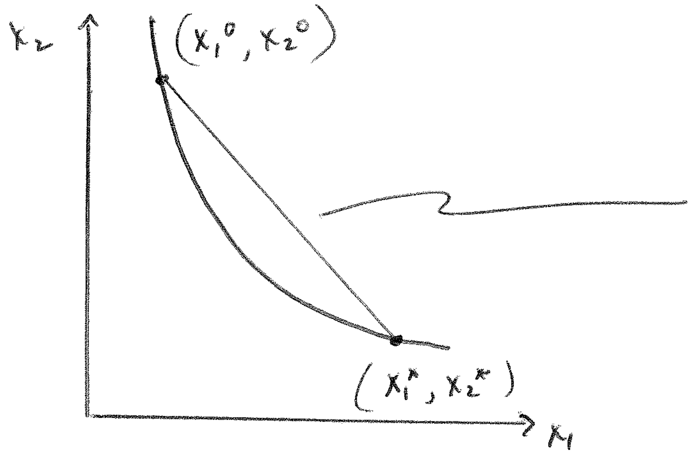
Note: If your utility function is U(x,y) and mine is V(x,y) = a + b·U(x,y) then we have the same preferences. Why? You can check that we have exactly the same indifference curves, with different labels on them. The same argument works for V(x,y) = F[U(x,y)] if F is increasing.

You may be familiar with the concept of diminishing marginal rates of substitution (DMRS).



DMRS Constant MRS Increasing MRS

Along an indifference curve (holding utility constant) the MRS is decreasing on x_1. As you get more x_1, relative x_2, you value additional units of x_1 less in terms of x_2. DMRS implies that people always prefer averages. Suppose I have two bundles on the same indifference curve: (x_1^0,x_2^0) and (x_1^*,x_2^*). Then the bundle that is the average of these two is preferred to either:



"average" bundle preferred

Equivalently, the set of points preferred to (x_1^0,x_2^0) is convex.

It is important to know that DMRS is not related to, nor is it the same thing as, "diminishing marginal utility". If $U(x_1,x_2)$ is the utility function, the marginal utility of x_1 is $U_1(x_1,x_2)$. We could say that U exhibits diminishing marginal utility if $U_{11}(x_1,x_2) < 0$. The sign of U_{11} says nothing about the MRS.

Check the following examples:

i) $$U(x_1,x_2) = (x_1^2 + x_2^2)^{1/4}$$

$$U_1(x_1,x_2) = 1/2\ x_1\ (x_1^2 + x_2^2)^{-3/4} < 0$$

$$U_{11}(x_1,x_2) = -3/4\ x_1^2\ (x_1^2 + x_2^2)^{-7/4} < 0$$

but the indifference curves are circles, implying an increasing marginal rate of substitution.

ii) $$U(x_1,x_2) = x_1^3 x_2^3$$

$$U_1(x_1,x_2) = 3\ x_1^2 x_2^3 > 0$$

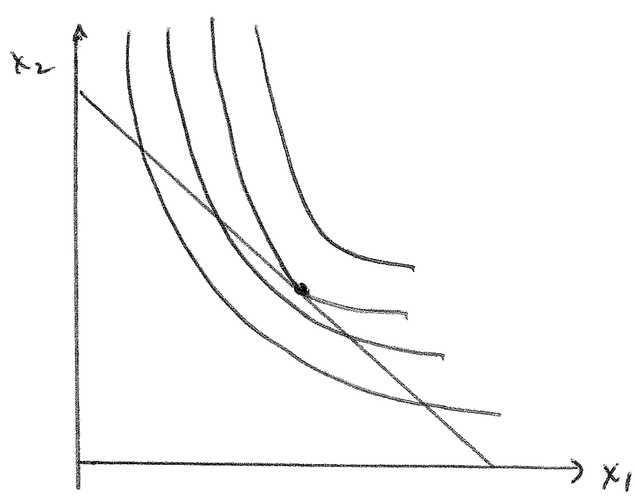
$$U_{11}(x_1,x_2) = 6x_1\ x_2^3 > 0$$

but the indifference are hyperbolas, implying a diminishing marginal rate of substitution.

<u>Consumer</u> <u>Optimum</u>

Analytically, the consumer's problem can be expressed as finding the solution to

$$\max U(x_1,x_2) \quad \text{s.t.} \quad p_1x_1 + p_2x_2 = I.$$

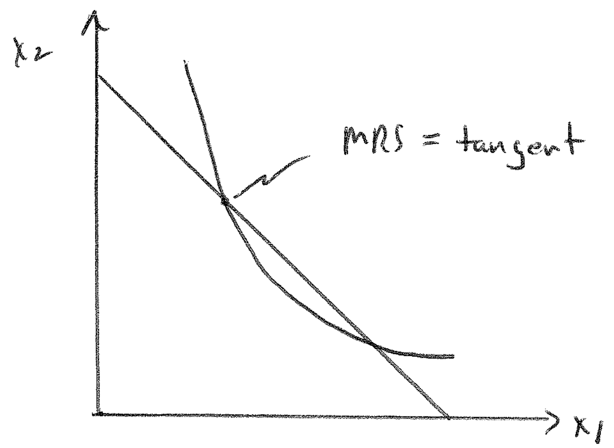


Clearly, we have at an optimum at (x_1^0, x_2^0) if two things are true:

(1) $p_1 x_1^0 + p_2 x_2^0 = I$ The budget constraint is satisfied exactly. (Why?)

(2) $MRS(x_1^0, x_2^0) = \frac{p_1}{p_2}$

Condition (2), the tangency condition, expresses the simple idea that there are no more gains from trade with the market. If MRS $>p_1/p_2$, the consumer values x_1 more than the market (in terms of x_2), so it would pay to sell x_2 back to the market and buy more x_1. Here is the picture when MRS > relative prices in the market:



$$\text{MRS} > \frac{p_1}{p_2}$$

On the margin, the consumer values x_1 in terms of x_2 more than the market and there is room for a profitable trade! What happens if MRS $<p_1/p_2$?

To proceed analytically, let's use the Lagrangian technique:

$$L(x_1,x_2,\lambda) = U(x_1,x_2) - \lambda(p_1x_1 + p_2x_2 - I)$$

i) $\frac{\partial L}{\partial x_1} = U_1(x_1,x_2) - \lambda p_1 = 0$

ii) $\dfrac{\partial L}{\partial x_2} = U_2(x_1,x_2) - \lambda p_2 = 0$

iii) $\dfrac{\partial L}{\partial \lambda} = -p_1x_1 - p_2x_2 + I = 0$

Divide the first-order condition (i) by the first-order condition (ii):

$\dfrac{U_1(x_1,x_2)}{U_2(x_1,x_2)} = \dfrac{p_1}{p_2}$, the tangency condition.

Also

$$\lambda = \frac{U_1(x_1,x_2)}{p_1} = \frac{U_2(x_1,x_2)}{p_2}$$

If I had an extra dollar to spend, I could either

(a) buy $1/p_1$ units of x_1 → net increase in utility = $\dfrac{U_1(x_1,x_2)}{p_1} = \lambda$

(b) buy $1/p_2$ units of x_2 → net increase in utility = $\dfrac{U_2(x_1,x_2)}{p_2} = \lambda$.

For this reason, λ is sometimes called the marginal utility of income.

Example $U(x_1,x_2) = x_1\, x_2 \Rightarrow L(x_1,x_2,\lambda) = x_1x_2 - \lambda(p_1x_1 + p_2x_2 - I)$

FONC: i) $x_2 - \lambda p_1 = 0$

ii) $x_1 - \lambda p_2 = 0$

iii) $-p_1x_1 - p_2x_2 + I = 0$

Therefore $x_1 = \lambda p_2$, $x_2 = \lambda p_1$. Substituting into budget constraint

$$p_1 x_1 + p_2 x_2 = \lambda p_1 p_2 + \lambda p_1 p_2 = I$$

$$\rightarrow \quad \lambda = I / (2\, p_1\, p_2)$$

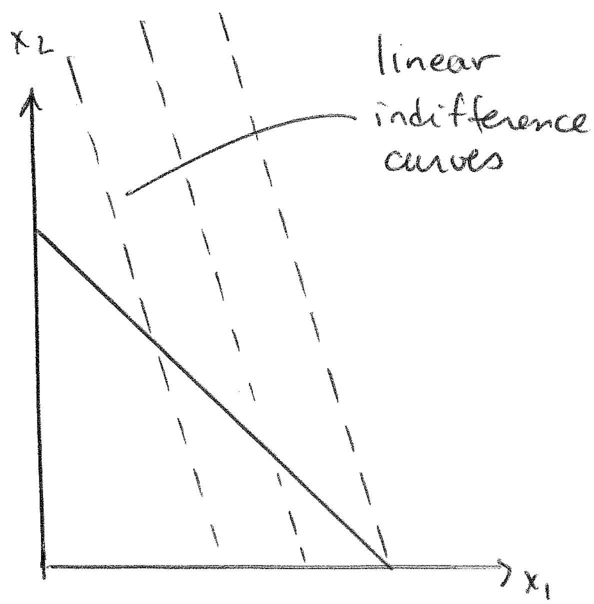
$$\rightarrow x_1 = \frac{I}{2p_1} \qquad x_2 = \frac{I}{2p_2}$$ these are the "demand functions".

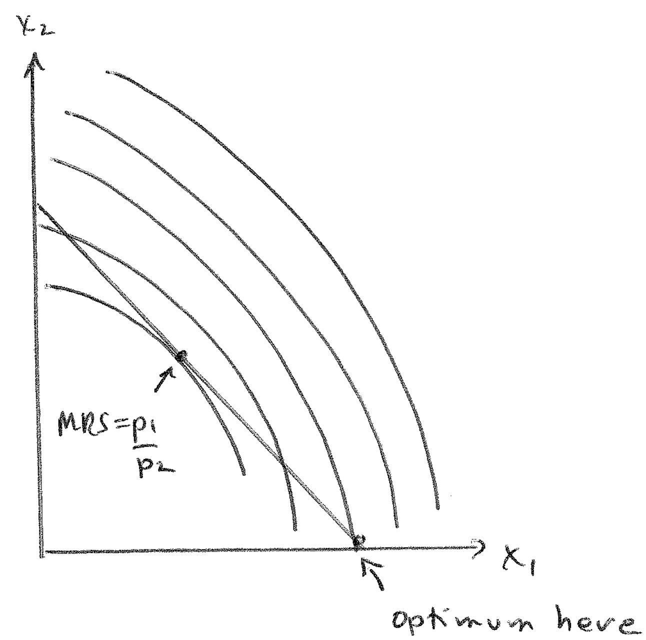
Notice that $x_1 p_1 = I/2$ and $x_2 p_2 = I/2$. So the consumer spends 1/2 of income on each good!

As an exercise: re-do the analysis with $U(x_1,x_2) = x_1^{\alpha}\ x_2^{\beta}$ for different choices of α and β.

Special Problems

(i) Preferences don't satisfy DMRS



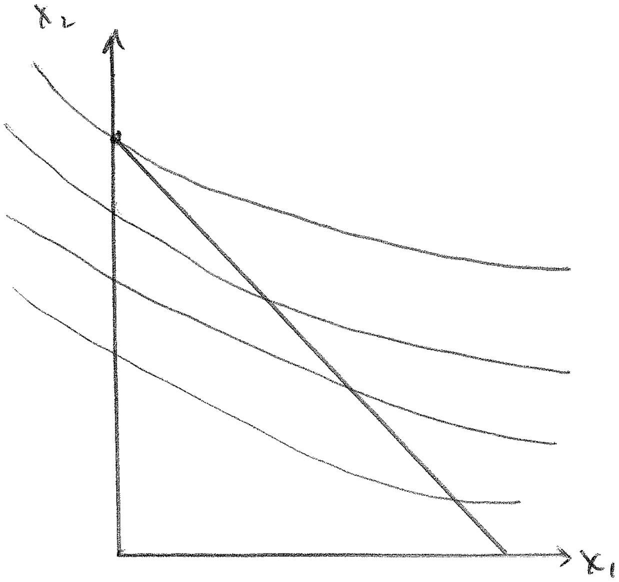


MRS is constant. There is no point where MRS = p_1/p_2.

At this point MRS=p_1/p_2 but this is not an optimum. What's wrong?

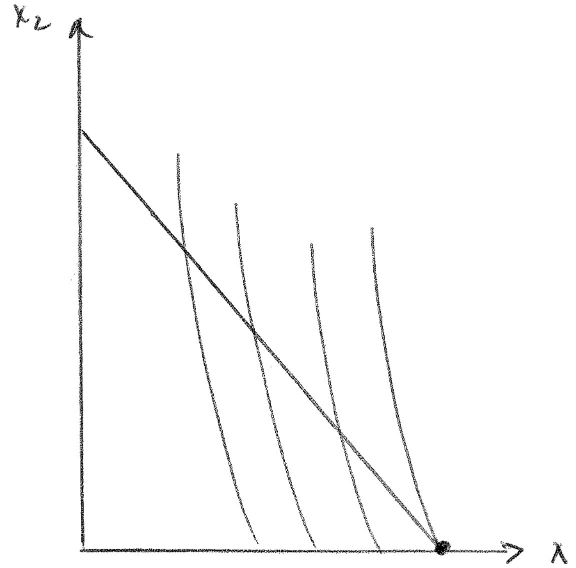
Often, we restrict preferences by requiring that the indifference curves are convex to the origin. A function that has this property is "quasi-concave". Mathematically, a quasi-concave function u(x) has the property that { x : u(x) ≥ k } is a so-called "convex set".

(ii) Even with quasi-concave utility (or "convex" indifferences curves) we can still have problems:



Endpoint optimum

MRS < p_1/p_2 ; $x_1 = 0$



Endpoint Optimum

MRS > p_1/p_2 ; $x_2 = 0$

Most consumers consume 0 amounts of most goods. So the "endpoint problem" is potentially one that economists have to deal with. The problem is much worse, the more narrowly goods are defined (e.g., Coke vs. Pepsi), and becomes less serious, the broader the definition (e.g., beverages). A lot of applied research on consumer demand that looks at individual choices uses a so-called "discrete choice" approach, focusing on whether consumers buy some or none of a commodity. Daniel McFadden won the Nobel prize for his research showing how to link up the "buy – don't buy" decision to underlying utility functions.

Lecture 3

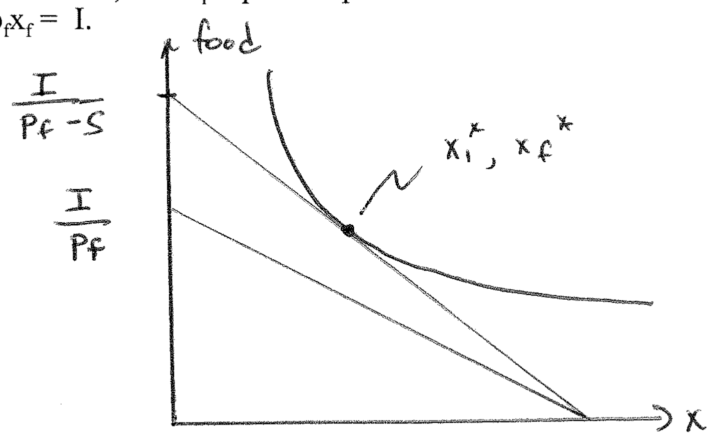
Two Applications of Indifference Curve Analysis

We have seen that the consumer's optimum is represented by a tangency between an indifference curve and the budget constraint. This condition expresses the simple economic idea that the consumer, on the margin, cannot adjust her consumption bundle to spend the same amount and achieve a higher "utility". Recall that the tangency condition is only true when the indifference curves exhibit diminishing marginal rates of substitution, and we don't have an endpoint optimum.

Application 1: Analysis of a Subsidy

In many economies, certain commodities are subsidized by the government. A subsidy is a negative tax which is usually introduced to aid low income consumers. Economists generally argue that subsidies are inefficient. Why?

There are two commodities: food and "other things". The price of other things is p_1, and the price of food is p_f. A typical consumer has income I and the usual kind of preferences. Let x_f represent purchases of food, and x_1 represent purchases of all other "stuff". The budget constraint is $p_1x_1 + p_fx_f = I$.



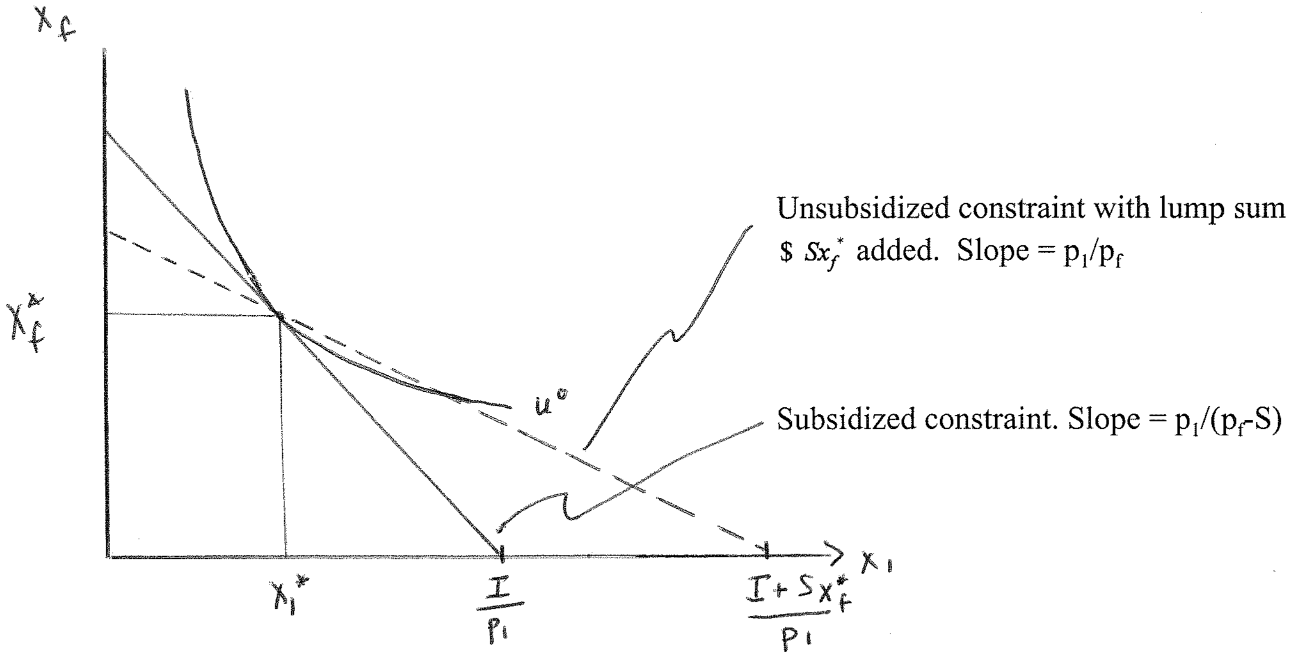
Suppose now that a subsidy of $S per unit is introduced on food. The budget constraint becomes: $p_1x_1 + (p_f - S)\, x_f = I$. If the consumer chooses the bundle (x_1^*, x_f^*) the cost of the subsidy program to the government (for this one consumer) is $\$\, S\, x_f^*$. Most economists would argue that you should give the consumer $\$\, S\, x_f^*$ directly and leave the price of food alone! To see the argument, suppose the subsidy amount is given to consumers directly, but they are forced to pay market (unsubsidized) prices. In this case, the budget constraint is

$$(*) \qquad p_1\, x_1 + p_f\, x_f = I + S\, x_f^* .$$

Notice that the bundle (x_1^*, x_f^*) satisfies this budget constraint, since originally

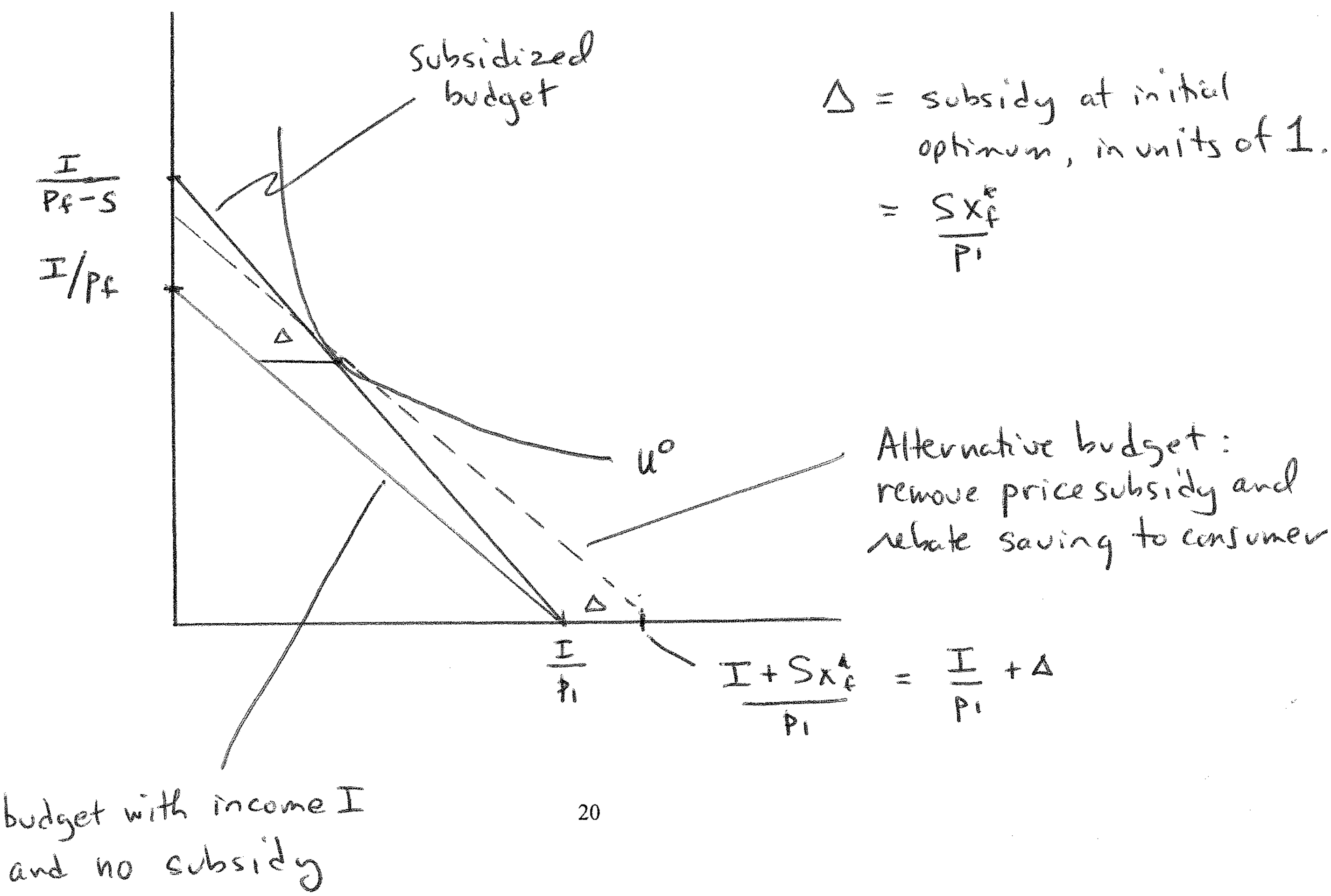
$$p_1\, x_1^* + (\, p_f - S)\, x_f^* = I .$$

In other words, if I give the consumer $\$\, S\, x_f^*$ he/she can still afford (x_1^*, x_f^*). But she can do even better, as shown in the following diagram:



The reason is that the budget constraint (*) with the lump sum subsidy is flatter than the budget constraint with the subsidy on food. But they both pass through the point (x_1^*, x_f^*). So the constraint (*) cuts through an indifference curve and allows access to a bundle with higher utility.

The following diagram shows another way to see the same point:



Application 2: The Consumer Price Index

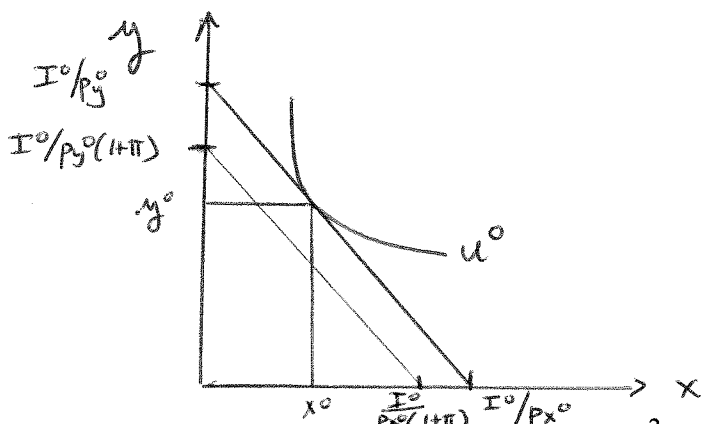
The CPI tells how much it costs today (in today's dollars) to buy a fixed bundle of commodities. Currently we use a 1982-84 = 100 bases, which means that the CPI is calculated by finding the cost of the bundle, relative to its cost in 1982-84.

Suppose that the CPI is 177.5 (which was its value in July 2001). That means that it now costs 1.775 times as much to purchase the "standard bundle" as it did on average in 1982-84. If someone (your parents) earns 1.78 times as much as they did in the early 1980s they are as well off as they were then.

Does your nominal income necessarily have to rise as fast as the CPI? Suppose that in 1983 you purchased (x^0, y^0) at prices (p_x^0, p_y^0) . Your incomes was I^0, and

$$x^0 p_x^0 + y^0 p_y^0 = I^0$$

Now suppose (in 2001) prices are $p_x = p_x^0(1 + \pi)$, $p_y = p_y^0(1 + \pi)$. In this case, both prices increased at a rate π. How much would your income have to increase by?



In order to get back to the original indifference curve, income will have to increase by π!

On the other hand, suppose p_x rises by $\frac{3}{2}\pi$ and p_y rises by $\frac{1}{2}\pi$: i.e.

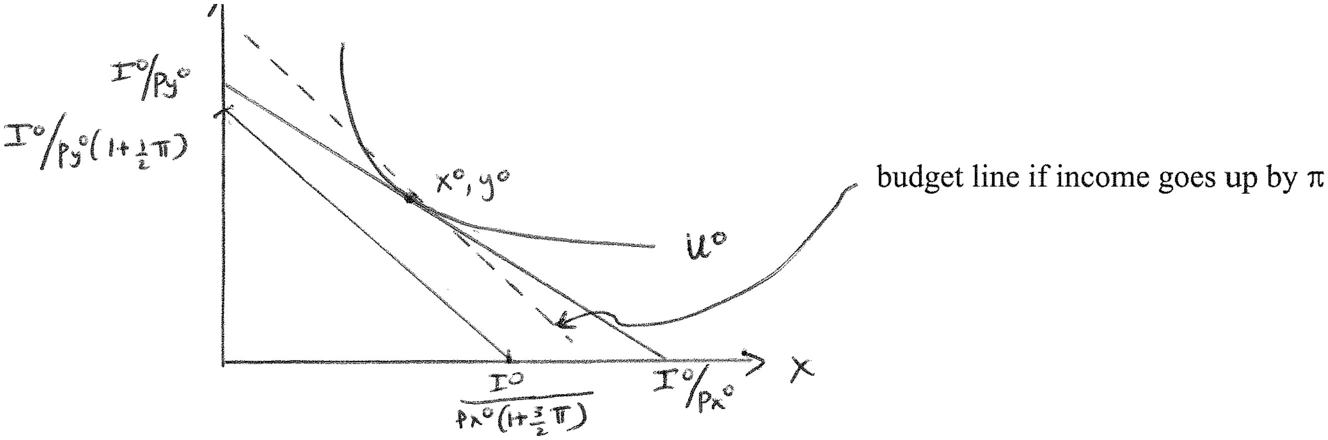
$p_x = p^0(1 + \frac{3}{2}\pi)$; $p_y = p^0(1 + \frac{1}{2}\pi)$. The increase in the cost of living is represented by the increase in the cost of the original or "reference" bundle x^0,y^0 :

$$p_x^0(1 + \frac{3}{2}\pi)x^0 + p_y^0(1 + \frac{1}{2}\pi)y^0 - p_x^0 x^0 - p_y^0 y^0 = \frac{3}{2}\pi p_x^0 x^0 + \frac{1}{2}\pi p_y^0 y^0.$$

If initially you spent one-half of your income on each x and y , $p_x^0 x^0 = p_y^0 y^0 = I^0/2$, and the increase in the cost of living is

$$\frac{3}{2}\pi\ [\frac{I^0}{2}] + \frac{1}{2}\pi\ [\frac{I^0}{2}] = \pi I^0$$

– a proportional increase of π. But, if your income goes up by π, you are <u>better off</u>!



The reasoning is as follows: If your income goes up by enough to allow you to buy (x^0,y^0) you will get the dashed budget line. But with that budget line, you will not consumer (x^0,y^0): You'll consume a bundle with more y, less x, and <u>higher utility</u>. You respond to the change in relative prices by altering your consumption.

<u>Example</u>: The CPI is really a weighted average of prices for a fixed set of purchases. Here are some of the major categories and their weights:

Category	Weight	Price Index (Dec. 2000)
All	100.0	174.1
Food&Beverages	16.3	169.5
Housing	39.6	171.6
Apparel	4.7	131.8
Transportation	17.5	155.2
Medical	5.8	264.1
Recreation	6.0	103.7 *
Education	2.7	115.4 *
Communication	2.7	92.3 *
Other Items	4.7	276.2

Price Indexes are 1982-84 = 100 except the ones noted by *, which are Dec. 1997.

Note the slow growth of apparel prices (usually attributed to the very rapid rise in cheap imports) and the very rapid rise in medical prices.

The difference between the rate of increase in the average price of the reference bundle and the minimum increase in income you'd need to maintain the original level of utility is called the "substitution bias" in the CPI. Note that it depends on 2 things: how unequally prices are rising for different things, and how much "curvature" there is in indifference curves. The more curvature, and the more dispersion in relative prices changes, the bigger is the substitution bias. The "Boskin Commission" estimated that on average, substitution bias was about ½ percent per year in the U.S. over the past couple of decades.

There are lots of other (bigger) sources of bias in the CPI. One that is very hard to measure is "quality bias": over time, the things that people buy change and its often hard to hold constant the reference bundle. Some new inventions since the early 1980s: CD and DVD players, air bags and anti-lock brakes, the internet, laser printers, portable PC's, cell phones, the X-files. Roughly speaking, quality changes are handled in the CPI by attempting to subtract off the part of any price change that is due to quality, measured at the time of the introduction of the higher quality product. So for example, when air bags first became available manufacturers charged about $500 extra for them. Thus, when we compare the price of new car in 2001 (that has air bags) to a "similar model" in 1990 (that didn't), we subtract off $500 from the 2001 price before we compute the price ratio, assuming that the $500 is for a higher quality car.

Lecture 3: Additional Material

A. Indirect Utility

We characterized the solution to the problem

$$\max U(x_1, x_2) \quad \text{s.t.} \quad p_1 x_1 + p_2 x_2 = I$$

as an optimal pair (x^0_1, x^0_2) that satisfy the first order conditions (tangency condition and budget constraint). Note that (x^0_1, x^0_2) vary with (p_1, p_2, I). We call the optimal choices at a given level of prices and incomes the "demand functions", and write:

$$x_1 = x^0_1 (p_1, p_2, I)$$

$$x_2 = x^0_2 (p_1, p_2, I)$$

Note that $p_1 x^0_1 (p_1, p_2, I) + p_2 x^0_2 (p_1, p_2, I) = I$, so the demand functions have to satisfy the budget constraint *as we vary prices.* This puts some restrictions on the functions.

The highest level of utility that can be achieved at (p_1, p_2, I) is $U(x^0_1 (p_1, p_2, I), x^0_2 (p_1, p_2, I))$, which is the utility of the optimal choices for those budget parameters. We introduce a new function

$$v(p_1, p_2, I) \quad = \max U(x_1, x_2) \quad \text{s.t.} \quad p_1 x_1 + p_2 x_2$$

$$= U(x^0_1 (p_1, p_2, I), x^0_2 (p_1, p_2, I))$$

This is called the **indirect utility function**. In tutorial this week we will discuss the derivatives of v. It should be clear that v is decreasing in p_1 and p_2, and increasing in I.

Example $\quad U(x_1, x_2) = x_1^{\alpha} x_2^{\beta} \quad$ where $\alpha+\beta=1$

From last lecture, $x^0_1 (p_1, p_2, I) = \alpha / (\alpha+\beta) \; I / p_1 = \alpha \; I / p_1$; $x^0_2 (p_1, p_2, I) = \beta \, I / p_2$.
Note the special form of these demand functions, which do not depend on other prices.

$v(p_1, p_2, I) = \alpha^{\alpha} \beta^{\beta} \; p_1^{-\alpha} p_2^{-\beta} I$.

B. Expenditure Function

B. Expenditure Function

Instead of maximizing utility subject to a budget constraint, one could minimize spending, subject to a utility constraint:

$$\min \; p_1 x_1 + p_2 x_2 \quad \text{s.t. } U(x_1, x_2) = u^0.$$

The lagrangean is

$$L(x_1, x_2, \mu) = p_1 x_1 + p_2 x_2 - \mu (U(x_1, x_2) = u^0)$$

The foc are:

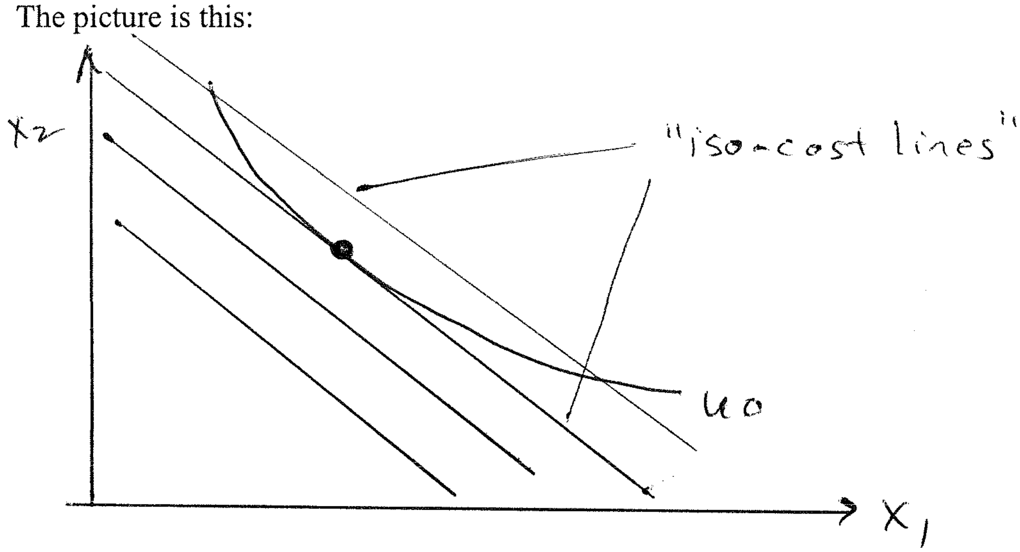
a) $p_1 - \mu U_1(x_1, x_2) = 0$

b) $p_2 - \mu U_2(x_1, x_2) = 0$

c) $U(x_1, x_2) = u^0$

Note that (a) and (b) give back the tangency condition $p_1/p_2 = U_1(x_1, x_2) / U_2(x_1, x_2)$.

The picture is this:



The parallel lines represent "iso-cost lines": combinations such that $p_1 x_1 + p_2 x_2$ is constant. These are just the "contours" of the objective function. Their slope is $- p_1/ p_2$. (why?)

The U-max and E-min problems are called "dual" problems, since they flip the objective and constraints.

What are the solution functions to the expenditure minimization problem?

The choices (x_1, x_2) that minimize spending subject to the utility constraint are like "demand" functions, *except that they take utility as given rather than income.* We will call these the "compensated demands":

$$x_1 = x^C_1(p_1, p_2, u^0)$$

$$x_2 = x^C_2(p_1, p_2, u^0)$$

Sometimes these are called the "Hicksian" demands, after John Hicks, an English economist who defined them (and won the second Nobel prize in economics).

The total spending at the expenditure minimizing choice is:

$$p_1 x^C_1(p_1, p_2, u^0) + p_2 x^C_2(p_1, p_2, u^0)$$

We define a new function (like the "indirect utility function") that gives the minimum spending when you have solved the expenditure minimization problem

$$e(p_1, p_2, u^0) = \min\ p_1 x_1 + p_2 x_2 \ \text{ s.t. } U(x_1, x_2) = u^0$$

$$= p_1 x^C_1(p_1, p_2, u^0) + p_2 x^C_2(p_1, p_2, u^0)$$

Note that $e(p_1, p_2, u^0)$ tells you the minimum amount of money you need to achieve utility u^0 when prices are (p_1, p_2).

Example: $U(x_1, x_2) = x_1^\alpha x_2^\beta \qquad \alpha + \beta = 1$

$$L = p_1 x_1 + p_2 x_2 - \mu(x_1^\alpha x_2^\beta + u^0)$$

$$\frac{\partial L}{\partial x_1} = +p_1 - \mu \alpha x_1^{\alpha-1} x_2^\beta = 0$$

$$\frac{\partial L}{\partial x_2} = p_2 - \mu \beta x_1^\alpha x_2^{\beta-1} = 0$$

$$\Rightarrow \frac{p_1}{p_2} = \frac{\alpha}{\beta}\frac{x_2}{x_1}$$

$$\Rightarrow x_2 = \frac{\beta}{\alpha}\frac{p_1}{p_2} x_1$$

subst. into constraint: $x_1^\alpha x_2^\beta = u^0$

$$x_1^\alpha \left[\beta/\alpha \ \frac{p_1}{p_2} x_1\right]^\beta = u^0 \Rightarrow x_1 = u^0 p_1^{-\beta} p_2^{\beta} \beta^{-\beta} \alpha^{\beta}$$

$$x_2 = u^0 p_2^{-\alpha} p_1^{\alpha} \alpha^{-\alpha} \beta^{\alpha}$$

Lecture 4

"Comparative Statics" of Consumer Choice

In this lecture, we characterize the changes in consumer demands that occur as income and prices vary. Our goal is to describe the consumers' <u>demand</u> <u>functions</u>. Analytically, the demand functions for the goods x and y are <u>functions</u>

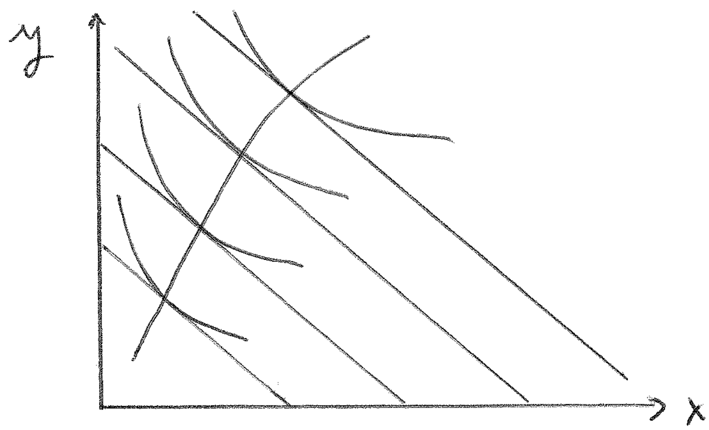
$x = f(p_x, p_y, I)$

$y = g(p_x, p_y, I)$

that describe the consumers' optimal choices of x and y, given prices and incomes. You can imagine that the nature of these functions is important for a wide variety of applications.

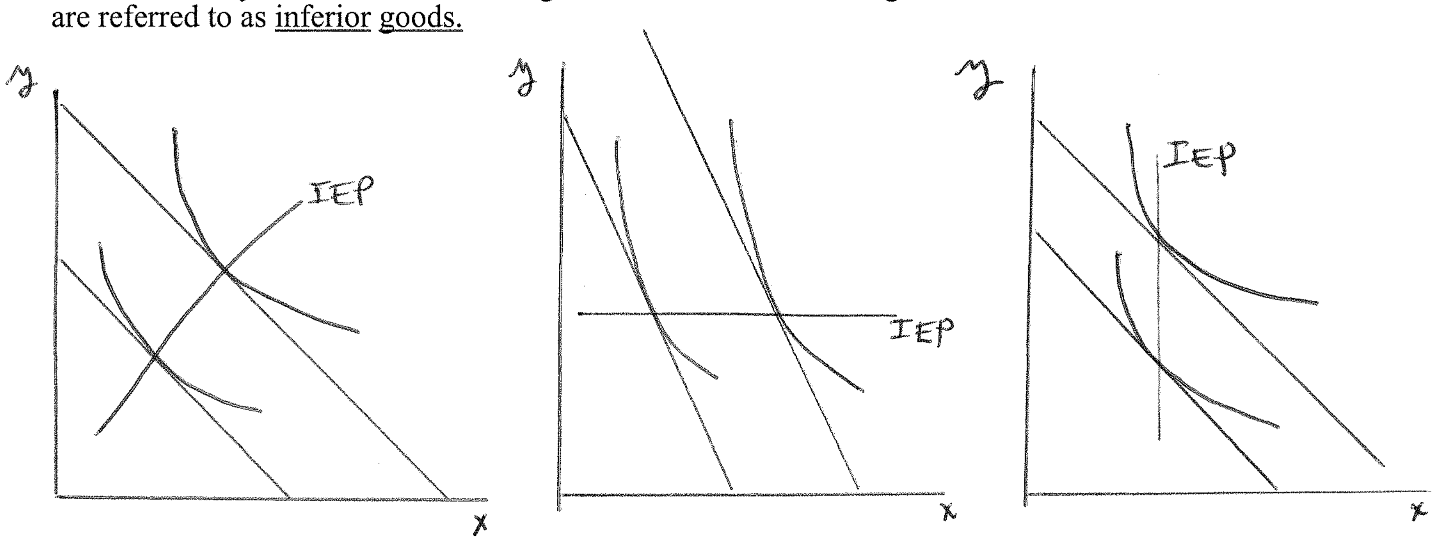
<u>(1) Changes in Demand with Respect to Income</u>

As income changes, the budget constraint shifts in a parallel fashion: inward if I decreases; outward if I increases.



income expansion path

In commodity space [(x,y) space] the tangencies of the budget constraints with higher and higher indifference curves trace out the income expansion path. If purchases of x and y increases with income, x and y are said to be <u>normal</u> <u>goods</u>. Purchases of some goods fall with income--these are referred to as <u>inferior</u> <u>goods.</u>



x,y normal

x normal
y "borderline inferior"

x inferior
y normal

There are several interesting implications of the budget constraint for the changes in x and y with income:

(a) Using the fact that income is always exhausted,

$$p_x \cdot x + p_y \cdot y = I$$

$$p_x \cdot dx + p_y \cdot dy = dI$$

$$\Rightarrow \quad p_x \frac{dx}{dI} + p_y \frac{dy}{dI} = 1$$, so both goods cannot be inferior.

(b) Starting from the previous equation, we get

$$\frac{xp_x}{I} \cdot \frac{I}{x}\frac{dx}{dI} + \frac{yp_y}{I} \cdot \frac{I}{y}\frac{dy}{dI} = 1$$

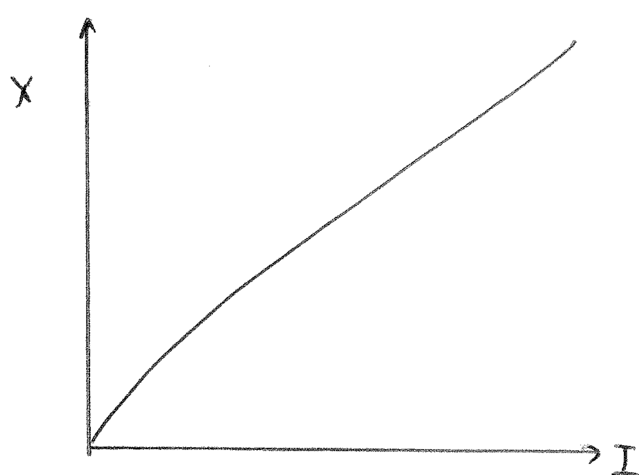
which can be written as:

$$s_x \cdot e_x + s_y \cdot e_y = 1 ,$$

where s_x, s_y are the expenditure shares (s_x is the fraction of income spent on good x) and e_x, e_y are the income elasticities. This equation can be stated in words as "the expenditure-weighted sum of income elasticities is unity".

Engel Curves

The relation between x and I , holding constant prices, is called the Engel curve:

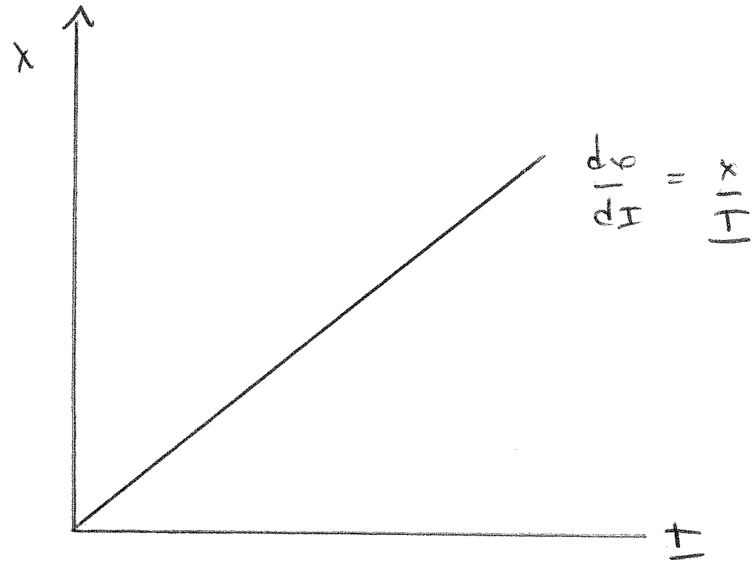


The Engel curve starts at the origin if x = 0 when I = 0.

The Engel curve is positively sloped if x is a normal good.

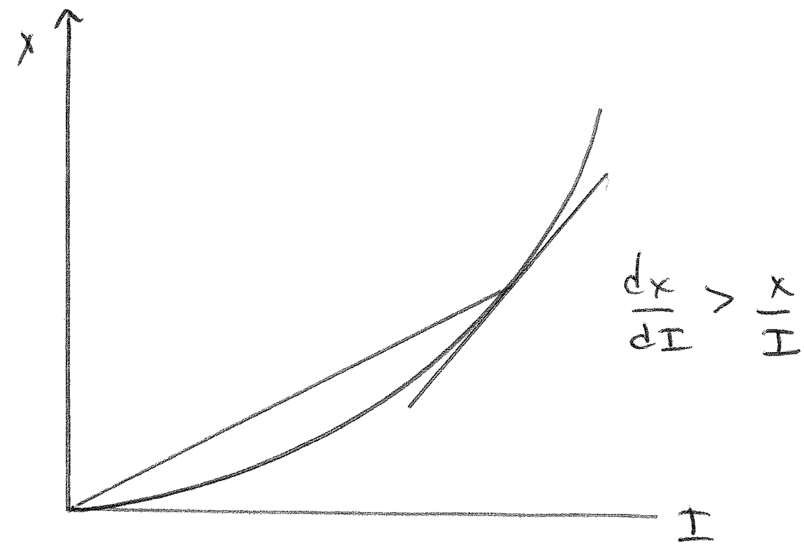
(1) Linear Engel curves

$$e_x = \frac{\frac{dx}{dI}}{\frac{x}{I}} = I$$



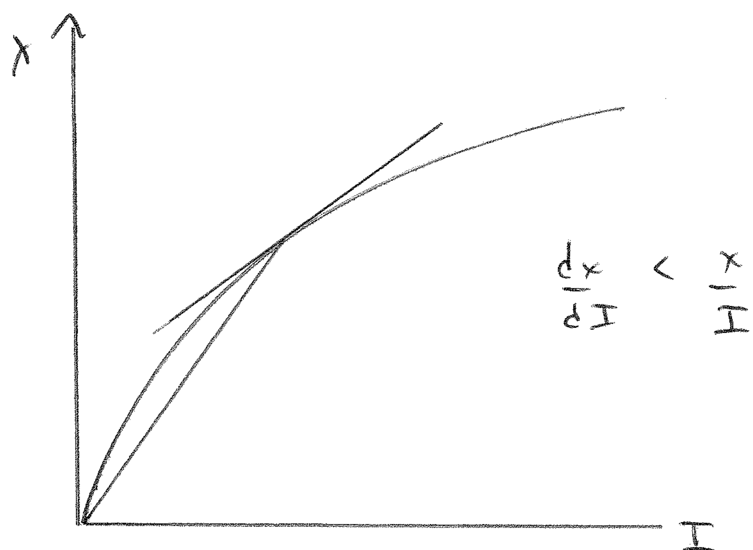
(2) Convex Engel curve

$$e_x > 1 \rightarrow \frac{dx}{dI} < \frac{x}{I}$$



(3) Concave Engel curve

$$e_x < 1 \rightarrow \frac{dx}{dI} < \frac{x}{I}$$



The data in the following table confirm Engel's Law. As income increases, a smaller share of income is devoted to food. The implication is that the income elasticity for food is less than unity.

Year	Share of Food in Standard Budget*
1935-39	35.4
1952	32.2
1963	25.2
1992	19.6
2000	16.3

*Budget used in calculation of the CPI.

Why?

Let $s_x = \frac{xp_x}{I}$, the share of income spent on x.

$$\frac{ds_x}{dI} = \frac{p_x \frac{dx}{dI}}{I} - \frac{1}{I^2} xp_x$$

$$= \frac{\frac{xp_x}{I} \frac{I}{x} \frac{dx}{dI}}{I} - \frac{1}{I} \frac{xp_x}{I}$$

$$= \frac{s_x}{I} [e_x - 1] \quad \text{or} \quad \frac{I}{s_x} \frac{ds_x}{dI} = e_x - 1$$

So, if $e_x < 1$ then the food share is declining with income. An alternative proof uses the favorite "trick" of economists: take natural logs.

$$\log (s_x) = \log x + \log p_x - \log I$$

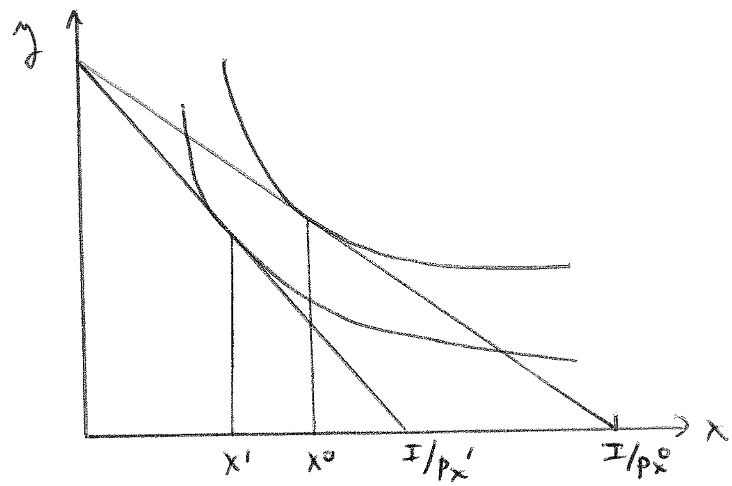
$$\frac{d \log s_x}{d \log I} = \frac{d \log x}{d \log I} - 1,$$

or $\frac{I}{s_x} \frac{ds_x}{dI} = e_x - 1$

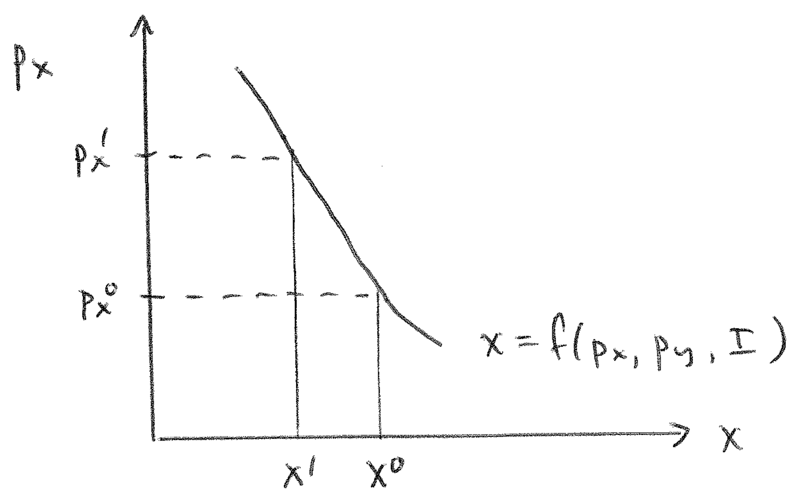
In some contexts, the food share is used as an indicator of welfare. It has been proposed that families in different countries with the same food share are as well off.

(2) Changes in Price

A change in price causes the budget constraint to rotate. As the budget line rotates, the tangencies with higher and higher indifference curves trace out the price consumption path.



You are familiar with the "demand curve", which graphs the demand function, holding constant income and prices of other commodities.



Note that we traditionally draw the amount demanded (the dependent variable) on the horizontal axis and the price (the independent variable) on the vertical axis.[1] The negative slope of the demand curve reflects the idea that consumption of a commodity falls as its price increases. However, demand curves are not necessarily downward sloping! We turn now to a decomposition of the change in demand resulting from a change in price. We show that there are two factors:

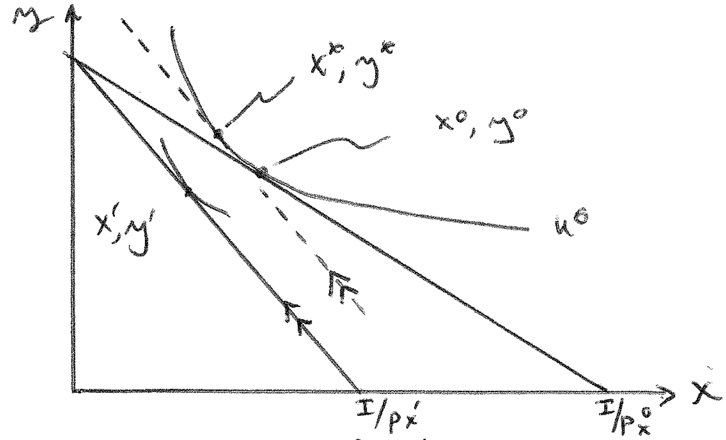
(1) the curvature of indifference curves
(2) the nature of income effects on demand

---

[1]We owe this convention to Alfred Marshall. As a result of this, steep demand curves are "inelastic", whereas flat demand curves are "elastic".

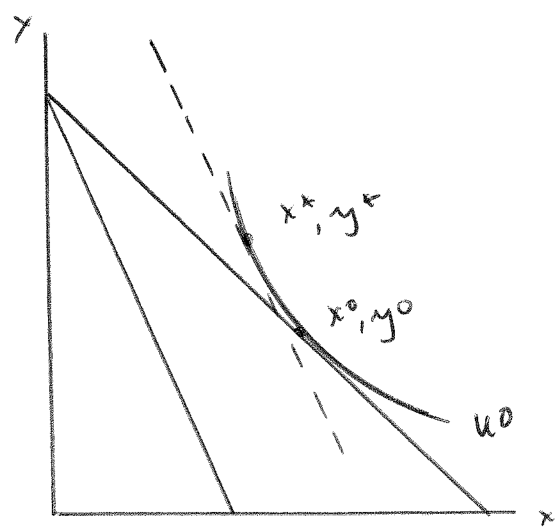
<u>A Graphical Decomposition of a Change in Demand</u>

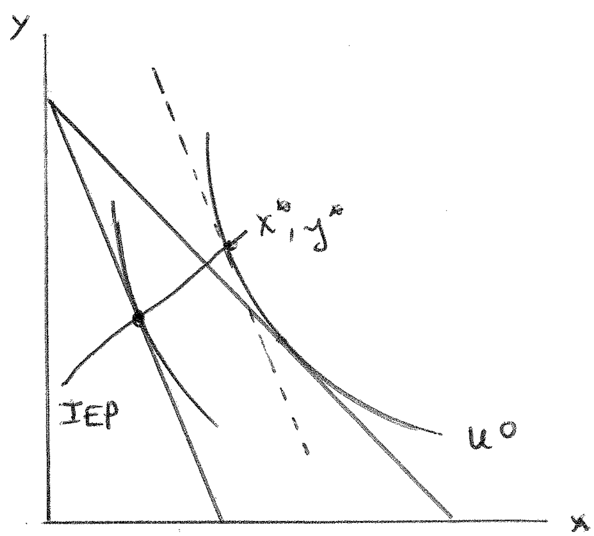
Suppose p_x increases from p_x^0 to p_x^1 : demand changes from (x^0,y^0) to (x^1, y^1).



We can decompose the change from x^0 to x^1 as follows:

(i) First, think of the change in x that arises purely from the fact that x costs more. Draw a budget line with slope px^1/p_y that still allows the consumer to reach the indifference curve that yields utility u^0. Note that, since it's steeper than the old budget line, it has a tangency with u^0 to the left of (x^0,y^0) (DMRS). This "artificial" budget constraint is shown as a dashed line in the above diagram.





Step 1: move to an new tangency old indifference curve

Step 2: move along an income expansion path to the new optimum.

(ii) Second, move from this intermediate point to the final optimum. Observe that this movement is a movement <u>along an income expansion path</u>, since the "intermediate" optimum occurs where the <u>old</u> indifference curve has a tangency with a budget line with slope p_x^1/p_y.

Analytically: $\Delta x = x^1 - x^0$

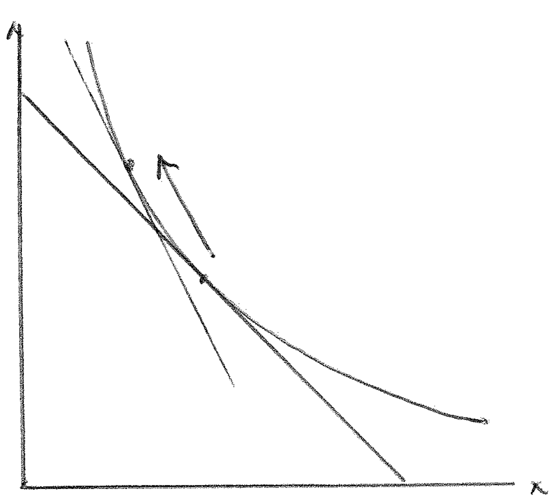
$$= (x^1 - x^*) + (x^* - x^0), \quad \text{where } x^* \text{ is the intermediate optimum}$$

We refer to the first change $(x^1 - x^*)$, which holds utility constant, as the SUBSTITUTION EFFECT. We refer to the second change $(x^* - x^0)$, as the INCOME EFFECT.

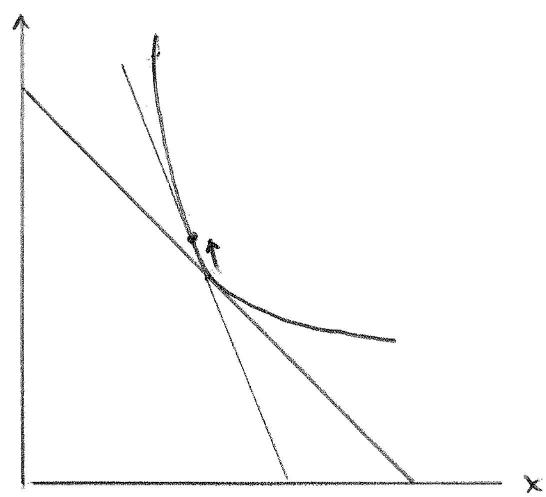
Thus, we write $\Delta x = \Delta x^s + \Delta x^I$

How big is the substitution effect?

The substitution effect represents a movement along an indifference curve. It tells you how far you have to go to get the slope of the indifference curve equal to the slope of the new budget constraint. Obviously, then, if the indifference curves are flat, you have to go a long way to re-equate the MRS with the new price ratio, and the substitution effect is larger. If indifference curves are highly curved ("more convex") the MRS changes very quickly and you don't have to move very far: the substitution effect is smaller.



Larger substitution effect:
u^0 flat

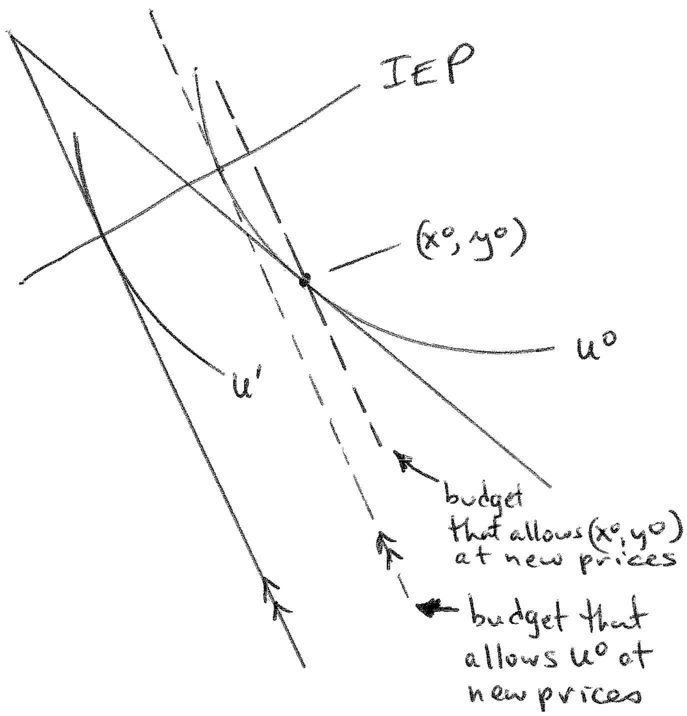


Smaller substitution effect:
u^0 very convex

Notice that if $\Delta p_x > 0$, the substitution effect on x is negative. (Why?) What about the substitution effect of Δp_x on y?

How big is the income effect?

Intuitively, you probably think that the income effect is larger, the more of good x you initially purchased. If initially, you bought very little x, the income effect will be small. Take a look at the following graph:



Notice that the "intermediate" budget constraint almost passes through (x^0, y^0) (it always cuts a little below).

So approximately, the income effect is proportional to the change in income represented by the change in income from (x^0,y^0) to the final budget constraint.

What is the change in income? The final budget constraint has $I of income available, the same as the initial constraint. Therefore $I = p_x^0 x^0 + p_y y^0$. In order to be able to afford (x^0,y^0) under the new prices you would need $\$p_x^1 x^0 + p_y y^0$. The difference is $(p_x^1 - p_x^0)x^0$ or $\Delta I = \Delta p_x \cdot x^0$. Thus, the difference in income between the intermediate budget constraint and the final one is (approximately) $\Delta p_x \cdot x^0$. (The approximation becomes exact as $\Delta p_x \to 0$.)

This confirms our intuition: the movement along the income expansion path from the intermediate to the final position (the income effect) will be larger, the larger was x^0, our initial purchase of x^0 (and the larger is Δp_x).

Economics 101A – Spring 2005

A Revised Version of the Slutsky Equation Using the Expenditure Function
or, the expenditure function is our friend!

Brief review...

$$e(p_1, p_2, u^0) \quad = \quad \min \; p_1 x_1 + p_2 x_2 \;\; \text{s.t.} \;\; U(x_1, x_2) = u^0$$

$$= \quad p_1 x^C_1 (p_1, p_2, u^0) \; + \; p_2 x^C_2 (p_1, p_2, u^0)$$

where x^C_1 and x^C_2 are the "compensated demands": the choices you would make to get utility level u^0 as cheaply as possible at prices (p_1, p_2).

Remember that the Lagrangean for the exp-min problem is:

$$L(x_1, x_2, \mu) = p_1 x_1 + p_2 x_2 \; - \mu (U(x_1, x_2) - u^0)$$

The foc are:

a) $p_1 - \mu U_1(x_1, x_2) = 0$
b) $p_2 - \mu U_2(x_1, x_2) = 0$
c) $U(x_1, x_2) = u^0$

What are the derivatives of the expenditure function w.r.t. (p_1, p_2)?

From the definition

$e(p_1, p_2, u^0) \; = \; p_1 x^C_1 (p_1, p_2, u^0) \; + \; p_2 x^C_2 (p_1, p_2, u^0)$, and so:

(†) $\partial e(p_1, p_2, u^0) / \partial p_1 \; = \; x^C_1 (p_1, p_2, u^0) + p_1 \partial x^C_1 (p_1, p_2, u^0) / \partial p_1 + p_2 \partial x^C_2 (p_1, p_2, u^0) / \partial p_1$.

In tutorial we discussed the "envelope theorem" which says that the 2nd and 3rd terms cancel out. A quick way to prove that:

Use the constraint: $U(x^C_1 (p_1, p_2, u^0) , \; x^C_2 (p_1, p_2, u^0) \; = \; u^0$, and differentiate w.r.t. p_1 to get

$$U_1 \; \partial x^C_1 (p_1, p_2, u^0) / \partial p_1 \; + U_2 \, \partial x^C_2 (p_1, p_2, u^0) / \partial p_1 \; = \; 0$$

But $U_1(x_1, x_2) = p_1/\mu$ and $U_2(x_1, x_2) = p_2/\mu$ from the f.o.c. Substituting we get

$$p_1/\mu \;\; \partial x^C_1 (p_1, p_2, u^0) / \partial p_1 \; + \; p_2/\mu \;\; \partial x^C_2 (p_1, p_2, u^0) / \partial p_1 = 0$$

which means that $p_1 \partial x^C_1 (p_1, p_2, u^0) / \partial p_1 + \; p_2 \partial x^C_2 (p_1, p_2, u^0) / \partial p_1 = 0$.

Thus we have:

$$\partial e(p_1, p_2, u^0) / \partial p_1 \quad = \quad x^C{}_1 (p_1, p_2, u^0) .$$

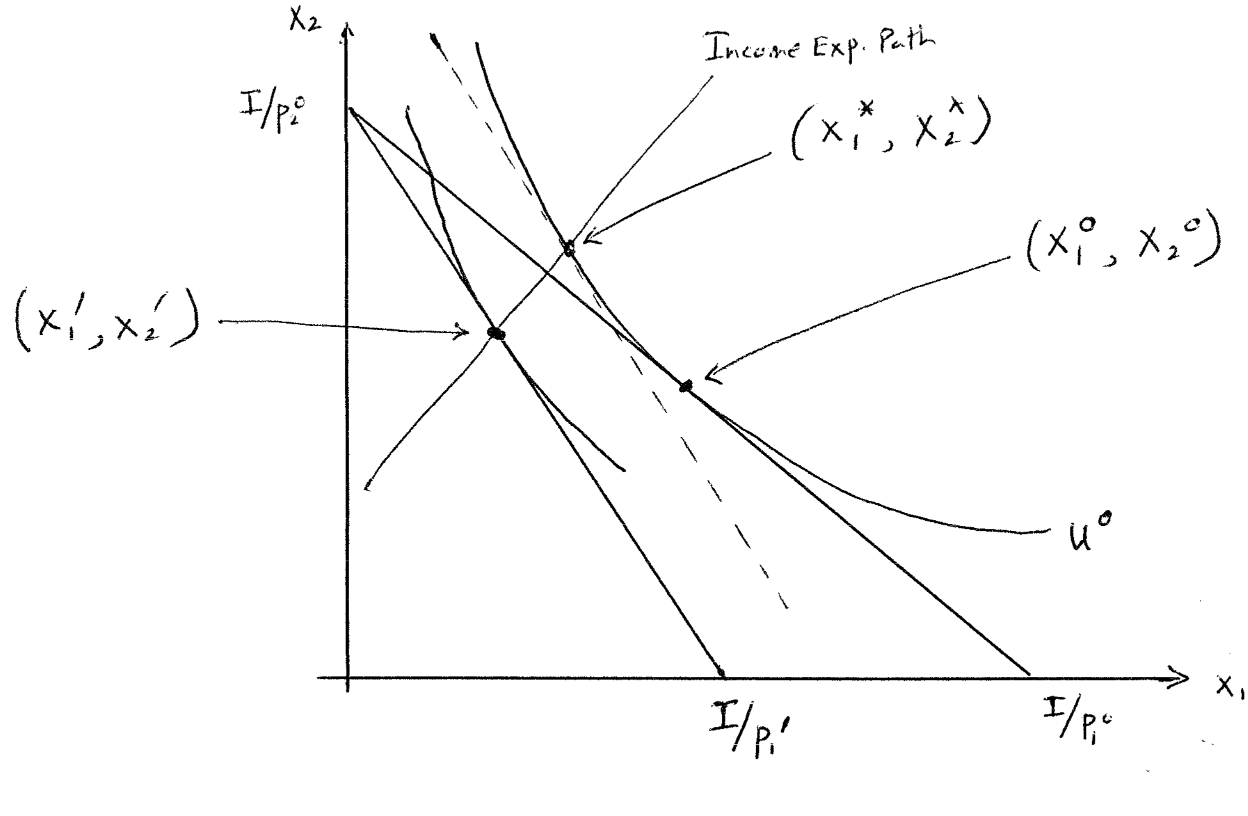
There is a story we tell to go along with this. If you are initially minimizing expenditure, and the price of good 1 goes up, what do you do? Your "first order" adjustment is to simply continue buying the old bundle – that will increase your spending by $x^C{}_1 \times \Delta p_1$. That is the first term in (†). But then you would like to re-adjust your choices of goods 1 and 2 to reflect the new prices. Those adjustments are the second and third terms in (†). But because your initial choice was optimal (satisfying the f.o.c) when you try and re-adjust x_1 and x_2 you don't save any more.

Now we are ready to analyze what happens to the **uncompensated or regular** demand when prices rise. Suppose we start at an intial situation with prices $(p_1{}^0, p_2{}^0)$ and income I^0. The initial choices are $x_1{}^0 = x_1 (p_1{}^0, p_2{}^0, I^0)$, $x_2{}^0 = x_2 (p_1{}^0, p_2{}^0, I^0)$ where the x_1 ( ) and x_2 ( ) functions *without superscripts* are the regular demand functions

We decompose the effect of a change in price $\Delta p_1 = p_1' - p_1{}^0$ as follows:

a) starting from $x_1{}^0$, $x_2{}^0$, think of the adjustment you would make if you could keep utility constant (remain on the old indifference curve). This gets you to a new position $x_1{}^*$, $x_2{}^*$. Since prices have risen this position costs more than you were initially spending. This move is called the "substitution effect" of the price increase.

b) then from $x_1{}^*$, $x_2{}^*$ think of the adjustment you make to get back to spending only the amount of income that you actually have. This is a movement inward along an income expansion path (IEP). You end up at x_1' , x_2' . This move is called the "income effect" of the price increase.



Note that the total change in x_1 is

$$\begin{aligned} \Delta x_1 &= x_1' - x_1^0 \\ &= (x_1' - x_1^*) + (x_1^* - x_1^0) \\ &= \Delta x_1^I + \Delta x_1^S . \end{aligned}$$

How big are these two parts? To begin, notice that (x_1^0, x_2^0) and (x_1^*, x_2^*) are on the u^0 indifference curve.

$$x_1^0 = x_1(p_1^0, p_2^0, I^0)$$

But it is also true that

$$x_1^0 = x_1^C(p_1^0, p_2^0, u^0).$$

Also,

$$x_1^* = x_1^C(p_1', p_2^0, u^0)$$

So $$\Delta x_1^S = x_1^* - x_1^0 = x_1^C(p_1', p_2^0, u^0) - x_1^C(p_1^0, p_2^0, u^0)$$

$$\approx \partial x^C_1(p_1^0, p_2^0, u^0) / \partial p_1 \times \Delta p_1$$

The substitution effect depends on the rate at which compensated demands change: this is purely a function of the curvature of the indifference curve.

How about the income effect?

$$\Delta x_1^I = x_1' - x_1^*$$

First note that $x_1' = x_1(p_1', p_2^0, I^0)$: its just your regular demand choice at (p_1', p_2^0, I^0).

But what is x_1^*? It is the choice when you have enough income to get to the old indifference curve at the new prices. How much money do you need to get there? That's just $e(p_1', p_2^0, u^0)$! So

$$x_1^* = x_1(p_1', p_2^0, e(p_1', p_2^0, u^0))$$ Make sure this makes sense to you!

Thus

$$\begin{aligned} \Delta x_1^I &= x_1(p_1', p_2^0, I^0) - x_1(p_1', p_2^0, e(p_1', p_2^0, u^0)) \\ &\approx \partial x_1(p_1^0, p_2^0, I^0) / \partial I \times (I^0 - e(p_1', p_2^0, u^0)) \end{aligned}$$

So the income effect depends on the income derivative of demand *times* the size of the income change $\Delta I = I^0 - e(p_1', p_2^0, u^0)$. Note that $\Delta I < 0$, since you need more than I^0 to get to the u^0 indifference curve when prices are (p_1', p_2^0).

But how big is ΔI? We have to use one last trick.

We know that $I^0 = e(p_1^0, p_2^0, u^0)$.

So we can write

$$\begin{aligned}
\Delta I &= I^0 - e(p_1', p_2^0, u^0). \\
&= e(p_1^0, p_2^0, u^0) - e(p_1', p_2^0, u^0) \\
&\approx \partial e(p_1^0, p_2^0, u^0) / \partial p_1 \times (p_1^0 - p_1') \\
&= \partial e(p_1^0, p_2^0, u^0) / \partial p_1 \times (-\Delta p_1) \\
&= - \partial e(p_1^0, p_2^0, u^0) / \partial p_1 \times \Delta p_1
\end{aligned}$$

Note that this is negative for a rise in the price of good 1. Finally (almost done) we have

$$\begin{aligned}
\partial e(p_1^0, p_2^0, u^0) / \partial p_1 &= x^C_1 (p_1^0, p_2^0, u^0) \quad \textit{from the first page of this lecture note} \\
&= x_1^0 \quad \text{because } x_1^0 \text{ is also the compensated demand choice}
\end{aligned}$$

All this together means that

$$\Delta I \approx - x_1^0 \Delta p_1 .$$

Note that the size of the income effect depends on how much x_1 you were buying.

Pulling it all together,

$$\begin{aligned}
\Delta x_1^I &= \partial x_1 (p_1, p_2, I^0) / \partial I \times \Delta I \\
&= - \partial x_1 (p_1, p_2, I^0) / \partial I \times x_1^0 \Delta p_1
\end{aligned}$$

Thus

$$\begin{aligned}
\Delta x_1 &= \Delta x_1^I + \Delta x_1^S \\
&= - \partial x_1 (p_1, p_2, I^0)/\partial I \times x_1^0 \Delta p_1 + \partial x^C_1 (p_1^0, p_2^0, u^0) / \partial p_1 \times \Delta p_1
\end{aligned}$$

or $\Delta x_1 / \Delta p_1 = - x_1^0 \; \partial x_1 (p_1^0, p_2^0, I^0)/\partial I + \partial x^C_1 (p_1^0, p_2^0, u^0) / \partial p_1$.

Now take the limit as Δp_1 gets small and the ratio $\Delta x_1 / \Delta p_1$ tells us the derivative of the regular demand function. We have established:

$$\partial x_1 (p_1^0, p_2^0, I^0) / \partial p_1 = -x_1^0 \; \partial x_1 (p_1^0, p_2^0, I^0)/\partial I \; + \; \partial x^C_1 (p_1^0, p_2^0, u^0) / \partial p_1$$

This is called Slutsky's equation, after the Russian economist who first proved it about 100 years ago. Slutsky's equation says that the derivative of the regular demand with respect to p_1 is a combination of the income and substitution effects. The income effect depends on the derivative of demand w.r.t income, *times* the amount of x_1 you initially consume. The substitution effect depends on the derivative of the compensated demand for good 1.

A neat thing about the Slutsky equation is that it gives us a way to recover information about indifference curves from the derivatives of demand w.r.t. prices and incomes. In principle, we can observe $\partial x_1 (p_1^0, p_2^0, I^0) / \partial p_1$ and $\partial x_1 (p_1^0, p_2^0, I^0)/\partial I$. So we can infer:

$$\partial x^C_1 (p_1^0, p_2^0, u^0) / \partial p_1 = \partial x_1 (p_1^0, p_2^0, I^0) / \partial p_1 \; + \; x_1^0 \; \partial x_1 (p_1^0, p_2^0, I^0)/\partial I$$

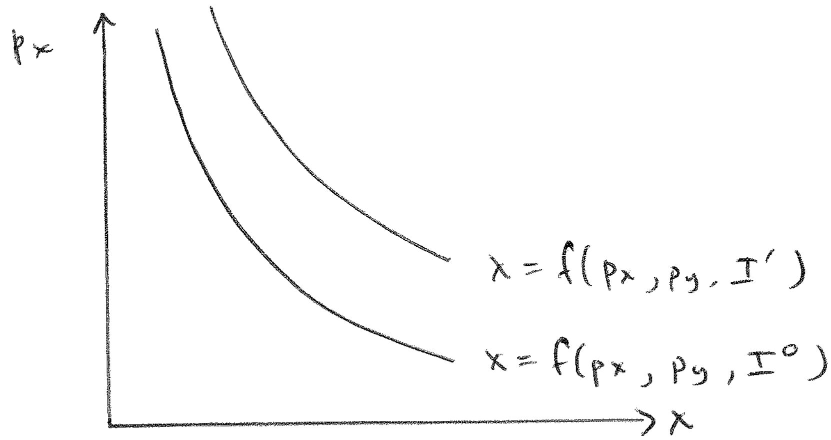
Suppose we get an estimate of $\partial x^C_1 (p_1^0, p_2^0, u^0) / \partial p_1$ that is close to 0. That means indifference curves must be almost like "right angles".

Lecture 5

Using Market Level Demand Curves

Shifts in Demand Curves

Since the demand curve graphs $x = f(p_x, p_y, I)$, if p_y or I changes, the demand curve shifts. For example, if income increases by $dI > 0$, at each price, demand is now increased by $dx = (\frac{\partial x}{\partial I})dI$. If x is a normal good, the demand curve shifts right with the income.



If the elasticities of demand are approximately constant,

$$\frac{dx}{x} \cong (\frac{\partial x}{\partial I}\frac{I}{x}) \times \frac{dI}{I} = e_x \frac{dI}{I} = e_x \, d \log I$$

where e_x is the income elasticity of demand for x. Similarly, if p_y changes, the demand curve shifts, unless $\frac{\partial x}{\partial p_y} = 0$ (as in the Cobb-Douglas case). If $\frac{\partial x}{\partial p_y} < 0$ an increase in the price of y shifts the demand curve right.

For the purposes of evaluating the effect of relatively small changes in prices and income, we often assume that the demand function has constant elasticities, i.e.,

$$\frac{\partial x}{\partial p_x}\frac{p_x}{x} = \frac{\partial \log x}{\partial \log p_x} = \eta_{xx} \text{, constant} \qquad \frac{\partial x}{\partial p_y}\frac{p_y}{x} = \frac{\partial \log x}{\partial \log p_y} = \eta_{xy} \text{, constant}$$

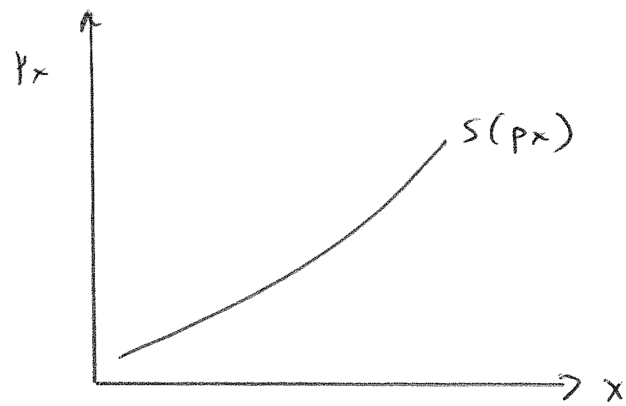
$$\frac{\partial x}{\partial I}\frac{p_x}{x} = \frac{\partial \log x}{\partial \log I} = e_x \text{, constant}$$

You could convince yourself that this is the same thing as assuming that the demand function is log-linear:

$$\log x = \text{constant} + \eta_{xx} \log p_x + \eta_{xy} \log p_y + e_x \log I$$

Note that homogeneity implies that $\eta_{xx} + \eta_{xy} + e_x = 0$: if prices and incomes all rise by 1 percent, x stays constant.

As you recall from introductory economics, the "market" is completed by introducing a supply curve of the form $x = S(p_x)$.



It is usually assumed that supply is upward sloping. [We will discuss the derivation of market supply curves later in the course]. For now, we will assume that the elasticity of supply is constant:

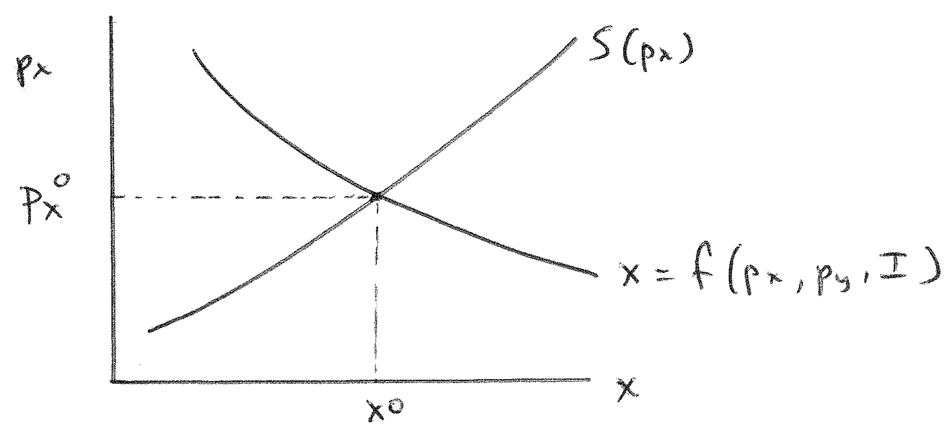
$$\frac{dS(p_x)}{dp_x} \frac{p_x}{S(p_x)} = \sigma_x .$$

where σ_x is the elasticity of supply. We can now combine supply and demand curves to analyze the effects of exogenous shocks to income or other prices. We have

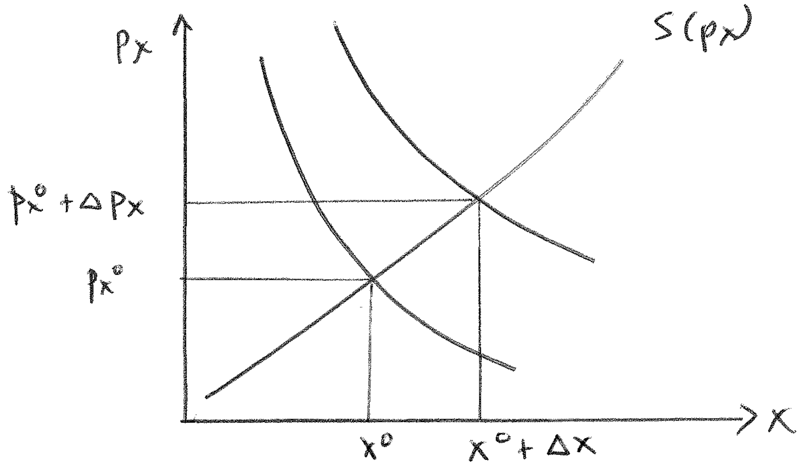
$x = S(p_x)$

$x = f(p_x, p_y, I)$,

which is a system of 2 equations in 2 unknowns (price and quantity of x), given incomes and other prices.



Application No. 1: An increase in income



Obviously, p_x increases and x increases. By how much? Starting at equilibrium at (x^0, p_x^0), the changes in demand and supply are:

demand: $$\frac{\Delta x}{x} = \eta_{xx}\frac{\Delta p_x}{p_x} + e_x\frac{\Delta I}{I}$$

supply: $$\frac{\Delta x}{x} = \sigma_x\frac{\Delta p_x}{p_x}$$

The proportional changes in supply and demand have to be equal to restore equilibrium. Therefore

$$\eta_{xx}\frac{\Delta p_x}{p_x} + e_x\frac{\Delta I}{I}\frac{\Delta p_x}{p_x} = \sigma_x\frac{\Delta p_x}{p_x}$$

implying that

$$\frac{\Delta p_x}{p_x} = [\frac{e_x}{\sigma_x - \eta_{xx}}]\frac{\Delta I}{I}$$

Note that $\eta_{xx} < 0$ so the denominator term $(\sigma_x - \eta_{xx}) > 0$.

Furthermore $\dfrac{\Delta x}{x} = \sigma_x \dfrac{\Delta p_x}{p_x} = \dfrac{\sigma_x\, e_x}{(\sigma_x - \eta_x)} \times \dfrac{\Delta I}{I}$

Example: suppose $\sigma_x = .60$ (short run)

$\eta_{xx} = -1.40$

$e_x = .40$

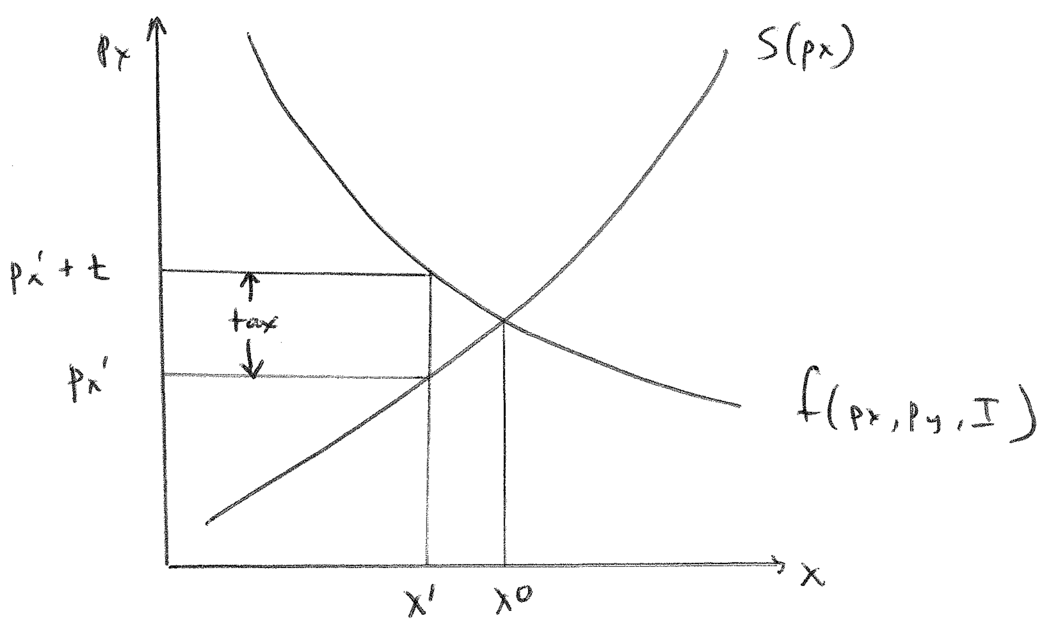
If $\dfrac{\Delta I}{I} = .10$ (10% increase)

$$\frac{\Delta p_x}{p_x} = (.40)\,(.10) \cong .02\;; \qquad \frac{\Delta x}{x} \cong .012$$

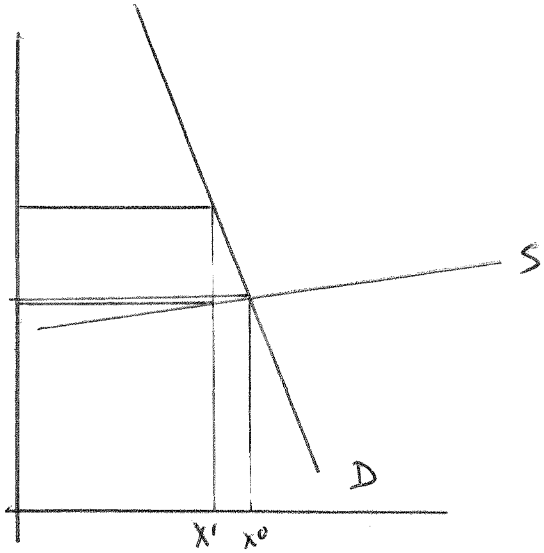
Exercise: Calculate the effect of a 10% cut in the price of a substitute good (good y) on the market for x. Use an estimate for the cross-price elasticity between x and y of 0.67 ($\eta_{xy} = 0.67$).

Application Number 2: Tax Incidence

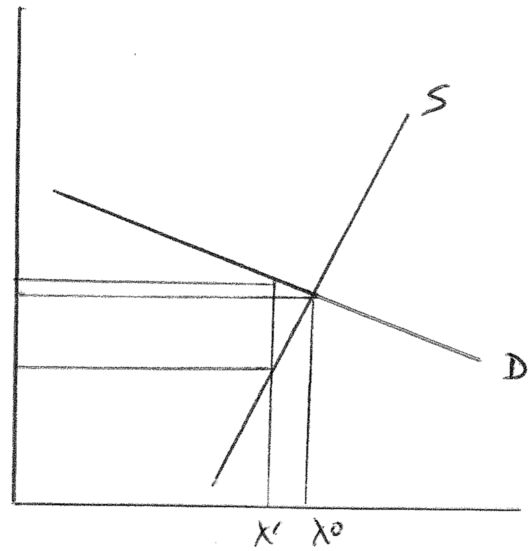
If a tax of t \$/unit is imposed on x, it opens up a gap between the price that consumers pay and the price that producers receive of \$t per unit. You are familiar with the diagram:



Starting from an equilibrium at (p_x^0, x^0), prices received by producers falls to p_x^1, the price paid by consumers risese to p_x^1+t, and the quantity falls to x^1. Consider two alternative markets, each with the same tax:



Demand inelastic
Supply elastic



Demand elastic
Supply inelastic

Obviously, the effect of the tax on the prices paid/received by the two sides depends on the **relative** elasticities of supply and demand. To see this more formally, we proceed using the assumption that elasticities are (approximately) constant. Assuming p_x is the price received by producers, the change in supply is:

$$\frac{\Delta x}{x} = \sigma_x \left[\frac{\Delta p_x}{p_x}\right].$$

The change in prices for consumers is t + Δp_x. Therefore, the change in quantity demanded is:

$$\frac{\Delta x}{x} = \eta_{xx} \left[\frac{\Delta p_x + t}{p_x}\right].$$

Market equilibrium requires that change in demand equals change in supply.

$$\eta_{xx} \; \frac{\Delta p_x + t}{p_x} = \sigma_x \; \frac{\Delta p_x}{p_x}$$

Solving for the equilibrium change in prices, we have

$$\eta_{xx} \frac{t}{p_x} = \frac{\Delta p_x}{p_x} [\sigma_x - \eta_{xx}]$$

$$\frac{\Delta p_x}{p_x} = [\frac{n_{xx}}{\sigma_x - n_{xx}}] \frac{t}{p_x} .$$ *Note that* $\frac{t}{p_x}$ *is the proportional tax rate.*

Since $\sigma_x > 0$ and $\eta_{xx} < 0$ the denominator term is positive and $\Delta p_x < 0$. On the quantity side,

$$\frac{\Delta x}{x} = \sigma_x \frac{\Delta p_x}{p_x} = [\frac{\sigma_x \eta_{xx}}{\sigma_x - \eta_{xx}}] \frac{t}{p_x} < 0 .$$

For producers, the change in price is $\frac{\Delta p_x}{p_x} = (\frac{\eta_{xx}}{\sigma_x - \eta_{xx}}) \frac{t}{p_x}$

For consumers, the change in price is $\frac{\Delta p_x + t}{p_x} = (\frac{\eta_{xx}}{\sigma_x - \eta_{xx}}) \frac{t}{p_x} + \frac{t}{p_x} = (\frac{\sigma_x}{\sigma_x - \eta_{xx}}) \frac{t}{p_x} > 0$

Notice that the ratio of the changes in prices for producers versus consumers is η_{xx}/σ_x. So, if demand is very inelastic$|\eta_{xx}|$ is small (e.g. $\eta_{xx} = -0.1$) and supply is moderately elastic (e.g. σ_x=1.0) producer prices don't fall by much relative to producer prices. On the other hand, if demand is very elastic, $|\eta_{xx}|$ is large (e.g. $\eta_{xx} = -3.0$) producer prices are more effected.

Another example: a per unit subsidy of $S on the price of commodity x. (For example, prior to the recent rise in electricity rates,electricity prices in the most parts of California were subsidized). The change in price received by producers is Δp_x, the change in price paid by consumers is $\Delta p_x - s$. The proportional changes in quantity are:

Demand: $$\frac{\Delta x}{x} = \eta_{xx} [\frac{\Delta p_x - s}{p_x}]$$

Supply: $$\frac{\Delta x}{x} = \sigma_x \frac{\Delta p_x}{p_x}$$

Setting these equal, we get

$$\frac{\Delta p_x}{p_x} = \left(\frac{-\eta_{xx}}{\sigma_x - \eta_{xx}}\right) \frac{s}{p_x} > 0$$

which implies that some of the effect of the subsidy is mitigated by a rise in prices. In fact, the change in price to consumers is:

$$\frac{\Delta p_x - s}{p_x} = \left(\frac{-\eta_{xx}}{\sigma_x - \eta_{xx}}\right) \frac{s}{p_x} - \frac{s}{p_x} = \left(\frac{-\sigma_x}{\sigma_x - \eta_{xx}}\right) \frac{s}{p_x} < 0$$

Lecture 6

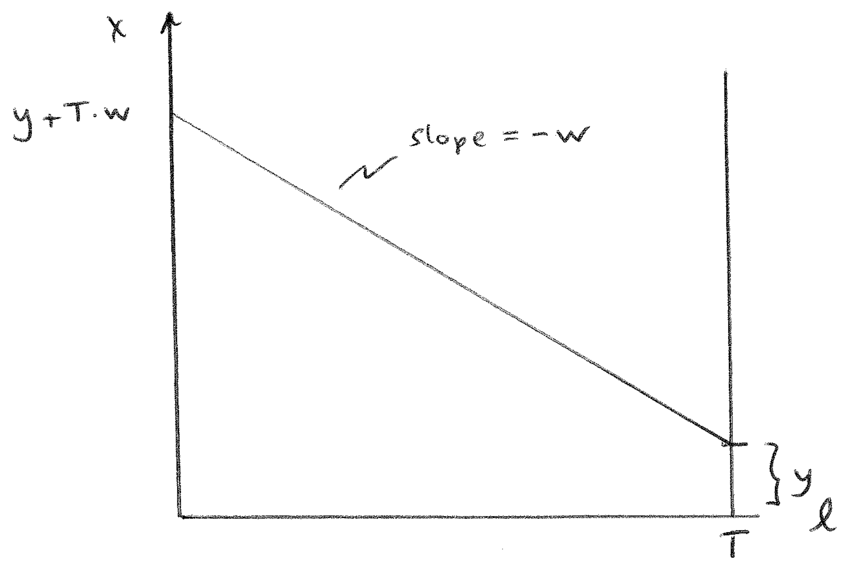
Labor Supply Analysis

In this lecture we continue the analysis of consumer demand theory with an application to the problem of labor supply. We set up the problem as follows. There are two goods of interest to consumers: a general consumption commodity (x) ; and leisure (ℓ). Leisure is defined as the difference (in time units) between total available time (T) and hours of work (h). Preferences are represented by the utility function u(x, ℓ). The budget constraint is

$$px = wh + y$$

where
p = price of consumption
w = nominal wage rate
y = other income.

For convenience, we set p = 1 (you should think of w and y as "real" wages and "real" income). Graphically, the labor supply problem is represented as



Every hour of reduced leisure allows the purchase of an additional w units of consumption. If ℓ=T (no labor supplied) maximum consumption is y. If ℓ=0 (labor supply = T) maximum consumption is y + wT
.

Analytically the problem is

$$\max_{x,\ell} \quad U(x,\ell) \quad \text{s.t. } x = wh + y$$

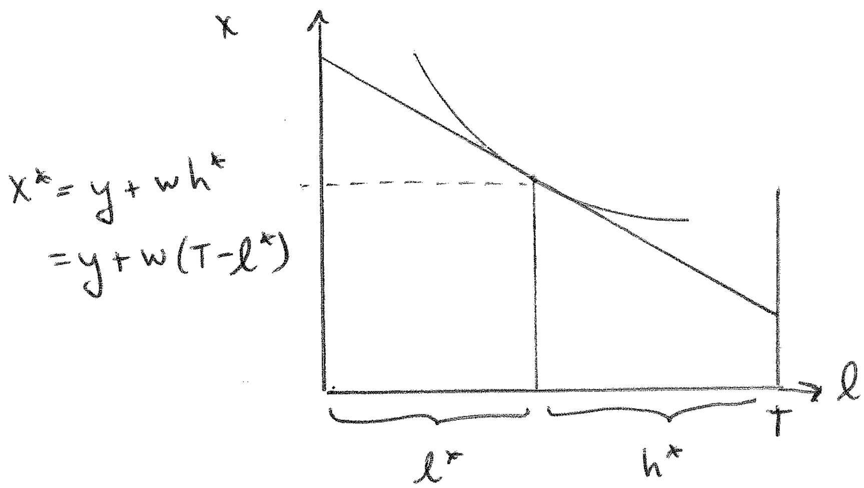
Re-write the budget constraint as

$$x = w(t - 1) + y ,$$

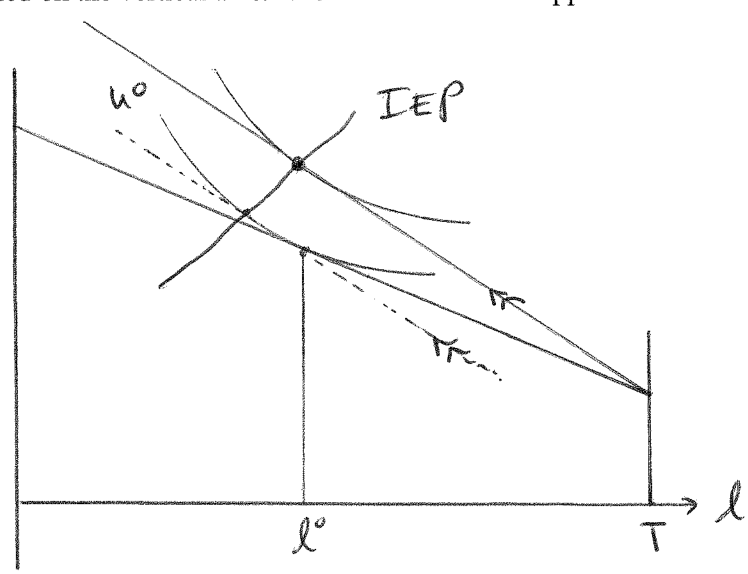
or $$x + w \cdot \ell = wT + Y$$

Notice that the total cost of "goods consumed", $x + w\ell$, equals total available income, $y + wT$. The latter is sometimes called "full income". It is important to note that "full income" varies with w . Thus (as will see) changes in w have a slightly different income effect than in the case where total income I is fixed and independent of prices.

The consumer's optimum is described by the following picture:

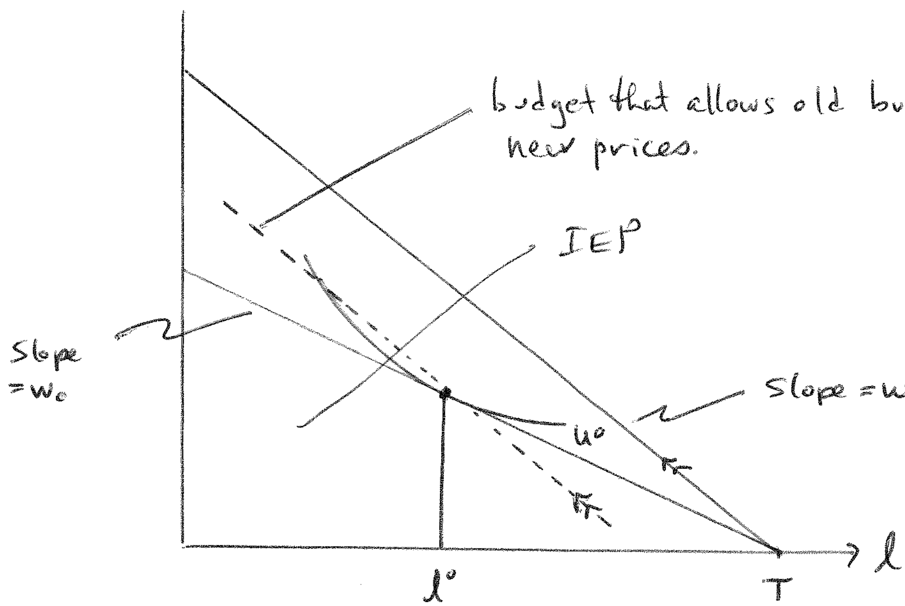


Notice that hours of work are represented by the horizontal distance between ℓ^* and T. Measured income $(w \cdot h^* + y)$ is represented on the vertical axis. Now consider what happens if w rises:



The budget constraint rotates through the (T,y) point. The substitution effect causes an increase in work but the income effect causes a decrease in work (provided leisure is a normal good). The net effect is ambiguous (as in the case of an increase in interest rates for net savers). The implication is that a reduction in tax rates on earned income may lead to more or less work, on average, depending on income and substitution effects.

Let's take a close look at the income effect in the labor supply case (for simplicity, set y = 0). Suppose wages increase from w_0 to w_1 .



As before, the income effect is approximately proportional to the change in income between the budget constraint passing through the old optimum (ℓ_0 , x_0) with slope $-w_1$ and the new budget constraint. The budget constraint through (x_0 , ℓ_0) has expenditure $x_0 + w_1\ell_0$. The new budget constraint has full income w_1T . Thus the difference between them in dollars is

$$\begin{aligned}\Delta I &= w_1 T - (x_0 + w_1\ell_0) \\ &= w_1(h_0 + \ell_0) - x_0 - w_1\ell_0 \\ &= w_1 h_0 - x_0 \quad .\end{aligned}$$

But recall that $x_0 = w_0h_0$ (from the original budget constraint). Thus

$$\begin{aligned}\Delta I &= w_1 h_0 - w_0 h_0 \\ &= \Delta w \cdot h_0\end{aligned}$$

Notice that the income effect of an increase in w is positive. This makes sense: a higher wage gives you more income. But it complicates the analysis of labor supply because the negative substitution effect and the positive income effect have to fight it out.

Example

Let $u(x, \ell) = x^{\alpha} \ell^{1-\alpha}$, where $0 < \alpha < 1$ and assume $y = 0$.

The Lagrangean is $L(x, \ell, \lambda) = x^{\alpha}\ell^{1-\alpha} - \lambda(x - w (T-\ell))$.

FONC:

(1) $$\frac{\partial L}{\partial x} = \alpha x^{\alpha-1}\ell^{1-\alpha} - \lambda = 0$$

(2) $$\frac{\partial L}{\partial \ell} = (1-\alpha)x^{\alpha}\ell^{-\alpha} - \lambda = 0$$

(3) $$\frac{\partial L}{\partial \lambda} = -x + w(T - \ell) = 0$$

Solving (1) and (2):

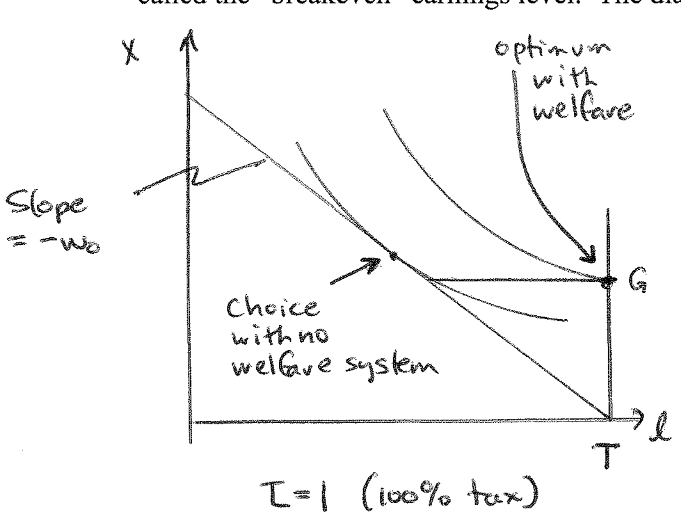
$$\frac{\alpha\ \ell}{(1-\alpha)\ x} = \frac{\ell}{w} \Rightarrow w\ell = \frac{(1-\alpha)\ x}{\alpha}$$

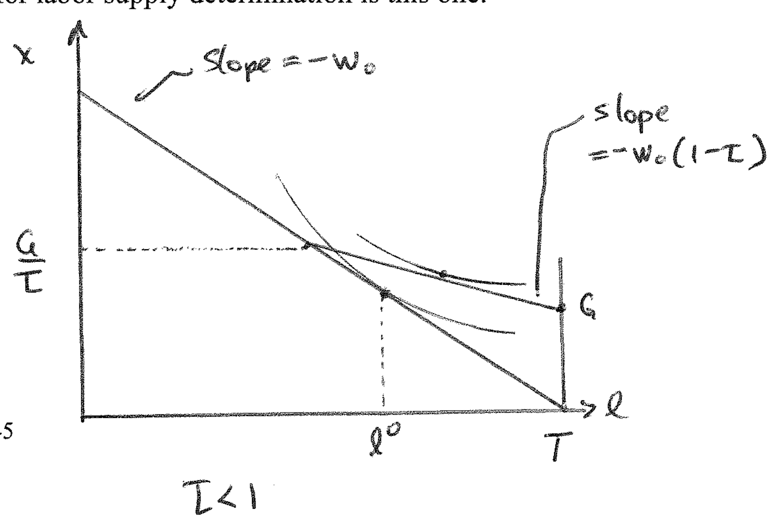
Substitute into (3) to get $-x + wT - \frac{1-\alpha}{\alpha}x = 0$ => $x = \alpha wT$; $\ell = (1-\alpha) T$; $h = T - \ell = \alpha T$

So, with Cobb-Douglas preferences and y=0, labor supply does not depend on wages: the income and substitution effects "cancel out". This is obviously a special case.

Effects of a Welfare System

Suppose that there is a "means-tested" benefit system that provides income support for low income people. One common type of program is characterized by 2 parameters: G=welfare amount for people with no other income; and τ = "tax rate". If you have earnings E, your welfare payment is reduced by τE. In the US welfare system in the 1980s and early 1990s, τ=1: each dollar of earnings reduced payments by \$1. More recently, some states such as CA have reformed their systems to allow τ<1. Note that with a tax rate of τ, people who earn less that G/τ are still eligible for (reduced) payments. This amount is often called the "breakeven" earnings level. The diagram for labor supply determination is this one:





Lecture 7

Consumption Over Time and Savings

An important issue in analyzing the rate of growth of the economy is the supply of savings. Other things equal, if more savings are available, the rate of interest that clears the market for investment funds market is lower, and more investment projects are undertaken (at least in principle). It is therefore interesting to ask: What determines an individual's saving and investment decisions?

For simplicity, we assume there are only two periods available: "today" and "tomorrow". We also abstract from consumption choices within each period and concentrate instead on consumption choices between periods.

Let

C_1 = consumption today

C_2 = consumption tomorrow

$U(C_1, C_2)$ = utility function over consumption plans

Y_1 = total income available today

Y_2 = total income available tomorrow

R = interest rate

Analytically, the problem is

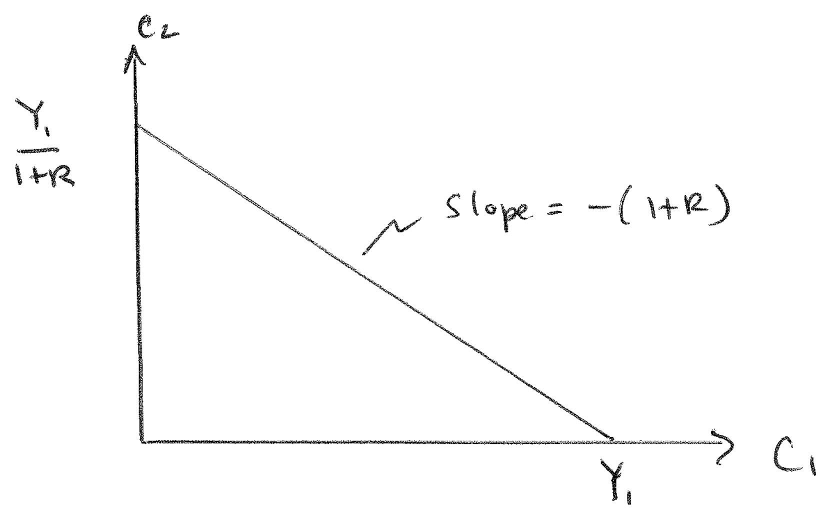
$$\text{Max } U(C_1, C_2) \quad \text{subject to} \quad C_1 + C_2/(1+R) = Y_1 + Y_2/(1+R)$$

Note that future income and expenditure are discounted to current dollars using R.

1. Suppose $Y_2 = 0$. This could be a stylized model to represent a decision about savings: period 1 is ages 25-65, period 2 is ages 65 and older. The budget constraint is

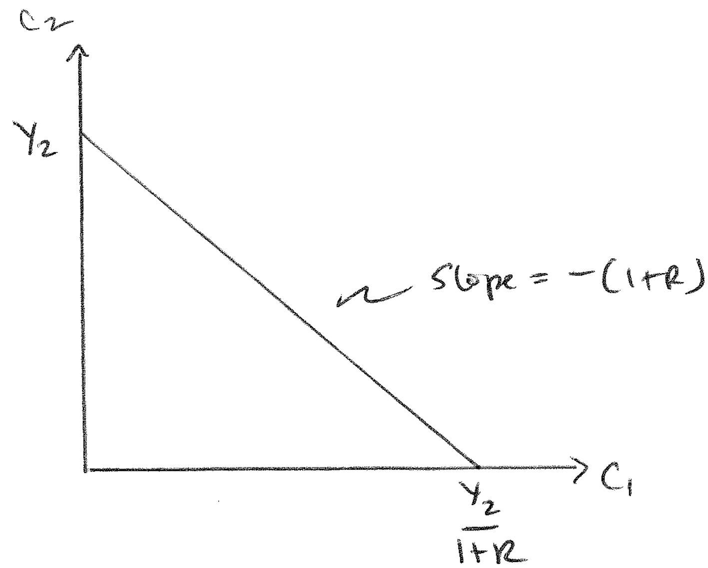
$$C_1 + C_2/(1+R) = Y_1 .$$

This looks like the "usual" problem, with $p_1 = 1$ and $p_2 = 1/(1+R)$.



2. Suppose $Y_1 = 0$. This could be a stylized model to represent a decision about how much to borrow while in school: period 1 is ages 19-25, period 2 is ages 26 and older. The budget constraint is

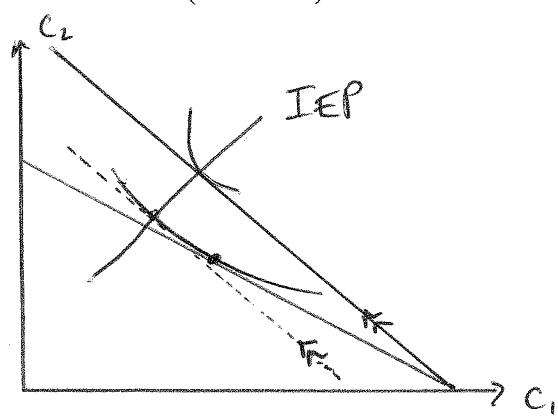
$$C_1 + C_2/(1+R) = Y_2/(1+R) .$$



Effects of Changes in Interest Rates

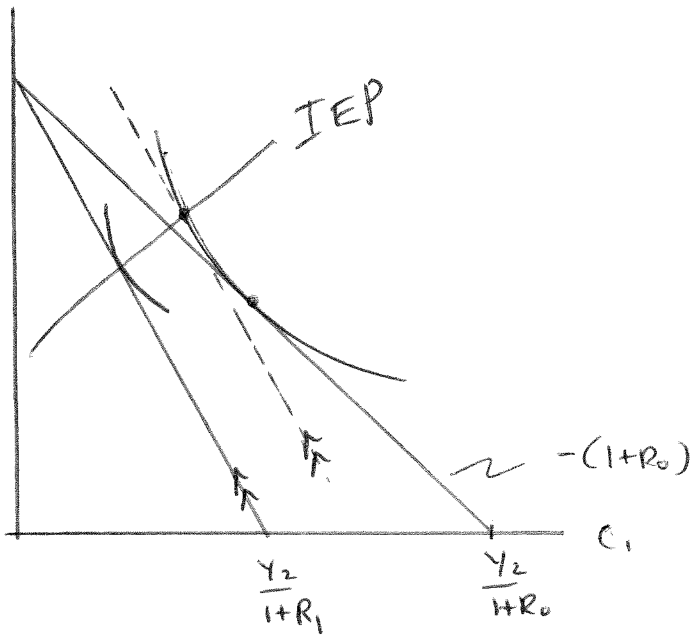
You often hear that higher interest rates will increase savings. The empirical support for this notion is very weak, however. Is the evidence inconsistent with the theory or is the theoretical prediction ambiguous?

(1) Consider a consumer with $Y_2 = 0$. An increase in R rotates the budget line around the point $(Y_1,0)$. As usual, there is a substitution effect (save more) and an income effect (consume more).



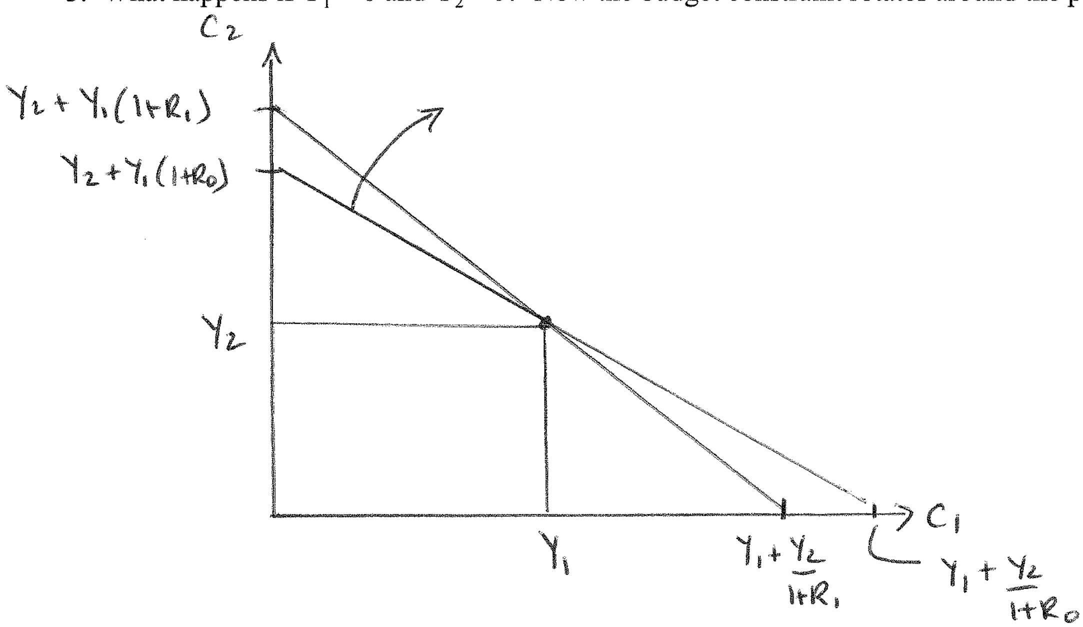
For consumers with $Y_2 = 0$, the IE and SE work in <u>opposite directions</u>. On the substitution side, the relative price of current consumption (1+R) goes up with R . Hence the substitution effect reduces current consumption. On the income side, an increase in R creates to a positive income effect.

(2) Consider consumers with $Y_1 = 0$. An increase in R rotates the budget line around the point $(0, Y_2)$.



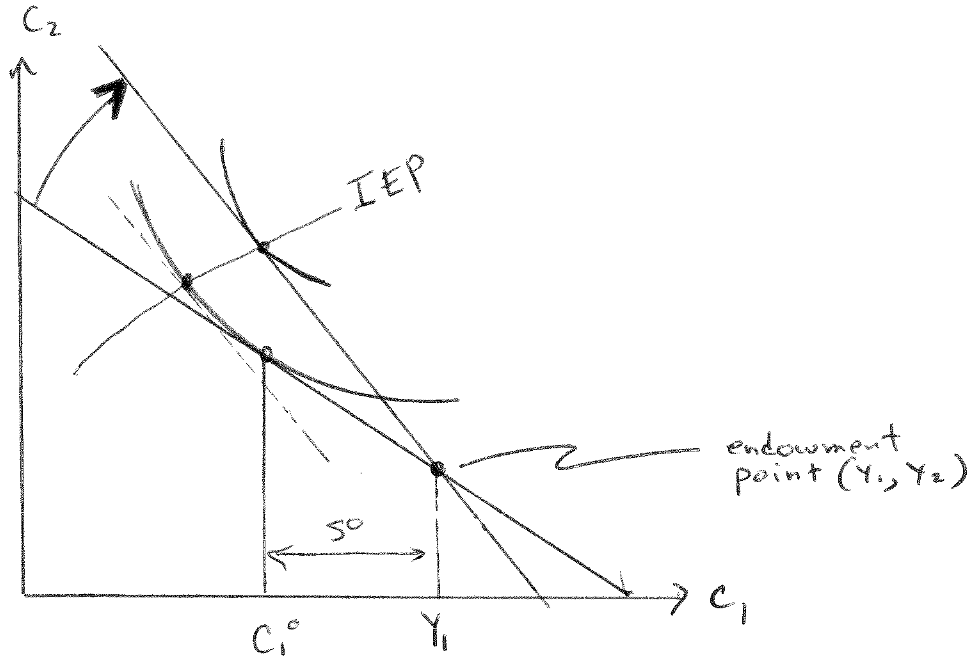
For those kinds of consumers (e.g. college students) an increase in R has an unambiguously negative effect on current consumption (they borrow less).

3. What happens if $Y_1 > 0$ and $Y_2 > 0$? Now the budget constraint rotates around the point (Y_1, Y_2).



If you were originally a <u>borrower</u> ($C_1 > Y$) there is a negative income effect. If you were originally a <u>saver</u> ($C_1 < Y_1$) their is a positive income effect.

Consider a saver:



The change in income (in second period units) is [income of new budget] - [cost of old bundle at new prices]

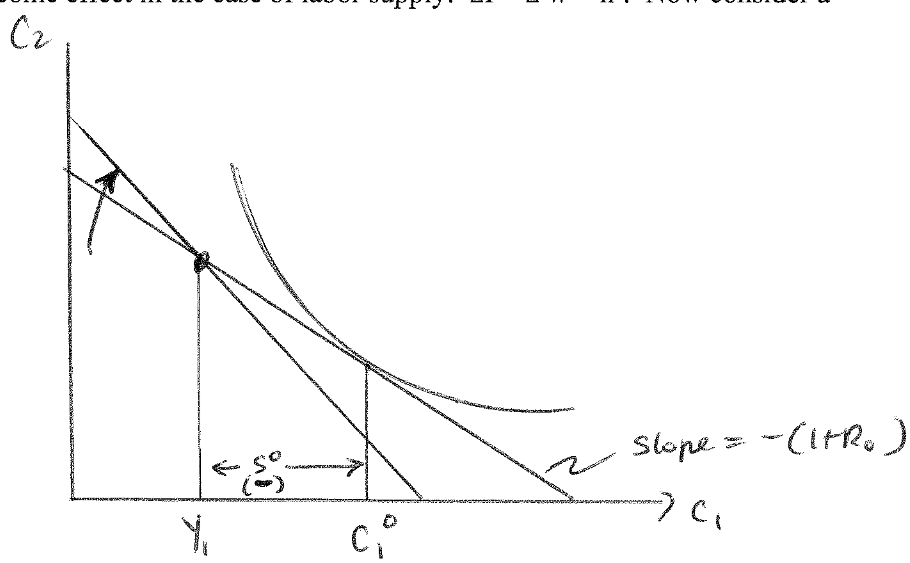
$$= Y_1 (1 + R_1) + Y_2 - [C_1^0 (1 + R_1) + C_2^0]$$

$$= Y_1 (1 + R_0 + R_1 - R_0) + Y_2 - C_1^0 (1 + R_0 + R_1 - R_0) - C_2^0$$

$$= Y_1 (1 + R_0) + Y_2 - C_1^0 (1 + R_0) - C_2^0 + (Y_1 - C_1^0) (R_1 - R_0)$$

$$= (R_1 - R_0) (Y_1 - C_1^0) = \Delta R \cdot S^0 \quad \text{where } S^0 = Y_1 - C_1^0 = \text{initial savings.}$$

Note the similarity to the income effect in the case of labor supply: $\Delta I = \Delta w \times h^0$. Now consider a borrower:



The change in income is [income of new budget] - [cost of old bundle at new prices]

$$= (R_1 - R_0) (Y_1 - C_1^0) = \Delta R \cdot S^0 .$$ But for a borrower, $S^0 < 0$.

Lecture 8

PRODUCTION AND COST - I

The technology available to a given firm is summarized by its production function . This function gives the quantities of output "produced" by various combinations of inputs. For example, an airline uses labor inputs, fuel, and machinery (airplanes, loading equipment, etc.) to produce the output "passenger seats." We write y = f(a, b) to signify that with inputs a and b, the output level y is available.

Examples: 1) (one input) $y = a$

$y = a^{\gamma} \quad (0 < \gamma < 1)$

$$y = \begin{cases} 0 & a < \bar{a} \\ 1 & a > \bar{a} \end{cases}$$

2) (two inputs) $y = a^{\alpha} b^{\beta}$ (Cobb-Douglas)

$y = \min[a,b]$ (Leontief)

$y = a + b$ (additive)

For two (or more) inputs, production functions are a lot like utility functions. The important difference is that output is measurable and has natural units (e.g., "seats"). It's as if the "indifference curves" have numbers attached to them that matter.

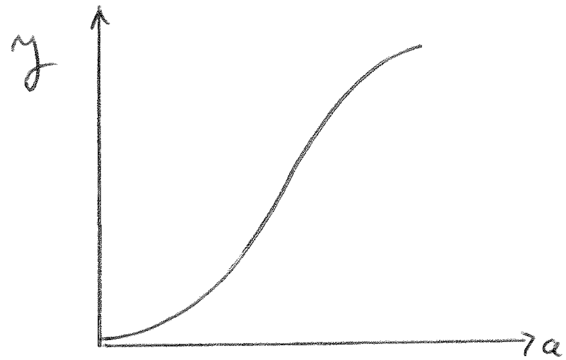
A second, less obvious, way to summarize technology is to compute the cost associated with producing a given output level y , at fixed prices for the inputs. In principle, it is easy to find the cost function if you know the production function using two steps:

1) find all possible ways of producing y .
2) find the cheapest one, and evaluate its cost.

Most of the economic behavior of firms is studied via the cost function. In the following lectures, we show how to derive cost functions and how to relate the properties of the cost function to the properties of the production function.

A. One-Factor Production and Cost Functions

Suppose that there is only one input (apart from, perhaps) a "set-up cost". Then we have a picture along the following lines:



$y = f(a)$.

Note that $f(0) = 0$, by convention.

Definitions and Facts

(i) The marginal product of factor a is the increase in output for a unit increase in input a.

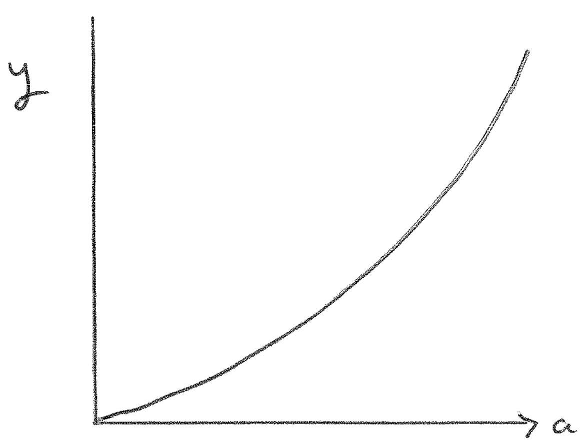
$$MP_a = \frac{\partial f(a)}{\partial a} = f'(a)$$

If factor a is <u>useful</u>, $f'(a) > 0$.

(ii) The average product of a factor a is the ratio of total output to total usage of the factor:

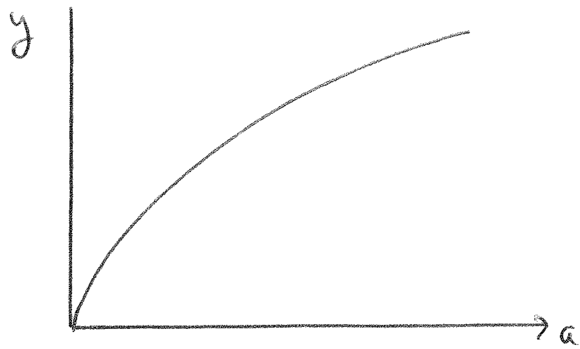
$$AP_a = \frac{f(a)}{a}.$$

(iii) If the MP of factor a is increasing, then $f''(a) > 0$ and we say that there are "increasing marginal returns": as the scale of output is expanded, each additional unit of input contributes more.



$f''(a) > 0$ (increasing returns)

If the MP is decreasing, the f''(a) < 0 and we say there are diminishing marginal returns.

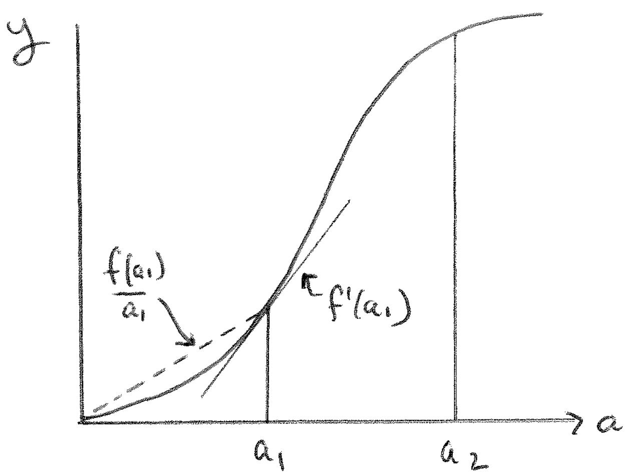


f''(a) < 0 (decreasing returns)

(iv) If $MP_a > AP_a$, then AP_a is increasing.

If $MP_a < AP_a$, then AP_a is decreasing.

Think of a baseball hitter: AP = lifetime average; MP = season average. A hitter who is having a season with an average that is better than his average so far will raise his lifetime average. Graphically:



slope of tangent = MP

slope of chord to origin = A

At $a = a_1$, AP = $\dfrac{f(a_1)}{a_1} < f'(a_1)$, AP is increasing.

At $a = a_2$, AP = $\dfrac{f(a_2)}{a_2} > f'(a_2)$, AP is increasing.

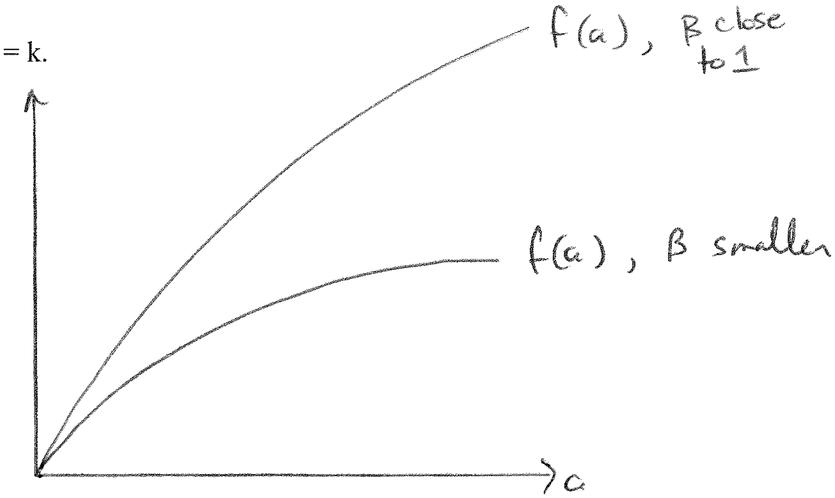
In general:

$$AP = \frac{f(a)}{a} \quad ; \quad \frac{dAP(a)}{da} = \frac{af'(a) - f(a)}{a^2} = \frac{1}{a}\left[f'(a) - \frac{f(a)}{a} \right] .$$

there is no page 52

Examples:

- $f(a) = k \times a$ with $k > 0$ (linear). $AP_a = MP_a = k$.
- $f(a) = a^{\beta}$ with $0 < \beta < 1$ (concave).



- $f(a) = 9a^2 - a^3$ for $a < 6$.

  For this function:

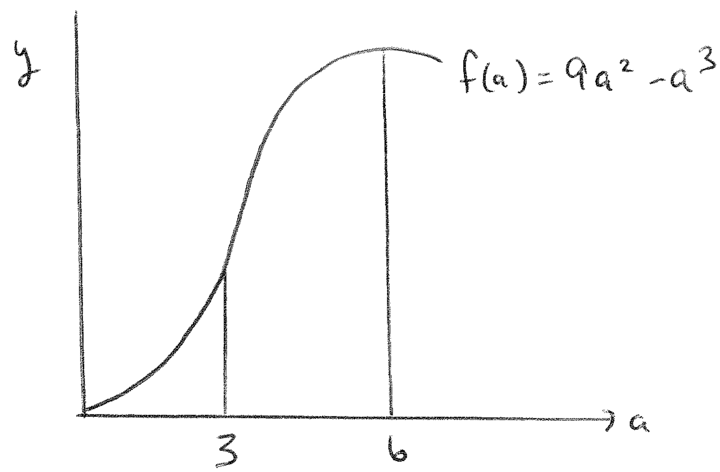
  $f'(a) = 18a - 3a^2$

  $f'(a) \geq 0 \Leftrightarrow a \leq 6$

  $f''(a) = 18 - 6a$

  $f''(a) > 0$ for $a < 3$

  & $f''(a) < 0$ for $a > 3$



COST

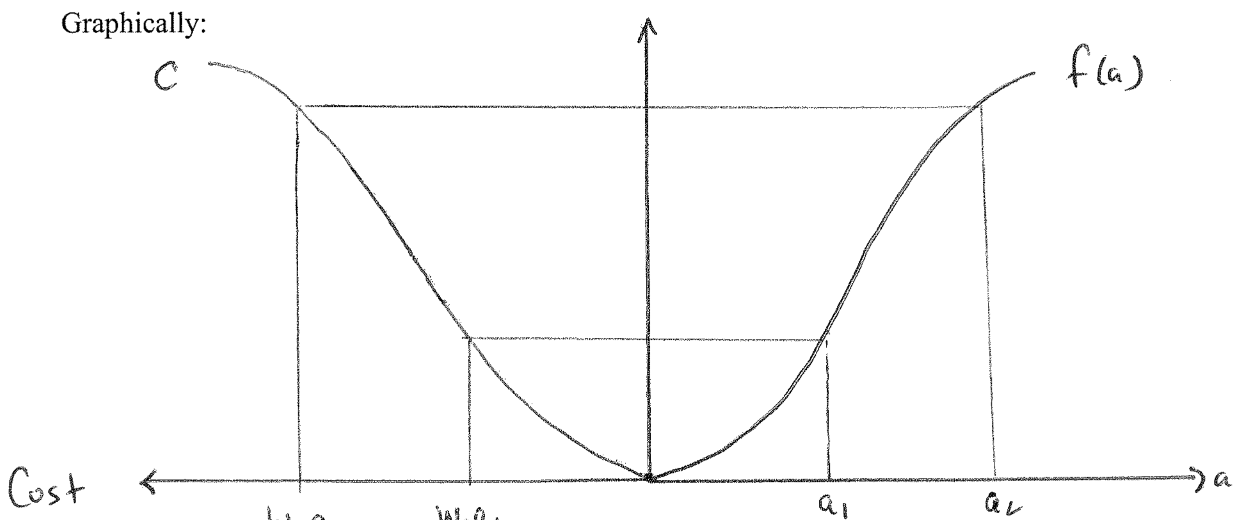
What is the cost function for a one-factor production function?

Let w = price per unit of a. Then:

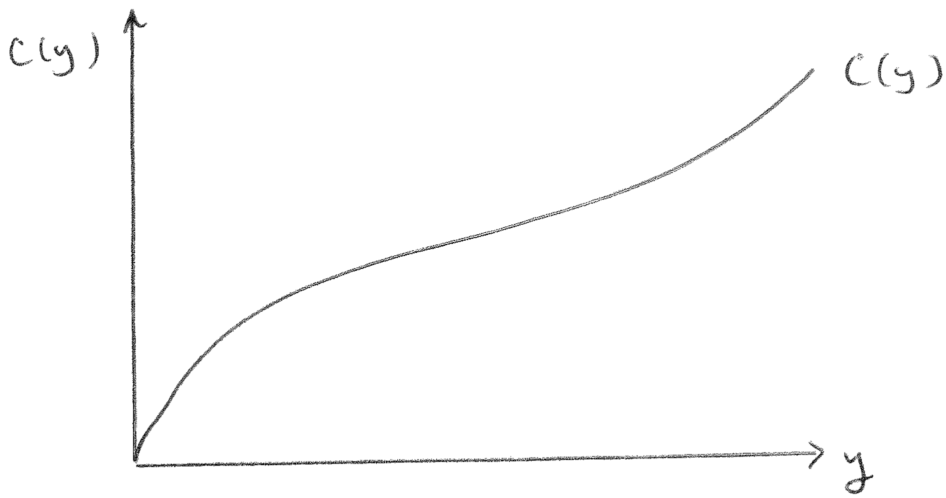
$$C(y,w) = \min w\, a \quad \text{s.t. } y = f(a)$$

For $y = f(a)$, we need $a = f^{-1}(y)$. Therefore $C(y,w) = w\, f^{-1}(y)$.

Graphically:



Rotate the left hand panel of this graph 90°

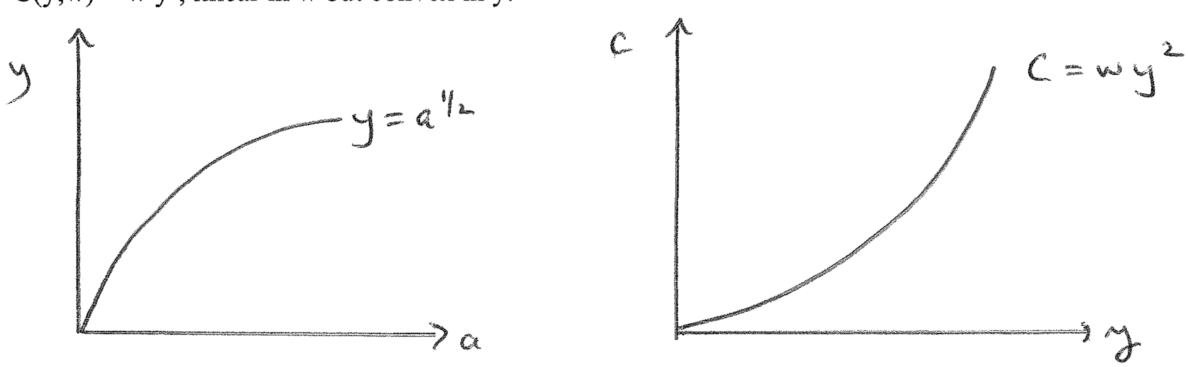


If w is fixed we often write the cost function as a function only of output, C(y) . In the more general case we write C(y,w). Define marginal cost MC(y) = C'(y) , and average cost AC(y) = C(y)/y.

Examples:

1) $y = f(a) = 2a$ (linear) => $a = f^{-1}(y) = \frac{1}{2}y$ (Linear "input requirement function")
$C(y,w) = w(\frac{1}{2}y) = \frac{1}{2}\, w\, y$, linear in both y and w.

2) $y = f(a) = a^{1/2}$ => $a = f^{-1}(y) = y^2$ (Convex input requirement function).
$C(y,w) = w\, y^2$, linear in w but convex in y.



Relations between MC and MP

The MC of output is the amount it costs to produce an extra unit.

By definition of MP_a one unit of input adds $MP_a = f'(a)$ units to output.

=> $1/MP_a = 1/f'(a)$ units of a are needed to produce 1 unit of y

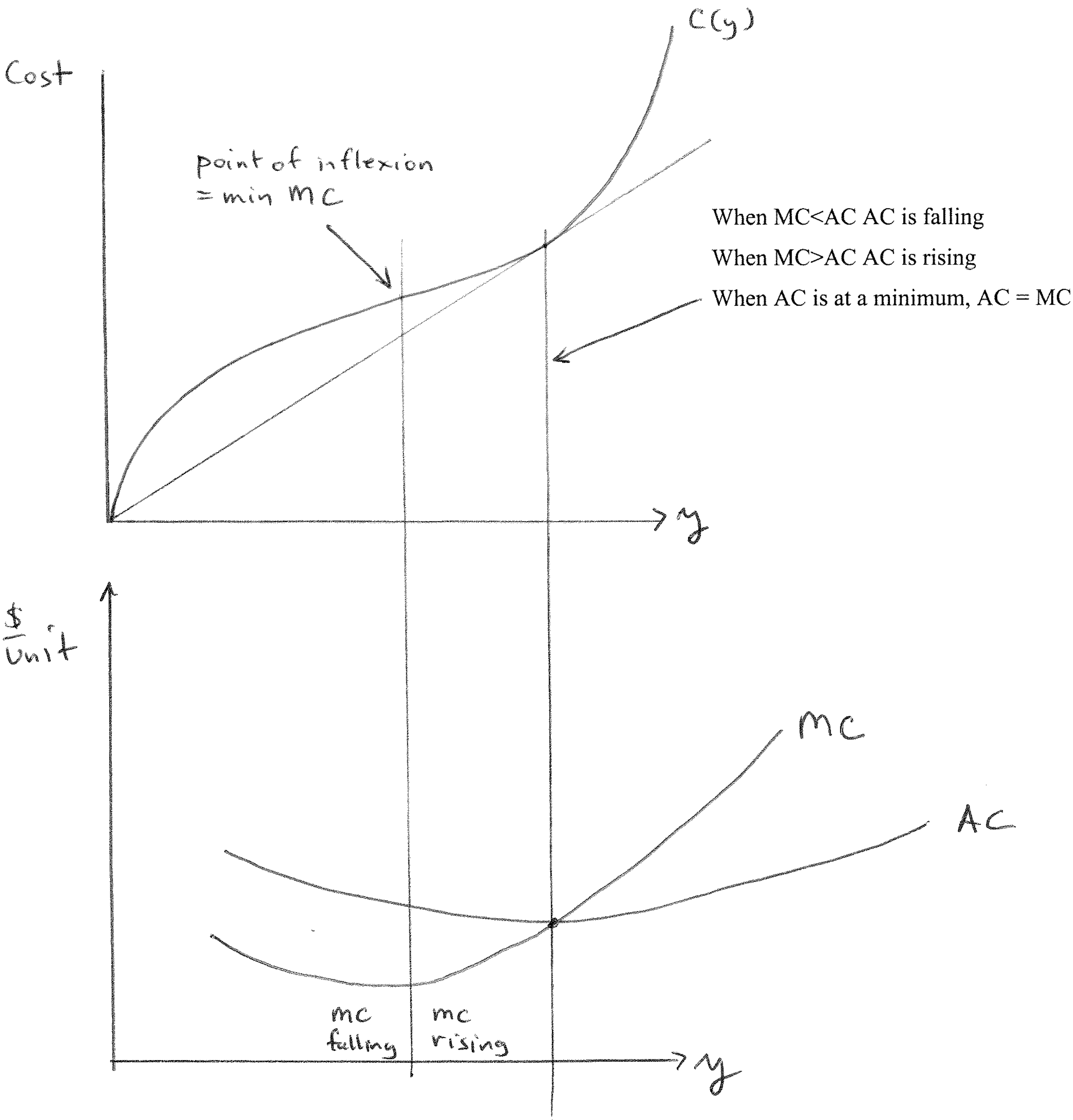
=> the marginal the cost of an extra unit is $MC(y) = w/f'(a)$, when the production function is $y = f(a)$.

Alternatively, $C(y) = wf^{-1}(y)$ using the input requirement $a = f^{-1}(y)$. Thus

$$C'(y) = w\,\frac{df^{-1}(y)}{dy} = \frac{w}{f'(a)}\ ,$$

using the fact that the derivative of an inverse function is the inverse of the derivative.

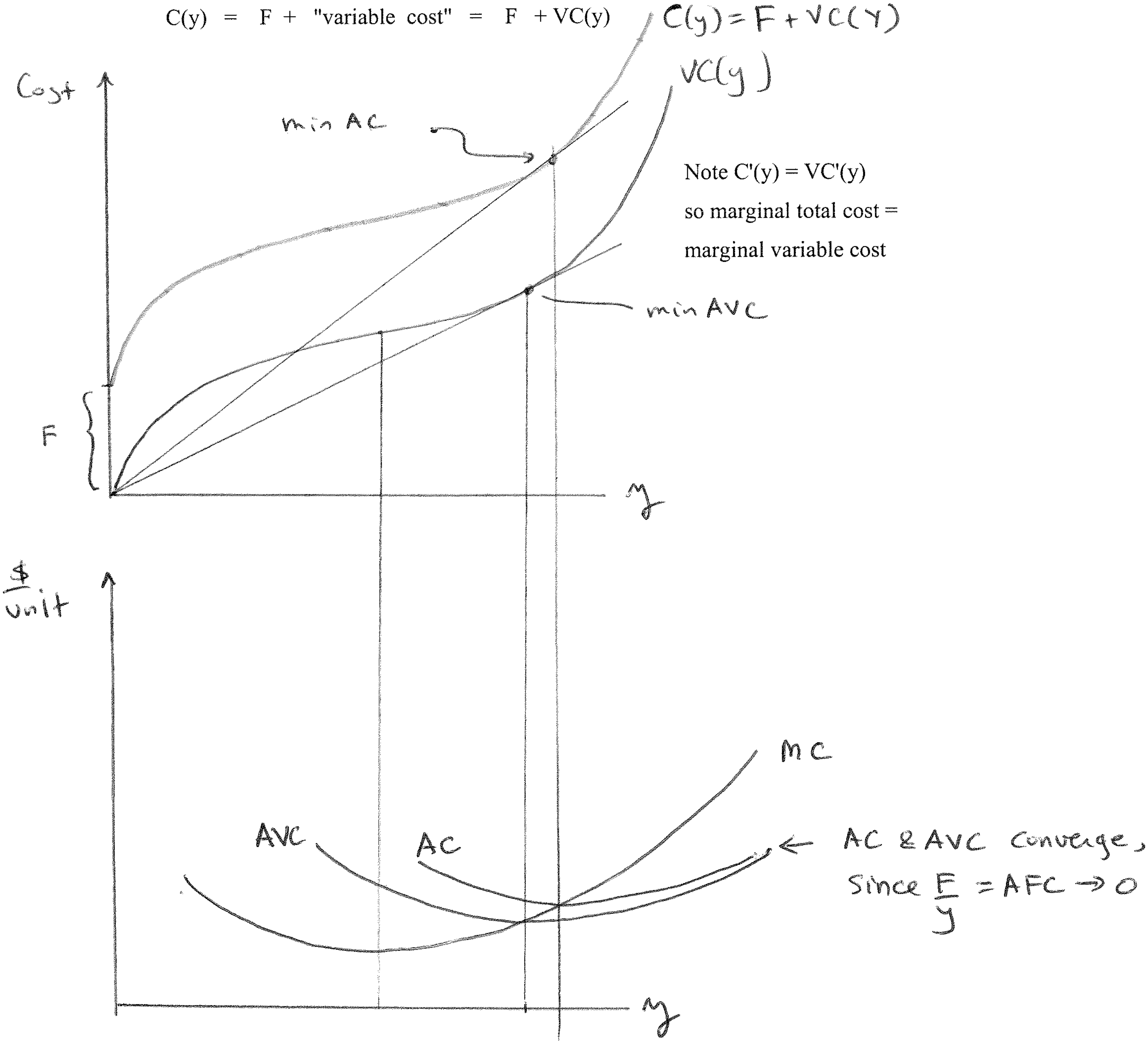
Geometry of C(y), AC(y) and MC(y)



Sometimes we add on a "set up" cost F (also called a fixed cost)

Then total cost is

$$C(y) = F + \text{"variable cost"} = F + VC(y)$$



Note C'(y) = VC'(y)

so marginal total cost =

marginal variable cost

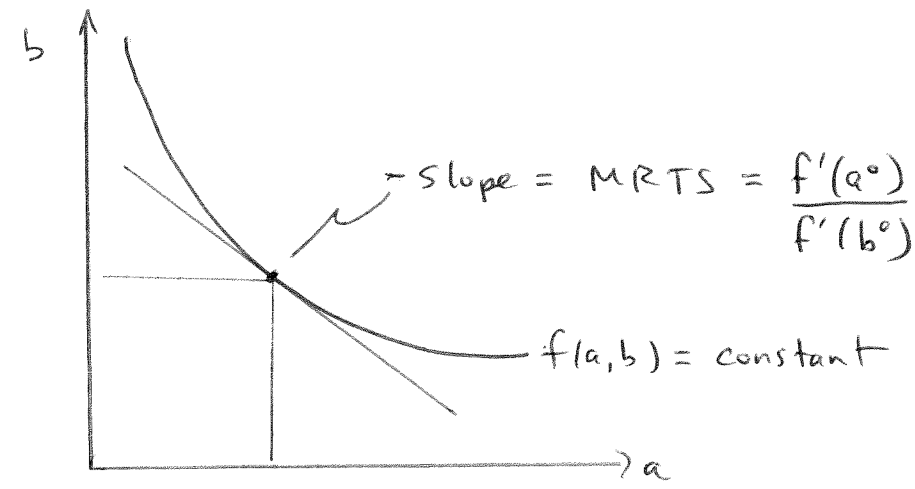
NOTES: 1. min AC occurs to the right of min AC. Why?

2. MC cuts through min of both AC and AVC. Why?

Lecture 9

Production and Cost II. Two-Factor Production and Cost Functions

The analysis of production and cost is more interesting when we can use combinations of two (or more) inputs to produce y. The production function is y = f(a,b). As in consumer theory, we begin by thinking about combinations of inputs that produce the same output. In the firm case these are called isoquants.



We define the marginal rate of technical substitution (MRTS) as the slope of an isoquant. It tells you how many units of b you need to add, per unit of a given up, to keep output constant.

Formally, suppose $y^0 = f(a^0,b^0)$, and consider changing a and b to keep output fixed at y^0,

$$dy = f_a\, da + f_b\, db = 0$$

$$\Rightarrow \left.\frac{db}{da}\right|_{y_0} = -\frac{f_a(a,b)}{f_b(a,b)} = -\frac{MP_a}{MP_b}$$

The MRTS is analogous to the marginal role of substitution (MRS) in consumer theory. When there are 2 or more inputs, the production function is characterized by both the degree of substitutability between inputs (the curvature of isoquants) and the extent to which output expands as inputs are expanded proportionately. The latter gives rise to the idea of returns to scale. For a production function y = f(a,b), we say f has *constant returns to scale* (CRS) if

$$f(\gamma a, \gamma b) = \gamma f(a,b) \quad \text{for any } \gamma > 0.$$

If y = f(a,b) we say f has *decreasing returns to scale* (DRS) if

$$f(\gamma a, \gamma b) < \gamma f(a,b) \quad \text{for any } \gamma > 1.$$

So if you double both inputs, you get less than a doubling of output. On the other hand, the same inequality implies that if you reduce inputs by some proportion, your output falls by a smaller proportion. So DRS says that smaller firms are always more "efficient". Conversely we say that f has *increasing returns to scale* (IRS) if

$$f(\gamma a, \gamma b) > \gamma f(a,b) \quad \text{for any } \gamma > 1.$$

Examples:

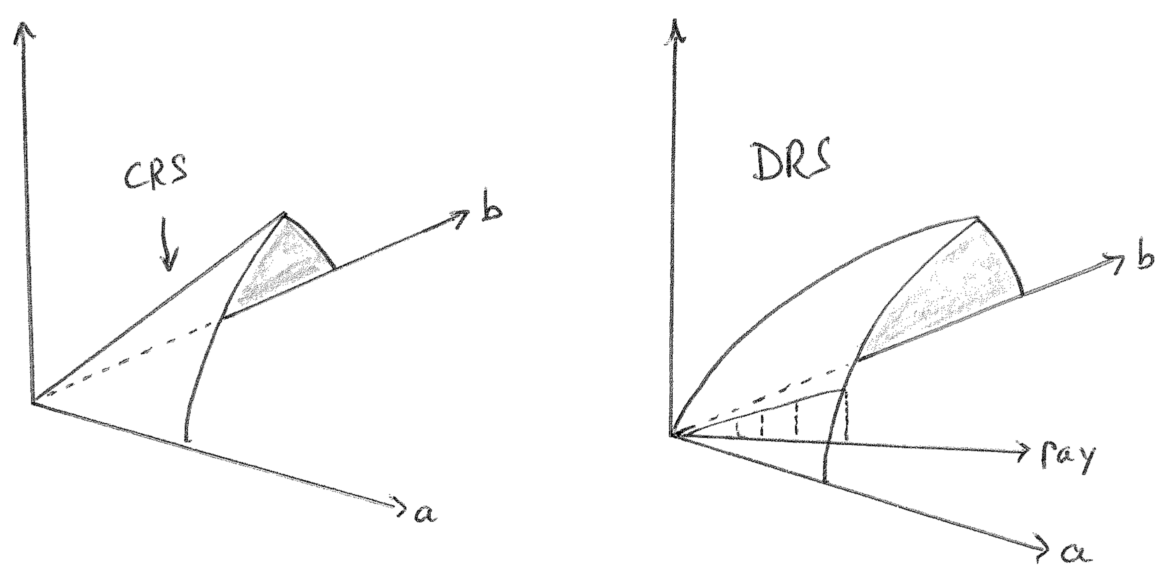
(i) one input $f(a) = a^{\theta}$ CRS if $\Theta = 1$
DRS if $\Theta < 1$
IRS if $\Theta > 1$.

(ii) Cobb Douglas $f(a,b) = a^{\alpha} b^{B}$. CRS if $\alpha + \beta = 1$.
DRS if $\alpha + \beta < 1$.
IRS if $\alpha + \beta > 1$.

Check:

$$f(\gamma a,\gamma b) = (\gamma a)^{\alpha} (\gamma b)^{\beta}$$
$$= \gamma^{\alpha+\beta} a^{\alpha} b^{\beta}$$
$$= \gamma^{\alpha+\beta} f(a,b)$$

Geometrically, returns to scale tells you whether the function f is concave or convex "over top" of a ray from the origin:



Derivation of the Cost Function

Given a production function f(a,b) and prices w_a, w_b, we can write

$$C(w_a, w_b, y) = \min\ w_a a + w_b b \quad \text{subject to } f(a,b) \geq y.$$

Set up the Lagrangean (using μ for the multiplier):

$$L = w_a a + w_b b - \mu (f(a,b) - y).$$

$$\frac{\partial L}{\partial a} = w_a - \mu f_a(a,b) = 0$$

$$\frac{\partial L}{\partial b} = w_b - \mu f_b(a,b) = 0$$

$$\frac{\partial L}{\partial \mu} = -f(a,b) + y = 0$$

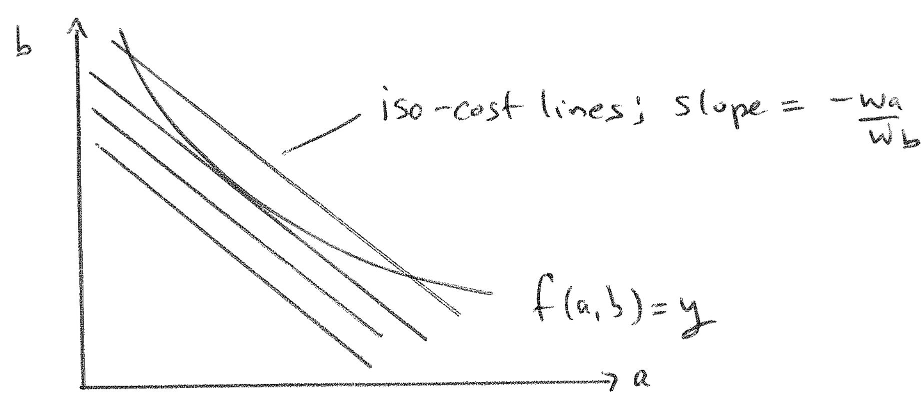
The ratio of the first 2 first-order conditions gives

$$\frac{Wa}{Wb} = \frac{fa(a,b)}{fb(a,b)} = \text{MRTS}.$$

Geometrically, we find the tangencies of "iso-cost" lines

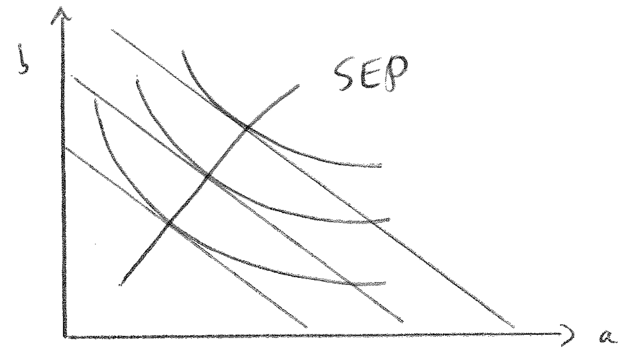
$$w_a a + w_b b = \text{constant}$$

with the isoquant corresponding to the desired level of output y:

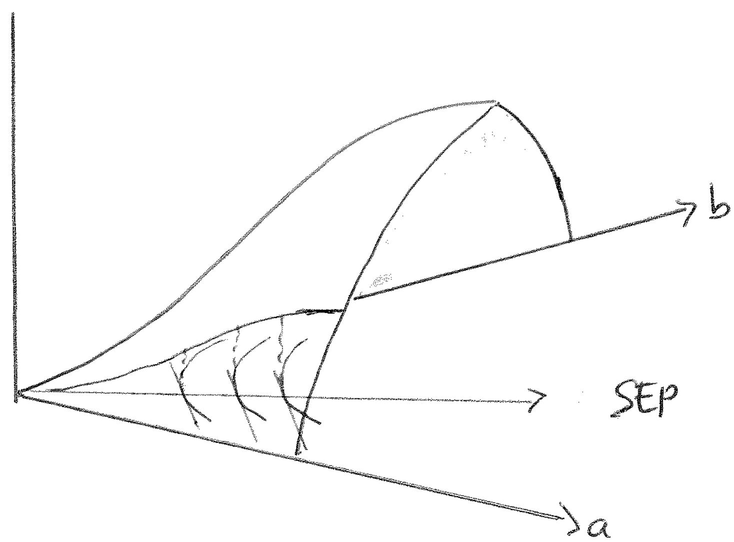


Notice that the problem is "reversed" relative to a consumer. In the cost problem, you are constrained to an isoquant and have to find the lowest "budget line" (iso-cost line). In the consumer problem, you are constrained to a budget line and have to find the highest "isoquant" (indifference curve).

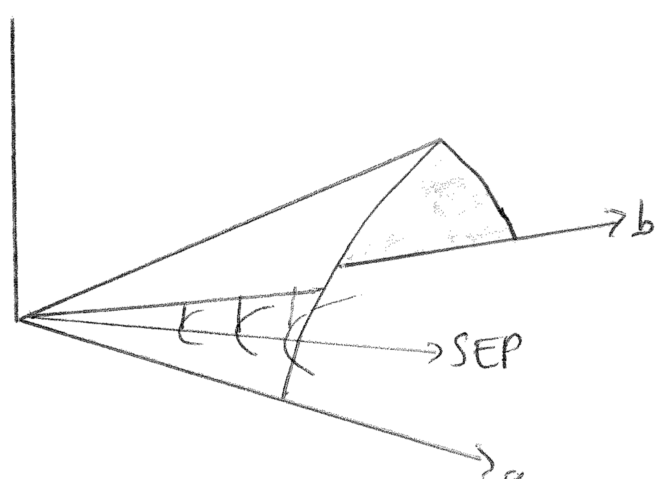
If we consider finding the cheapest way to get different levels of output at <u>constant</u> (w_a, w_b) we trace out the scale expansion path (SEP)



Note the similarity between a firm's SEP and a consumer's income expansion path (IEP). Geometrically, the shape of the cost function (as a function of y) depends on the shape of the production function "over top" of the SEP:



If the curve over the SEP is S-shaped (as illustrated) we get cost functions of the usual shape. If the curve is linear we get a <u>linear</u> cost function:



Marginal Cost

If we needed to produce 1 more unit of y, we could use input a, or input b (or both). If we use only a, we need $1/MP_a$ units of a for 1 unit of y. The marginal cost is w_a/MP_a (just as in the 1-input case). By symmetry, we could also use input b, at marginal cost w_b/MP_b. But from the first-order conditions

$$\frac{W_a}{W_b} = \frac{MP_a}{MP_b} \quad => \quad \frac{W_a}{MP_a} = \frac{W_b}{MP_b} \quad .$$

So on the margin, you are indifferent to expanding output via increases in a or increases in b. This reflects that fact that a and b were "optimally chosen" to begin with. Note also that

$$\mu = \frac{w_a}{f_a(a,b)} = \frac{w_a}{MP_a} = \frac{W_b}{MP_b} .$$

Thus the Lagrange multiplier in the cost-minimization problem gives marginal cost.

Examples:

(i) $f(a,b) = \min[a,b/2]$. At a cost minimum we must have $a=b/2=y$.

=> $C(w_a,w_b, y) = y\,(w_a + 2w_b)$. Note this production function has CRS.

(ii) $f(a,b) = a + 2b$. These are linear isoquants, with $f_a/f_b = 1/2$.

If $w_a/w_b > 1/2$, should use only b. Then $y = 2b$ => $b = y/2$, and $C(w_a, w_b,y) = ½w_b\, y$.

But, if $w_a/w_b < 1/2$, should use only a. Then $y = a$, and $C(w_a,w_b,y) = w_a\, y$.

Combining these results, for any w_a, w_b, we have $C(w_a, w_b, y) = \min[w_a, w_b/2]\, y$.

These two examples illustrate the what is called the "duality" relationship between cost and production functions. Leontief production functions imply linear cost functions. Linear cost functions imply "Leontief-like" cost functions.

(iii) $f(a,b) = a^\alpha b^\beta$. (This was in problem set number 4).

The Lagrangean is

$$L(a,b,\mu) = w_a\, a + w_b\, b - \mu(a^\alpha b^\beta - y).$$

$$\frac{\partial L}{\partial a} = w_a - \mu\alpha^{\alpha-1}b^\beta = 0$$

$$\frac{\partial L}{\partial b} = w_b - \mu\beta\alpha^{\alpha}b^{\beta-1} = 0$$

$$\frac{\partial L}{\partial \mu} = -\alpha^\alpha b^\beta + y = 0$$

Using the first two FOC's, we get:

$$\frac{w_a}{w_b} = \frac{\alpha a^{\alpha-1} b^{\beta}}{\beta a^{\alpha} b^{\beta-1}} = \frac{\alpha b}{\beta a}$$

$$\text{or} \quad b = \frac{\beta a w_a}{\alpha w_b}$$

Substitute into the constraint: $a^{\alpha} b^{\beta} = a^{\alpha} \left[\frac{\beta a w_a}{\alpha w_b}\right]^{\beta}$

$$= a^{\alpha+\beta} \beta^{\beta} w_a^{\beta} \alpha^{-\beta} w_b^{-\beta} = y$$

This gives the input requirement function for input a:

$$a = y^{\frac{1}{\alpha+\beta}} \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} w_a^{-\frac{\beta}{\alpha+\beta}} w_b^{\frac{\beta}{\alpha+\beta}}$$

and substituting back in (or using symmetry) we get $b = y^{\frac{1}{\alpha+\beta}} \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{-\frac{\alpha}{\alpha+\beta}}$

Finally $C(w_a, w_b, y) = w_a a + w_b b$ when a and b are set to the cost-minimizing input choices, so

$$C(w_a, w_b, y) = y^{\frac{1}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{\frac{\beta}{\alpha+\beta}} \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + y^{\frac{1}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{\frac{\beta}{\alpha+\beta}} \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}}$$

$$= y^{\frac{1}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{\frac{\alpha}{\alpha+\beta}} \left\{\left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}}\right\}$$

If $\alpha + \beta = 1$ (constant returns to scale) this simplifies to

$$C(w_a, w_b, y) = y\, w_a^{\alpha} w_b^{\beta} \left\{\left(\frac{\alpha}{\beta}\right)^{\beta} + \left(\frac{\beta}{\alpha}\right)^{\alpha}\right\}$$

$$= y\, w_a^{\alpha} w_b^{\beta} \left(\alpha^{-\alpha} \beta^{-\beta}\right)$$

So with CRS, cost is <u>linear</u> in output. In general the exponent on y in the cost function is $1/(\alpha+\beta)$, so if $\alpha + \beta > 1$, cost is concave in output (IRS) and if $\alpha + \beta < 1$, cost is convex in output (DRS).

Economics 101A
Handout

Cost Functions and Input Requirement Functions

Suppose we have a production function $f(x_1, x_2)$, and an associated cost function $C(y, w_1, w_2)$. We obtain the cost function by solving the "cost minimization problem":

$$\text{Min} \quad w_1 x_1 + w_2 x_2 \quad \text{s.t.} \quad f(x_1, x_2) = y .$$

We set up the Lagrangian:

$$L(x_1, x_2, \mu) = w_1 x_1 + w_2 x_2 - \mu (f(x_1, x_2) - y) .$$

The f.o.c.'s are:

$$L_1 = w_1 - \mu f_1(x_1, x_2) = 0$$

$$L_2 = w_2 - \mu f_2(x_1, x_2) = 0$$

$$L_3 = - f(x_1, x_2) + y = 0.$$

The first two of these imply the "tangency condition" $w_1/w_2 = f_1/f_2$, while the third just gives back the constraint. Solving these two equations in two unknowns we get the input requirement functions (IRF's):

$$x_1 = x_1^*(y, w_1, w_2)$$

$$x_2 = x_2^*(y, w_1, w_2) .$$

The IRF's are analogous the consumer's demand functions: they represent the optimal (cost-minimizing) input choices to produce y when input prices are (w_1, w_2). With these we can obtain the cost function

(*) $\quad C(y, w_1, w_2) = w_1 x_1^*(y, w_1, w_2) + w_2 x_2^*(y, w_1, w_2)$,

which is just the cost of the "cost minimizing" input combination.

<u>Sheppard's Lemma</u>

It turns out that if you know C, you can "recover" the IRF's by simple differentiation:

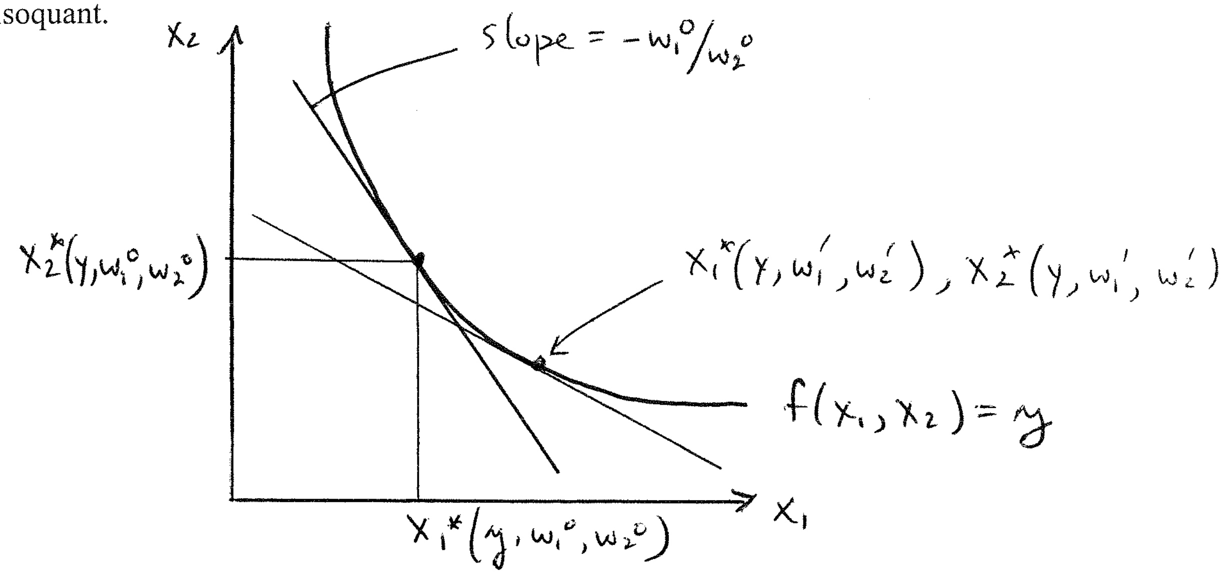
$$x_1^*(y, w_1, w_2) = \partial C(y, w_1,w_2)/ \partial w_1 .$$

At first glance, it looks like this is inconsistent with (*). Indeed, differentiating (*) with respect to w_1, we get **three terms**:

(□) $\quad \partial C(y, w_1,w_2)/ \partial w_1 = x_1^*(y, w_1, w_2) + w_1 \partial x_1^*(y, w_1, w_2)/\partial w_1 + w_2 \partial x_2^*(y, w_1, w_2)/\partial w_1$.

However, when we change an input price, $x_1^*(y, w_1, w_2)$ and $x_2^*(y, w_1, w_2)$ have to move in such a way as to keep total output constant. In fact, as w_1 changes, the combinations of (x_1^*, x_2^*)

trace out the isoquant.



In other words, we have that

$$f(\,x_1^*(y, w_1, w_2)\,,\ x_2^*(y, w_1, w_2)\,) \;=\; y\,.$$

This has to hold as w_1 varies, so we can differentiate w.r.t. w_1 (applying the chain rule) to get:

$$f_1\, \partial x_1^*(y, w_1, w_2)/\partial w_1 \;+\; f_2\, \partial x_2^*(y, w_1, w_2)/\partial w_1 = 0\,.$$

This says that

(†) $\qquad \partial x_2^*(y, w_1, w_2)/\partial w_1 = -\, f_1 / f_2\ \partial x_1^*(y, w_1, w_2)/\partial w_1$.

So, since x_1^* falls as w_1 rises, x_2^* has to rise, and the rates of change are in the ratio of f_1/f_2. (Note that the response of x_1^* to a change in w_1 is just like a "substitution effect" for a consumer. Since the isoquant has diminishing MRTS, a rise in w_1 must lead to a fall in x_1^*). Substituting (†) into (□), we get

$$\partial C(y, w_1,w_2)/\, \partial w_1 = x_1^*(y, w_1, w_2) + \ \partial x_1^*(y, w_1, w_2)/\partial w_1\ \{\, w_1 \;-\; w_2\ f_1 / f_2 \,\}\,.$$

But from the tangency condition, $w_1 \;-\; w_2\ f_1 / f_2 = 0$. So the second and third terms in (□) always sum to 0, leaving us with (*).

Equation (*) says that if w_1 rises, the first-order effect on cost is just proportional to the amount of input 1 that you were using to produce at minimum cost. Although the optimal choices of x_1 and x_2 also change, they have to change in such as way as to keep y constant, and because of the initial tangency condition the movements in the inputs leave cost constant.

Lecture 10

OUTPUT (SUPPLY) DETERMINATION

So far, we have studied cost, taking output as given. In this lecture, we consider the output or supply decision of individual competitive firms. By "competitive," we mean that the firm takes the prices for inputs and outputs as exogenous (i.e., beyond the firm's control). For any firm, profit is defined as the difference between revenue and cost. For a competitive firm that uses 2 inputs, 1 and 2, to produce a single output (y) which is sold at a price p, profit as a function of y is:

$$\pi(y) = p\,y - C(y, w_1, w_2) \ .$$

Note that revenue ($p \times y$) is a linear function of output, whereas the cost function is potentially non-linear. We assume that the firm chooses y to maximize profit:

$$\max \ p\,y - C(y, w_1, w_2)$$

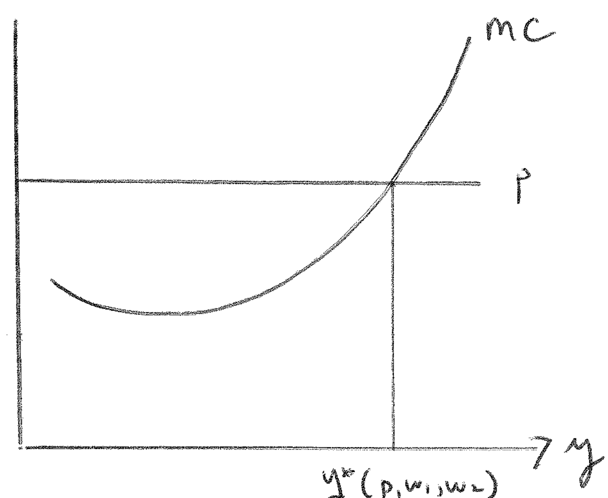
FONC:

$$\frac{d\Pi}{dy} = p - C_y(y^*, w_1, w_2) = 0, \quad \underline{\text{or}} \quad \text{price} = \text{marginal cost, when } y = y^*.$$

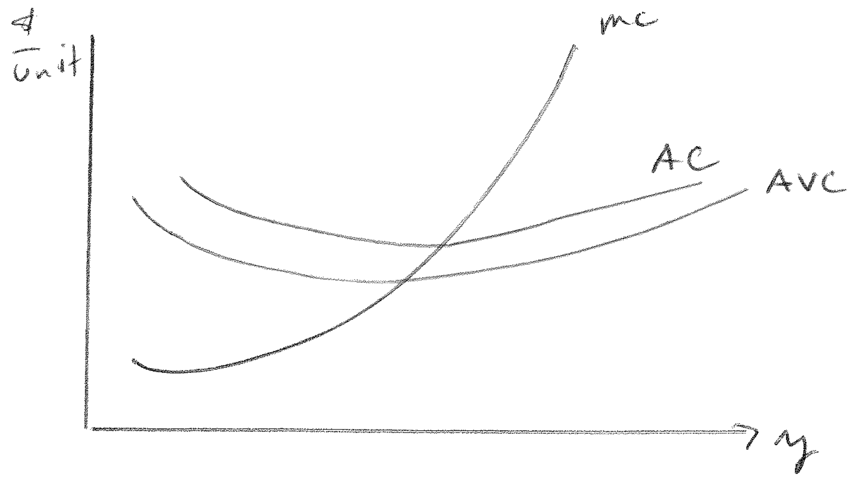
The second order condition for maximum profit is

$$\frac{d^2\Pi}{dy^2} < 0 \quad \Rightarrow \quad -C_{yy}(y^*, w_1, w_2) < 0 \quad \Rightarrow \quad C_{yy}(y^*, w_1, w_2) > 0 : \text{increasing MC at optimal y.}$$

Here is the diagram:

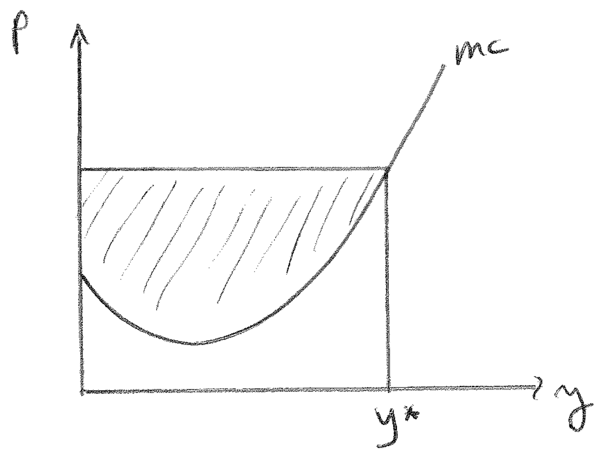


Note that y^* is a function of p, and all input prices. We write $y = y^*(p, w_1, w_2)$ as the "supply function". What happens if $\pi < 0$ at $y = y^*(p, w)$?



1) If $p < AVC$ then $y^* = 0$. The firm is losing on both fixed an variable inputs: the best supply action is to shut down.
2) If $p > AC$ the firm is earning profit, so y^* is defined by $p = MC(y^*)$.
3) If $AVC < p < AC$, the firm is turning a loss, but it is covering its "operating" costs, and only failing to pay off its fixed costs. It may well stay in business and hope for better times.

The following diagram is a useful representation of the firm's optimal choice.



The rectangle py^* represents revenue.

The area under MC represents costs (<u>not</u> including fixed costs).

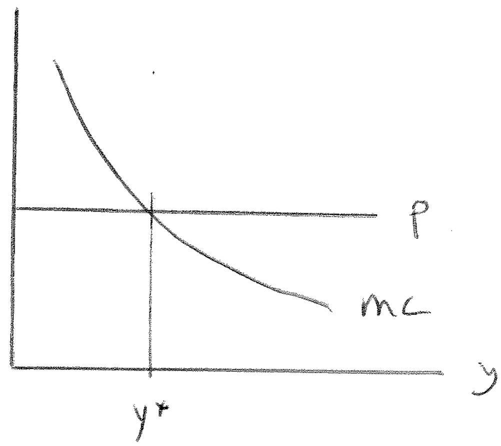
Thus the <u>shaded</u> <u>area</u> represents profit <u>not</u> including fixed cost payments.

Here, we are using the fact that $C(y^*) = \int MC(y)\,dy$ + constant, where the limits of integration are 0, y^*. This is just the **area** under the MC function.

<u>Observations</u>

1) If MC is constant (e.g., Cobb-Douglas with exponents that add to 1) then either $y^*= 0$ or $y^* = \infty$ because either profit is negative or profit is infinite (if there are no fixed costs).

2) If MC is always decreasing, then supply is not defined (or equals 0).



At y^* defined by $p = MC(y^*)$, profit is not maximized.

Why? Consider a cut in output. Cost falls by MC, revenues falls by p. So π actually increases.

The second-order conditions are not satisfied, since $C_{yy} < 0$.

Examples:

(1) $y = x^a \quad 0 < a < 1$ (one input, decreasing returns)

The input requirement function is $x^*(y) = y^{1/a}$ (this does not depend on input prices). Thus,

$C(w,y) = w\,x^*(y) + F$ (F = fixed costs)

$\quad = w\,y^{1/a} + F$

$$MC(y) = \frac{w}{a}\,y^{\frac{1-a}{a}}$$

$$AC(y) = \frac{C(w,y)}{y} = \frac{F}{y} + w\,y^{\frac{1-a}{a}}$$

Output supply choice y^* solves $p = MC(y)$, implying

$$p = \frac{w}{a}\,y^{\frac{1-a}{a}} \quad \text{or} \quad y^*(p,w) = \left[a\,\frac{p}{w}\right]^{\frac{a}{1-a}} .$$

NOTE: (a) y^* is homogenous of degree 0 in p.w.

(b) y^* increases with p, decreases with w.

(2) $y = x_1^{\alpha}\,x_2^{\beta} \quad \alpha+\beta < 1$ Cobb-Douglas with decreasing returns to scale.

Recall that $C(y, w_1, w_2) = k_1\, w_1^{\frac{\alpha}{\alpha+\beta}}\, w_2^{\frac{\beta}{\alpha+\beta}}\, y^{\frac{1}{\alpha+\beta}}$, for $k_1 > 0$ some constant. Therefore:

$$MC(y) = k_2\, y^{\frac{1-\alpha-\beta}{\alpha+\beta}}\, w_1^{\frac{\alpha}{\alpha+\beta}}\, w_2^{\frac{\beta}{\alpha+\beta}} \quad \text{for some constant } k_2 .$$

Setting p = MC and solving for y => $y^* = k_3\, p^{\frac{\alpha+\beta}{1-\alpha-\beta}}\, w_1^{\frac{-\alpha}{1-\alpha-\beta}}\, w_2^{\frac{-\beta}{1-\alpha-\beta}}$ for some constant k_3.

<u>or</u> $$\log y^* = \text{constant} + \frac{\alpha+\beta}{1-\alpha-\beta}\log p - \frac{\alpha}{1-\alpha-\beta}\log w_1 - \frac{\beta}{1-\alpha-\beta}\log w_2 .$$

Again y^* is homogeneous of degree 0 in all prices, increasing in p, decreasing in w_1 and w_2 .

<u>Exercise</u>:

For a general cost function, prove that the competitive supply response is homogeneous of degree 0 in all prices (input <u>and</u> output prices).

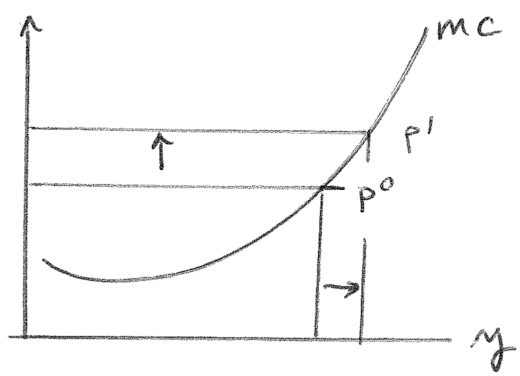
Hint: Recall that the cost function is homogeneous of degree 1 in all input prices.

<u>The Law of Supply</u>

The "law of supply" states that competitive supply functions always slope upward:

$$\frac{\partial y^*}{\partial p} > 0.$$

Why? At a supply optimum, p = MC <u>and</u> MC is increasing from the second-order condition. Therefore, if p increases, the new optimum supply response is higher: we simply move along the MC schedule.



Formally: y^* is defined as the solution to the equation

(*) $p - C_y(y^*(p, w_1, w_2),\ w_1, w_2) = 0.$

This first-order condition must hold if we move p (or either of w_1, w_2) . Therefore, we can differentiate equation (*) w.r.t. p:

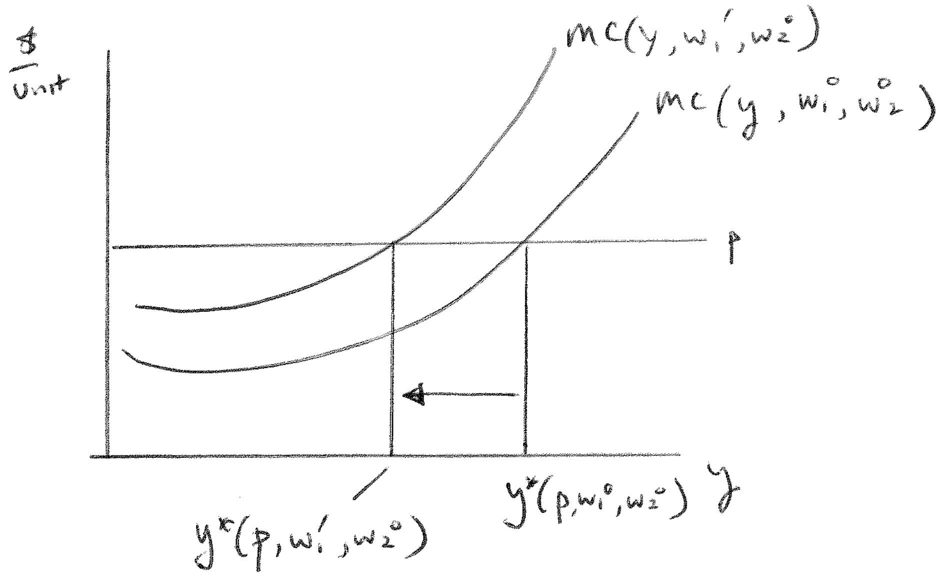
$$dp - C_{yy}(y^*(p, w_1, w_2), w_1, w_2) \times dy^* = 0 \qquad \text{(using the chain rule).}$$

Hence $\dfrac{dy^*}{dp} = \dfrac{1}{C_{yy}(y^*, w_1, w_2)}$,

and, from the second-order conditions, $C_{yy}(y^*(p, w_1, w_2), w_1, w_2) > 0$, implying that $\dfrac{dy^*}{dp} > 0$!

<u>Changes in Input Prices</u>

What is the effect of an increase in input prices on the firm's output decisions? Graphically, a rise in w_1 causes the MC curve to shift. Intuitively, an increase in input prices (say w_1) is associated with a shift in MC. So we have the following diagram:



In the case where MC rises as w_1 rises, we have $\dfrac{dy^*}{dw_1} < 0$. Is that always true? See the next lecture!

Economics 101A
Input Demand for a Competitive Firm

In this lecture we describe the determination of input demands for a competitive firm that sells output y at a price p. Its production function is $y=f(x_1, x_2)$. Inputs 1 and 2 have prices w_1 and w_2.

The firm's optimal choices of x_1, and x_2 are determined in two steps. First, the firm constructs its cost function $C(w_1, w_2, y)$. This implicitly defines the optimal input demands x_1, and x_2 for each level of y, and given input prices.

$$C(w_1, w_2, y) = w_1 x_1 + w_2 x_2 \quad \text{s.t.} \quad y=f(x_1, x_2).$$

$$= w_1 x_1^c(w_1, w_2, y) + w_2 x_2^c(w_1, w_2, y)$$

where $x_1^c(w_1, w_2, y)$ and $x_2^c(w_1, w_2, y)$ are the "conditional factor demands". The word "conditional" denotes that fact that these input demands are conditioned on the output choice. Note that $x_1^c(w_1, w_2, y)$ and $x_2^c(w_1, w_2, y)$ are very much like the compensated demands for the consumer. In particular, setting up the Lagrangean for the cost-min problem:

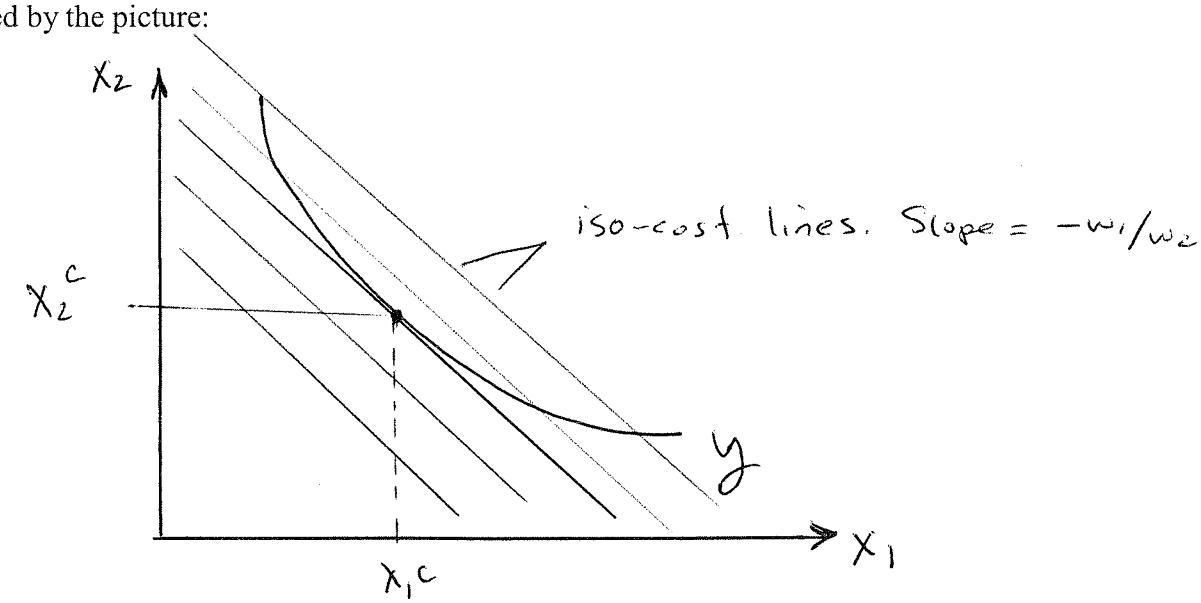
$$L(x_1, x_2, \mu) = w_1 x_1 + w_2 x_2 - \mu(y - f(x_1, x_2))$$

$$\partial L/\partial x_1 = w_1 - \mu f_1(x_1, x_2) = 0$$

$$\partial L/\partial x_2 = w_2 - \mu f_2(x_1, x_2) = 0$$

$$\partial L/\partial \mu = -y + f(x_1, x_2) = 0$$

The ratio of the first two implies that $w_1 / w_2 = f_1(x_1, x_2) /f_2(x_1, x_2)$. Recall that $f_1(x_1, x_2)$ is the "marginal product" of input 1. The ratio $f_1(x_1, x_2) /f_2(x_1, x_2)$ is called the "marginal rate of technical substitution (MRTS). This is the firm's equivalent of the consumer's MRS. It gives the slope of an isoquant at a point (w_1, w_2). So the first order conditions for cost-min are described by the picture:



For future reference, recall from earlier lectures that

$x_1^c(w_1, w_2, y) = \partial C(w_1, w_2, y)/\partial w_1$ and $x_2^c(w_1, w_2, y) = \partial C(w_1, w_2, y)/\partial w_2$.

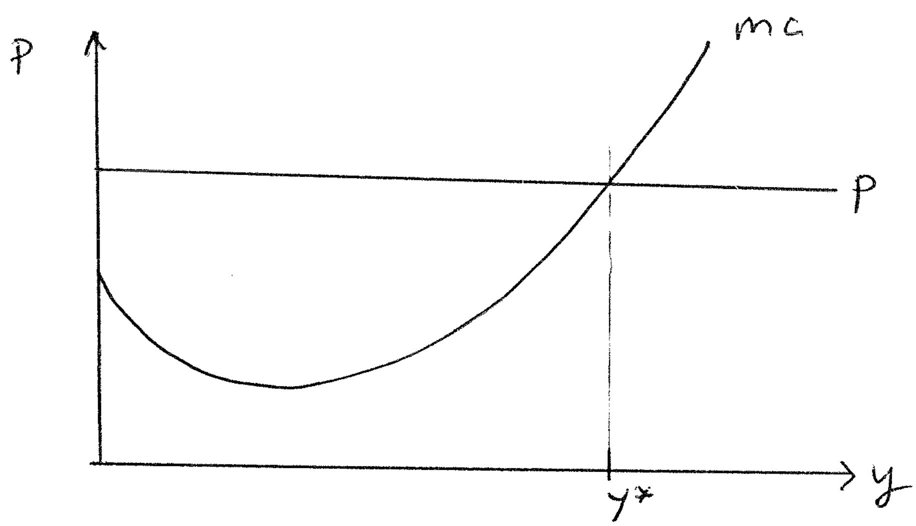
The second step is for the firm to decide what level of output to choose.
Having found C( ), the firm's decision is to maximize profit by choosing y:

$\max \; p \cdot y - C(w_1, w_2, y)$

The f.o.c. is $p - \partial C(w_1, w_2, y)/\partial y = 0$ or $p = MC(w_1, w_2, y)$ = "marginal cost"

the s.o.c is $-\partial^2 C(w_1, w_2, y)/\partial y^2 < 0$ or $\partial MC(w_1, w_2, y)/\partial y > 0$ "rising marginal cost"

This gives us the picture:



The optimal choice of y for given (p, w_1, w_2) is $y^*(p, w_1, w_2)$ which is implicitly defined by:

$$p = \partial C(w_1, w_2, y^*(p, w_1, w_2))/\partial y$$

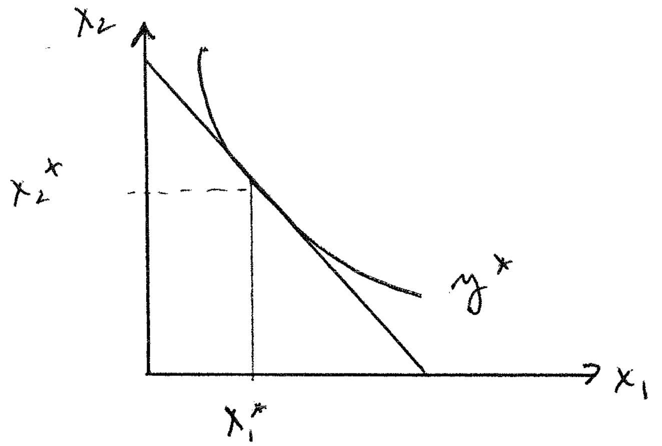
$$= MC(w_1, w_2, y^*(p, w_1, w_2)) .$$

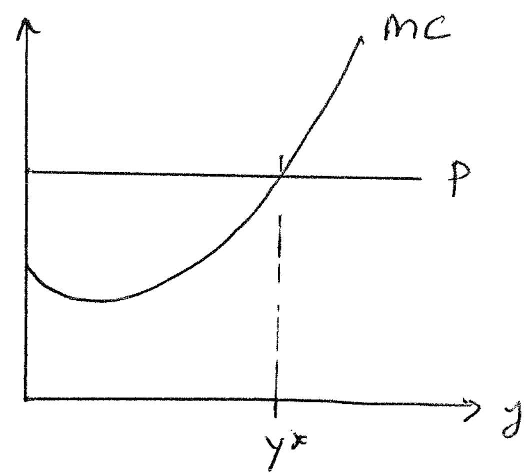
In other words, y^* is the value of y that equates MC to p.

Now we are ready to define the firm's **unconditional** input choices. The firms unconditional input demands are simply:

$x_1(p, w_1, w_2) = x_1^c(w_1, w_2, y^*(p, w_1, w_2))$
$x_2(p, w_1, w_2) = x_2^c(w_1, w_2, y^*(p, w_1, w_2))$.

These equations say that the unconditional input demands are the conditional demands for the "optimized" choice of y. We can think of the problem of finding optimal input demand choices as one of solving two problems simultaneously: cost-min and p=MC



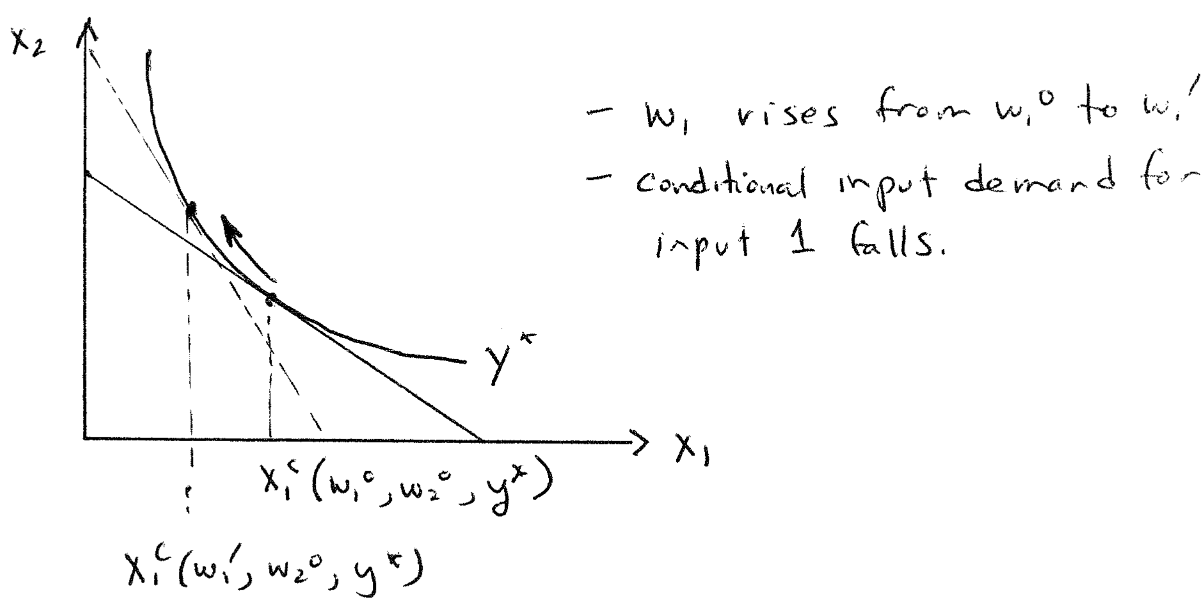


What happens when w_1 rises? Since

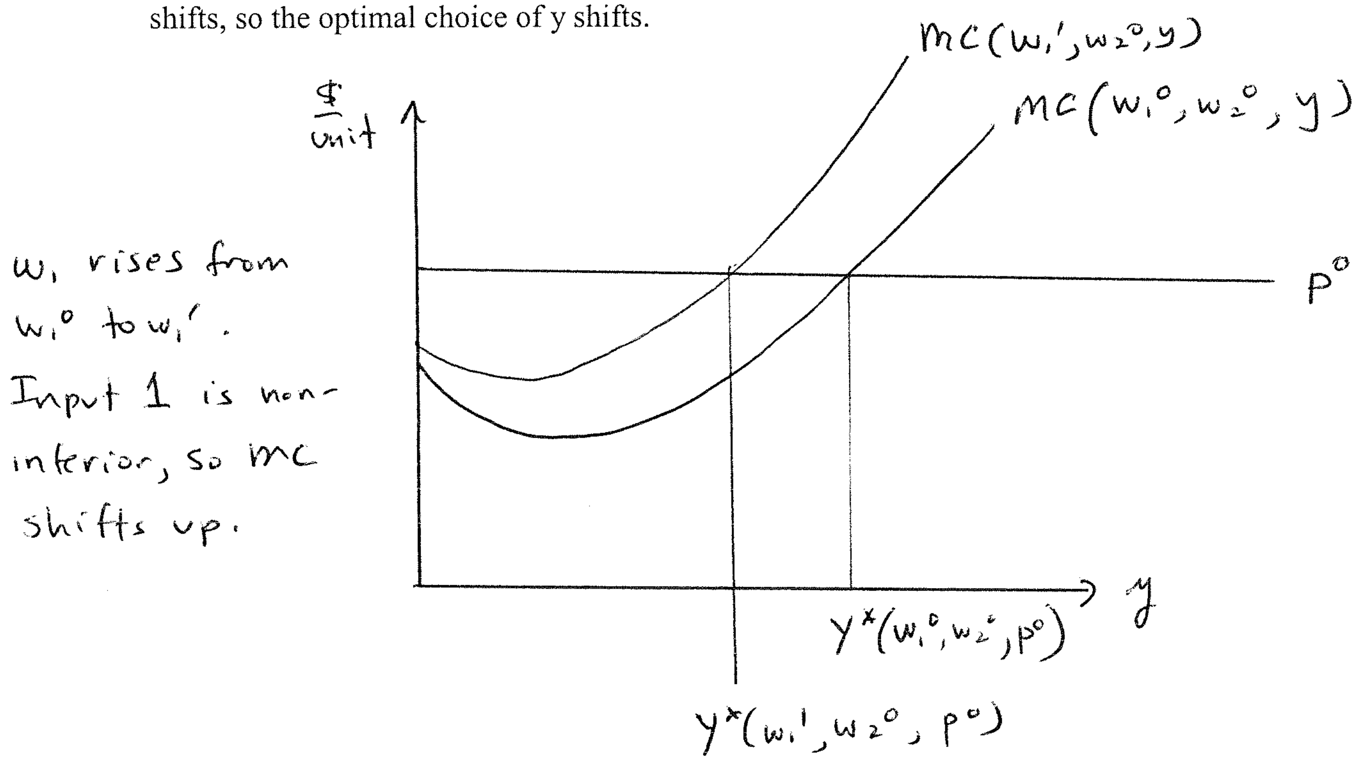
$$x_1(p, w_1, w_2) = x_1^c(w_1, w_2, y^*(p, w_1, w_2))$$

$$\partial x_1(p, w_1, w_2)/\partial w_1 = \partial x_1^c/\partial w_1 + \partial x_1^c/\partial y \times \partial y^*(p, w_1, w_2)/\partial w_1$$

The first term is the response of optimal input demand, **holding constant y**. This is called the substitution effect. It is just like the consumer's substitution effect, which is defined as the change in demand holding utility constant. For the firm, however, the substitution effect holds y constant, giving a movement along an isoquant.



The second term is called the "scale effect". It has some similarity to the consumer's income effect, but the analogy can be misleading. It reflects the fact when w_1 rises, the firm's MC curve shifts, so the optimal choice of y shifts.



Recall from earlier lectures that if input 1 is a "non-inferior" input, when w_1 rises MC shifts up. Why?

$$\partial MC(w_1, w_2, y)/\partial w_1 = \partial/\partial w_1 \{ \partial C(w_1, w_2, y)/\partial y \}$$

$$= \partial^2 C(w_1, w_2, y)/\partial w_1 \partial y$$

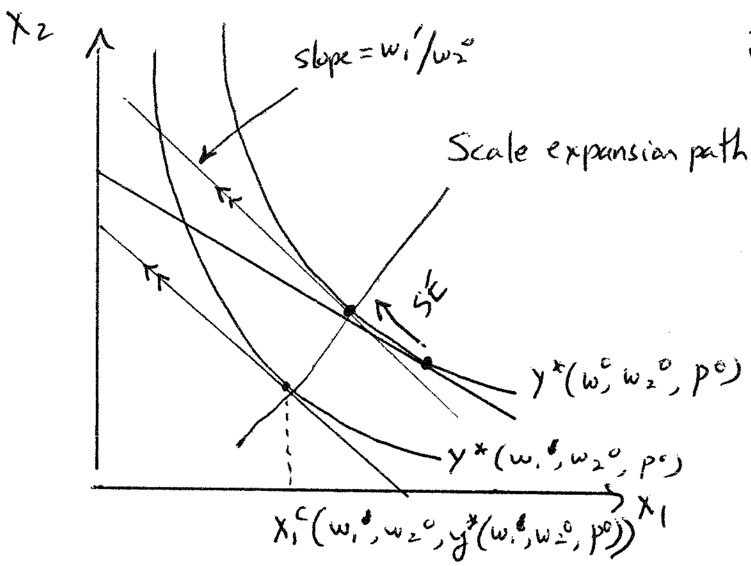
$$= \partial^2 C(w_1, w_2, y)/\partial y \partial w_1 \quad \text{since we can interchange order}$$

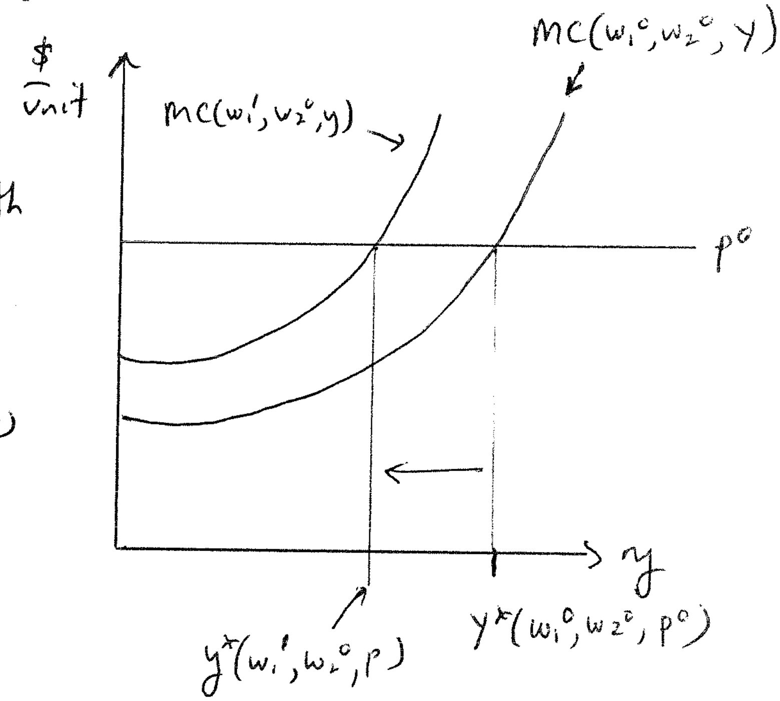
$$= \partial/\partial y \{ \partial C(w_1, w_2, y)/\partial w_1 \}$$

$$= \partial x_1^c(w_1, w_2, y)/\partial y$$

Thus, the derivative of MC w.r.t. w_1 **is the same thing** as the derivative of conditional input demand w.r.t. y. If input 1 is non-inferior, $\partial x_1^c/\partial y > 0$, so when w_1 rises MC shifts up.

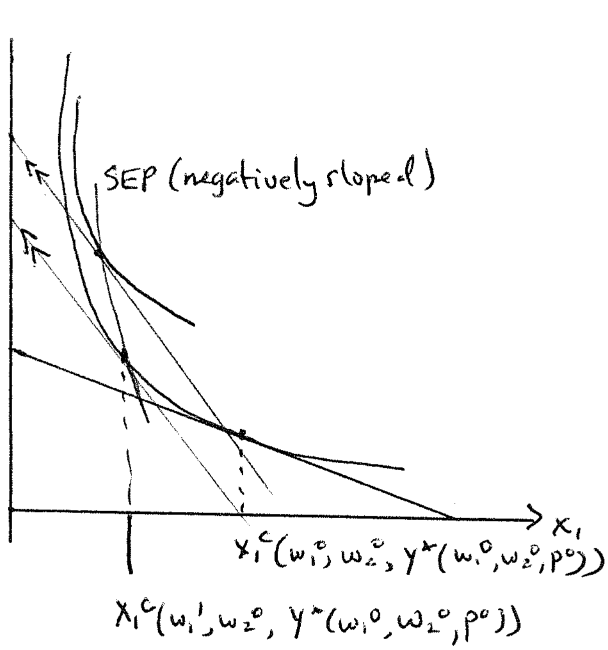
In this case we have the following pair of pictures. When w_1 rises, the substitution effect and scale effect both cause a reduction in demand for input 1.

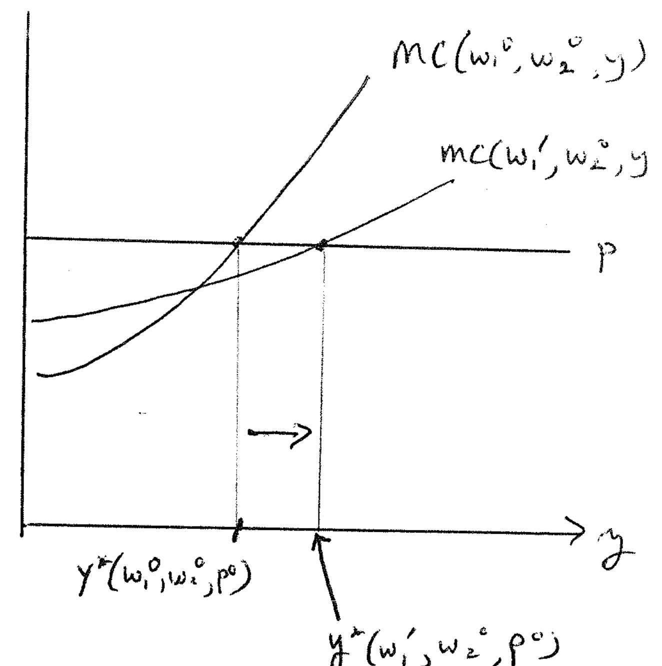




- substitution effect (SE) causes x_1 ↓
- Scale effect (movement along scale expansion path) also causes x_1 ↓

If we have an inferior input, when w_1 rises MC shifts down. (When shovels rise in prices the marginal cost of holes goes down). Then the scale effect is also negative, because although the rise in w_1 makes the firm want to increase output, input 1 is inferior so the expansion in output lowers demand!





For inferior input, w_1↑ ⇒ y^* rises!

There is another way to look at the problem of input demands – a so-called "direct approach". Suppose the firm simply chose x_1 and x_2 to maximize:

$$\pi = p \cdot f(x_1, x_2) - w_1 x_1 - w_2 x_2 \ .$$

This is an unconstrained problem so the f.o.c. are:

a) $\quad p \cdot f_1(x_1, x_2) - w_1 = 0$

b) $\quad p \cdot f_2(x_1, x_2) - w_2 = 0$

Note that the ratio of a) to b) gives the "tangency condition" $w_1 / w_2 = f_1(x_1, x_2) / f_2(x_1, x_2)$.

Also, the firm sets $p = w_1 / f_1(x_1, x_2) = w_2 / f_2(x_1, x_2)$.

What do these mean? If the firm had to increase output by 1 unit, it could do it by increasing input 1 or input 2. If it used input 1, it would need $1/f_1(x_1, x_2)$ units of input 1 to add 1 unit of output. The marginal cost of this would be $w_1 / f_1(x_1, x_2)$. Similarly if it used input 2 the marginal cost would be $w_2 / f_2(x_1, x_2)$. The tangency condition implies that these are equal. So we can interpret the optimum conditions as p=MC plus the tangency condition.

Looking back at the Lagrangean for the cost min problem, notice that the f.o.c. are

$$w_1 = \mu \ f_1(x_1, x_2) \ \Rightarrow \ \mu = w_1 / f_1(x_1, x_2)$$

and

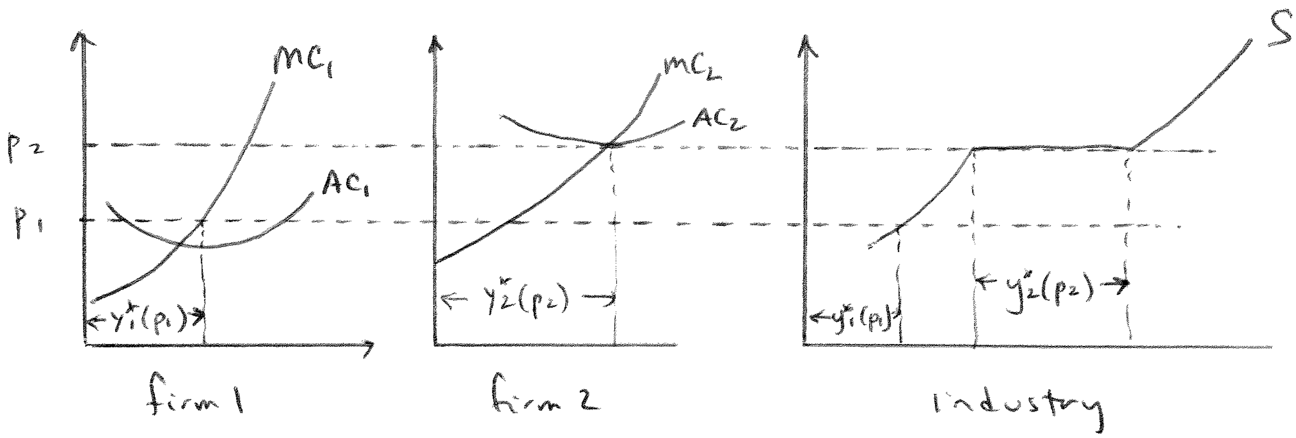
$$w_2 = \mu \ f_2(x_1, x_2) \ \Rightarrow \ \mu = w_2 / f_2(x_1, x_2)$$

Also, recall that μ is marginal cost. So when the firm solves the cost min **and** sets p=MC=μ it gets the same answer as the "direct approach".

Sometimes it is more convenient to work with the "cost-min & p=MC" approach. Other times it is easier to work with the "direct approach". They give the same answers.

Industry Supply

The supply curve for an industry consists of the "horizontal sum" of the supply curves of all individual firms:



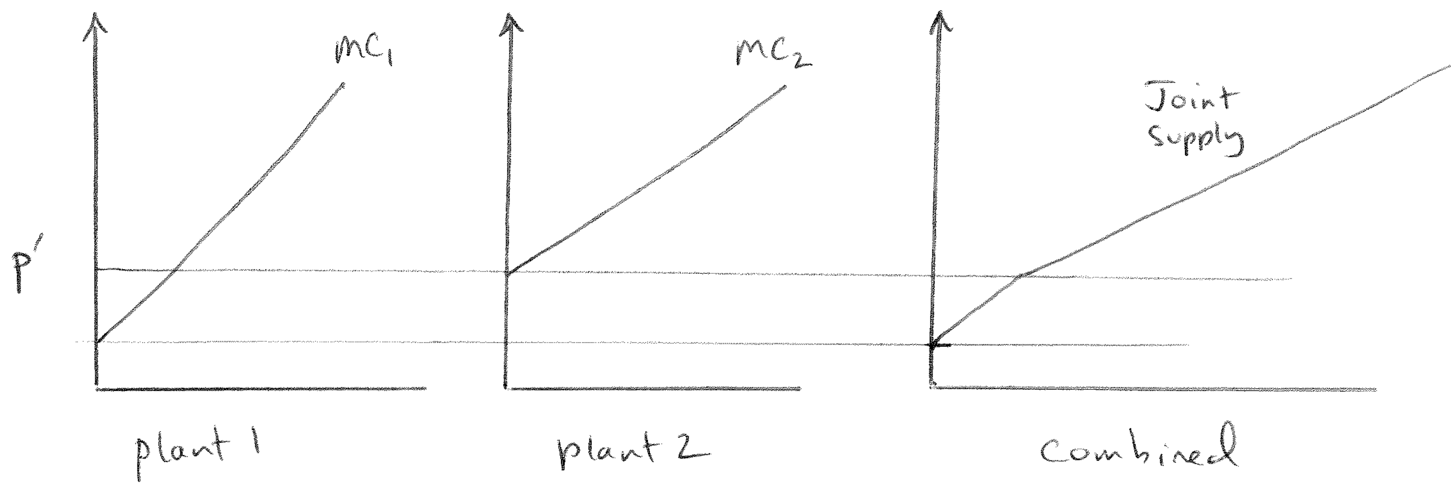
Notice that if firms vary in their costs, at any market price some firms are making profit, some are just on the margin of profitability, and others are "out of the market". A good example of this is the case of oil wells. Some wells have low production costs and are always profitable to operate. Others are high-cost, and only come "on-line" when crude prices are high. We usually call the profits earned by the infra-marginal suppliers "rents". Presumably, the lower costs of these firms arise from their control of a scarce resource.

A competitive market is in equilibrium if

(1) each existing firm has $p = MC$ and $\pi \geq 0$

(2) no other firms can enter and earn profits

These ideas give a nice way to think about a mutli-plant firm (as in problem set 4). If a firm runs 2 plants, with MC schedules $MC_1(y_1)$ and $MC_2(y_2)$, then the firm can operate efficiently by thinking of the plants as independent suppliers. For example consider the following 2-plant firm:



At prices below p′, plant 2 does not operate. At prices above p′, both plants operate where

$$p = MC_1(y_1^*) = MC_2(y_2^*) .$$

Notice that the firm can run "as if" each plant was a separate entity. This is the so-called "principle of decentralization".

Lecture 12

MONOPOLY

A monopolist is the sole supplier to a given market. The critical feature of monopolistic behavior is the fact that a monopolist "sets price" (or sets quantity). Monopolies arise

(a) through exclusive control of inputs or resources: e.g. DeBeers monopoly of diamond marketing

(b) through exclusive legal rights: e.g., public utilities; drug companies with patents

Suppose the demand for output is represented by the function y = D(p). Then we can invert this to p = p(y), which is usually referred to as the "inverse demand function. A monopolist's profit is

$$\pi(y, w_1, w_2) = y\,p(y) - C(y, w_1, w_2)$$

The first order condition for profit maximization is

$$p(y) + y\,p'(y) - C_y(y, w_1, w_2) = 0$$

$$p(y) + y\,p'(y) = C_y(y, w_1, w_2)\,.$$

The expression on the left hand side represents "marginal revenue", MR(y) = p(y) + yp'(y). If demand is downward sloping p'(y) < 0 , so **MR(y) is less than p**. This is the key point about a monopoly. Since a monopoly controls the market, it cannot treat price as exogenous. Rather, it has to take account of the fact that a rise in sales will necessarily come at the expense of a reduction in price. Note that there may be close substitutes for a product. But as long as a firm is the sole supplier of the product, it has monopoly power.

Define the elasticity of demand:

$$\eta = \frac{\partial y}{\partial p} \cdot \frac{p}{y}$$

$$= \frac{1}{p'(y)} \cdot \frac{P}{y} \Rightarrow p'(y) = \frac{1}{\eta}\frac{P}{y}\,.$$

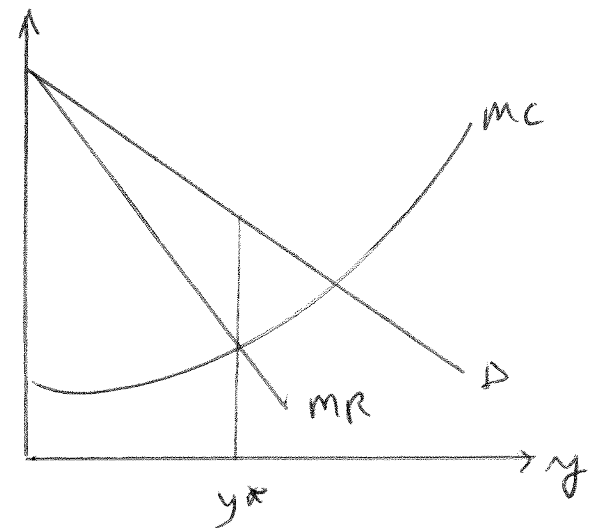
$$MR(y) = p(y) + yp'(y) = p(y) + y\left[\frac{1}{\eta}\frac{P}{y}\right] = p(y)\left[1 + \frac{1}{\eta}\right],$$

For a monopolist, then,

$$p(y)\left[1 + \frac{1}{\eta}\right] = MC\,.$$

As the market demand becomes closer and closer to a horizontal line, η goes to minus infinity,demand becomes perfectly elastic, and p = MC. So in the limiting case monopoly becomes perfect competition.

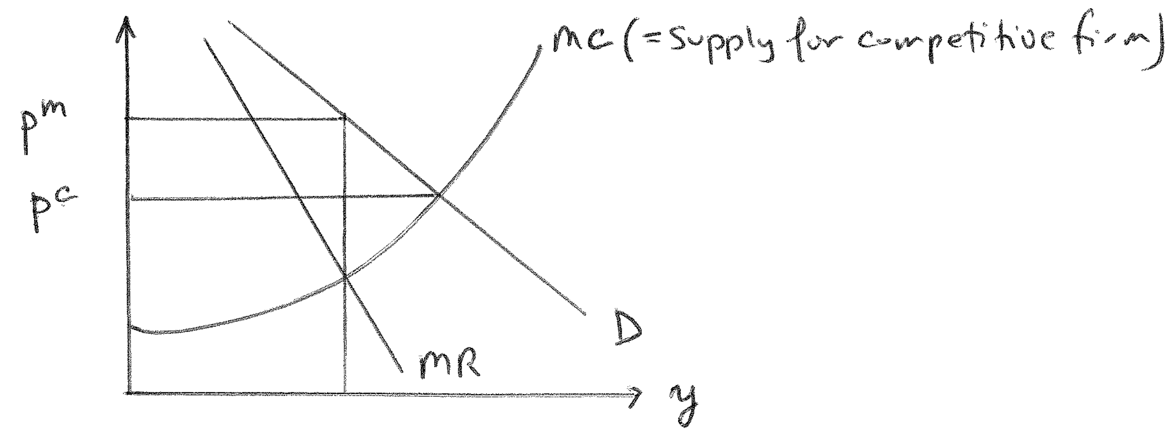
The picture associated with monopoly is the following:



NOTES

(1) A monopolist always sets MR=MC . Since MR = p (1 + 1/η) and $\eta < 0$, MR < p . If $|\eta| < 1$ then $1/\eta < -1$ and MR is <u>negative</u>. It follows that a monopolist never operates in a region where demand is inelastic. The intuition is that if demand is <u>inelastic</u>, you can increase revenue by raising price! This is a very powerful result. It says that some markets cannot be considered "monopoly" markets: namely those with measured elasticities of demand less than 1 in absolute value.

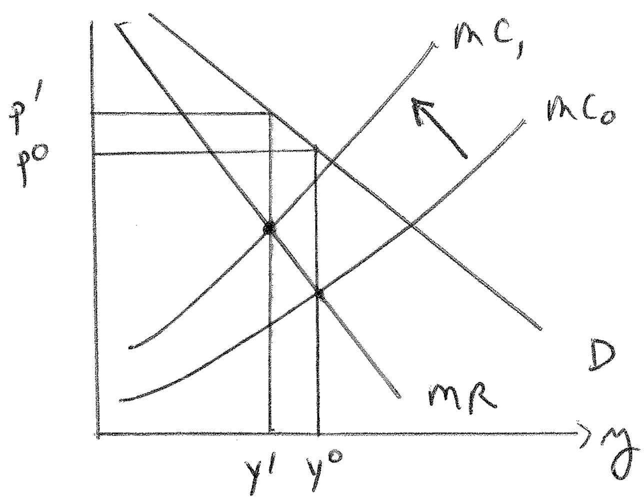
(2) If the monopolist's MC schedule was the MC schedule for a price taker, or a set of price takers (i.e. a competitive industry supply function) then equilibrium would occur where p = MC . This point would entail higher output and lower price, but lower profit to the industry as a whole.



(3) A monopolist does not have a supply schedule, per se. First, the monopolist looks at the demand function. <u>Then</u> she establishes a price. There is no schedule of price/quantity combinations.

(4) The second order condition for profit maximum is $\frac{\partial}{\partial y}[MR - MC] < 0$. or slope of MR < slope of MC. Even if MC is downward sloping, there may exist an equilibrium for the monopolist.

Comparative Statistics for a monopolist



(a) If MC increases (say because an input becomes more expensive) output will fall if MR is negatively sloped.

(b) Factors that shift marginal revenue will cause output to increase or decrease along the MC schedule. A constant elasticity demand function gives:

$$y = A\, p^{\eta}\, p_z^{\gamma}\, I^{e} \;,$$

where z is another good, I is income, p = price of y, p_z = price of z. Inverse demand is

$$p = A^{-1/\eta}\, y^{1/\eta}\, p_z^{-\gamma/\eta}\, I^{-e/\eta} \;.$$

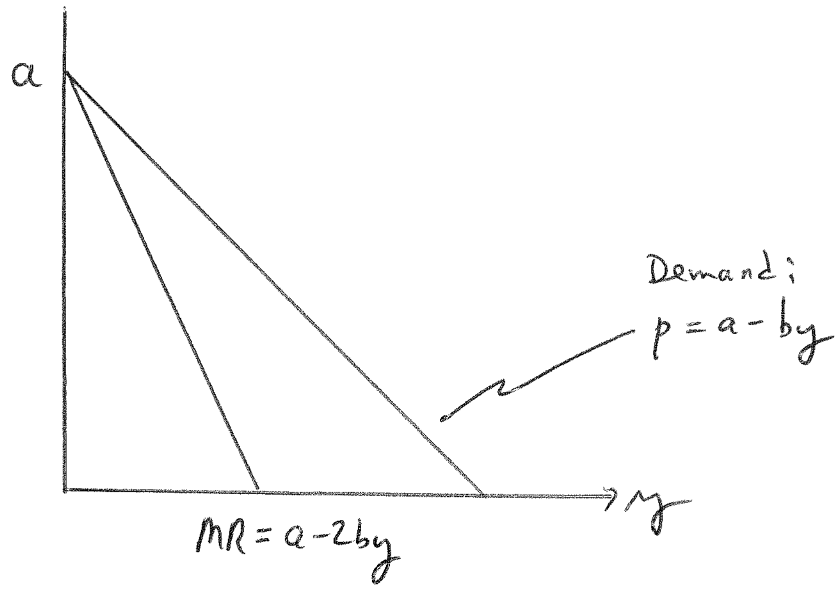
Marginal revenue is $p(1 + \frac{1}{\eta}) \;=\; (1 + \frac{1}{\eta})\, A^{-1/\eta} y^{1/\eta} p_z^{-\gamma/\eta}\, I^{-e/\eta}$

Thus increases in income or p_z shift marginal revenue.

Some examples:

(1) The "linear" case. Suppose demand is linear: y = A - By, and costs are linear: $C(y) = \alpha + \beta y$. To find inverse demand, note p = A/B - 1/B y. Let a = A/B and b = 1/B. Inverse demand can be written as $p = a - by$. Revenue is : $R(y) = p(y) \cdot y \;=\; ay - by^2 \;\Rightarrow\; MR(y) = a - 2by$.

The picture associated with a linear demand function is the following:



If $C(y) = \alpha + \beta y$, then $MC(y) = \beta$. Equating MC=MR, we get $a - 2by = \beta$ or

$$y^* = \frac{a-\beta}{2b}$$

$$p^* = a - by^* = \frac{a+\beta}{2}$$

(2) The "exponential" case. Suppose demand is exponential (constant elasticity):

$y = A\,p^{\eta}$ (for $\eta < -1$).

Inverse demand is $p = (A/y)^{-1/\eta} = ay^{1/\eta}$ where $a = A^{-1/\eta}$. Revenue $= p\,y = a\,y^{(1+1/\eta)}$

$MR(y) = (1 + 1/\eta)\,ay^{1/\eta}$

Suppose cost is also exponential: $C(y) = \alpha y^{\beta}$, $\beta > 0 \Rightarrow MC(y) = \alpha\beta y^{\beta-1}$.

The firm's profit is π = Revenue - Cost.

FONC for max: MR - MC = 0

$\Rightarrow (1 + 1/\eta)\,ay^{1/\eta} - \alpha\beta y^{\beta-1} = 0.$

SOC: $1/\eta\,(1 + 1/\eta)\,a\,y^{(1/\eta-1)} - \alpha\beta(\beta-1)y^{\beta-2} > 0$

The S.O.C. is always satisfied if $\beta > 1$. (Why?)

Solving the FONC we have

$(1+1/\eta)\,a/(\alpha\beta) = y^{(\beta-1-1/\eta)} \Rightarrow y = \{\,(1+1/\eta)\,a/(\alpha\beta)\,\}^{\eta/(\eta(\beta-1)-1)}$.

Note that the optimal choice depends on the parameters of the demand and cost functions. A change in the elasticity of demand will lead to a shift in the optimum.

Monopoly with 2 or more markets

Suppose a monopolist can sell in two markets

Market 1: $p_1 = p_1(y_1)$

Market 2: $p_2 = p_2(y_2)$

If trade is restricted between the two markets, then p_1 and p_2 can differ. The firms profits are

$$\pi = p_1 y_1 + p_2 y_2 - C(y_1 + y_2)$$

The FONC for profit maximization are:

$$p_1 + y_1\, \partial p_1/\partial y_1 \;-\; C'(y_1+y_2)\,, \quad \text{or} \quad MR_1 = MC$$

and

$$p_2 + y_2\, \partial p_2/\partial y_2 \;-\; C'(y_1+y_2)\,, \quad \text{or} \quad MR_2 = MC$$

Since $MR_1 = p_1(1 + 1/\eta_1)$, $MR_2 = p_2(1 + 1/\eta_2)$, we get the prediction that $MR_1 = MR_2 = MC$, or

$$p_1(1 + 1/\eta_1) = p_2(1 + 1/\eta_2)$$

$$\Rightarrow \quad p_1/p_2 = (1 + 1/\eta_2)/(1 + 1/\eta_1)\,.$$

E.g., $\eta_1 = -1.5$, $\eta_2 = -2.5$, $\Rightarrow$ $p_1/p_2 = 0.60/0.33 = 1.82$

The monopolist charges a higher price in the more inelastic market. This is known as "price discrimination".

Lecture 13

More on Monopoly

We have shown how a monopolist would prefer to distinguish between markets and charge a higher price to consumers with more inelastic demand. This phenomenon is called price discrimination. Sellers have strong incentives to try and separate consumers according to their demand elasticities, and charge them discriminatory prices. Consumers, on the other hand, have strong incentives to try and act like "high-elasticity" consumers. There are many devices to separate consumers.

+advanced purchase versus regular coach fares on airlines. Here, the airlines discriminate against customers who book at the last minute (typically business travellers) and charge lower prices to consumers who are willing to "shop around".

+single tokens versus monthly passes on public transit. Presumably, commuters have more elastic demand for public transit than out-of-town or occasional passengers.

+discount coupons. Here, retailers are trying to charge lower prices to consumers who are better informed (and more price responsive), while continuing to charge high prices to consumers who "can't be bothered" with coupons (and therefore reveal themselves as inelastic-demand consumers).

+after-season sales; special "Monday and Tuesday only" sales. Again, retailers are trying to separate high elasticity consumers from those who will buy in-season or in-fashion items at the peak of their popularity.

In each case, the key to price discrimination is to impose a cost on the low-price consumers (those with more elastic demands), in order to prevent high price (inelastic demand) consumers from masquerading as low price consumers. The cost must be "too high" for inelastic-demand consumers, but not high enough to discourage others from buying altogether.

An Example:

Suppose that, across people, individual demand elasticities are negatively correlated with wage rates. Those with highest wage have the most inelastic demand; those with lowest wages have elastic demand. Then a firm can use a queue or "line up" to price discriminate: charge a high price with no waiting time; and a low price to those who will line up in a queue for a while. (An example is the difference in prices between buying a ticket at the ticket counter or through a phone service). For a consumer with wage w_i the "full" price is

p	if she buys in the fast line (no time cost)
$p + w_i - d$	if she buys in the slow line (time cost of 1 hour, price discount d).

For an individual consumer:

if $w_i > d$, you buy in the fast line and pay price p

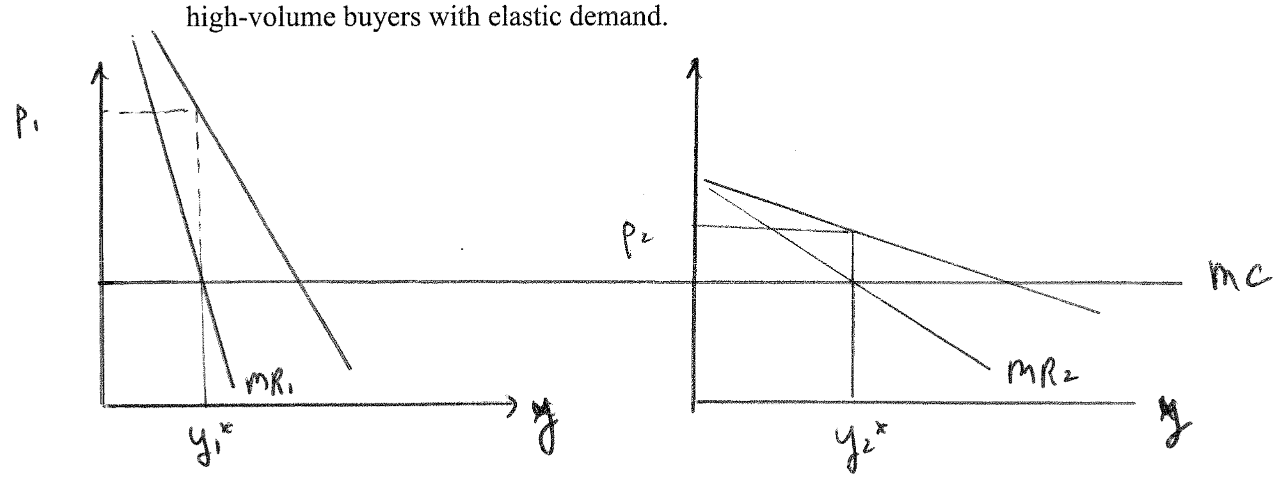
if $w_i < d$, you buy in the slow line and pay price p-d .

The firm has succeeded in charging 2 prices!

Another way to price discriminate is to charge different prices according to how much you buy. Suppose, for example, that there are two kinds of buyers:

low-volume buyers with inelastic demand,

high-volume buyers with elastic demand.

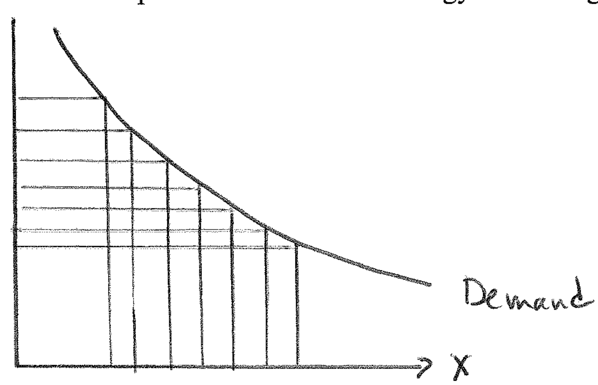


The monopolist can choose a y^0 in between y_1^* and y_2^* and offer a two-part price system:

if you buy $y < y^0$ pay price p_1 per unit

if you buy $y > y^0$ pay price p_2 per unit.

Note that it has to true that $p_1\, y_1^* < p_2\, y^0$ or the low-volume consumers will buy y^0 units and throw away what they don't need. The "ultimate" price discrimination strategy is to charge a separate price for each unit sold:



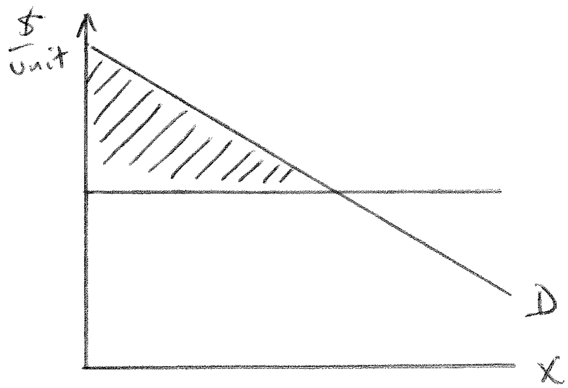
For the first unit sold, charge price p_1; for the 20th unit, charge price p_{20} , and so on.

Notice that in this case, the marginal revenue of the next unit sold is its price, since the seller doesn't have to lower prices on the infra-marginal (previously sold) units to sell an additional unit. In this case, the MR curve is the same as the demand function.[1] Thus, under "perfect price discrimination":

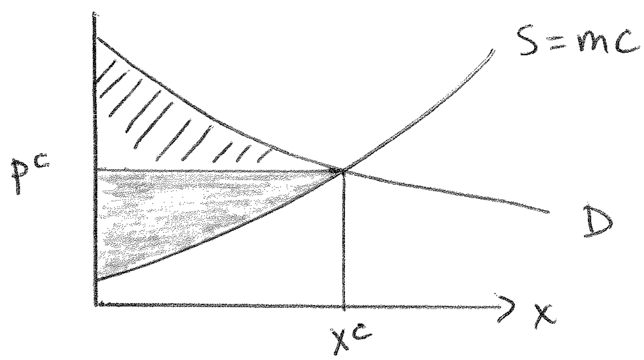
(1) quantity is equal to its level under "perfect competition".

(2) monopolist revenue = area under demand curve.

Relative to a perfect price discrimination scheme, consumers benefit when all consumers pay the same price. The "saving" to all consumers, relative to perfect price discrimination, is the shaded area below the demand curve and above the price line.



This area is sometimes called consumer's surplus (CS). By analogy, the area over the MC curve and up to the price line is called producer's surplus (PS). We have previously noted that this area gives the difference between revenue and total variable cost, so PS = profit + fixed costs. Also, we noted that if an industry is competitive, the supply schedule is just the MC schedule of the combined set of firms (using the horizontal sum of all firms' MC's). The area between supply and demand (or MC and demand) represents the sum of producers and consumer's surplus. Also note that both consumer surplus (CS) and producer surplus (PS) are measured in units of $.

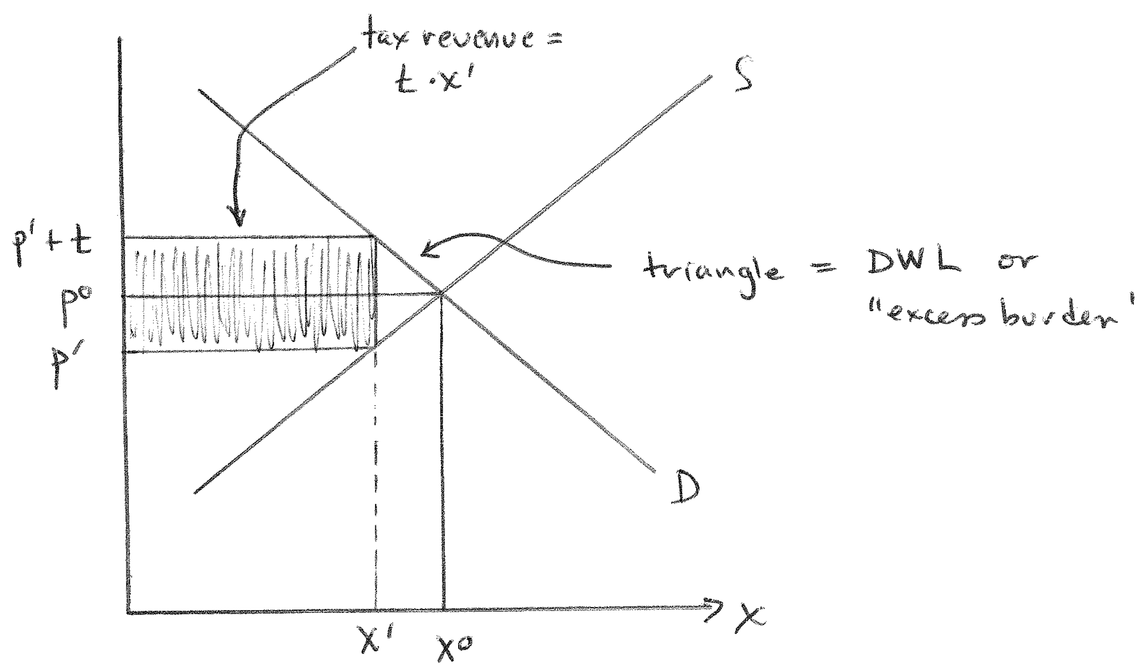


[1]Actually, as discussed on the next page, this statement is untrue. The "demand function" represents the number of units demanded when all units are sold at some price.

Applied economists often evaluate the effect of an intervention in a particular market by calculating the sum of charges in consumer surplus and producer surplus and government cost. The appeal of this exercise is obvious: it gives a dollar value for the inefficiency costs of "monopolization" or "imposition of a tax" or "imposition of a subsidy".

Nonetheless, there is a problem. Recall that if y = D(p) is a consumer's demand function, D (the demand function) tells me how much is purchased when y cost p per unit. The demand function really doesn't tell you demand when price for the next unit is p , but prices for all previous units are higher. In general, if you charge higher prices for the "inframarginal" units, the consumers who purchased them have less income to purchase additional units. Higher prices on all the previous units have an income effect that is not captured by the "ordinary" demand function. In fact, the only case where consumer's surplus analysis is completely legitimate is the case where demand for the commodity under study is independent of income (the vertically-parallel indifference curves case), or in cases where each consumer only buys 0 or 1 units of the commodity (so higher prices on the first unit purchased don't lower subsequent demands for additional units). In the discussion sections we will go through the arguments in more detail.

In spite of this problem, consumer and producer surplus analysis gives a rough way to evaluate the costs of a market intervention. For example, suppose that from an initial supply and demand equilibrium at price p^0, quantity x^0 , a tax of \$t is imposed (per units):



The quantity falls to x′ and the price received by supplies falls to p′. Total tax revenue is tx′. The loss in combined consumer and producer surplus, however, is greater than the amount of tax revenues by the shaded triangle. The size of this triangle is referred to as the "deadweight loss" from the tax. It gives a rough measure of the inefficiency of the tax.

Exercises

(1) Calculate the deadweight loss in terms of the elasticities of supply and demand.

(2) Prove that CS + PS is maximized where demand intersects marginal cost (or supply).

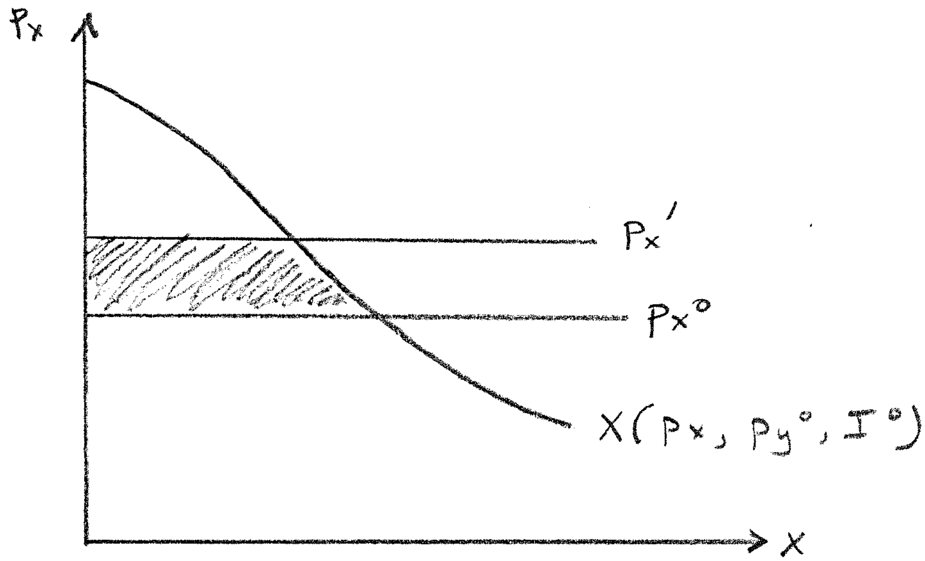
(3) Calculate the lost in CS and PS together when a competitive industry is monopolized.

Economics 101A – Spring 2005

Consumer Surplus

In introductory economics you were introduced to the concept of "consumer's surplus" (CS). Consider a consumer who is choosing between two goods, x and y. Denote the demand function for good x by $x(p_x, p_y, I)$. At some initial situation prices are (p_x^0, p_y^0) and income is I^0. Now suppose the price of good x rises to $p_x' > p_x^0$. The change in consumer's surplus is the shaded area on the graph, which can be written as

$$\Delta CS = \int_{p_x^0}^{p_x'} x(p_x, p_y, I)\, dp_x \quad .$$



As we noted in last lecture, there is a problem with CS: though the vertical height of the inverse demand seems to represent the maximum you'd be willing to pay for each additional unit, if someone actually charged you different prices for each unit, your demand would NOT be given by the conventional demand curve (since that is derived assuming that you pay the same price for all units of x). There is, however, a measure of welfare that makes sense. In fact there are two.

Let u^0 represent utility at (p_x^0, p_y^0, I^0) and let u' represent utility at (p_x', p_y^0, I^0). Note $u^0 > u'$, since a rise in prices makes you worse off.

Also, note that $I^0 = e(p_x^0, p_y^0, u^0)$. This says that I^0 is the minimum amount of money needed to get to u^0 at prices (p_x^0, p_y^0). That follows from the fact that you weren't wasting money when you got to u^0.

Likewise $I^0 = e(p_x', p_y^0, u')$, since at the new prices you get to u' with income I^0.

Consider the amount of money

$$EV = e(p_x', p_y^0, u') - e(p_x^0, p_y^0, u') = I^0 - e(p_x^0, p_y^0, u')$$

This is the amount of cash someone could take away from you from your initial situation, leaving prices at p_x^0, p_y^0, so that you would just be indifferent to facing the new prices. This is called the "equivalent variation" – hence the EV. It is "the income equivalent" of the price rise. This seems like a pretty natural measure of the welfare loss caused by the price rise.

Alternatively, consider the amount of money

$$CV = e(p_x', p_y^0, u^0) - e(p_x^0, p_y^0, u^0) = e(p_x', p_y^0, u^0) - I^0 .$$

This is the amount of cash you'd need **to get back to u^0 at the new prices**. The bigger this is, the worse off you are when prices rise. This is called the "compensating variation" – hence the CV. This also seems to be a plausible measure of the welfare loss

Now we use Sheppard's lemma to link CV and EV to the area under certain compensated demand curves.

Specifically, start from the fact that

$\partial e(p_x, p_y^0, u^0)/\partial p_x = x^c(p_x, p_y^0, u^0)$ — the derivative of expenditure is the compensated demand

Using the fundamental theorem of calculus,

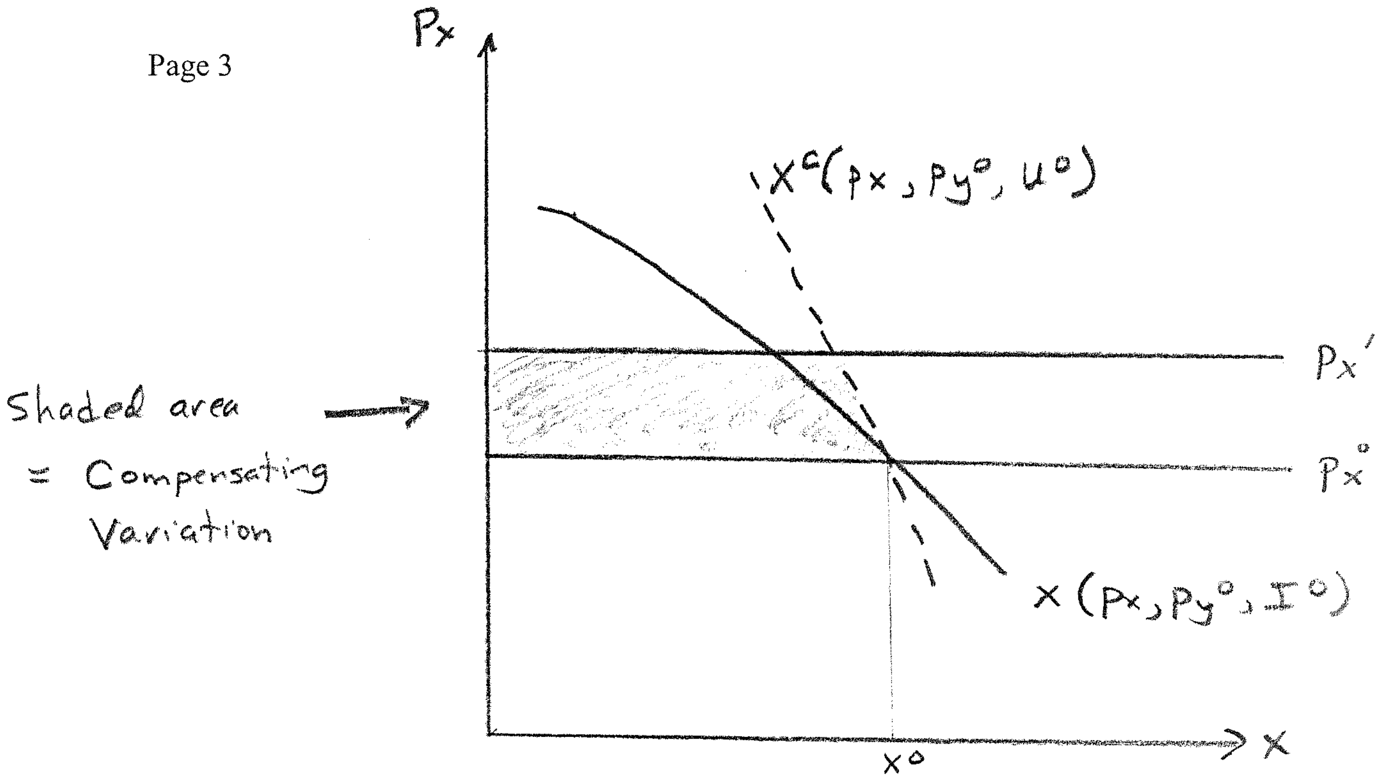
$$e(p_x', p_y^0, u^0) = e(p_x^0, p_y^0, u^0) + \int_{p_x^0}^{p_x'} \{ \partial e(p_x, p_y^0, u^0)/\partial p_x \} dp_x$$

and substituting

$$e(p_x', p_y^0, u^0) = e(p_x^0, p_y^0, u^0) + \int_{p_x^0}^{p_x'} x^c(p_x, p_y^0, u^0)\, dp_x$$

So $\quad CV = \int_{p_x^0}^{p_x'} x^c(p_x, p_y^0, u^0)\, dp_x$

which is the area "under"the compensated demand curve between p_x^0 and p_x':



Note that $x^c(p_x^0, p_y^0, u^0) = x(p_x^0, p_y^0, I^0)$, which says that the compensated demand with $u=u^0$ and the uncompensated demand with $I=I^0$ are the same. So the regular demand curve (with $I=I^0$) and the compensated demand (with $u=u^0$) meet at (x^0, p_x^0). But the regular demand curve is flatter. Why? Recall from Slutsky that

$$\partial x(p_x^0, p_y^0, I^0)/\partial p_x = \partial x^c(p_x^0, p_y^0, u^0)/\partial p_x - x^0\, \partial x(p_x^0, p_y^0, I^0)/\partial I$$

If x is a “normal good”, $\partial x(p_x^0, p_y^0, I^0)/\partial I > 0$ and a given rise in prices causes the regular demand to fall off **faster** than the compensated demand because of the “income effect” – hence it appears flatter. All of this means that $CV > \Delta CS$ if x is a normal good.

We can also show the EV.

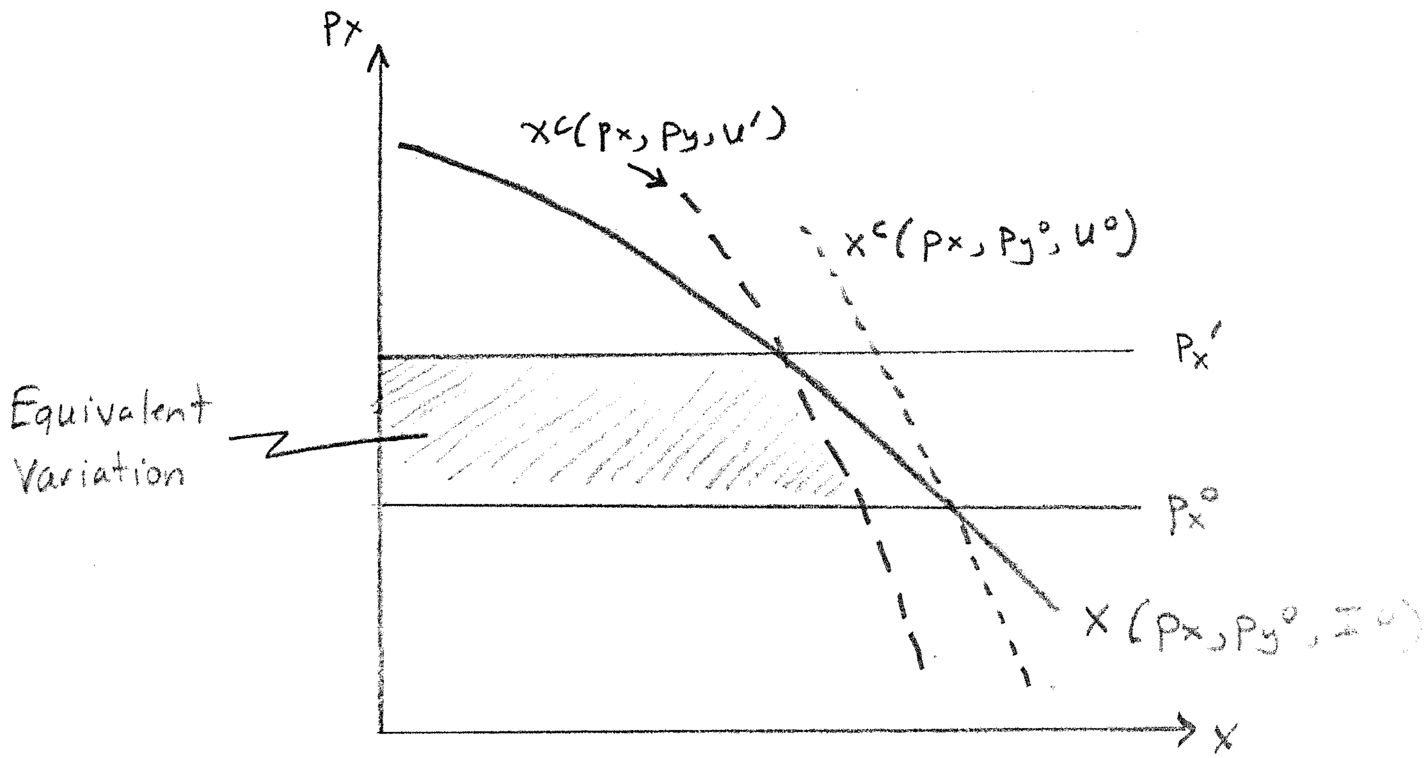
$$e(p_x', p_y^0, u') = e(p_x^0, p_y^0, u') + \int_{p_x^0}^{p_x'} \{ \partial e(p_x, p_y^0, u')/\partial p_x \} dp_x$$

$$= e(p_x^0, p_y^0, u') + \int_{p_x^0}^{p_x'} x^c(p_x, p_y^0, u')\, dp_x$$

So $EV = \int_{p_x^0}^{p_x'} x^c(p_x, p_y^0, u')\, dp_x$

which is the area “under”the compensated demand curve between p_x^0 and p_x'. But this is the compensated demand curve for $u=u'$, which meets up with regular demand curve at (x', p_x').

So we have the following picture, if x is a normal good:



This shows that CV > ΔCS > EV. So you can think of ΔCS as "approximating" one or the other of these.

Lecture 14

DUOPOLY

The simplest market to analyze in between the extremes of monopoly and perfect competition is one with two suppliers (hence the term duopoly). In particular, suppose that there are two suppliers of a homogeneous good (we will ignore for the moment the possibility that the firms can differentiate their products). Let

x_1 = supply of firm 1

x_2 = supply of firm 2

$p(x1 + x_2)$ = inverse demand

Note that inverse demand depends on the sum of x_1 and x_2, reflecting the assumption that the outputs of the two firms are perfect substitutes. For simplicity, we'll assume

(a) $p(x_1 + x_2) = a - b(x_1 + x_2)$ (linear demand)

(b) constant marginal costs of \$c per unit for both firms

The decision problems of the two firms are very simple:

Firm 1: choose x_1 to maximize $\pi(x_1, x_2) = x_1 p(x_1 + x_2) - cx_1$

Firm 2: choose x_2 to maximize $\pi(x_2, x_1) = x_2 p(x_1 + x_2) - cx_2$

Note that the decision problem for firm 1 depends on the choice of firm 2, and vice versa.

Monopolization

What would a monopolist do? Suppose you owned both firms. You would choose x_1 and x_2 to solve:

$$\max_{x_1,x_2} \quad x_1p(x_1 + x_2) + x_2p(x_1 + x_2) - cx_1 - cx_2$$

$$= \max_{x_1,x_2} \quad (x_1 + x_2)\, p(x_1 + x_2) - c(x_1 + x_2) ,$$

$$= \max_{x} \quad xp(x) - c\,x ,$$ there is no distinction between x_1 or x_2, so you only care about their sum, $x = x_1 + x_2$.

For this problem, $p(x) = a - bx$, so $xp(x) = ax - bx^2$, and

$$xp(x) - cx = (a - c)\,x - bx^2 .$$

FONC: $(a - c) - 2bx = 0 \Rightarrow x = (a - c)/2b = x^m$ (m for monopoly).

This implies that $p = p^m = a - bx^m = a - b\left[\frac{a-c}{2b}\right] = \frac{a}{2} + \frac{c}{2}.$

Now, suppose $p = p^m$, and the firms were separate, each producing $\frac{1}{2}x^m$ (This would be a perfect cartel.) Is that an equilibrium? Probably not. For firm 1,

$$MR_1 = \frac{\partial}{\partial x_1}\left[x_1\, p(x_1 + x_2)\right] .$$

Suppose firm 1 could increase x_1 and x_2 didn't change. Then

$$MR_1 = p(x_1 + x_2) + x_1 \frac{\partial p}{\partial x_1}$$

$$= \frac{a+c}{2} + \frac{1}{2} x^m \,[b]$$

$$= \frac{a+c}{2} - \frac{a-c}{4} = \frac{3}{4}c + \frac{1}{4}a > MC = c$$

Thus, at the "joint monopoly" output (both firms producing ½ of the monopoly output) each firm has an incentive to cheat. For both firms together,

$$R = a(x_1 + x_2) - b(x_1 + x_2)^2 \quad \Rightarrow \quad MR = a - 2b(x_1 + x_2).$$

For one of the firms on its own, however,

$$R = x_1[a - b(x_1 + x_2)] , \quad \Rightarrow MR_1 = a - bx_2 - 2bx_1 > MR$$

This is a fundamental problem with a cartel. Each firm has an incentive to "cheat" if it believes the other will hold its output constant. The reason is that when firm 1 increases its output, it only considers how this additional output will lower the price of the units it sells. It ignores the fact that price will fall for the units sold by its rival. A monopolist, by comparison, takes account of the full effect of a change in price on all units sold.

Duopoly Equilibrium

How does the duopoly market equilibrate? The answer really depends on how each duopolist thinks that the other will react to a change in output. The simplest assumption is that firm 1 thinks that firm 2 won't change its output (and vice versa). This assumption was introduced by Cournot (a 19th century French economist). Let's look at firm 1's choice problem under this assumption, for a fixed value of x_2, say $x_2 = \bar{x}_2$. The profit maximization problem becomes:

$$\max_{x_1} \quad x_1\, p(x_1 + \bar{x}_2) - cx \;=\; x_1[a - b(x_1 + \bar{x}_2)]$$

FONC: $a - b\bar{x}_2 - 2bx_1 - c = 0.$ (Note that the SOC is fine).

This leads to an optimal choice:

$$x_1 = \frac{a - c - b\bar{x}_2}{2b} ,$$

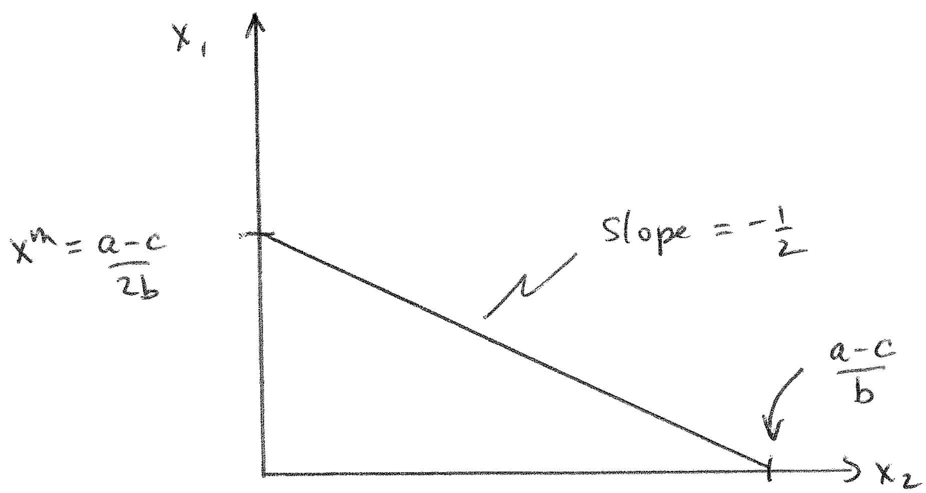
which we can write in a more general notation as $x_1 = x_1^*(\bar{x}_2)$. The function $x_1^*(\,)$ is called 1's "reaction function". It gives the optimal choice of 1, as a function of 2's output, under the Cournot assumption that 2's output is fixed.

Observations:

(1) if $\bar{x}_2 = 0$, firm 1 acts like a monopolist. Check that $x_1^*(0) = \frac{a-c}{2b} = x^m$

(2) if $\bar{x}_2 \geq \frac{a-c}{b} = 2x^m$, $x_1 = 0$. If $\bar{x}_2$ is too large, firm 1 is driven out of the market.

(3) the slope of the reaction function is −½. Each additional unit produced by 2 causes 1 to reduce output by ½ units.



By the same token, there is a reaction function for 2, taking 1's output as given. Following the same algebra (or using symmetry), we get

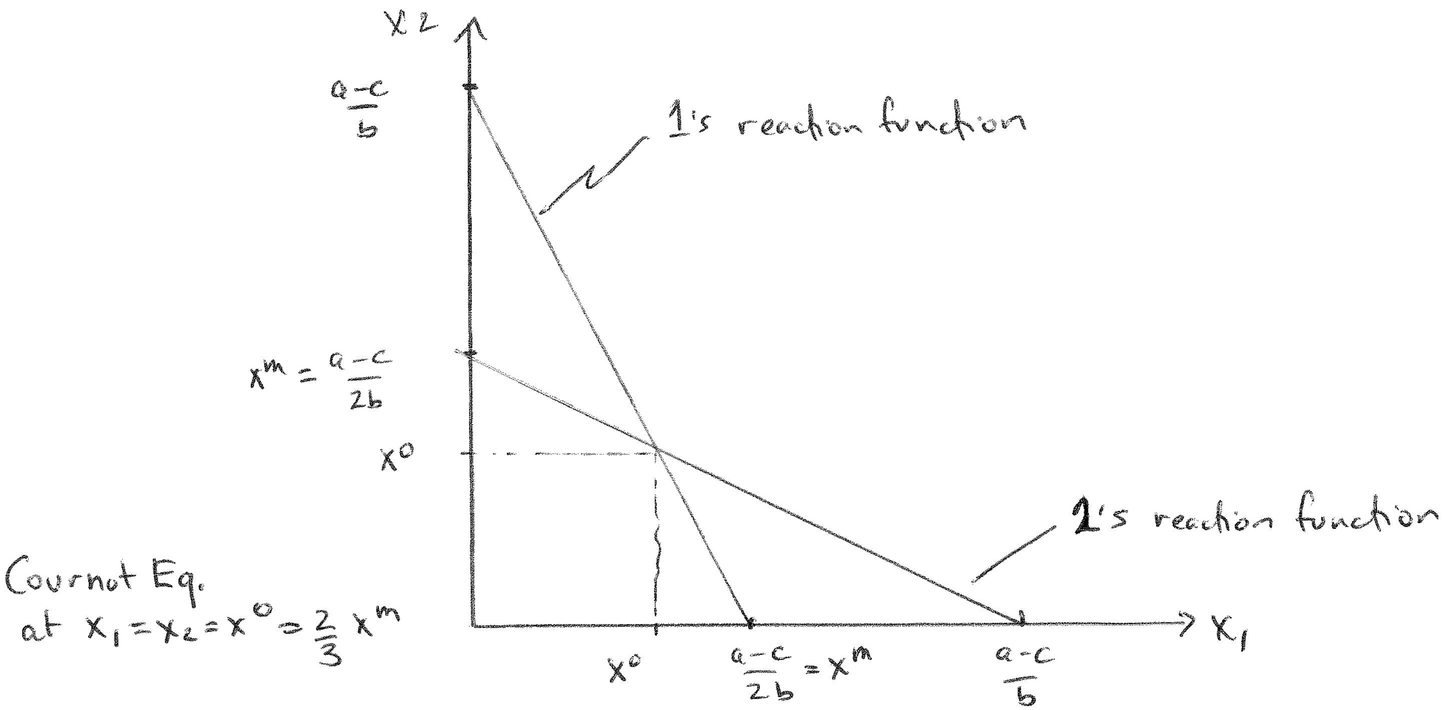
$$x_2^*(\bar{x}_1) = \frac{a-c}{2b} - \frac{1}{2}\bar{x}_1$$

Now, 1 sets x_1 , taking $x_2 = \bar{x}_2$, and 2 sets x_2 , taking $x_1 = \bar{x}_1$. Where does all this end? Presumably, it ends when 1's choice, taking 2's output as given, is just the right one to cause 2 to produce the output that 1 though he would produce (think of holding 2 mirrors face to face). Formally, equilibrium requires

$$x_1 = x_1^*(x_2^*(x_1)) .$$

If x_1 is an equilibrium choice, it has the property that when 1 chooses x_1, 2 chooses $x_2^*(x_1)$, and the optimal response of 1 to 2's choice is $x_1^*(x_2^*(x_1))$ which takes us back to x_1. Mathematically, x_1 is a so-called "fixed point" of the combined function $x_1^*(x_2^*(\ ))$.

Fortunately, there's a nice picture. All we do is graph the two reaction functions (remembering which is which):



Equilibrium occurs where $x_1 = x_1^*(x_2) = x_1^*(x_2^*(x_1)) = (a-c)/2b - ½ \{ (a-c)/2b - ½ x_1 \}$.

This implies

$$¾ x_1 = ½ \{ (a-c)/2b \}$$

or $$x_1 = ⅔ \{ (a-c)/2b \} = ⅔ x^m .$$

Clearly, if each of the two firms is producing ⅔ of the monopoly output, industry output as a whole is 4/3 as big as it would be if the two firms were controlled by a single monopoly.

Price Setting vs. Quantity Setting

The previous example analyzes the outcome when two duopolists takes each others quantities as given. A similar analysis can be used when duopolists set prices. For example, consider the problem of freight rate setting for two end-to-end railroads. The first railroad hauls from point A to point B at price $p_1$per ton. The second hauls from B to C at price p_2 . Demand for transport services from A to C depends on the total price $p_1 + p_2$. For simplicity, let's assume demand is linear:

$$x = A - b(p_1 + p_2).$$

Note that this means that the two segments are perfect complements. (They are only consumed together, so demand only depends on the sum of their prices). Let's also assume the cost per ton on the first segment is c_1, and the cost is $c_2$per ton on the second segment. Suppose a single firm owned both rail segments. It would set a combined price p in order to maximize

$$\Pi(p) = (A - bp)(p - c_1 - c_2) \ .$$

The first-order condition for profit maximization implies:

$$A - bp - b(p - c_1 - c_2) = 0 \ ,$$

or

$$\frac{A + b(c_1 + c_2)}{2b} = p^M \ ,$$

where p^M is the monopolist's price.

Now suppose that the two railroads act as duopolists, each taking the others price as fixed. For the first railroad, profit is:

$$\prod (p_1, \bar{p}_2) = (A - b\,(p_1 + \bar{p}_2))\,(p_1 - c_1).$$

The first-order conditions implies

$$A - b(p_1 + \bar{p}_2) - b(p_1 - c_1) = 0 \ ,$$

or $$\frac{A - b\bar{p}_2 + bc_1}{2b} = p_1 \ .$$

The first duopolist's reaction function is:

$$p_1^*(\bar{p}_2) = \frac{A - b\bar{p}_2 + bc_1}{2b} \ .$$

which looks a lot like the reaction function in the quantity-setting case. In particular, the slope is $-½$. By symmetry, the second duopolist's reaction function is

$$p_2^*(\bar{p}_1) = \frac{A - b\bar{p}_1 + bc_2}{2b} \ .$$

Again, at an equilibrium (now each duopolist takes the other's price as fixed) $p_1 = p_1^*(p_2^*(p_1))$.

Solving this equation, we get

$$p_1^* = \frac{2}{3}\left(\frac{A}{2b}\right) - \frac{1}{3}c_2 + \frac{2}{3}c_1 \quad .$$

And using the same algebra (or symmetry) we also get

$$p_2^* = \frac{2}{3}\left(\frac{A}{2b}\right) - \frac{1}{3}c_1 + \frac{2}{3}c_2 \quad .$$

For price-setting duopolists who sell perfectly complementary products, we get

$$p_1^* + p_2^* = \frac{4}{3}\left(\frac{A}{2b}\right) + \frac{1}{3}(c_1 + c_2) \quad .$$

Now note that

$$p_1^* + p_2^* - p^M = \frac{1}{3}\left[\frac{A - b(c_1 + c_2)}{2b}\right] \quad .$$

If the railroads charged the combined price $c_1 + c_2$, demand would be $x^0 = A - b(c_1 + c_2)$, which must be positive. Thus the duopolists actually charge a price that is even higher than the monopoly price. This is a special result of the assumed perfect complementarity.

Lecture 15

Symmetric Cournot Equilibrium

There is an easy way to solve the duopoly problem when firms are identical and the equilibrium satisfies the property that each firm produces the same share of industry output (symmetric equilibrium). Recall that for firm 1,

$$\text{Profit} = p(x_1 + x_2)\, x_1 - cx_1$$
$$= ax_1 - b(x_1 + x_2)\, x_1 - cx_1$$

Thus, the first order condition for x_1, given x_2, is

$$a - bx_2 - c = 2\, bx_1 \ .$$

Now suppose we're at a symmetric equilibrium where $x_1 = x_2 = x^s$ (s for symmetric). Then at that equilibrium, each firm is on its reaction function, so the FONC must hold when $x_1 = x_2 = x^s$. Thus:

$$a - bx^s - c = 2bx^s$$

or

$$x^s = (a-c)/3b = ⅔\, x^m \ .$$

This same appeal to symmetry allows us to solve the equilibrium of an "n-opoly" (where there are n identical firms, each of which assumes, when it sets quantity, that the other (n-1) firms will hold their outputs constant. For the n-firm problem

$$\text{Profit of firm 1} = p(x_1 + x_2 + \ldots + x_n)x_1 - cx_1$$
$$= ax_1 - b(x_1 + \ldots + x_n)\, x_1 - cx_1$$

The first order condition is

$$a - b(x_2 + x_3 + \ldots x_n) - 2bx_1 - c = 0$$

Now impose symmetry ($x_1 = x_2 = \ldots = x_n$) at the equilibrium. Then

$$a - b(n-1)\, x^s - 2bx^s - c = 0$$

or

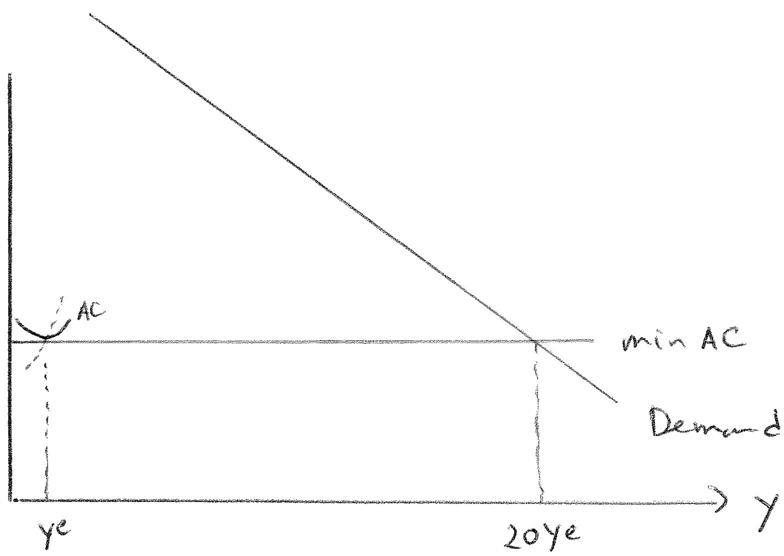
$$x^s = \frac{a-c}{b(n+1)}$$ Note: formula checks for n=1 , n=2.

and $$p = a - b(nx^s) = a - \frac{n}{n+1}(a-c) = (1 - \frac{n}{n+1})a + \frac{n}{n+1}c \ .$$

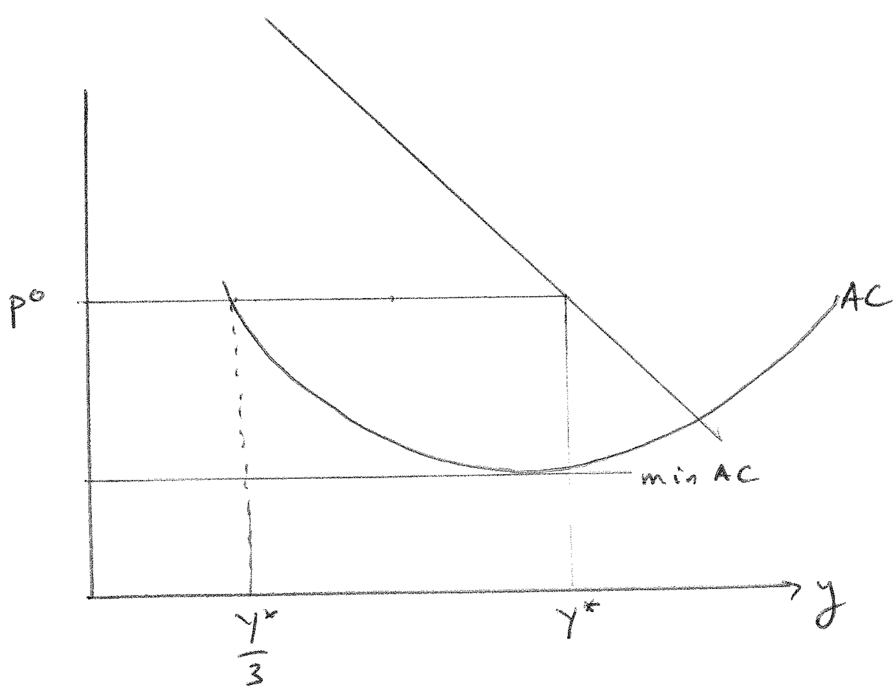
This result says (roughly) that as the number of firms increases (relative to the "size" of the market) the symmetric Cournot equilibrium has each firm producing less and less, and price converging to the competitive price (c).

As a practical matter, the presence of fixed costs often prevents us from having a "large number" of firms in each industry. With fixed costs, there is a social cost of having more firms -- namely, that total fixed cost associated with the industry rise -- as well as a benefit from lith less monopolistic behavior. In our example, since costs are constant, there is no inefficiency as output per firm falls.

The problem is easily illustrated by two alternative pictures. In each case, firms have $C(y) = F + \frac{1}{2} k y^2$, so MC = ky, and AC = F/y + ½ k y (which is U-shaped). Minimum AC is achieved at output such that MC=AC, or $y^e = \{ 2F/k \}^{1/2}$ ("e" for efficient scale).



"Competitive Paradigm." Min.AC is achieved at small scale, relative to the size of the market



"Non-Competitive Paradigm." Min. AC is achieved at an output which is "large", relative to the size of the market.

In the second example, if we want to have 3 firms in the industry, we will have to have p > p0 for each firm to be able to cover its costs. In some cases p^0 will exceed even the price that a monopolist would charge.

Alternatives to the Cournot Assumption

(1) Suppose that we are back in the duopoly model with linear demand and constant costs. Recall that firm 1 tries to maximize

$$\Pi = p(x_1 + x_2)\, x_1 - c(x_1)$$

Under the Cournot assumption, firm 1 chooses $x_1^*(x_2)$ by maximizing Π, assuming x_2 is fixed. Suppose however that 1 thinks 2 will respond to his choice by setting $x_2 = \Psi(x_1)$. What does 1 do in that case? The FONC becomes:

$$\frac{\partial \Pi}{\partial x_1} = p + \frac{\partial p}{\partial x} x_1 + \frac{\partial p}{\partial x} \psi'(x_1) x_1 - c'(x)$$

Suppose for example that 2 announces the rule "I will increase my output in constant proportion to yours". In this case, 2's announced reaction function has the property that

$$\frac{dx_2}{x_2} = \frac{dx_1}{x_1} \quad \text{or} \quad \psi'(x_1) = \frac{dx_2}{dx_1} = \frac{x_2}{x_1} .$$

In that case,

$$\frac{\partial \Pi}{\partial x_1} = p + \frac{\partial p}{\partial x} \cdot x_1 + \frac{\partial p}{\partial x} \left[\frac{x_2}{x_1} \right] \cdot x_1 - c$$

$$= p + \frac{\partial p}{\partial x} (x_1 + x_2) - c = 0$$

But this is the F.O.C. for joint profit maximization (look back at the monopoly lecture). Therefore: if each party announced to the other the rule that

$$\frac{dx_i}{dx_j} = \frac{x_i}{x_j} \ (j = 1,2)$$

The two parties could maintain the joint-monopoly output (provided each believed the other's threat).

(2) A second class of alternatives to the Cournot assumption considers a duopoly where one firm is "smart" and one is "naive." Suppose for example that firm 2 always takes firm 1's output as given. Firm 2 is therefore following the Cournot assumption. Firm 1, on the other hand, is smart, and recognizes that 2 has the Cournot reaction function $x_2 = x_2^*(x_1)$. In this case, we say that firm 1 is a "Stackelberg leader" while firm 2 is a "Stackelberg follower". (Stackelberg was a German economist of the early 20th century.) In the problem set you are asked to analyze the duopoly under the Stackelberg assumption. You will be able to show that: (1) the leader does better than the follower; (2) the leader does better than either party in a symmetric Cournot model; (3) the follower does worse than either party in a symmetric Cournot model.

Lecture 16 **Game Theory - I**

In the last two lectures we considered a duopoly model with

linear demand $p = a - b\,(x_1 + x_2)$

constant marginal costs $c(x_1) = c\,x_1$, $c\,(x_2) = c\,x_2$.

We identified three possible strategies

(a) Co-operate. Each firm produces ½ x^m , where x^m = monopoly output

=> $x = x^m/2 = (a - c) / 4b$; $p = p^m = ½\,a + ½\,c$

$\pi = \pi^m / 2 = ⅛\,(a-c)^2 / b$; one-half of the monopoly profit

(b) Jointly "non-cooperate": Each firm produces the Cournot output

=> $x = x^0 = ⅔\,x^m$

$p = p^0 = ⅓\,a + ⅔\,c$

$\pi = \pi^0 = 1/9\,(a-c)^2 / b$

The Cournot output is the "jointly" non-cooperative output in the sense that, each firm is acting in its own narrowly-defined best interest, given what the other firm is doing.

Given that 1 produces x^0 , 2 will do best by producing x^0 .

(c) Cheat given that your opponent is co-operating.

If 1 has $x_1 = x^m / 2$ (co-operative)

2's "best response" is

$x_2 = x^m - ½\,x_1$ (from the reaction function)

=> $x_2 = ¾\,x^m$

p^c = price if 1 player cheats given the other is co-operating

$= a - b\;5/4\;x^m = a - b\;5/4\;(a-c / 2b) = ⅜a + ⅝c$

π^c = profit to cheater = $¾\,x^m\,(p^c - c) = 9/64\,(a-c)^2 / b$

π^L = profit to "loser" when his/her opponent cheats

$= ½\,x^m\,(p^c - c) = 3/64\,(a-c)^2 / b$

Notice that $\pi^c > \pi^m/2 > \pi^0 > \pi^L$. So co-operating is better than "jointly non-cooperating", but given that your opponent is cooperating, you should cheat.

We can illustrate the "dilemma" in a box showing each sides actions and payoffs. Each entry in the box gives the payoffs to the pair.

	Player #2's Action	
	co-operate	don't co-operate
Player #1 Action		
co-operate	$\frac{1}{2}\pi^m$, $\frac{1}{2}\pi^m$	π^L , π^c
don't co-operate	π^c , π^L	π^0 , π^0

E.g., if 1 co-operates, but 2 doesn't, the payoff pair is (π^L , π^c), where the first entry in the ordered pair shows that #1 gets π^L , and the second entry shows that #2 gets π^c .

Our "game box" looks like:

	2's Strategy Choice:	
	C	N
1's Choice:		
C	1/8 , 1/8	3/64 , 9/64
N	9/64 , 3/64	1/9 , 1/9

where we have ignored the $(a-c)^2/b$ factor that is common to all the payoffs.

Given a box like this you can figure out what strategies each player will follow. E.g., suppose 2 thinks 1 will play C (co-operative). 2 then looks at the record entry in each payoff pair for row 1:

C	N	
1/8 , [1/8]	3/64 , [9/64]	The second column is bigger. So 2 should choose N if 1 chooses C.

If 2 thinks 1 will play N , 2 looks at row 2:

C	N	
9/64 , [3/64]	1/9 , [1/9]	Again, the second column is bigger. So, 2 should choose N if 1 chooses N.

We can also evaluate 1's choices. If 1 thinks 2 will play C , 1 looks at the first entry of payoff pairs in column 1:

C	1/8 , 1/8	
N	9/64 , 3/64	The second row is bigger, so 1 should choose N if 2 is choosing C

Notice that in this game there is always an incentive for either player to play N (regardless of what the other does). If one action is always the best response, we call it the dominant strategy.

The following game doesn't have a unique dominant strategy:

	Player 2: N	Player 2: C	
Player 1: N	1 , 1	1 ½ , ½	← 2 chooses N if 1 plays N .
Player 1: C	½ , 1 ½	2 , 2	← 2 chooses C if 1 plays C .

If 2 thinks 1 will play N , then 2's best response is N . If 2 thinks 1 will play C , then 2's best response is C . In this game, we would say that (N , N) and (C , C) are both "Nash equilibriums." A Nash equilibrium (with 2 players) has two things being true simultaneously:

+ given 1's choice, 2's is optimal,

+ given 2's choice, 1's is optimal.

The duopoly game has 1 unique Nash equilibrium: (N , N) . The second game has 2 Nash equilibria: One is "superior" to the other.

You may have seen other versions of the duopoly game. One common version is the "prisoners' dilemma." For example, suppose you and an ex-friend are involved in a legal dispute. You and he/she will appear before a judge to determine who gets custody of the cat you bought together. You can hire a lawyer (or not). Suppose you think the probability of winning custody is ½ if you don't hire a lawyer and she doesn't, or if you both hire a lawyer. But, if she has a lawyer and you don't, she has a ¾ probability of winning custody. Symmetrically, if she doesn't have a lawyer and you do, you win with probability ¾. We can set up a matrix of probabilities of success:

His / Her Choices

Your Choices	Lawyer	No Lawyer
Lawyer	½ , ½	3/4 , 1/4
No Lawyer	1/4 , 3/4	½ , ½

Notice that having a lawyer is a dominant strategy. The problem is that lawyers cost money. So your true payoff if you hire a lawyer is lower. Both parties would be better off agreeing not to have lawyers. But this is not a Nash equilibrium. The next page shows some real data on child custody suits in California in the early 1980s. The entry in each table shows the probability that the mother wins (the father wins with the complementary probability).

Percentage of Mothers Awarded Child Physical Custody in San Mateo and Santa Clara Counties, California, 1984-84

	Mother Uses: No Lawyer	Lawyer
Father Uses:		
No Lawyer	75%	86%
Lawyer	49%	65%

Source: Mnookin, Maccoby, Depha, and Albiston (1989).

Sometimes it may be possible to "induce" co-operation if the game is played over and over in the future. For example, consider the following idea; If player 1sees that the price last time was p^m , she produces ½ x^m this time. If player 1 sees that the price last time was p^c , she infers that her rival cheated, and she then adopts a "punishment strategy": for the next K periods. She produces x^0 (the Cournot output). After K periods of punishment, she reverts to ½ x^m .

Some questions to consider about this strategy:

1. Will the threat of punishment prevent #2 from cheating?
2. Is the threat "credible"?

To answer Question 1, consider the costs and benefits of cheating for this period.

Benefit: $\pi^C - ½\, \pi^m = 9/64 - 1/8 = 1/64$

For 1 period, the cheater gets π^C instead of $\pi^m/2$.

Cost: $K [½\, \pi^m - \pi^0] = K [1/8 - 1/9] = K / 72$

For the next K (after the initial period of cheating), the other player will "punish" you by producing x^0 . Given her choice, the best you can do is produce x^0 too, giving profit $\pi^0 = 1/9$. Your foregone profit is ($½\, \pi^m - \pi^0$) for each of the K periods. Looking at this, its clear that if $K \geq 2$, the treat will deter cheating!

Credibility?

Notice that the "punishment" is to produce x^0 . This is not "too crazy". But is it credible? After #2 has cheated, he could just plead ignorance and promise not to do it again. #1 could then by-pass punishment (does this sound familiar?) and save herself $K [\frac{1}{2} \pi^m - \pi^0]$ too! So she has a pretty strong incentive not to carry out the threat of punishment!

This is an example of a "dynamic inconsistency". #1 would like to "commit herself" to always punishing a deviation from co-operative play. But once a deviation has occurred, she has an incentive to bail out early, since punishment hurts her too!

Lecture 17 **Game Theory - II**

In the last lecture we described a "punishment strategy" for th repeated Cournot game. A player's strategy is to choose output today on last period's price. If $p_{t-1} < p^m$ = monopoly price, you infer that your opponent cheated and respond by "punishing" him/her for K periods by producing the Cournot output x^0. Afterwards, you revert to co-operative play and produce $\frac{1}{2} x^m$. We showed that this strategy will deter your opponent from cheating, if she believes you will actually carry it out. But should she believe you?

The same issue arises in many contexts:

• the "Cold War": The U.S. threatened to start a nuclear war with Russia if Russian invaded Europe. But many Europeans believed that if the Russians invaded, the U.S. would simply cut its losses.

• flood relief: The government would like to discourage home owners from living in flood-prone areas such as the NJ shore. But when a flood strikes, the government inevitably offers "disaster relief."

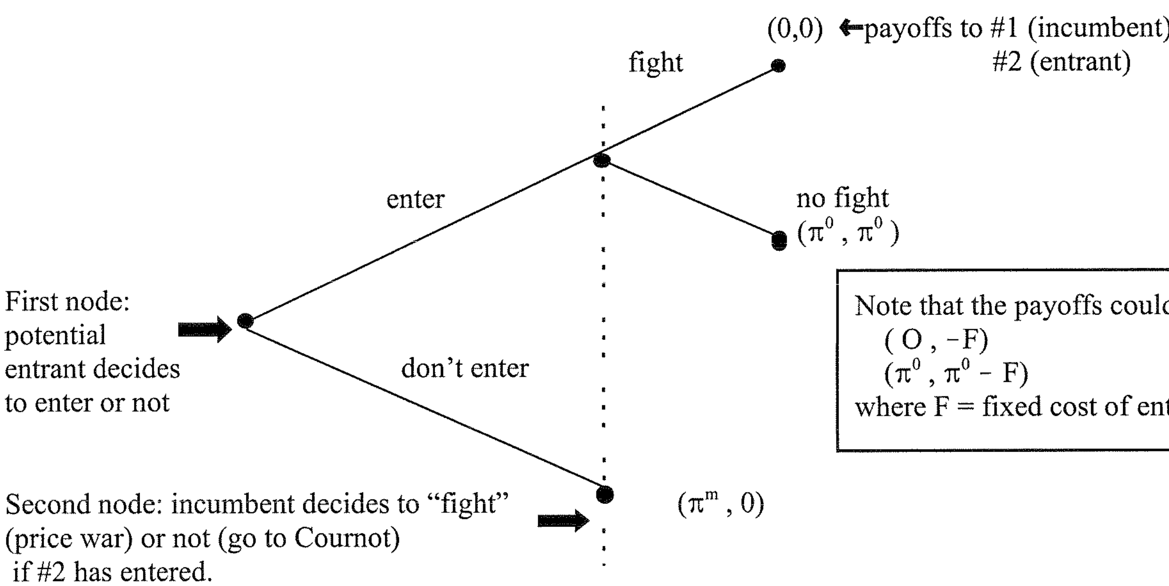
• entry deterrence: a grocery store currently has a monopoly in a certain town. Another chain is thinking of building a new store to compete. The incumbent threatens to keep prices very low if the new chain opens up.

We can analyze simple dynamic games with the aid of a "tree diagram" showing the parties' moves. Consider the entry deterrence game. First, the potential entrant decided whether to enter or not. Then, the incumbent decides whether to engage in a "price war" or not.

Let π^m = profit (per year) for incumbent without entry

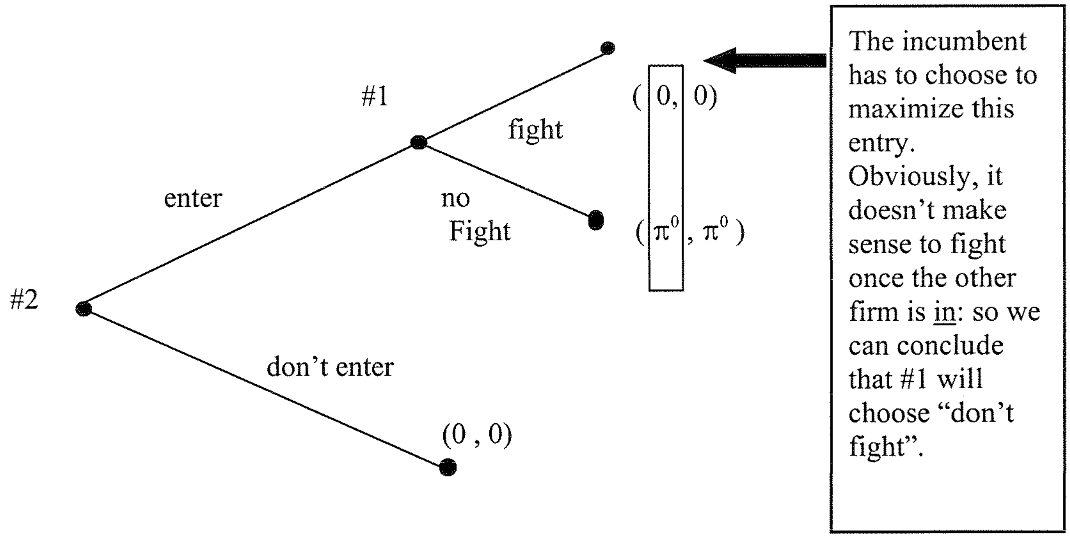
π^0 = profit (per year) for each firm if the new entrant enters and they subsequently act like a Cournot duopoly ($\pi^0 < \pi^m$).

In a "price war", the incumbent charges p = c and earns no profit:



Note that the payoffs could be:
(O , -F)
(π^0 , π^0 - F)
where F = fixed cost of entry

Notice that once the potential entrant has acted, it's up to the incumbent to decide what to do. Suppose that #2 has entered.



Now #2 looks at the ultimate payoffs to entry. If she enters, she gets π^0 since she knows that #1 will choose don't fight. So she always enters.

The solution method here is called "backward induction". At the last stage, depending on whose turn it is to move, we deduce his/her action by looking at his/her payoffs. Then, we back up the tree to the previous move.

Notice that (Enter , don't fight) is the "dynamically consistent" equilibrium. (Enter , fight) is not "dynamically consistent" because even though #1 would like to threaten to fight, once #2 has entered #1's payoffs are maximized by not fighting.

Implications:

• in the "Cold War", it was not a credible threat to promise all-out nuclear war if Russia invaded Europe.

• in hostage situations, it is not a credible threat to claim that you "never negotiate" with terrorists.

• in the entry game, it is not a credible threat to threaten a price war if the new chain enters.

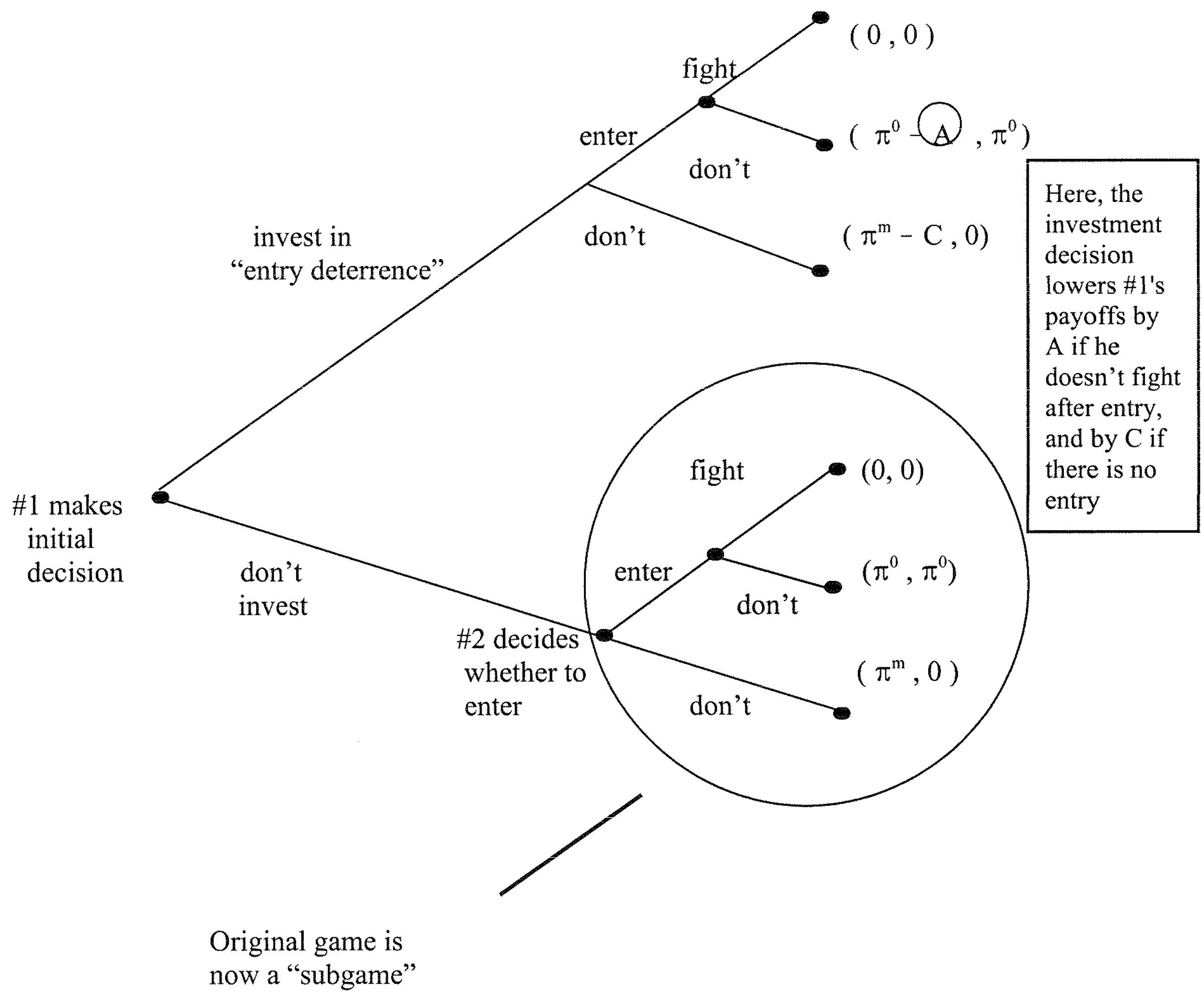
• in the dynamic duopoly, it is not a credible threat to threaten K periods of punishment.

Problem / Interpretation:

This analysis is predicated on the notion that players behave "rationally" even if they threaten to do something, once the time comes to carry out the threat they will always do what is in their self-interest form then on, regardless of what happened earlier. This is a common notion in economics: e.g., that "bygones are bygones" or that "sunk costs don't count."

Notice that in our entry game, the incumbent would like to be able to "commit" herself to behaving irrationally. If #2 knows that th incumbent will always fight, #2 won't enter (especially if there is a fixed cost of entry).

Suppose that there is an earlier decision that #1 can make to alter the payoffs of fighting when #2 enters.



Here, the prior decision might involve some investment in "overhead" that increases the cost of operation for the incumbent, so payoffs are

$\pi^m - C$ if #2 doesn't enter

$\pi^0 - A$ if #2 enters and #1 doesn't fight .

Will #1 invest in this strategy? Again, we find an answer by backward induction.

<u>Path #1</u> (#1 doesn't invest)

We know that if #2 enters, #1 won't fight
=> #2 gets π^0 if he enters, 0 if not => #2 will enter => #1 will earn π^0

<u>Path #2</u> (#1 invests)

We know that if #2 enters, #1 will fight if $\pi^0 - A < 0$ <u>or</u> $A > \pi^0$.
Assuming this is true, #2 knows that #1 will fight
=> #2 won't enter => #1 earns $\pi^m - C$

This is worthwhile if $\pi^m - C > \pi^0$ or $C < \pi^m - \pi^0$

Thus 1's payoffs boil down to:
don't invest in entry deterrence: earn π^0
invest in entry deterrence: earn $\pi^m - C$

<u>Conclusion</u>: #1 may make an investment in "overhead costs" provided

(i) it lowers the profit from not fighting after #2 has entered ($A > \pi^0$)
(ii) it doesn't cost "too much" if #2 doesn't enter ($C < \pi^m - \pi^0$)

Notice the key to entry deterrence is that once you've made the decision to invest, it has to affect your payoffs. You are "committing" to fighting by changing your payoffs in the later subgame.

There is an interesting extension of this model to a case where potential entrants don't know who they are "up against". Suppose there are two types of incumbents:

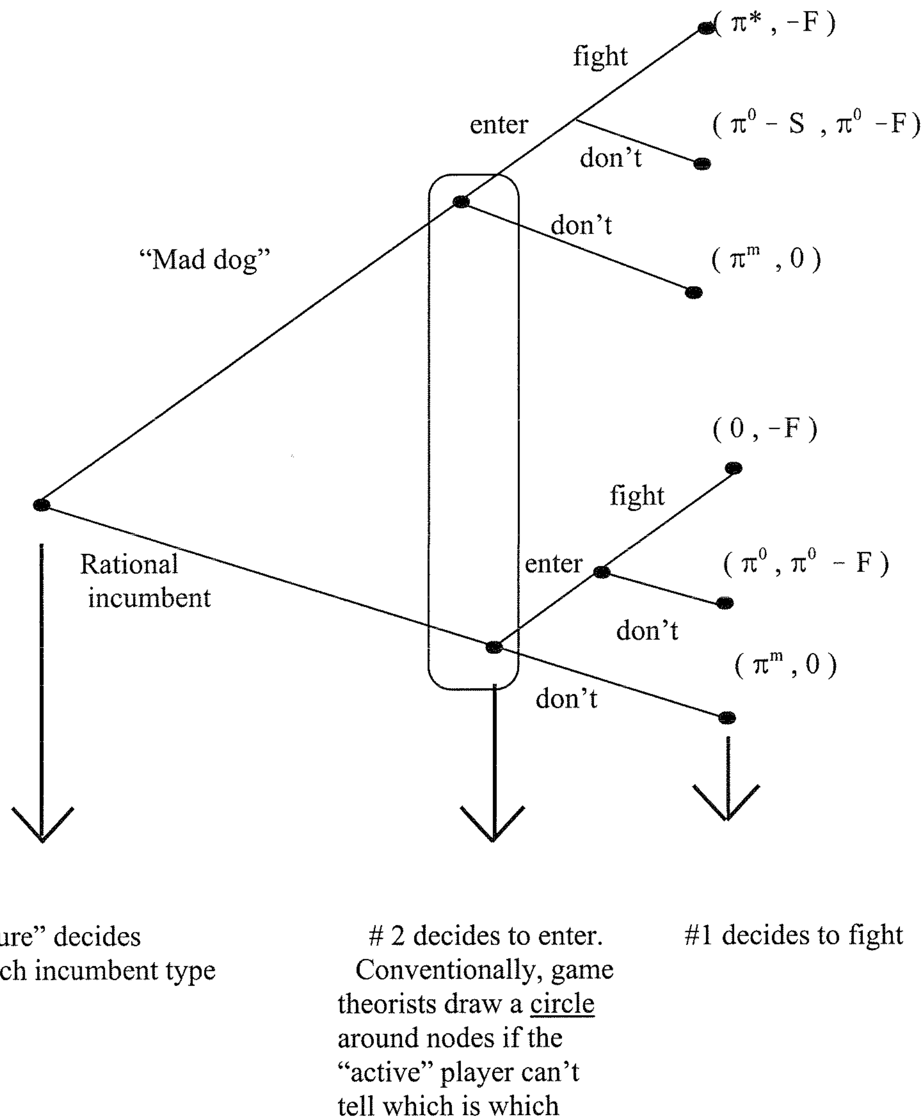
- rational incumbent with payoffs:
  - 0 to a price war
  - π^0 to a duopoly with the entrant

- "mad dog" incumbent with payoffs
  - π^* to a price war
  - $\pi^0 - S$ to a duopoly

The possibility that $\pi^* > 0$ reflects the idea that the "mad dog" likes to fight. $S > 0$ is the "shame" that a "mad dog" feels for not fighting.

If $\pi^* > \pi^0 - S$, the "mad dog" will fight. "Mad dog" <u>really</u> likes to fight <u>or</u> gets a lot of shame from being a "wimp".

Now #2 faces two different potential opponents, although he doesn't know which is actually the incumbent. Suppose also that there is a fixed cost of entry F.



If #2 enters and the incumbent is "mad-dog" then a fight occurs and the entrant earns $-F$.

If #2 enters and the incumbent is rational, the incumbent won't fight and the entrant will earn $\pi^0 - F$.

=> Expected profit is (prob. of mad-dog) $\times$ ($-F$) + (prob. rational) $\times$ ($\pi^0 - F$) .

As the incumbent, you should try and raise #2's belief that you're crazy.

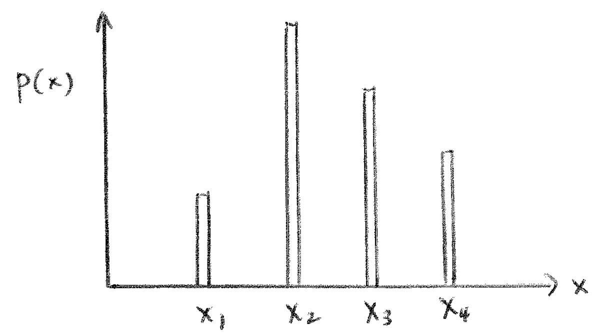
Lecture 18

Introduction to Uncertainty

In the next four lectures we show how to extend the theory of consumer choice to handle choice under uncertainty. For simplicity, we deal mainly with uncertainty over income. Assuming that prices are fixed, alternative realizations of random income translate directly into alternative utility levels. We begin with a brief review of statistics.

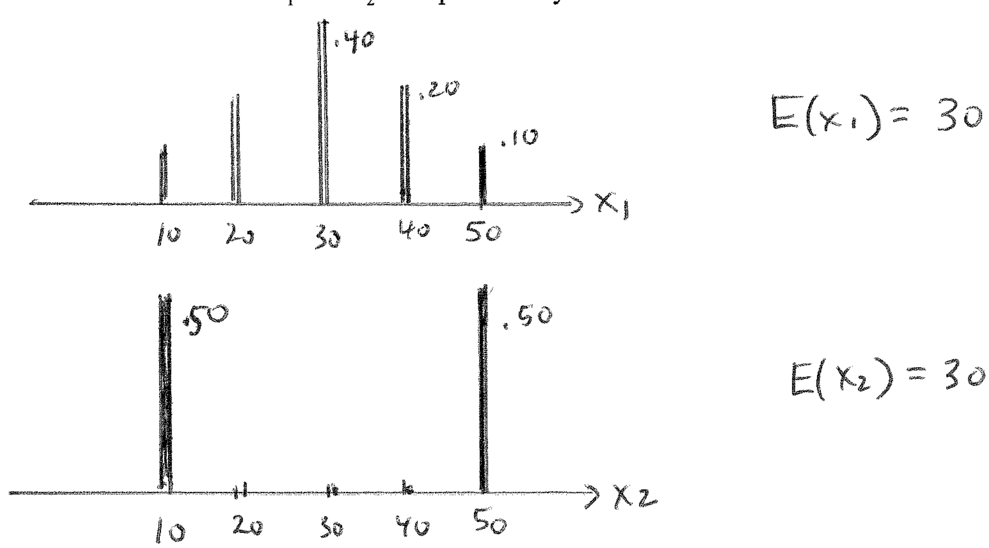
Review of Basic Statistical Concepts

Suppose that x is a random variable that takes on values x_1, x_2 ,..., x_n with probabilities p_1 , p_2 , ..., p_n . Note that $\Sigma j\ p_j = 1$ if the alternatives x_1, ..., x_n are exhaustive and mutually exclusive. We can draw the probability distribution of x as follows:



The height of the bar above each point indicates its probability. We define the mean or expected value of x, denoted by E(x) (or sometimes by $\bar{x}$) as $E(x) = \Sigma_j\ p_j\ x_j$. The mean E(x) is just a weighted average of the alternative realizations of x. The weights are the probabilities associated with each realization.

Consider the two random variables x_1 and x_2 with probability distributions as shown:



Note that these distributions have the same mean, but x_2 is "more disperse". One way to describe the dispersion of a random variable is its variance, denoted Var (x):

$$Var(x) = \Sigma_j p_j (x_j - E(x))^2 .$$

The variance of x is the average of the squared difference between x_j and the mean of x. As an exercise, calculate Var (x_1) and Var(x_2) from the previous page. We say that x is a "degenerate" random variable if x=E(x) all the time, in which case Var(x) = 0.

We can also consider functions of random variables. If g is a function (from the real numbers) then the variable z = g(x) is a random variable created by taking $z_j = g(x_j)$ in each realization of x. We can therefore define the mean of g(x) as

$$E(g(x)) = \Sigma_j p_j g(x_j) \quad .$$

In the special case where g is linear: g(x) = a + bx , we have

$$\begin{aligned} E(g(x)) &= \Sigma_j (a + bx_j) p_j \\ &= \Sigma_j (ap_j + b x_j p_j) \\ &= a \Sigma_j p_j + b \Sigma_j x_j p_j \end{aligned}$$

But if the p's sum to 1 the first term is a. And the second term is just b E(x). So we have

$$E(a + bx) = a + b E(x) \quad .$$

Exercise: Show var(a + bx) = b^2var(x) .

Choices Over Uncertain Incomes

We now suppose that individuals are asked to make choices between alternative "income lotteries". Each lottery is a probability distribution over income. In ranking two alternative lotteries, we hold constant income in the absence of either lottery (which may itself be random). In some cases, the covariation between the payoffs of the lottery we are considering and income in the absence of the lottery that pays some positive amount if you have an accident, and some negative amount otherwise. The policy is attractive, however, because in the event of an accident and no insurance you lose a lot of money!

Now with no uncertainty individuals always prefer more income to less. So the following utility functions are all "equivalent" in terms of the rankings they give over amounts of "income (y):

$$U(y) = a + by \quad b > 0$$

$$U(y) = e^y$$

$$U(y) = y^3$$

Since each function is increasing, each indicates a preference of more over less. That's all we need to know, if all we want to know is how to rank different amounts of income.

On the other hand, suppose that we want to rank different "lotteries" on income. For example, consider:

lottery 1 Payoff= \$100 with p=½
Payoff= 0 with p=½P

lottery 2 Payoff = \$70 with p=½
Payoff = \$30 with p=½

In the 1940's von-Neumann and Morgenstern asked: is there some way of assigning a utility number to each possible outcome, in such a way that if U(100) = U_{100} ; U(70) = U_{70} ; U(30) = U_{30} ; and U(0) = U_0, then we can compare these lotteries by comparing "expected utility:"

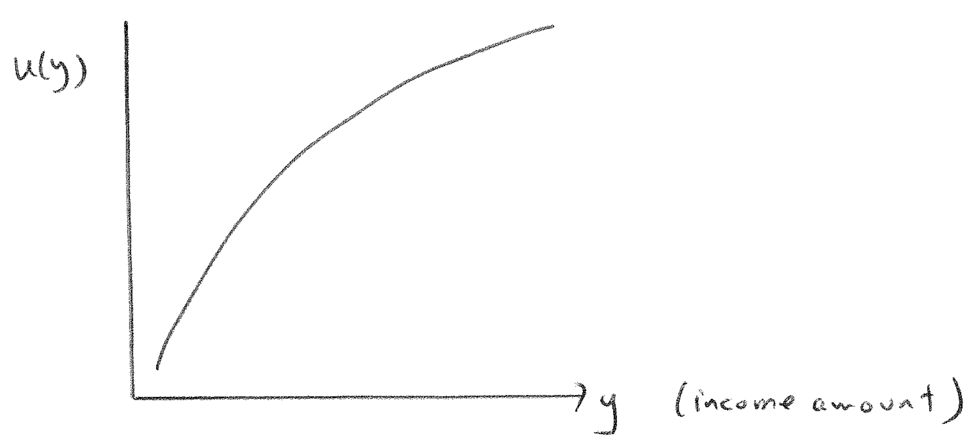
$\frac{1}{2} U_{100} + \frac{1}{2} U_0$ in case of Lottery 1

$\frac{1}{2} U_{70} + \frac{1}{2} U_{30}$ in case of Lottery 2

The answer is yes, under some assumptions, although we won't prove it. Thus, if preferences satisfy some assumptions, there is a utility function defined only over income amounts that we can use to compare both certain incomes (which is trivially easy anyway) but also lotteries over incomes. The idea is that if we get the utility differences between different incomes just right, then we can use the expected utility criterion to compare lotteries or uncertain income prospects.

NOTE: Normally we don't care about the gauge of utility. If U(x) is a utility function, then we could use $[U(x)]^3$, although the cubing operation distorts the gauge.

How do you feel about Lottery 1 vs. Lottery 2? Chances are, you would take Lottery 2. That says something about the shape of your expected utility function.



The expected utility function is always increasing, since more money is always better than less (if you're an economist, at least).

Suppose your utility function (the special one that you use when you compare lotteries -- call it your "expected utility function") is linear: e.g.

$$U(y) = a + by.$$

Then

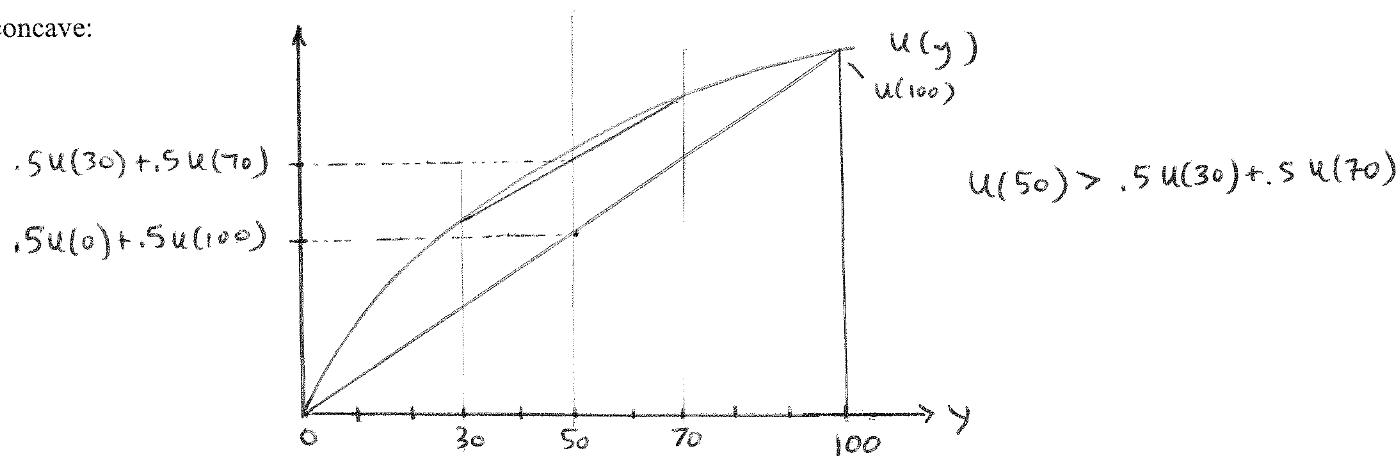
$$U(0) = a$$
$$U(30) = a + 30b$$
$$U(70) = a + 70b$$
$$U(100) = a + 100b$$

It's easy to see that $\frac{1}{2}U(70) + \frac{1}{2}U(30) = \frac{1}{2}(0) + \frac{1}{2}U(100)$.

This gives us our first result:

If the expected utility is linear in income, then lotteries of equal expected value are ranked equally.

On the other hand, if you prefer Lottery 2 over Lottery 1, your expected utility function must look concave:



Conversely, if you prefer Lottery 1 over Lottery 2, your expected utility function must be convex.

Generally, it seems useful to assume that people dislike "risk" (although gambling is an exception). We say that someone is risk-averse if they prefer $\bar{y}$ dollars for sure to the lottery which is created by adding a random error ϵ (with mean zero) to $\bar{y}$:

$$E\,U(\bar{y}) \geq E\,U(\bar{y} + \epsilon) \;,\; E(\epsilon) = 0$$

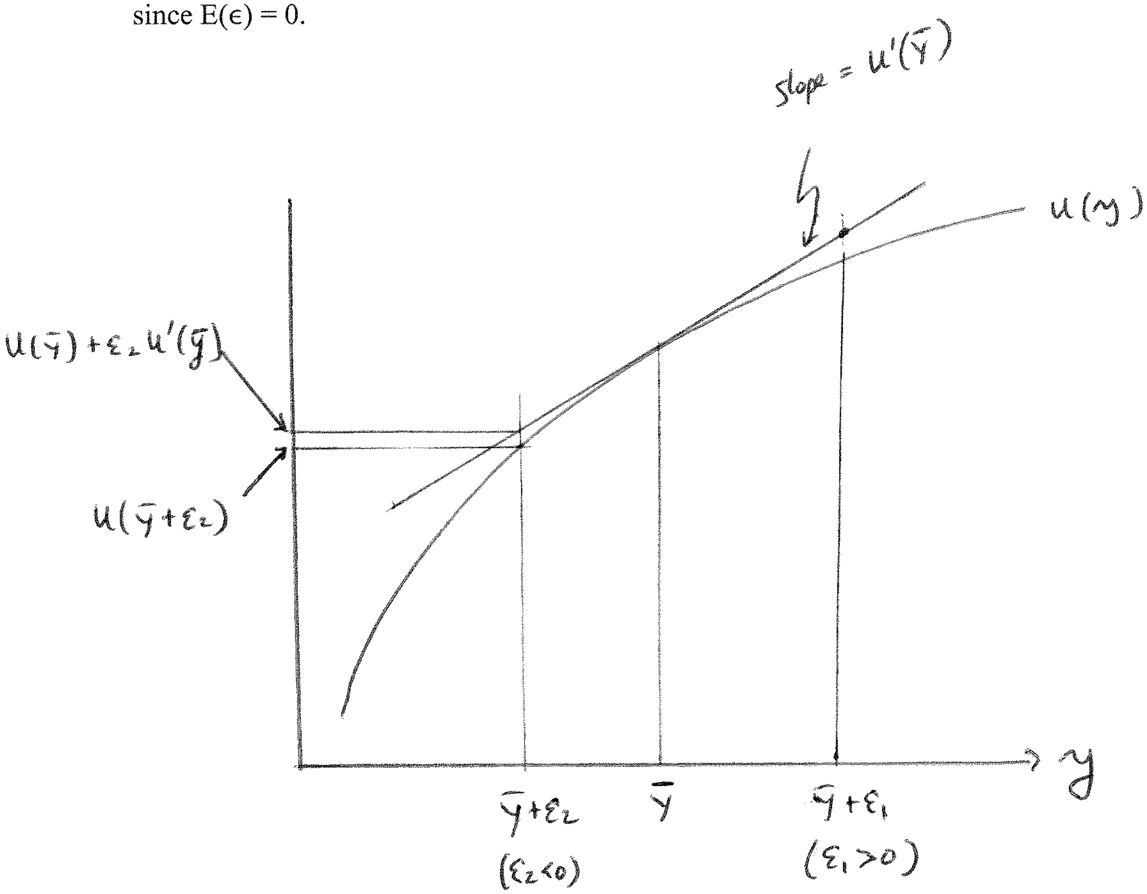
If U is concave, this equation is true. Why? For any realization of the random variable ϵ (say $\epsilon = \epsilon_1$), if U is concave then

$$U(\bar{y} + \epsilon_1) \leq U(\bar{y}) + \epsilon_1 U'(\bar{y}).$$

So, taking expectations over all realizations of ϵ ,

$$\begin{aligned} E\,U(\bar{y}+\epsilon) &\leq E\,U(\bar{y}) + E\ (\epsilon U'(\bar{y})), \\ &= U(\bar{y}) + U'(\bar{y})E(\epsilon), \\ &= U(\bar{y}) \end{aligned}$$

since $E(\epsilon) = 0$.



For any concave function, with slope $u'(\bar{y})$ at a point $\bar{y}$:

$$u(y) \leq u(\bar{y}) + (y-\bar{y})\,u'(\bar{y})$$

This expresses the result that a concave function always lies <u>below</u> its tangent planes

Lecture 19

Uncertainty II

Last lecture we introduced the idea that people might use a special utility function, defined over nonrandom income amounts, with the "right" degree of curvature so that it allows a consumer to rank income lotteries. In particular, if an income lottery is available that pays the amount y_j with probability p_j (with $p_1+p_2+...+p_n=1$), then the lottery can be ranked against any other lottery by the expected utility criterion:

$$\sum_j \; p_j \, U(y_j) = E \, U(y) \quad .$$

The function that you use to rank income lotteries is called your "von-Neumann Morgenstern" utility function, or sometimes just your "expected utility" function.

Examples:

(1) linear: $U(y) = a + b\,y$. Gives rise to an expected value ranking

(2) power function: $U(y) = y^{\alpha}$ for $0 < \alpha < 1$. This is a concave function, so people with preferences like this exhibit risk aversion.

(3) exponential: $U(y) = -\exp(-Ry)$ for $R > 0$. Note that this function is increasing and concave, and ranges from minus infinity to 0. This particular function is often used in finance because if all income lotteries are normally distributed, we get a nice ranking.

For $y \sim N(\mu,\sigma^2)$ one can show that[1] $E[\,-\exp(-Ry)\,] = -\exp\,(\,-R\mu + \frac{1}{2}R^2\,\sigma^2\,)$

$$= -\exp\,(-R\,(\,\mu - \tfrac{1}{2}R\sigma^2\,)\,) .$$

Therefore a lottery with mean μ and variance σ^2 is assigned a value based on $\mu - \frac{1}{2}R\sigma^2$.

This is pretty nice, since it means that people with higher R assign a bigger discount to a lottery with higher variance (for a given mean).

We know that expected utility functions are not invariant to an arbitrary transformation. If your expected utility function is $U(y) = \alpha y$ then you are "risk neutral", and only care about expected value. If mine is

$$V(y) = \{\,U(y)\,\}^{1/2}$$

then my utility function is concave ($V(y) = \alpha^{1/2}\, y^{1/2}$). So I exhibit risk aversion. Thus you and I will not evaluate lotteries in the same way.

---

[1]The trick here is to look up in a stats. book the "moment generating function" for a normal random variable, and then play with that function.

It turns out that expected utility functions are invariant to increasing linear transformations. In other words, if your utility function is U(y) and mine is V(y) = a + bU(y) for b > 0, then we rank lotteries exactly the same. To see this, consider lotteries y_1 and y_2. Suppose you prefer 1 to 2. Then

$$E[\ U(y_1)\ ] > E[\ U(y_2)\ ].$$

If this is true, then its also true that

$$a + b\ E[\ U(y_1)\ ]\ >\ a + b\ E[\ U(y_2)\ ]\quad <=> \quad E[\ a + b\ U(y_1)\ ] > E[\ a + b\ U(y_2)\ ]$$

so I also prefer 1 to 2.

This is a useful fact, because it means that you can "rescale" your expected utility function so that the worst income realization (among a set of lotteries under consideration) is assigned a utility of 0 and the best is assigned a utility of 1. To see this, imagine that we have several lotteries to compare, and the worst outcome is −10,000 and the best outcome is +250,000. Suppose that U(−10,000) = u_0 and U(250000) = u_1. Consider re-scaling to V(y) = a + b U(y), and choose

$b = 1/\ (u_1 - u_0)$ and

$a = -u_0/(u_1 - u_0)$.

You can check that V(−10,000)=0 and V(250000)=1. And we've seen that V gives the same ranking over lotteries. So we might as well use V.

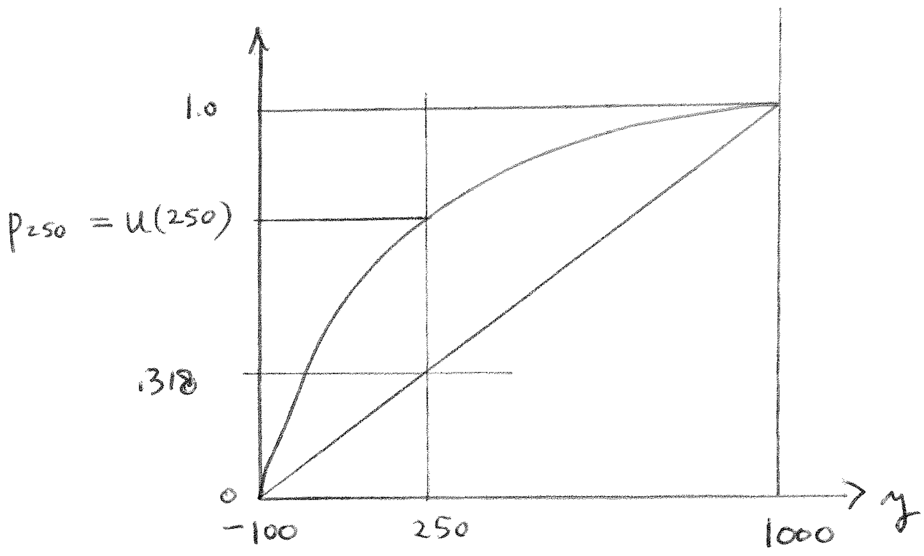
We are now in a position to describe how to derive your own expected utility function. Lets assume the best possible outcome among the lotteries we are considering is $1000, and the worst is −$100. We want to assign utility numbers to all possible incomes from −100 to 1000. Begin by assigning U(−100) = 0 and U(1000)=1. For any intermediate income level (e.g., 250) ask yourself the following question:

"If I had $250 for sure, or a lottery with probability p of $1000 and (1−p) of −100, what p would make me indifferent?"

Call that p_{250}. Its clear that $0 < p_{250} < 1$. Also, $p_{251} > p_{250}$ (but probably not much bigger). Now assign the income $250 the utility number $p_{250.}$ Why does this work? By definition of p_{250}:

$$p_{250}\ U(1000) + (1-p_{250})\ U(-100) = U(250)$$

and we've normalized U(1000)=1 and U(−100)=0. So U(250) = p_{250} . Experimental economists use this idea in the lab to figure out who is more or less risk averse. As the following diagram shows, the more curvature in your preferences, the bigger is p_{250}, and the higher the relative chance of winning 1000 has to be to make you indifferent.



- Risk neutral person has $p(1000) + (1-p)(-100) = 250$ $\Rightarrow p = .318$
- Risk averse person has $p_{250} > .318$

Applications of Expected Utility: the Demand for Insurance

We will use the exp. utility function to show that if you are risk averse, and if you can buy "actuarily fair" insurance, then you will always buy full insurance for any risk. For concreteness, suppose that you have an income of $30,000 and face the risk of an accident with p=0.05 (i.e., 5%). If the accident occurs you will face a medical bill of $10,000. Your exp. utility function is U(y). If you can't buy any insurance, your expected utility is

$$(1-p)\, U(30{,}0000) + p\, U(20{,}000).$$

How would insurance work in this simple world? An insurance contract for $1 of coverage is a promise by the ins. company to pay you $1 if you have an accident (and nothing if you don't). Suppose you can buy $1 of coverage for a premium π. For the ins. company, the expected value of the contract is:

$$(1-p)\,\pi + p\,(\pi - 1)\,.$$

With probability (1−p), you pay the premium and nothing happens. With probability p, you pay the premium, but there is an accident and they have to pay out a claim of $1. If insurance companies are risk neutral, they will compete for business by lowering π until the point where

$$(1-p)\,\pi + p\,(\pi - 1) = 0$$

which implies that $\pi = p$. This is so-called "actuarily fair" insurance: $1 of coverage is available at a premium equal to the probability of a claim.

Suppose you buy C units of coverage at premium π. Your expected utility is:

$$\Phi(C) = (1-p)\, U(30{,}0000 - \pi C) + p\, U(20{,}000 - \pi C + C) \quad ,$$

where the function Φ captures the value of different levels of coverage. If you choose C to maximize Φ, we get the FONC

$$\Phi'(C) = -\pi\,(1-p)\,U'(30{,}0000 - \pi C\,) + (1-\pi)\,p\,U'(20{,}000 - \pi C + C) = 0\,,$$

where U′( ) is the derivative of your exp. utility function. Note that the SOC is

$$\Phi''(C) = \pi^2\,(1-p)\,U''(30{,}0000 - \pi C) + (1-\pi)^2\,p\,U''(20000 - \pi C + C)$$

which is always negative if $U'' < 0$.

Consider the FONC carefully when $\pi = p$. In this case it says:

$$-p\,(1-p)\,U'(30{,}0000 - pC\,) + (1-p)\,p\,U'(20{,}000 - pC + C) = 0\,,$$

or

$$U'(30{,}0000 - pC\,) \;=\; U'(20{,}000 + C(1-p)\,)$$

Now if $U''<0$ that says that the function $U'(y)$ is strictly decreasing. So

$$U'(y_1) = U'(y_2) \quad \Leftrightarrow y_1 = y_2\,.$$

So we'll satisfy the first order condition when

$$30{,}000 - pC = 20{,}000 + C(1-p)$$

or $\quad C = 10{,}000$!

An interesting exercise is to redo this analysis, assuming that if you buy any insurance at all, you have to pay an "underwriting fee" of \$F, but that the price per unit of coverage is $\pi = p$. In this case, you can show that if you buy insurance, you fully insure. But, if F is too high, you may decide not to buy any. Another interesting exercise is to re-do the analysis assuming that $\pi > p$ (i.e., that the premium is higher than the actual probability of an accident).

Lecture 20

UNCERTAINTY - III

One of the most interesting problems in markets with uncertainty is the problem of <u>moral hazard</u>. This phenomenon arises when somebody changes their behavior "inefficiently" because of a contract they have entered. The term arises from the insurance industry: to an insurance company, a policy-holder who fails to exercise due caution <u>because he is insured</u> is a "moral hazard." A nice example is a driver who has rented a car and purchased the "full insurance" option. Moral hazard can arise in other contexts. For example, it is often argued that the welfare system discourages people from looking for work. People who receive welfare payments have less incentive to look for work that they would if they were not receiving any payments. In this lecture, we will analyze the demand for insurance when policyholders, through their own effort, can influence the probability of an accident. We will show:

(1) with full insurance, policyholders have no incentive to avoid accidents.

(2) a "solution" to the moral hazard problem involves a deductibility clause.

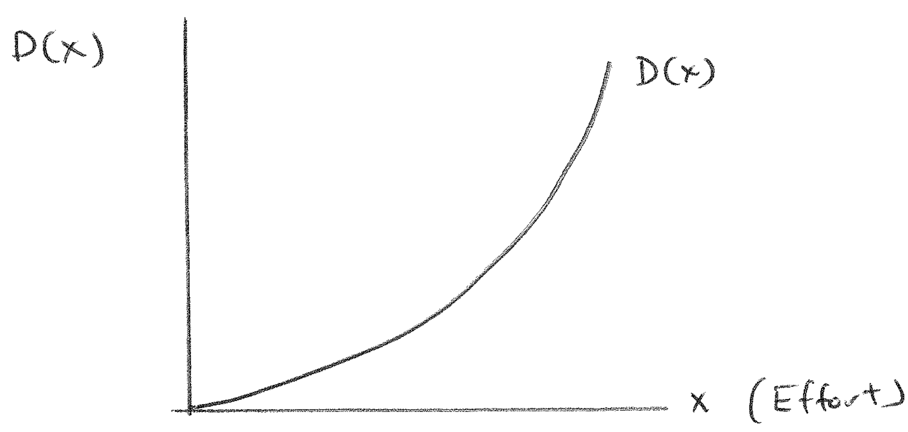
In particular, a higher deductible will generally induce a greater level of preventative care <u>at the cost</u> of inducing income variability in policy holders' income. Thus there is a tradeoff between insurance and efficiency.

The set up of the model is very simple. In each state of the world (accident/no-accident) the insured has $\$y_0$ to start. In the event of an accident, he loses \$L. The insurance company offers to pay \$C (coverage) in the event of an accident, and charges $\$\pi$ per unit of coverage. π is thus the premium rate. Expected utility depends on both ultimate wealth, and effort expended on accident prevention (x).

For simplicity, we assume that consumers evaluate income and effort packages using:

U(net income) - D(x)

where U is a standard expected utility function, and D(x) represents the cost of effort. We assume that D(x) is convex, with D(0)=0 and D′(0) = 0. This means that the D(x) function looks like:



The probability of an accident is p(x), where p(0) = .05 (say) and more effort reduces p. So this means

$$p(x) > 0 \text{ for all } x, \quad \text{and} \quad p'(x) < 0 .$$

Expected utility when a consumer buys C units of coverage and exerts effort level x is:

$$\Phi(C,x) = p(x)\{U(y_0 - \pi C - L + C) - D(x)\} + (1 - p(x))\{U(y_0 - \pi C) - D(x)\}$$

$$= p(x)\,U(y_0 - L + C(1-\pi)) + (1 - p(x))\,U(y_0 - \pi C) - D(x) .$$

Notice that since equal effort is expended in both accident and no accident cases, you end up subtracting D(x) from expected utility of income. Suppose that the insurance company (through long experience) knows p(x) (i.e., it knows how much effort you will expend). If it breaks even on insurance, it must be true that

$$(1-p(x))\,\pi - p(x)\,(1-\pi) = 0$$

using the same argument as we used in the last lecture. Thus it has to charge a premium $\pi = p(x)$, where x is the level of effort that consumers end up exerting.

A consumer takes π as a given and choose x and C to maximize

$$\Phi(x, C) = p(x)\,U(y_0 - L + C(1-\pi)) - (1 - p(x))\,U(y_0 - \pi C) - D(x) .$$

The first-order conditions are

(1) $\Phi_C = p(x)(1-\pi)\,U'(y_0 - L + C(1-\pi)) + \pi(1-p(x))\,U'(y_0 - \pi C) \geq 0$

(with > implying that the consumer is "rationed")

(2) $\Phi_x = p'(x)\{U(y_0 - L + C(1-\pi)) - U(y_0 - \pi C)\} - D'(x) = 0 .$

If the premium is set knowing the level of care that consumers will ultimately select, then $p(x)(1-\pi) = \pi(1-p(x))$ and equation (1) can be rewritten as:

(3) $U'(y_0 - L + C(1-\pi)) - U'(y_0 - \pi C) \geq 0$

Suppose that (3) is an <u>equality</u>, i.e., you get all the coverage you want. Then, as in the last lecture,

$$y_0 - L + C^*(1-\pi) = y_0 - \pi C^* \quad \Longleftrightarrow \quad C^* = L \quad \text{(i.e., full coverage).}$$

But, with full coverage, what is your incentive to take care? If you take extra care, p falls and you save

$$U(y_0 - L + C(1-\pi)) - U(y_0 - \pi C)$$

units of utility. But with full coverage, the saving is nil: you get none of the benefits of your action because the insurance company is taking all the risks! Therefore (if $D'(0) = 0$) the FONC are satisfied with $x^* = 0$ and $C^* = L$. Everybody takes <u>minimum care</u>. Insurance companies expect them to do so, and set their premiums accordingly.

This level of care is socially inefficient, because the marginal cost of care is 0 when x=0. If people took a bit more care, it would cost them almost nothing, yet it would generate a lot less accidents and end up reducing premiums. However, there is a breakdown in the usual argument about markets leading to socially efficient outcomes, because each individual consumer takes the premium rate as given, although ultimately $\pi = p(x)$, since in the long run, the insurance company will realize what's going on.

Solution with No Moral Hazard

Suppose that the insuree recognized the impact of his effort choice on the premium (this would be true if insurance companies could monitor effort). In this case the problem is:

$$\max \quad \Phi(x, C) = p(x)\, U(y_0 - L + C(1-p(x)) + (1-p(x))\, U(y_0 - p(x)C) - D(x).$$

Note that we've set the premium to p(x) – this assumes that as x rises the premium adjusts. The first-order conditions are:

(4) $\Phi_c = p(1\text{-}p)\, U'(y_0 - L + C(1-p)) - p(1-p)U'(y_0 - pC) = 0 \quad \Leftrightarrow C = L$

and

(5) $\Phi_x = p'(x)\, \{ U(y_0 - L + C(1-p)) - U(y_0 - pC)\} - D'(x)$

$$- p(x)\, p'(x)\, C\, U'(y_0 - L + C(1-p)) - (1-p(x))p'(x)\, C\, U'(y_0 - pc) = 0 .$$

Compare this to equation (2) and note that making premiums depend on effort adds the last set of terms. Now use the substitution $y_0 - L + C(1\text{-}p) = y_0 - pC$ and this becomes

(6) $- p'(x)\, U'(y_0 - pC)\; L \;=\; D'(x)$. (Recall that $p'(x) < 0$).

This has a very nice interpretation. If you add more effort, it costs $D'(x)$, which is the right hand side. If you exert more effort, it reduces the probability of an accident by $p'(x)$, and the saving is \$L times the marginal utility of income $U'(y_0 - pC)$, which is the left hand side. So the optimal level of care equates the marginal costs and marginal benefits of effort. Note that (6) generally implies a level of effort greater than 0, unless $p'(x) = 0$, in which case effort does not affect accident probabilities. Notice too that the "optimal" (no moral hazard) solution gives the marginal benefits of accident prevention as $U'(y_0\text{-}\, pC) \times L$.

A Partial Solution

How could we "induce" insured consumers to expend some effort in accident prevention, even if their level of care doesn't lead to a lower premium rate? Look at the first order conditions (1) and (2) from our original problem. Assuming $p = \pi$, these say:

(1*) $U'(y_0 - L + C(1-\pi)) - U'(y_0 - \pi C) \;\geq\; 0$

and

(2*) $p'(x)\, \{U(y_0 - \pi C) - U(y_0 - L + C(1 - \pi))\,\} \;=\; D'(x)$

Suppose that an ins. company only sells policies with less than full coverage (C < L). Then the level of utility in the accident state is less than the level of utility in the accident state:

$$U(y_0 - \pi C) > U(y_0 - L + C(1-\pi))$$

and there is some incentive to prevent accidents. Insured would prefer more coverage, since the first-order condition is positive. But insurance companies refuse to offer it! They demand a "deductible amount", that the insured has to pay in event of an accident. The amount of the deductible L−C, influences how much care is taken.

Let's write the model in terms of the loss L and the deductible amount A. Observe that C=L−A . Plug these two expressions into the equation (2*) to get:

(1**) $$-p'(x)\{U(y_0 - \pi L + \pi A) - U(y_0 - \pi L - A(1-\pi))\} = D'(x) .$$

Notice what a deductible does: it gives you more income in the no-accident state and less income in the event of an accident. Now we can show that a higher deductible causes the insured to take more care. Equation (1**) says that

$$\Delta U = U(y_0 - \pi L + \pi A) - U(y_0 - \pi L - (1-\pi)A) = -D'(x)/p'(x).$$

Suppose that the probability and cost functions are:

$$p(x) = p_0 - ax.$$

$$D(x) = \alpha x^2 .$$

Then $p'(x) = -a$; $D'(x) = 2\alpha x$, and equation (1**) requires that optimal effort x^* satisfies

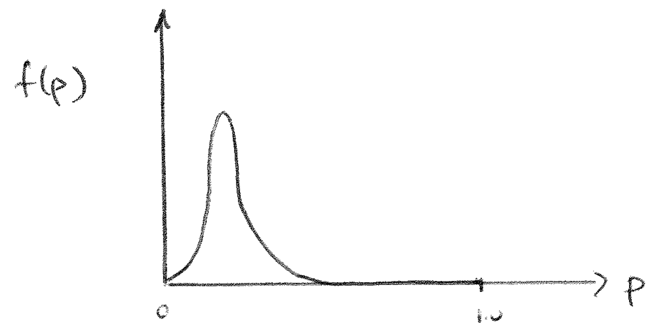
$$\Delta U = 2\alpha\, x^* / a$$

If ΔU is bigger (i.e., A is bigger) then optimal effort is higher.

Lecture 21

UNCERTAINTY IV: Adverse Selection

In this lecture we continue our discussion of insurance markets. As in the last couple of lectures, we consider individuals with initial wealth y_o who face the risk of a loss L. Now we introduce heterogeneity: individual i faces a probability p_i of the risk, where p_i is a random variable, with a distribution across the population. For example, p_i could have the following (continuous) distribution:



The problem facing insurance companies is that they cannot tell who is who: although individuals have different p_i's, they all look the same. On the other hand, we assume that each person knows his or her own risk probability. [This may or may not be appropriate, depending on the application. In fact, all that is really necessary for our story is that individuals can make a better "guess" of their own accident rate]. A natural assumption is that insurance companies set a premium rate π so that on average the premium equals the probability of an accident (i.e., $\pi = E(p)$). What we will show in this lecture is that this will lead to a problem: the high-risk people will buy a lot of insurance, the low-risk people will buy only a little or even no insurance. As a result, the probability of an accident among the group who buy insurance is greater than E(p). This generic problem is referred to as adverse selection.

Let's begin by looking at the decision of person i when insurance is available at a premium rate π. Let C be the amount of coverage i buys. The outcomes are :

* if an accident occurs (probability = p_i): income = $y_o - L + C - C\pi$
* if there is no accident (probability = $(1- p_i)$): income = $y_o - C\pi$

Expected utility is

$$(1 - p_i)\, U(y_o - C\pi) + p_i\, U(y_o - L + C - C\pi)$$

If person i can choose any amount of coverage, the FONC is:

$$-\pi(1-p_i)\, U'(y_o - \pi C) + p_i(1-\pi)\, U'(y_o - L + C(1-\pi)).$$

First question: Should i buy *any* insurance?

To answer this question, look at the net marginal benefit of a unit of insurance when C = 0: this is the value of the FONC. at C = 0:

$$\text{marginal value} = p_i (1-\pi) U'(y_o - L) \quad - \pi (1-p_i) U'(y_o)$$

The first term is the marginal benefit of a \$1 more coverage when C=0. This gives you \$1 in the event of an accident, but costs π, for a net payoff of $(1-\pi)$. The probability of an accident is p_i, and the marginal utility of income is $U'(y_0 - L)$. The second term is marginal cost of coverage -- if there is no accident, you give up π, and your marginal utility is $U'(y_o)$. If a person is risk averse $U'(y_o - L) > U'(y_o)$, since U' is a *decreasing function of income*. Notice that for given π, people with higher p_i have higher net benefit. There is a "critical value" of p_i such that this net benefit is just on the margin of being positive. This critical value p^* is defined by:

$$p^* (1-\pi) \ U'(y_o - L) - \pi(1-p^*) U'(y_o) = 0 ,$$

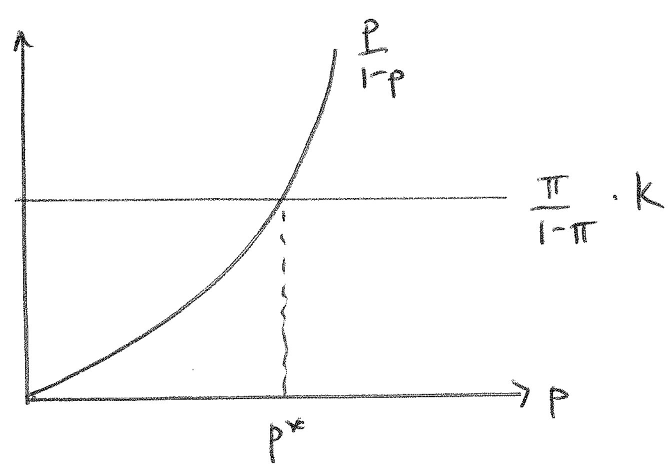
or

$$p^*/(1-p^*) = \pi/(1-\pi) \times U'(y_0) / U'(y_o - L) .$$

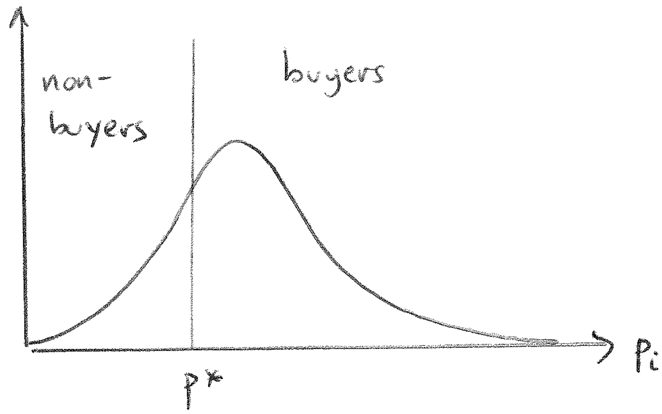
Let $U'(y_o) \div U'(y_o - L) = K \leq 1$, which depends on the degree of curvature of U. (For a linear U, K=1. For a concave U, K < 1). Then p^* is defined by

$$p^*/(1-p^*) = \pi/(1-\pi) \times K .$$

A useful graphical summary is the following picture:



For people with $p_i > p^*$, it pays to buy insurance. For people with $p_i \leq p^*$ the marginal value of the first unit of coverage is negative, so they don't buy. So if we started with some distribution of p_i's (and assumed everyone has the same K) we would have the "upper tail" of the distribution buying insurance.



Observations:

(1) In some cases, people don't know their own probability of an accident any better than the insurance company. In this case, adverse selection won't be a problem.

(2) If the insurance company sets any premium π, only people with

$$p_i/(1-p_i) > \pi/(1-\pi) \times K$$

will buy insurance. If K=1 (everyone has linear expected utility) then only those with $p_i > \pi$ will buy. So the covered pool **always** has a higher accident probability than the premium. This means that the insurance company will lose money for sure. If K varies across people, then the company will sell policies to two groups: those with high accident probabilities (who it doesn't want) and those with low K's (who it does want).

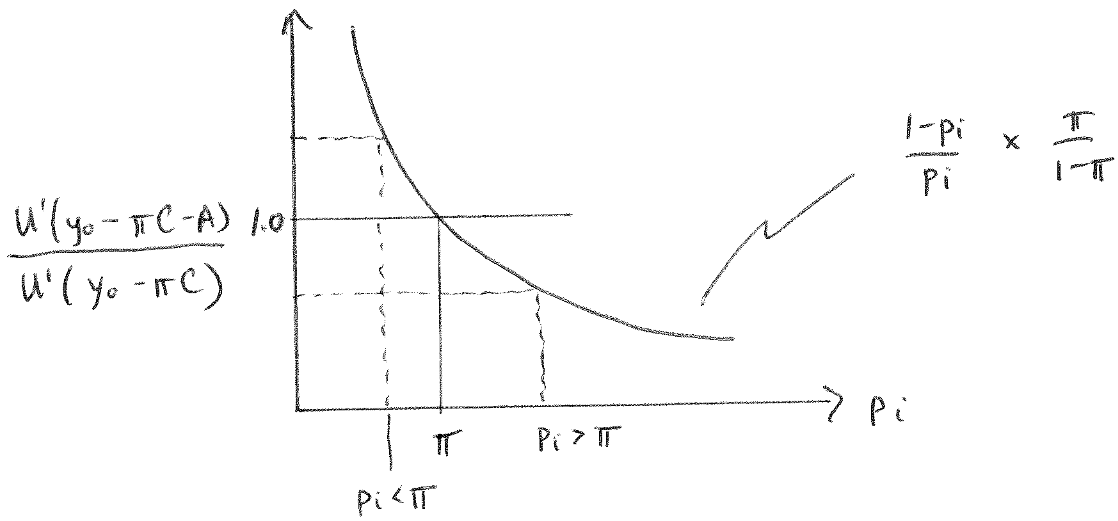
(3) We've been ignoring the fact that i may not buy as much insurance as j, even if i & j both buy some. In fact, if $p_i > p_j$, i will buy more. To see this, suppose i buys some insurance. Her optimal coverage solves

$$p_i(1-\pi)U'(y_o - L + C(1-\pi)) = \pi(1-p_i)U'(y_o - C\pi)$$

or (*) $$\frac{U'(y_o - \pi C - A)}{U'(y_o - \pi C)} = \frac{\pi(1-p_i)}{(1-\pi)p_i}$$

where A = L − C = the amount of uncovered loss. Notice that if $\pi = p_i$ (so individual i gets fair insurance), the r.h.s. of equation (*) is 1, which implies that A=0, or full insurance.

Equation (*) gives a nice picture:



We see from this that:

* for a person with $p_i < \pi$, the r.h.s. of (*) is bigger than 1 => at the optimal choice of C the uncovered loss (or deductible amount) A is greater than 0.

* for a person with $p_i > \pi$, the r.h.s. of (*) is less than 1 => at the optimal choice of C the uncovered loss (or deductible amount) A is less than 0. So these high-risk types want to over-insure! The reasoning is simple: for them, insurance is a "better-than-fair" bet. For π, you get a contract that pays you $1 in the event of an accident, which occurs with $p_i > \pi$. (By the way, this explains why insurance companies never want you to be over-insured).

Suppose there are only two kinds of people:

"low risk" $p_i = p_1$

"high risk" $p_i = p_2$

with $p_2 > p_1$. Suppose that a fraction α of people are high-risk (α = ½, e.g.). If an insurance company offers a an average premium $\pi = (1-\alpha)\, p_1 + \alpha p_2$ it will have a big problem: the low risk people will buy fewer (or even no) units of coverage, whereas the high risk people will all try to buy excess coverage. Furthermore, another company could potentially enter the market and "steal away" all the customers who buy the smallest amount of insurance -- these are the low risk guys. If the competitor succeeds (by offering a rate $\pi^c = p_1$ to anybody who used to buy just a little (or no insurance)the original company will be stuck with only the high-risk customers. But if the new competitor continues to offer coverage at premium $\pi^c = p_1$, eventually all the high-risk customers will show up at her door too! The solution to this problem can sometimes involve restricted policies that force different consumers to reveal their types by their choices.

Revealing Contracts

Suppose a company offers a policy that has a deductible (or uninsured loss) A and charges a premium π per unit of coverage. A policy package is (π, A). A consumer with $p = p_i$ values this package by looking at

$$V(\pi, A) = p_i\, U(y_o - A - \pi(L-A)) + (1-p_i)\, U(y_o - \pi(L-A))$$

(this is our usual expression, setting C = L - A).

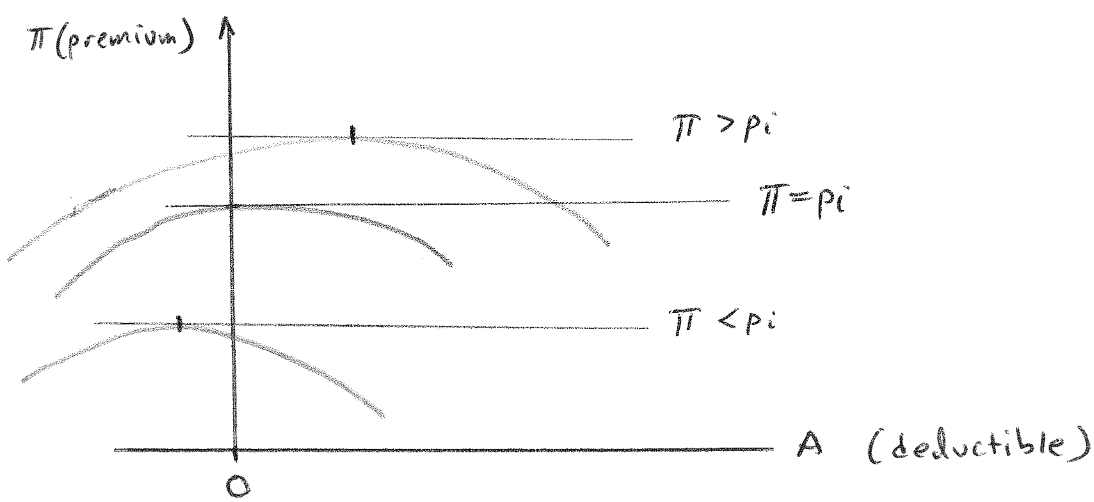
Notice that

$$V_\pi = -(L-A)\,\{\ p_i\, U'(y_o - A - \pi(L-A)) + (1-p_i)\, U'(y_o - \pi(L-A))\ \} < 0 \quad \text{if } L-A = C > 0$$

while

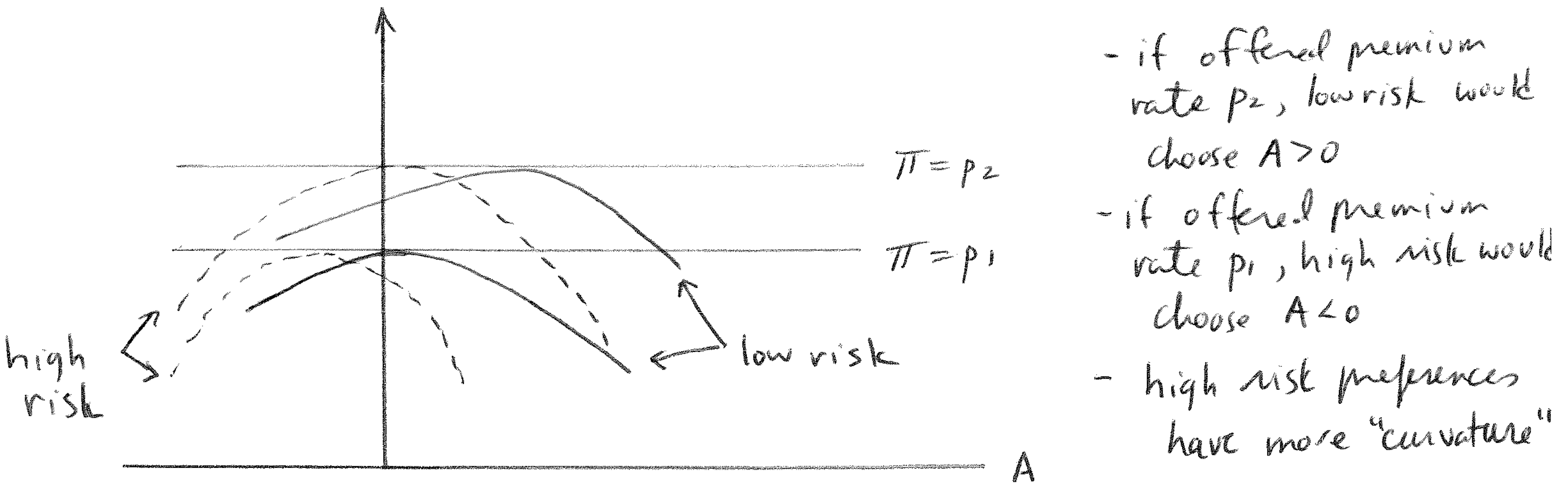
$$V_A = -\,p_i\,(1-\pi)\, U'(\text{ net income with accident }) + (1-p_i)\,\pi\, U'(\text{ net income with no accident}) \;.$$

If $p_i=\pi$ then $V_A = 0$ when A=0 (since that is the FONC for optimal coverage with fair insurance, which we know is true at A=0). If $p_i > \pi$, then $V_A = 0$ when A<0 (since that is the FONC for optimal coverage when the premium is less than your accident probability, which we know is true at A<0, the "over-insurance" case). Finally, if $p_i < \pi$, then $V_A = 0$ when A>0 (since that is the FONC for optimal coverage when the premium is greater than your accident probability, which we know is true at A>0, the "under-insurance" case). Using these facts, the indifference curves in (A, π) space look like this:



There is a way to think about what these curves are telling you. Observe first that higher utility is in the Southwest direction: a consumer always would prefer lower premium and lower deductible. Suppose now that a consumer can buy all the insurance he wants at the premium rate π. Then the consumer gets to choose any point along the horizontal line (i.e., any level of A). If π happens to be equal to p_i, we know that the optimal choice for A is A=0. So it must be true that the indifference curve is just tangent to the horizontal line at A=0. Similar arguments apply when $\pi > p_i$ and $\pi < p_i$.

Now with 2 kinds of customers there are 2 sets of indifference curves. High-risk consumers (shown with the dashed indifference curves) have accident probability p_2. Low risk consumers (shown with solid indifference curves) have an accident probability $p_1 < p_2$. Note that the high-risk types have a greater disutility of a higher deductible, at any premium rate.



An equilibrium could involve companies offering 2 kinds of policies:

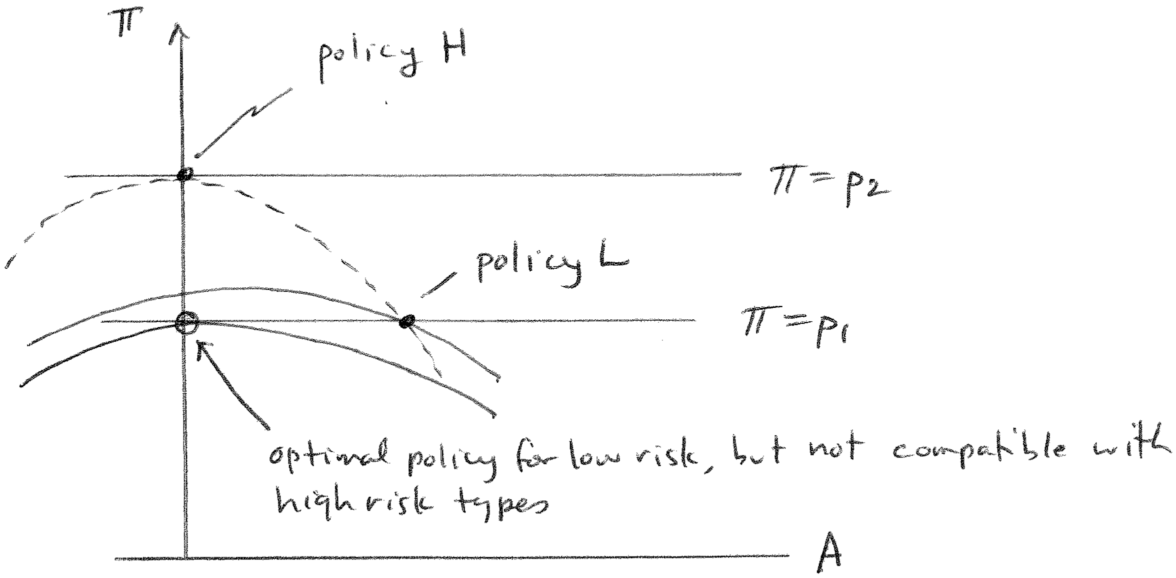
Policy H: full coverage (nondeductible) at premium $\pi = p_2$

Policy L: deductible A at premium $\pi = p_1$.

The companies have to set A big enough so the high-risk consumers choose policy H. Notice that both policies are "break even" for the insurance company.

To keep high-risk consumers from masquerading as low-risk, policy L has to lie on or outside the high risk indifference curve that passes through policy H. We say that the presence of high-risk consumers imposes an incentive compatibility constraint: the high risk types (who know who they are) have to be dissuaded from acting like low-risk consumers.

If low-risk customers didn't have to worry about the presence of high-risk types, they could get a policy L^* offering $\pi = p_1$, $A = 0$ and giving utility U^*. But this isn't incentive-compatible. The best that can be done for low-risk consumers is policy L (giving utility $< U^*$). A general lesson is the presence of the "less desirable" group (from the insurance company's point of view) forces the low risk group to signal their type by being willing to accept policy L, when policy H is available.



At policy L ① high risk consumers won't buy if H is available
② low risk prefer L to H

Economics 101A
Auctions

Introduction

Many items are sold by auction, including Treasury Bills, rights to use broadcasting "spectrum", real estate, livestock, fine art, and natural resources (e.g., timber lands, and oil fields). Large companies and governments also procedures that are equivalent to auctions to determine who will supply goods or services.

In these lectures we will overview how economists model auctions. Although auctions have existed for centuries, the basic theory is quite modern. One good reference (but somewhat advanced) is Paul Klemperer, "Auction Theory: A Guide to the Literature", in *Journal of Economic Surveys* Vol 13 (3), July 1999. This is available at http://www.nuff.ox.ac.uk/users/klemperer/Survey.pdf

I. Basic Types of Auctions

There are four basic types of auctions for a single good:

1. English auction. Also known as an "ascending bid" auctions. This is probably the one you are most familiar with. An "auctioneer" acts as a moderator, and asks for bids from a collection of bidders. If you bid $X, and no one bids higher, you win the auction, pay $X, and receive the good. Note that the auctioneer can be a computer. Ebay is essentially an English auction, though each auction has a time limit, which is not typically true of English auctions.

2. Dutch auction. Also known as a "descending bid" auctions. The auctioneer calls out a decending series of prices, starting from a price that is clearly "too high". The first bidder to call out that he/she will accept the current price wins the auction and pays that price.

3. First-price sealed bid auction. Bidders submit written bids. At a certain point the bidding is closed. The auctioneer then selects the highest bid, who is declared the winner, and pays his/her bid

4. Second-price sealed bid auction. (Also known as a "Vickery auction"). Bidders submit written bids. At a certain point the bidding is closed. The auctioneer then selects the highest bid, who is declared the winner, and pays **the second-highest bid**.

Models of how an auction works differ in their assumptions about how the value of the item at auction varies across people, and how much people know about their own, and other bidders', potential valuations. The value of the item to bidder i will be denoted v_i, i=1,2,...n.

We will focus on a couple of important cases.

A. Each v_i is independent, and known only to the bidder. This is called the "private values" case.

B. v_i is the same for all bidders, but the true value of v_i is uncertain. This is called the "common values" case. Examples might include an auction to sell the rights to drill for oil in a certain tract of land.

C. v_i varies across bidders, but the true value of v_i is uncertain (bidders do not know their own values with certainty) *and the valuations are positively correlated among bidders*. This is sometimes called the "affiliated values" case. Examples might include an auction for a house.

II. Some Important Results in the Private Values Case

1. A Dutch auction is equivalent to a first price sealed bid auction.

Why? In the Dutch auction there is no dynamic choice: you have to choose an "opt-in" price ex ante, and if the price falls to that level, opt in there and receive the good at that price. This is the *same problem* as deciding what bid to submit to a first price sealed bid auction.

We defer for a moment the choice of the optimal bidding strategy in these auctions.

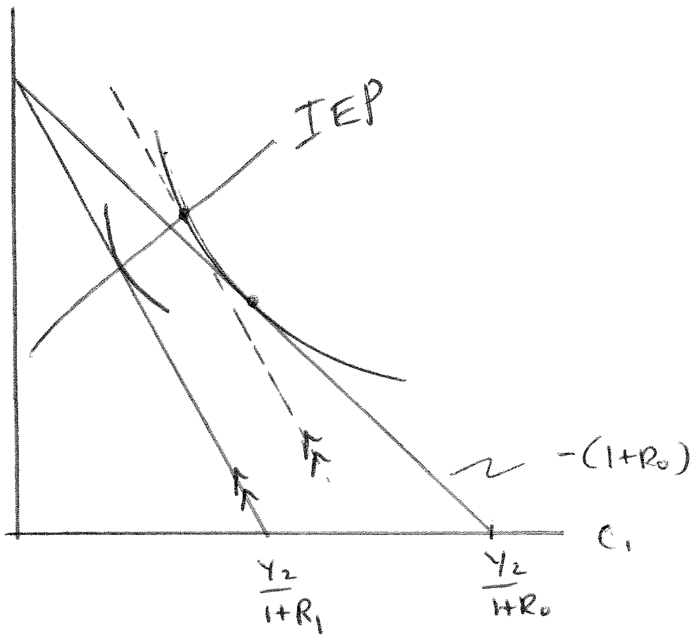
2. In an English auction the best thing to do is keep bidding until the highest bid, b, exceeds your valuation, v_i.

Why: - if $b > v_i$ you should walk away.
- if $v_i > b$, and you walk away, you leave a surplus $s = b - v_i$ "on the table"
- if you bid $b+\epsilon$, you become the highest bidder and retain the option to win s

So the dominant strategy is to stay in the bidding until $b > v_i$.

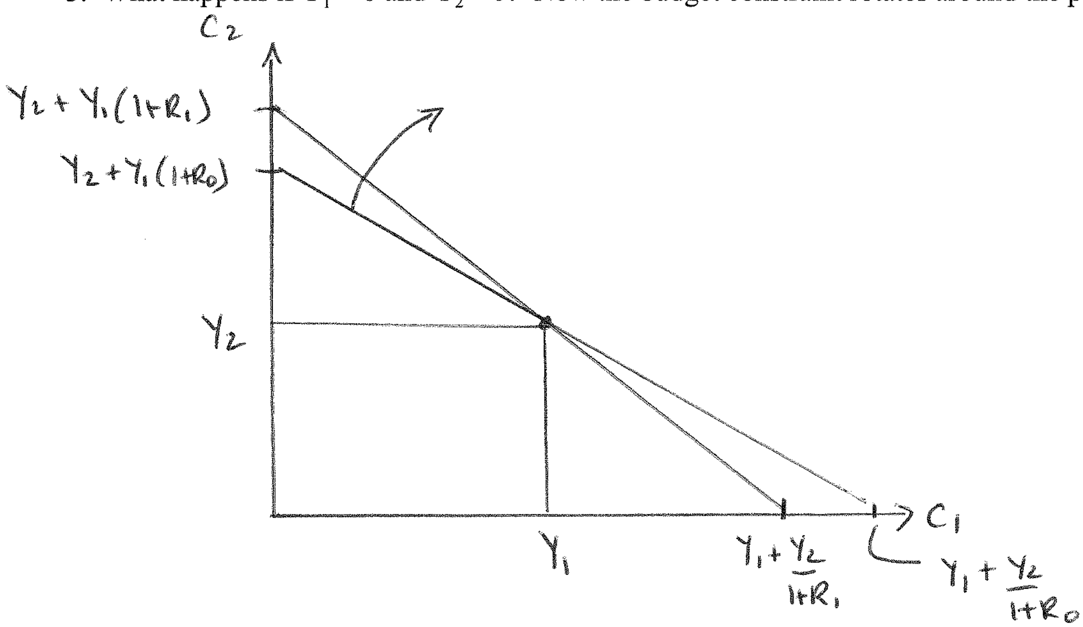
3. Based on result 2, in an English auction the bidder with the highest valuation will win, *and will pay the second highest valuation (plus ϵ needed to just pass the second highest bidder's valuation).*

(2) Consider consumers with $Y_1 = 0$. An increase in R rotates the budget line around the point $(0, Y_2)$.



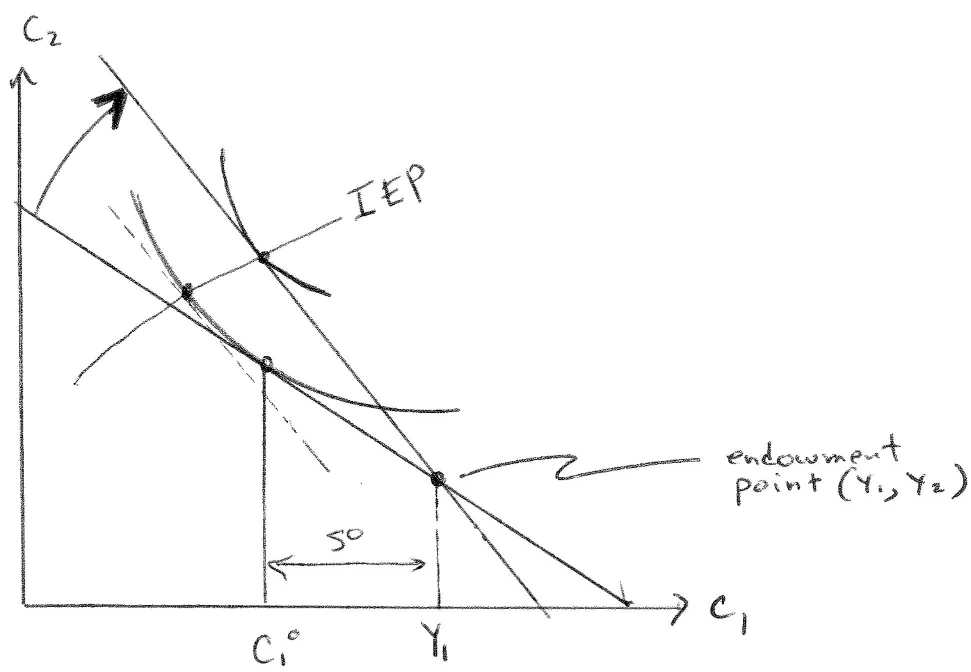
For those kinds of consumers (e.g. college students) an increase in R has an unambiguously negative effect on current consumption (they borrow less).

3. What happens if $Y_1 > 0$ and $Y_2 > 0$? Now the budget constraint rotates around the point (Y_1, Y_2).



If you were originally a <u>borrower</u> ($C_1 > Y$) there is a negative income effect. If you were originally a <u>saver</u> ($C_1 < Y_1$) their is a positive income effect.

Consider a saver:



The change in income (in second period units) is [income of new budget] - [cost of old bundle at new prices]

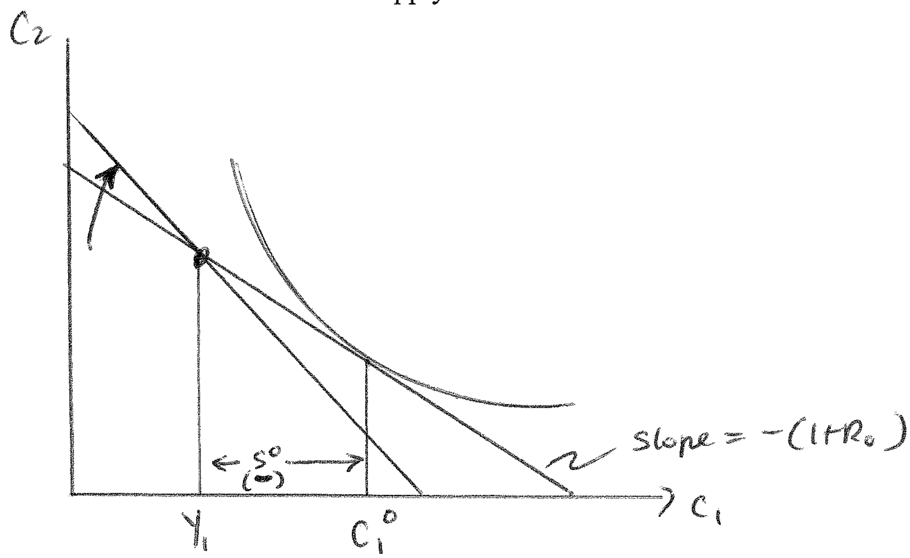
$= Y_1 (1 + R_1) + Y_2 - [C_1^0 (1 + R_1) + C_2^0]$

$= Y_1 (1 + R_0 + R_1 - R_0) + Y_2 - C_1^0 (1 + R_0 + R_1 - R_0) - C_2^0$

$= Y_1 (1 + R_0) + Y_2 - C_1^0 (1 + R_0) - C_2^0 + (Y_1 - C_1^0) (R_1 - R_0)$

$= (R_1 - R_0) (Y_1 - C_1^0) = \Delta R \cdot S^0$ where $S^0 = Y_1 - C_1^0$ = initial savings.

Note the similarity to the income effect in the case of labor supply: $\Delta I = \Delta w \times h^0$. Now consider a borrower:



The change in income is [income of new budget] - [cost of old bundle at new prices]

$= (R_1 - R_0) (Y_1 - C_1^0) = \Delta R \cdot S^0$. But for a borrower, $S^0 < 0$.

Lecture 8

PRODUCTION AND COST - I

The technology available to a given firm is summarized by its production function . This function gives the quantities of output "produced" by various combinations of inputs. For example, an airline uses labor inputs, fuel, and machinery (airplanes, loading equipment, etc.) to produce the output "passenger seats." We write $y = f(a, b)$ to signify that with inputs a and b, the output level y is available.

Examples: 1) (one input) $y = a$

$$y = a^{\gamma} \quad (0 < \gamma < 1)$$

$$y = \begin{cases} 0 & a < \bar{a} \\ 1 & a > \bar{a} \end{cases}$$

2) (two inputs) $y = a^{\alpha} b^{\beta}$ (Cobb-Douglas)

$y = \min [a,b]$ (Leontief)

$y = a + b$ (additive)

For two (or more) inputs, production functions are a lot like utility functions. The important difference is that output is measurable and has natural units (e.g., "seats"). It's as if the "indifference curves" have numbers attached to them that matter.

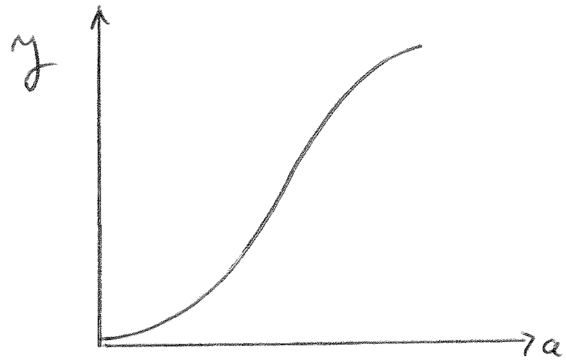
A second, less obvious, way to summarize technology is to compute the cost associated with producing a given output level y , at fixed prices for the inputs. In principle, it is easy to find the cost function if you know the production function using two steps:

1) find all possible ways of producing y .
2) find the cheapest one, and evaluate its cost.

Most of the economic behavior of firms is studied via the cost function. In the following lectures, we show how to derive cost functions and how to relate the properties of the cost function to the properties of the production function.

A. One-Factor Production and Cost Functions

Suppose that there is only one input (apart from, perhaps) a "set-up cost". Then we have a picture along the following lines:



y = f(a).

Note that f(0) = 0, by convention.

Definitions and Facts

(i) The marginal product of factor a is the increase in output for a unit increase in input a.

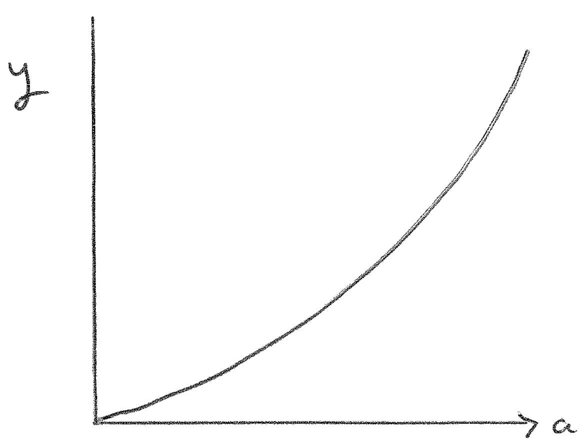
$$MP_a = \frac{\partial f(a)}{\partial a} = f'(a)$$

If factor a is useful, f'(a) > 0.

(ii) The average product of a factor a is the ratio of total output to total usage of the factor:

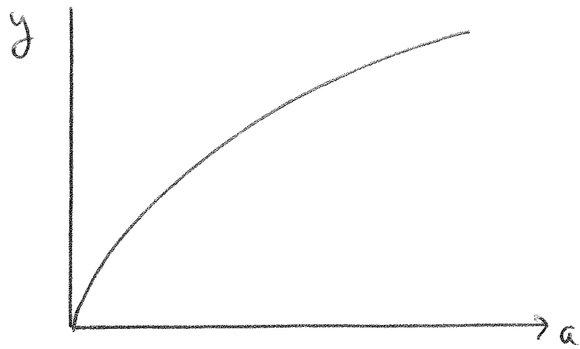
$$AP_a = \frac{f(a)}{a}.$$

(iii) If the MP of factor a is increasing, then f"(a) > 0 and we say that there are "increasing marginal returns": as the scale of output is expanded, each additional unit of input contributes more.



f"(a) > 0 (increasing returns)

If the MP is decreasing, the f"(a) < 0 and we say there are diminishing marginal returns.

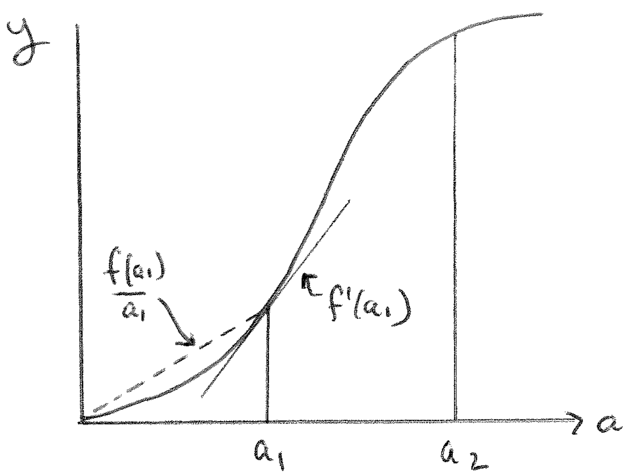


f"(a) < 0 (decreasing returns)

(iv) If $MP_a > AP_a$, then AP_a is increasing.

If $MP_a < AP_a$, then AP_a is decreasing.

Think of a baseball hitter: AP = lifetime average; MP = season average. A hitter who is having a season with an average that is better than his average so far will raise his lifetime average. Graphically:



slope of tangent = MP

slope of chord to origin = A

At $a = a_1$, $AP = \dfrac{f(a_1)}{a_1} < f'(a_1)$, AP is increasing.

At $a = a_2$, $AP = \dfrac{f(a_2)}{a_2} > f'(a_2)$, AP is increasing.

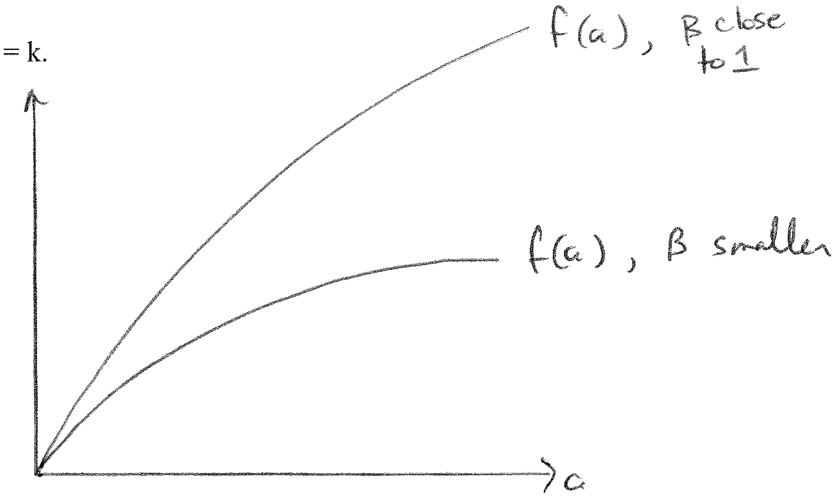
In general:

$$AP = \frac{f(a)}{a} \quad ; \quad \frac{dAP(a)}{da} = \frac{af'(a) - f(a)}{a^2} = \frac{1}{a}\left[f'(a) - \frac{f(a)}{a} \right] .$$

there is no page 52

Examples:

- $f(a) = k \times a$ with $k > 0$ (linear). $AP_a = MP_a = k$.
- $f(a) = a^\beta$ with $0 < \beta < 1$ (concave).



- $f(a) = 9a^2 - a^3$ for $a < 6$.

  For this function:

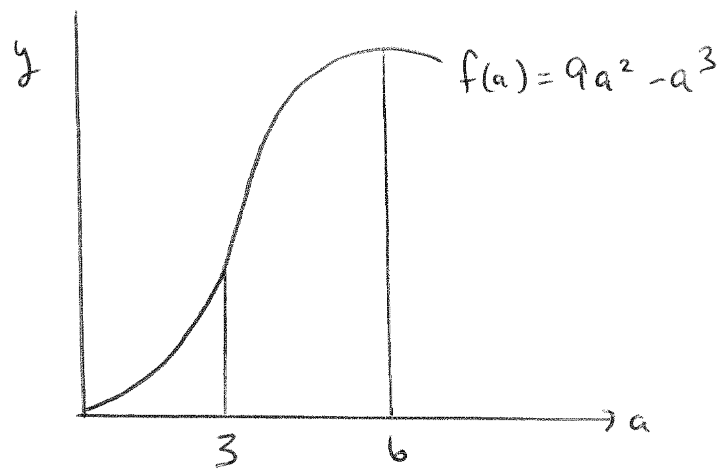
  $f'(a) = 18a - 3a^2$

  $f'(a) \geq 0 \Leftrightarrow a \leq 6$

  $f''(a) = 18 - 6a$

  $f''(a) > 0$ for $a < 3$

  & $f''(a) < 0$ for $a > 3$



COST

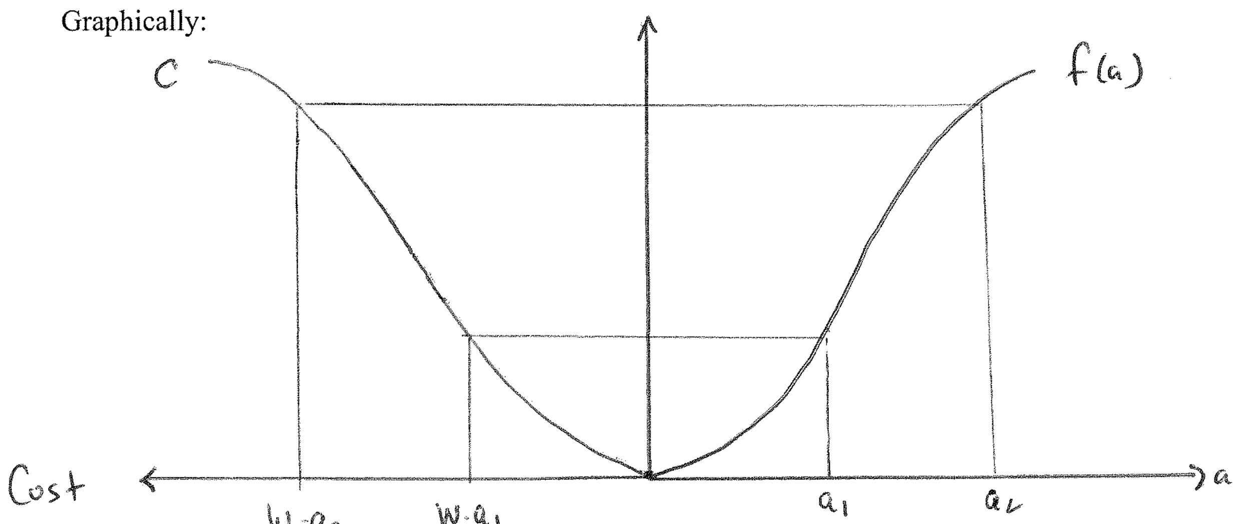
What is the cost function for a one-factor production function?

Let w = price per unit of a. Then:

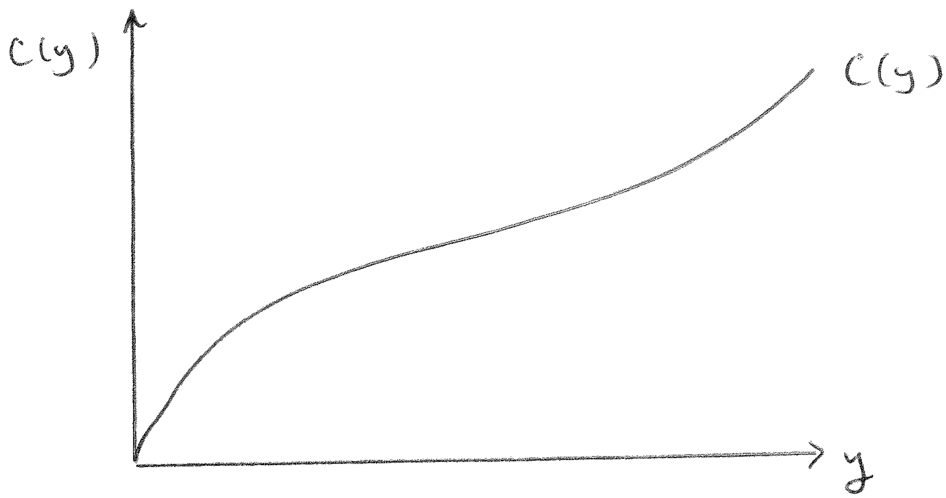
$$C(y,w) = \min w\, a \qquad \text{s.t. } y = f(a)$$

For $y = f(a)$, we need $a = f^{-1}(y)$. Therefore $C(y,w) = w\, f^{-1}(y)$.

Graphically:



Rotate the left hand panel of this graph 90°



If w is fixed we often write the cost function as a function only of output, C(y) . In the more general case we write C(y,w). Define marginal cost MC(y) = C'(y) , and average cost AC(y) = C(y)/y.

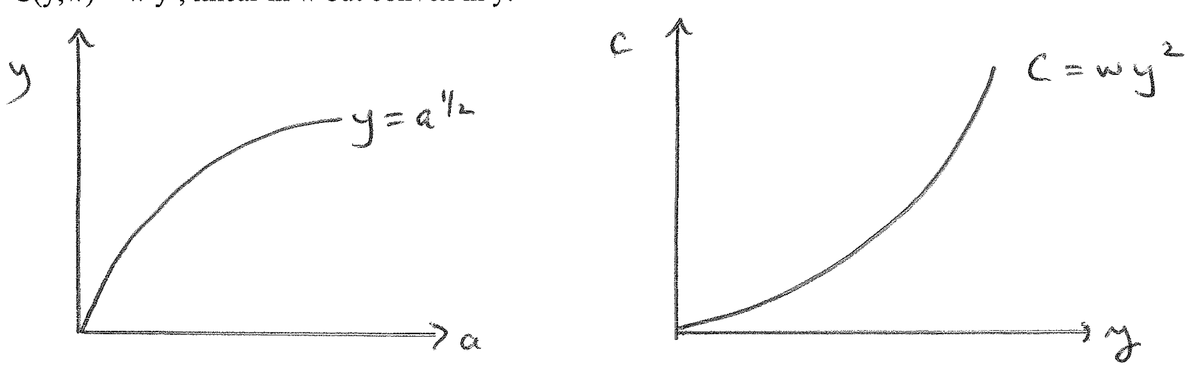
Examples:

1) $y = f(a) = 2a$ (linear) => $a = f^{-1}(y) = ½y$ (Linear "input requirement function")

$C(y,w) = w(½y) = ½\, w\, y$, linear in both y and w.

2) $y = f(a) = a^{½}$ => $a = f^{-1}(y) = y^2$ (Convex input requirement function).

$C(y,w) = w\, y^2$, linear in w but convex in y.



Relations between MC and MP

The MC of output is the amount it costs to produce an extra unit.

By definition of MP_a one unit of input adds $MP_a = f'(a)$ units to output.

=> $1/MP_a = 1/f'(a)$ units of a are needed to produce 1 unit of y

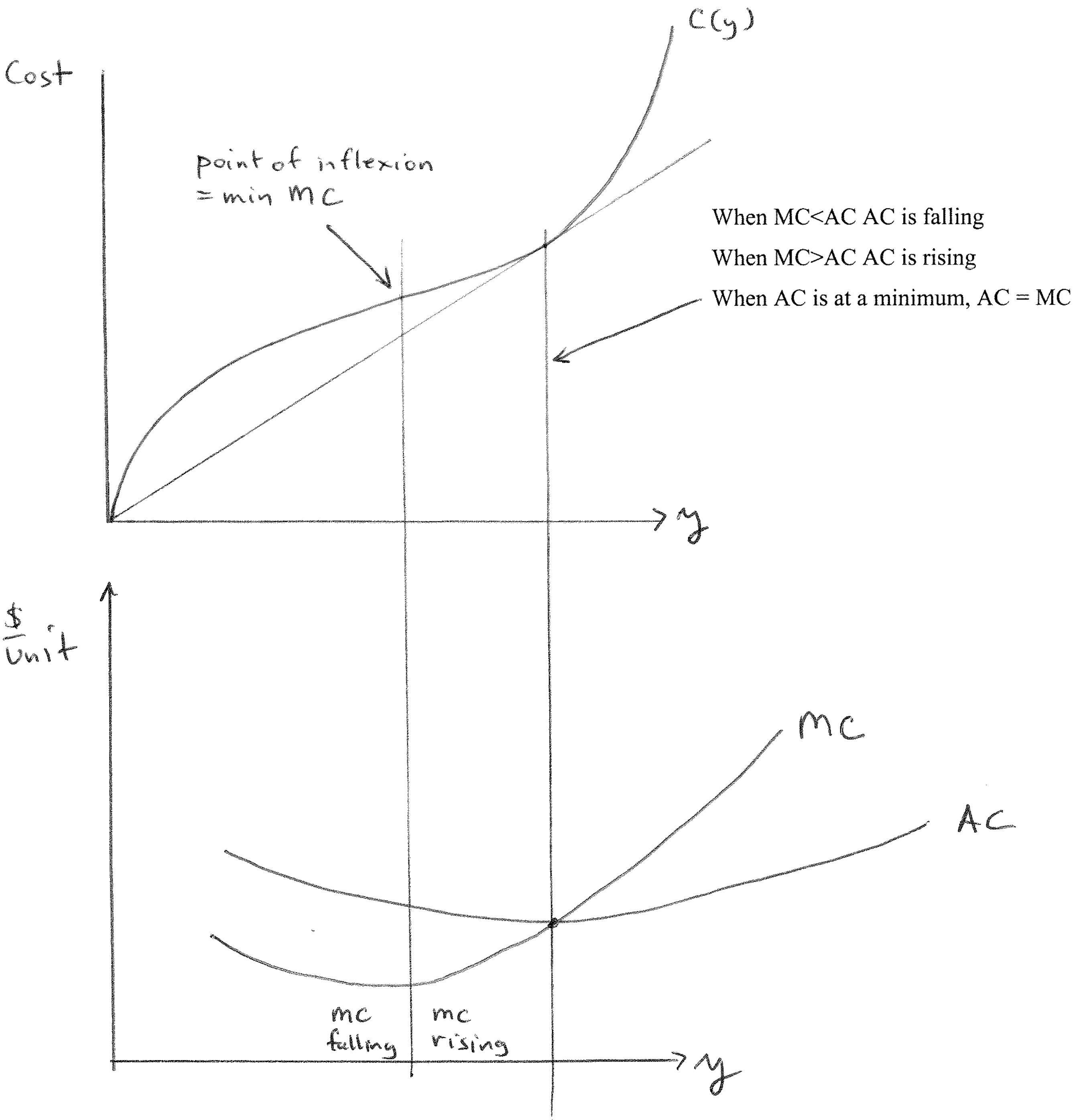
=> the marginal the cost of an extra unit is $MC(y) = w/f'(a)$, when the production function is $y = f(a)$.

Alternatively, $C(y) = wf^{-1}(y)$ using the input requirement $a = f^{-1}(y)$. Thus

$$C'(y) = w\,\frac{df^{-1}(y)}{dy} = \frac{w}{f'(a)} ,$$

using the fact that the derivative of an inverse function is the inverse of the derivative.

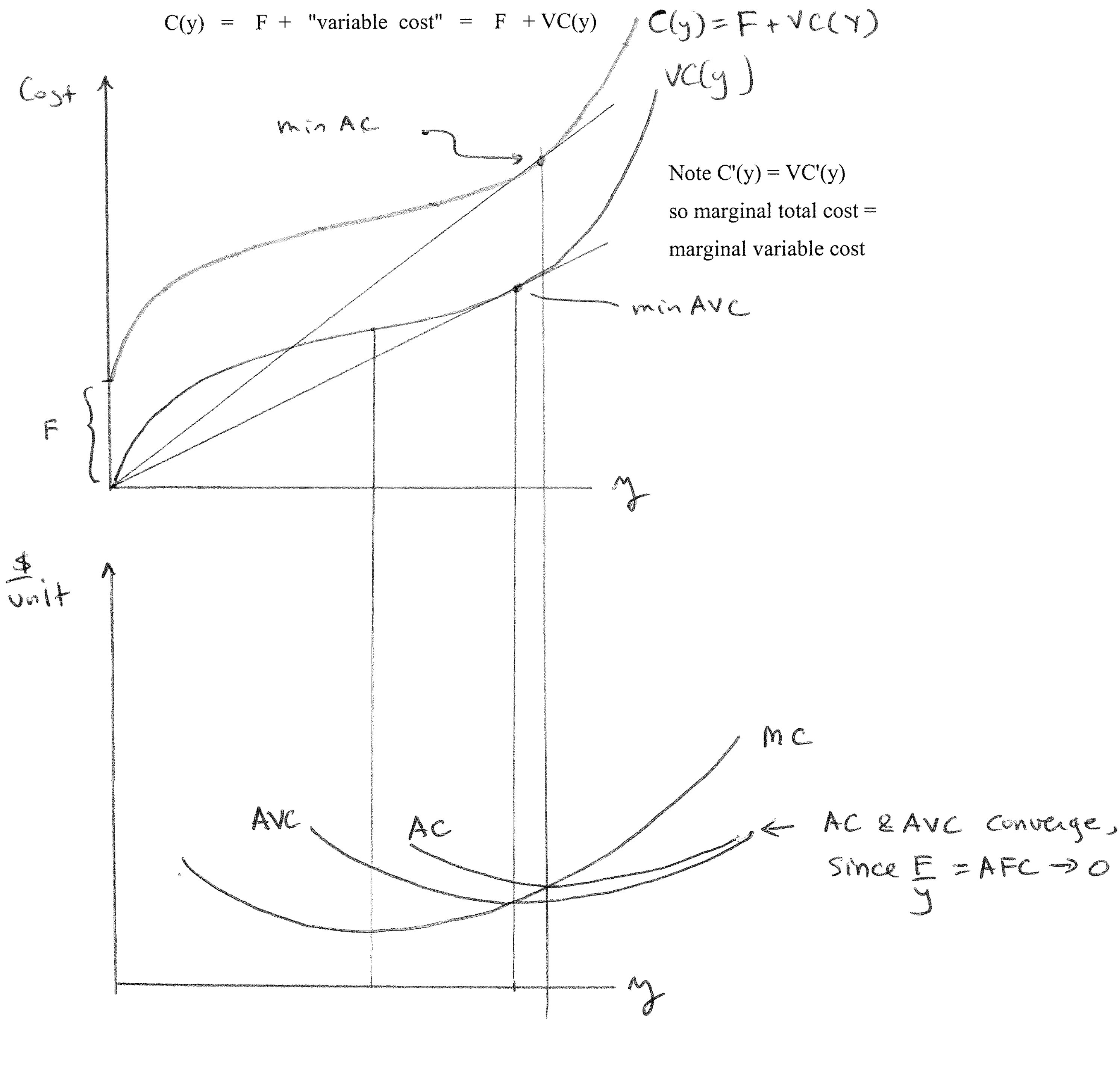
Geometry of C(y), AC(y) and MC(y)



Sometimes we add on a "set up" cost F (also called a fixed cost)

Then total cost is

$$C(y) = F + \text{"variable cost"} = F + VC(y)$$



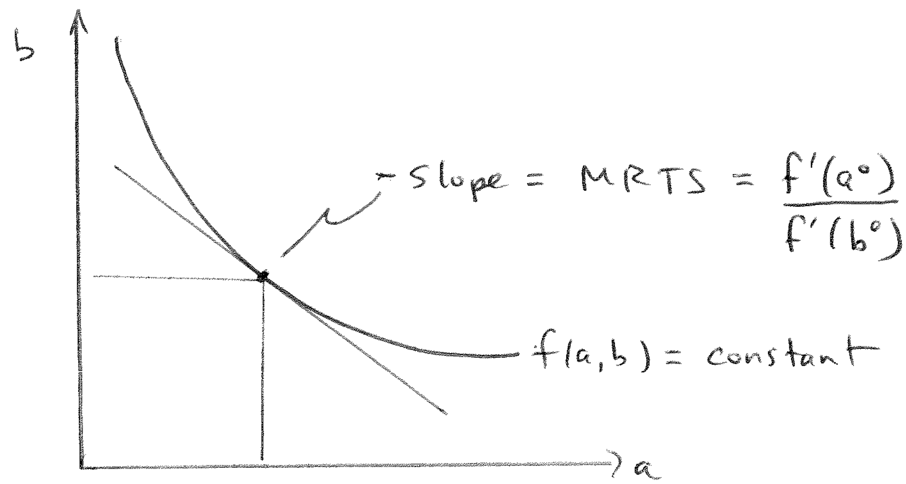
NOTES: 1. min AC occurs to the right of min AC. Why?

2. MC cuts through min of both AC and AVC. Why?

Lecture 9

Production and Cost II. Two-Factor Production and Cost Functions

The analysis of production and cost is more interesting when we can use combinations of two (or more) inputs to produce y. The production function is y = f(a,b). As in consumer theory, we begin by thinking about combinations of inputs that produce the same output. In the firm case these are called isoquants.



We define the marginal rate of technical substitution (MRTS) as the slope of an isoquant. It tells you how many units of b you need to add, per unit of a given up, to keep output constant.

Formally, suppose $y^0 = f(a^0,b^0)$, and consider changing a and b to keep output fixed at y^0,

$$dy = f_a\, da + f_b\, db = 0$$

$$\Rightarrow \left.\frac{db}{da}\right|_{y_0} = -\frac{f_a(a,b)}{f_b(a,b)} = -\frac{MP_a}{MP_b}$$

The MRTS is analogous to the marginal role of substitution (MRS) in consumer theory. When there are 2 or more inputs, the production function is characterized by both the degree of substitutability between inputs (the curvature of isoquants) and the extent to which output expands as inputs are expanded proportionately. The latter gives rise to the idea of returns to scale. For a production function y = f(a,b), we say f has *constant returns to scale* (CRS) if

$$f(\gamma a, \gamma b) = \gamma f(a,b) \quad \text{for any } \gamma > 0.$$

If y = f(a,b) we say f has *decreasing returns to scale* (DRS) if

$$f(\gamma a, \gamma b) < \gamma f(a,b) \quad \text{for any } \gamma > 1.$$

So if you double both inputs, you get less than a doubling of output. On the other hand, the same inequality implies that if you reduce inputs by some proportion, your output falls by a smaller proportion. So DRS says that smaller firms are always more "efficient". Conversely we say that f has *increasing returns to scale* (IRS) if

$$f(\gamma a, \gamma b) > \gamma f(a,b) \quad \text{for any } \gamma > 1.$$

Examples:

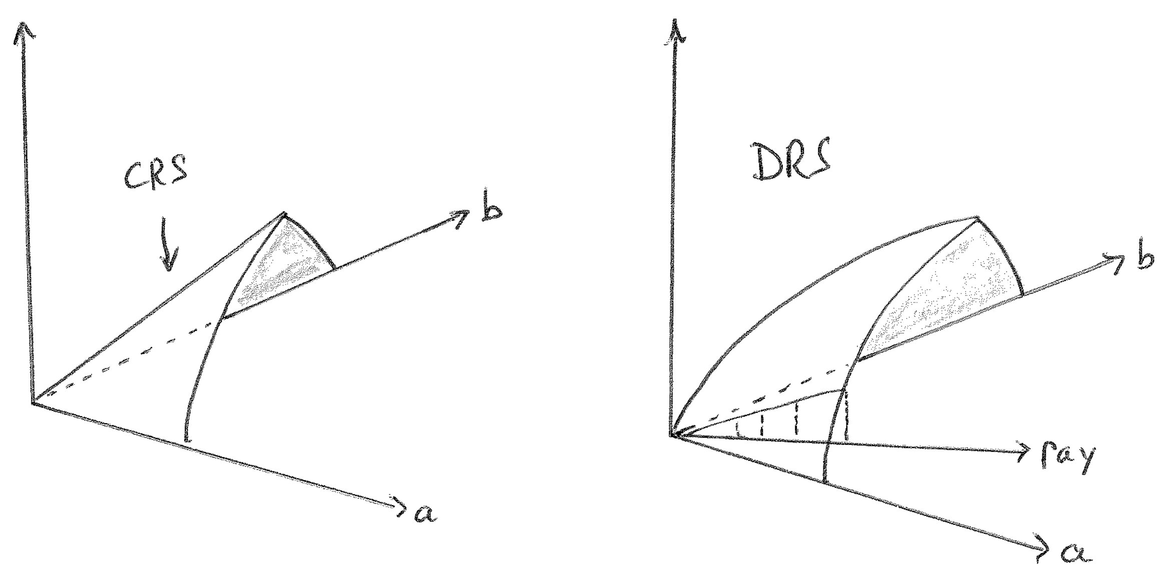
(i) one input $f(a) = a^{\theta}$ CRS if $\Theta = 1$
DRS if $\Theta < 1$
IRS if $\Theta > 1$.

(ii) Cobb Douglas $f(a,b) = a^{\alpha} b^{B}$. CRS if $\alpha + \beta = 1$.
DRS if $\alpha + \beta < 1$.
IRS if $\alpha + \beta > 1$.

Check:

$$\begin{aligned} f(\gamma a, \gamma b) &= (\gamma a)^{\alpha} (\gamma b)^{\beta} \\ &= \gamma^{\alpha+\beta} a^{\alpha} b^{\beta} \\ &= \gamma^{\alpha+\beta} f(a,b) \end{aligned}$$

Geometrically, returns to scale tells you whether the function f is concave or convex "over top" of a ray from the origin:



Derivation of the Cost Function

Given a production function f(a,b) and prices w_a, w_b, we can write

$$C(w_a, w_b, y) = \min\ w_a\, a + w_b\, b \quad \text{subject to } f(a,b) \geq y.$$

Set up the Lagrangean (using μ for the multiplier):

$$L = w_a\, a + w_b\, b - \mu\, (f(a,b) - y).$$

$$\frac{\partial L}{\partial a} = w_a - \mu\, f_a(a,b) = 0$$

$$\frac{\partial L}{\partial b} = w_b - \mu\, f_b(a,b) = 0$$

$$\frac{\partial L}{\partial \mu} = -f(a,b) + y = 0$$

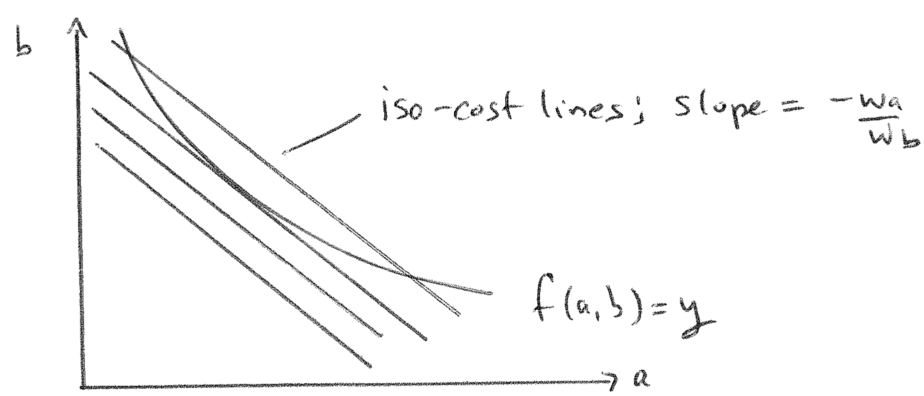
The ratio of the first 2 first-order conditions gives

$$\frac{Wa}{Wb} = \frac{fa(a,b)}{fb(a,b)} = \text{MRTS}.$$

Geometrically, we find the tangencies of "iso-cost" lines

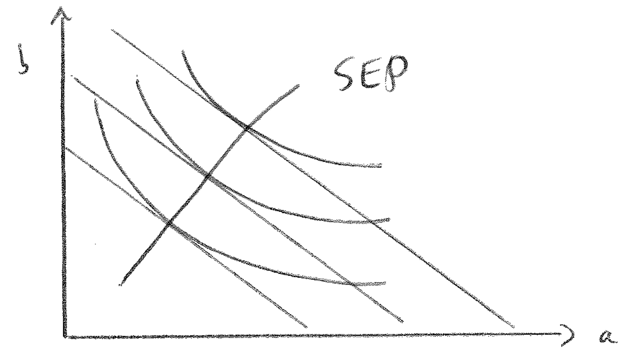
$$w_a\, a + w_b\, b = \text{constant}$$

with the isoquant corresponding to the desired level of output y:

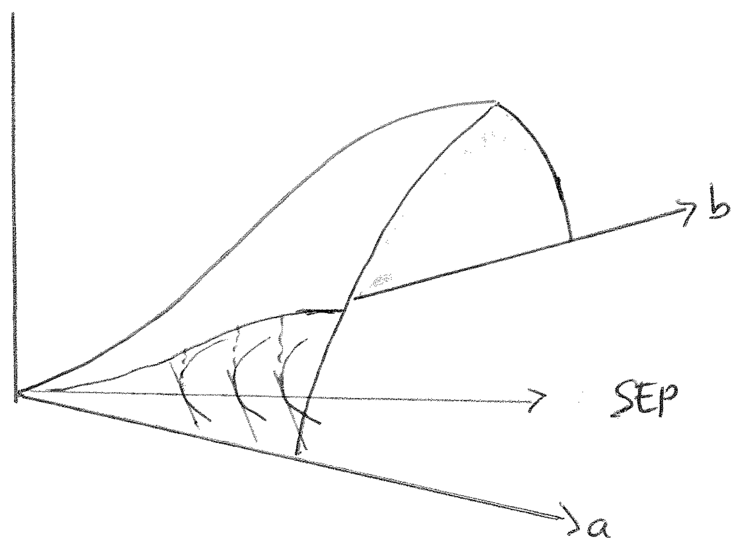


Notice that the problem is "reversed" relative to a consumer. In the cost problem, you are constrained to an isoquant and have to find the lowest "budget line" (iso-cost line). In the consumer problem, you are constrained to a budget line and have to find the highest "isoquant" (indifference curve).

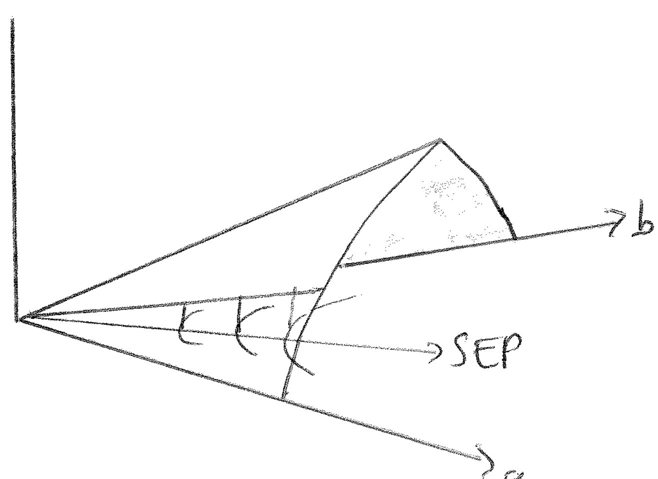
If we consider finding the cheapest way to get different levels of output at <u>constant</u> (w_a, w_b) we trace out the scale expansion path (SEP)



Note the similarity between a firm's SEP and a consumer's income expansion path (IEP). Geometrically, the shape of the cost function (as a function of y) depends on the shape of the production function "over top" of the SEP:



If the curve over the SEP is S-shaped (as illustrated) we get cost functions of the usual shape. If the curve is linear we get a <u>linear</u> cost function:



Marginal Cost

If we needed to produce 1 more unit of y, we could use input a, or input b (or both). If we use only a, we need $1/MP_a$ units of a for 1 unit of y. The marginal cost is w_a/MP_a (just as in the 1-input case). By symmetry, we could also use input b, at marginal cost w_b/MP_b. But from the first-order conditions

$$\frac{W_a}{W_b} = \frac{MP_a}{MP_b} \quad => \quad \frac{W_a}{MP_a} = \frac{W_b}{MP_b} \ .$$

So on the margin, you are indifferent to expanding output via increases in a or increases in b. This reflects that fact that a and b were "optimally chosen" to begin with. Note also that

$$\mu = \frac{w_a}{f_a(a,b)} = \frac{w_a}{MP_a} = \frac{W_b}{MP_b} .$$

Thus the Lagrange multiplier in the cost-minimization problem gives marginal cost.

Examples:

(i) $f(a,b) = \min[a,b/2]$. At a cost minimum we must have $a=b/2=y$.

$=> C(w_a,w_b, y) = y\ (w_a + 2w_b)$. Note this production function has CRS.

(ii) $f(a,b) = a + 2b$. These are linear isoquants, with $f_a/f_b = 1/2$.

If $w_a/w_b> 1/2$, should use only b. Then $y = 2b => b = y/2$, and $C(w_a, w_b,y) = ½w_b\ y$.

But, if $w_a/w_b < 1/2$, should use only a. Then $y = a$, and $C(w_a,w_b,y) = w_a\ y$.

Combining these results, for any w_a, w_b, we have $C(w_a, w_b, y) = \min[w_a, w_b/2]\ y$.

These two examples illustrate the what is called the "duality" relationship between cost and production functions. Leontief production functions imply linear cost functions. Linear cost functions imply "Leontief-like" cost functions.

(iii) $f(a,b) = a^{\alpha} b^{\beta}$. (This was in problem set number 4).

The Lagrangean is

$$L(a,b,\mu) = w_a\,a + w_b\,b - \mu(a^{\alpha}b^{\beta} - y).$$

$$\frac{\partial L}{\partial a} = w_a - \mu\alpha^{\alpha-1}b^{\beta} = 0$$

$$\frac{\partial L}{\partial b} = w_b - \mu\beta\alpha^{\alpha}b^{\beta-1} = 0$$

$$\frac{\partial L}{\partial \mu} = -\alpha^{\alpha}b^{\beta} + y = 0$$

Using the first two FOC's, we get:

$$\frac{w_a}{w_b} = \frac{\alpha a^{\alpha-1} b^{\beta}}{\beta a^{\alpha} b^{\beta-1}} = \frac{\alpha b}{\beta a}$$

$$\text{or} \quad b = \frac{\beta a w_a}{\alpha w_b}$$

Substitute into the constraint: $a^{\alpha} b^{\beta} = a^{\alpha} \left[\frac{\beta a w_a}{\alpha w_b}\right]^{\beta}$

$$= a^{\alpha+\beta} \beta^{\beta} w_a^{\beta} \alpha^{-\beta} w_b^{-\beta} = y$$

This gives the input requirement function for input a:

$$a = y^{\frac{1}{\alpha+\beta}} \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} w_a^{-\frac{\beta}{\alpha+\beta}} w_b^{\frac{\beta}{\alpha+\beta}}$$

and substituting back in (or using symmetry) we get $b = y^{\frac{1}{\alpha+\beta}} \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{-\frac{\alpha}{\alpha+\beta}}$

Finally $C(w_a, w_b, y) = w_a a + w_b b$ when a and b are set to the cost-minimizing input choices, so

$$C(w_a, w_b, y) = y^{\frac{1}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{\frac{\beta}{\alpha+\beta}} \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + y^{\frac{1}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{\frac{\beta}{\alpha+\beta}} \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}}$$

$$= y^{\frac{1}{\alpha+\beta}} w_a^{\frac{\alpha}{\alpha+\beta}} w_b^{\frac{\alpha}{\alpha+\beta}} \left\{ \left(\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} + \left(\frac{\beta}{\alpha}\right)^{\frac{\alpha}{\alpha+\beta}} \right\}$$

If $\alpha + \beta = 1$ (constant returns to scale) this simplifies to

$$C(w_a, w_b, y) = y\, w_a^{\alpha}\, w_b^{\beta} \left\{ \left(\frac{\alpha}{\beta}\right)^{\beta} + \left(\frac{\beta}{\alpha}\right)^{\alpha} \right\}$$

$$= y\, w_a^{\alpha}\, w_b^{\beta} \left(\alpha^{-\alpha} \beta^{-\beta}\right)$$

So with CRS, cost is linear in output. In general the exponent on y in the cost function is $1/(\alpha+\beta)$, so if $\alpha + \beta > 1$, cost is concave in output (IRS) and if $\alpha + \beta < 1$, cost is convex in output (DRS).

Economics 101A
Handout

Cost Functions and Input Requirement Functions

Suppose we have a production function $f(x_1, x_2)$, and an associated cost function $C(y, w_1, w_2)$. We obtain the cost function by solving the "cost minimization problem":

$$\text{Min} \quad w_1 x_1 + w_2 x_2 \quad \text{s.t.} \quad f(x_1, x_2) = y .$$

We set up the Lagrangian:

$$L(x_1, x_2, \mu) = w_1 x_1 + w_2 x_2 - \mu (f(x_1, x_2) - y) .$$

The f.o.c.'s are:

$$L_1 = w_1 - \mu f_1(x_1, x_2) = 0$$

$$L_2 = w_2 - \mu f_2(x_1, x_2) = 0$$

$$L_3 = - f(x_1, x_2) + y = 0.$$

The first two of these imply the "tangency condition" $w_1/w_2 = f_1/f_2$, while the third just gives back the constraint. Solving these two equations in two unknowns we get the input requirement functions (IRF's):

$$x_1 = x_1^*(y, w_1, w_2)$$

$$x_2 = x_2^*(y, w_1, w_2) .$$

The IRF's are analogous the consumer's demand functions: they represent the optimal (cost-minimizing) input choices to produce y when input prices are (w_1, w_2). With these we can obtain the cost function

(*) $\quad C(y, w_1, w_2) = w_1 x_1^*(y, w_1, w_2) + w_2 x_2^*(y, w_1, w_2)$,

which is just the cost of the "cost minimizing" input combination.

Sheppard's Lemma

It turns out that if you know C, you can "recover" the IRF's by simple differentiation:

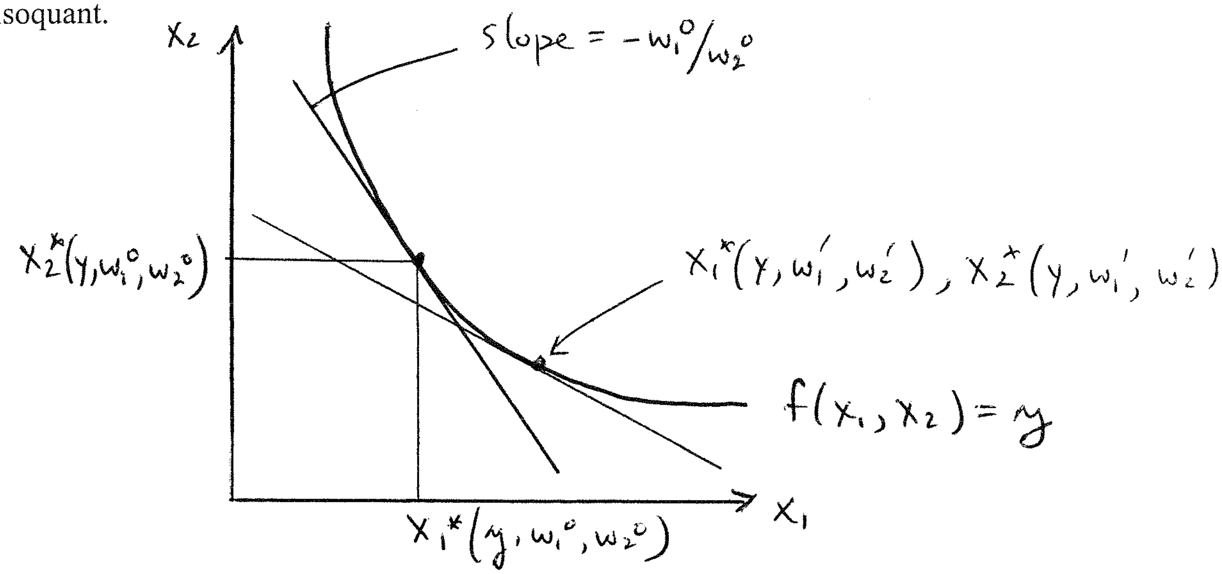
$$x_1^*(y, w_1, w_2) = \partial C(y, w_1, w_2) / \partial w_1 .$$

At first glance, it looks like this is inconsistent with (*). Indeed, differentiating (*) with respect to w_1, we get **three terms**:

(□) $\partial C(y, w_1,w_2)/ \partial w_1 = \quad x_1^*(y, w_1, w_2) \; + \; w_1 \,\partial x_1^*(y, w_1, w_2)/\partial w_1 \; + \; w_2 \,\partial x_2^*(y, w_1, w_2)/\partial w_1$.

However, when we change an input price, $x_1^*(y, w_1, w_2)$ and $x_2^*(y, w_1, w_2)$ have to move in such a way as to keep total output constant. In fact, as w_1 changes, the combinations of (x_1^*, x_2^*)

trace out the isoquant.



In other words, we have that

$$f\,(\,x_1^*(y, w_1, w_2)\,,\; x_2^*(y, w_1, w_2)\,) \;=\; y\,.$$

This has to hold as w_1 varies, so we can differentiate w.r.t. w_1 (applying the chain rule) to get:

$$f_1\, \partial x_1^*(y, w_1, w_2)/\partial w_1 \;+\; f_2\, \partial x_2^*(y, w_1, w_2)/\partial w_1 = 0\,.$$

This says that

(†) $\quad \partial x_2^*(y, w_1, w_2)/\partial w_1 = -\, f_1 / f_2\;\; \partial x_1^*(y, w_1, w_2)/\partial w_1$.

So, since x_1^* falls as w_1 rises, x_2^* has to rise, and the rates of change are in the ratio of f_1/f_2. (Note that the response of x_1^* to a change in w_1 is just like a "substitution effect" for a consumer. Since the isoquant has diminishing MRTS, a rise in w_1 must lead to a fall in x_1^*). Substituting (†) into (□), we get

$$\partial C(y, w_1, w_2)/\,\partial w_1 = \; x_1^*(y, w_1, w_2) + \;\; \partial x_1^*(y, w_1, w_2)/\partial w_1\; \{\, w_1 \;-\; w_2\;\; f_1 / f_2 \,\}\,.$$

But from the tangency condition, $w_1 - w_2\;\; f_1 / f_2 = 0$. So the second and third terms in (□) always sum to 0, leaving us with (*).

Equation (*) says that if w_1 rises, the first-order effect on cost is just proportional to the amount of input 1 that you were using to produce at minimum cost. Although the optimal choices of x_1 and x_2 also change, they have to change in such as way as to keep y constant, and because of the initial tangency condition the movements in the inputs leave cost constant.

Lecture 10

OUTPUT (SUPPLY) DETERMINATION

So far, we have studied cost, taking output as given. In this lecture, we consider the output or supply decision of individual competitive firms. By "competitive," we mean that the firm takes the prices for inputs and outputs as exogenous (i.e., beyond the firm's control). For any firm, profit is defined as the difference between revenue and cost. For a competitive firm that uses 2 inputs, 1 and 2, to produce a single output (y) which is sold at a price p, profit as a function of y is:

$$\pi(y) = p\,y - C(y, w_1, w_2)\ .$$

Note that revenue ($p \times y$) is a linear function of output, whereas the cost function is potentially non-linear. We assume that the firm chooses y to maximize profit:

$$\max\ p\,y - C(y, w_1, w_2)$$

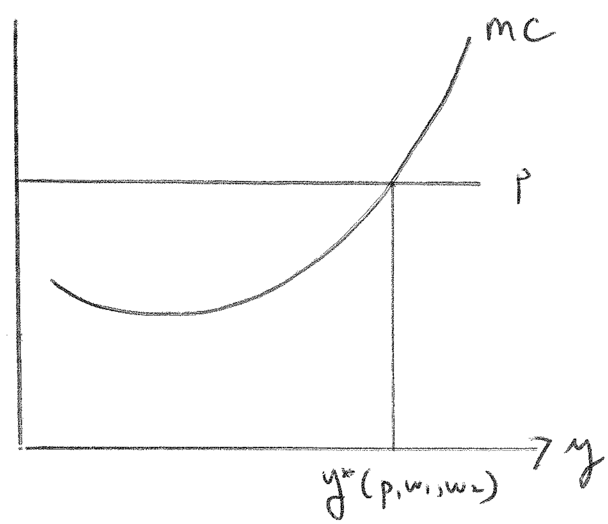
FONC:

$$\frac{d\Pi}{dy} = p - C_y(y^*, w_1, w_2) = 0, \quad \underline{\text{or}} \quad \text{price} = \text{marginal cost, when } y = y^*.$$

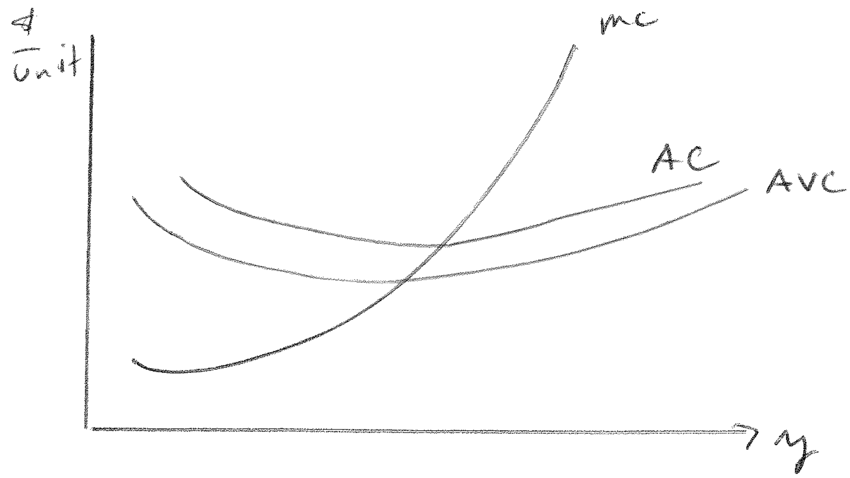
The second order condition for maximum profit is

$$\frac{d^2\Pi}{dy^2} < 0 \quad \Rightarrow \quad -C_{yy}(y^*, w_1, w_2) < 0 \quad \Rightarrow \quad C_{yy}(y^*, w_1, w_2) > 0 : \text{increasing MC at optimal y.}$$

Here is the diagram:



Note that y^* is a function of p , and all input prices. We write $y = y^*(p, w_1, w_2)$ as the "supply function". What happens if $\pi < 0$ at $y = y^*(p, w)$?

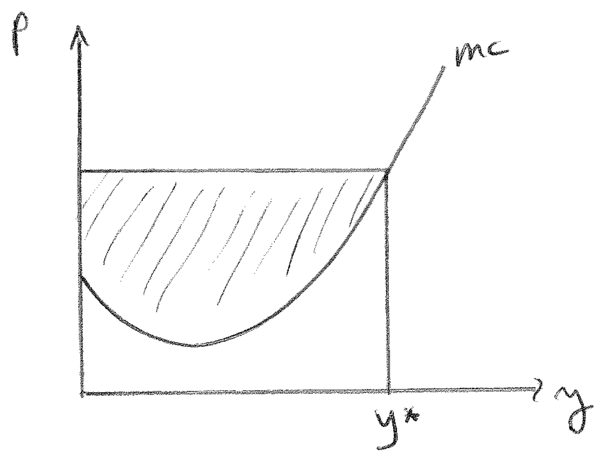


1) If $p < AVC$ then $y^* = 0$. The firm is losing on both fixed an variable inputs: the best supply action is to shut down.

2) If $p > AC$ the firm is earning profit, so y^* is defined by $p = MC(y^*)$.

3) If $AVC < p < AC$, the firm is turning a loss, but it is covering its "operating" costs, and only failing to pay off its fixed costs. It may well stay in business and hope for better times.

The following diagram is a useful representation of the firm's optimal choice.



The rectangle py^* represents revenue.

The area under MC represents costs (<u>not</u> including fixed costs).

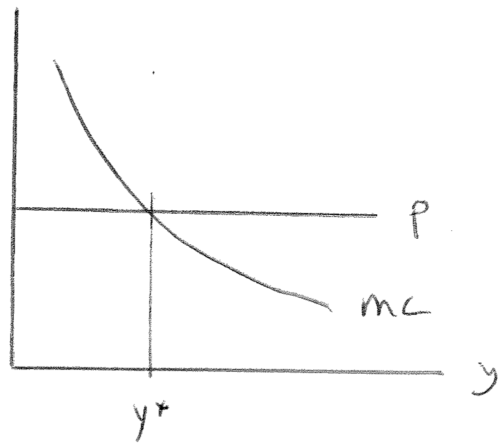
Thus the <u>shaded</u> <u>area</u> represents profit <u>not</u> including fixed cost payments.

Here, we are using the fact that $C(y^*) = \int MC(y)\, dy$ + constant, where the limits of integration are 0, y^*. This is just the **area** under the MC function.

<u>Observations</u>

1) If MC is constant (e.g., Cobb-Douglas with exponents that add to 1) then either $y^* = 0$ or $y^* = \infty$ because either profit is negative or profit is infinite (if there are no fixed costs).

2) If MC is always decreasing, then supply is not defined (or equals 0).



At y^* defined by $p = MC(y^*)$, profit is not maximized.

Why? Consider a cut in output. Cost falls by MC, revenues falls by p. So π actually increases.

The second-order conditions are not satisfied, since $C_{yy} < 0$.

Examples:

(1) $y = x^a \quad 0 < a < 1$ (one input, decreasing returns)

The input requirement function is $x^*(y) = y^{1/a}$ (this does not depend on input prices). Thus,

$$C(w,y) = w\, x^*(y) + F \qquad (F = \text{fixed costs})$$
$$= w\, y^{1/a} + F$$

$$MC(y) = \frac{w}{a}\, y^{\frac{1-a}{a}}$$

$$AC(y) = \frac{C(w,y)}{y} = \frac{F}{y} + w\, y^{\frac{1-a}{a}}$$

Output supply choice y^* solves $p = MC(y)$, implying

$$p = \frac{w}{a}\, y^{\frac{1-a}{a}} \quad \text{or} \quad y^*(p,w) = \left[a\, \frac{p}{w}\right]^{\frac{a}{1-a}} .$$

NOTE: (a) y^* is homogenous of degree 0 in p.w.

(b) y^* increases with p, decreases with w.

(2) $y = x_1^{\alpha}\, x_2^{\beta} \qquad \alpha + \beta < 1$ Cobb-Douglas with decreasing returns to scale.

Recall that $C(y, w_1, w_2) = k_1 \, w_1^{\frac{\alpha}{\alpha+\beta}} \, w_2^{\frac{\beta}{\alpha+\beta}} \, y^{\frac{1}{\alpha+\beta}}$, for $k_1 > 0$ some constant. Therefore:

$$MC(y) = k_2 \, y^{\frac{1-\alpha-\beta}{\alpha+\beta}} \, w_1^{\frac{\alpha}{\alpha+\beta}} \, w_2^{\frac{\beta}{\alpha+\beta}} \quad \text{for some constant } k_2 .$$

Setting p = MC and solving for y => $y^* = k_3 \, p^{\frac{\alpha+\beta}{1-\alpha-\beta}} \, w_1^{\frac{-\alpha}{1-\alpha-\beta}} \, w_2^{\frac{-\beta}{1-\alpha-\beta}}$ for some constant k_3.

<u>or</u> $$\log y^* = \text{constant} + \frac{\alpha+\beta}{1-\alpha-\beta} \log p - \frac{\alpha}{1-\alpha-\beta} \log w_1 - \frac{\beta}{1-\alpha-\beta} \log w_2 .$$

Again y^* is homogeneous of degree 0 in all prices, increasing in p, decreasing in w_1 and w_2 .

<u>Exercise</u>:

For a general cost function, prove that the competitive supply response is homogeneous of degree 0 in all prices (input <u>and</u> output prices).

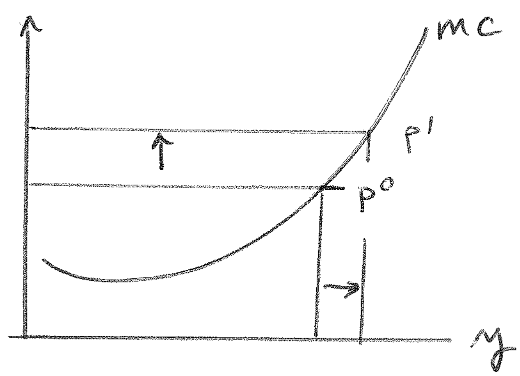
Hint: Recall that the cost function is homogeneous of degree 1 in all input prices.

<u>The Law of Supply</u>

The "law of supply" states that competitive supply functions always slope upward:

$$\frac{\partial y^*}{\partial p} > 0.$$

Why? At a supply optimum, p = MC <u>and</u> MC is increasing from the second-order condition. Therefore, if p increases, the new optimum supply response is higher: we simply move along the MC schedule.



Formally: y^* is defined as the solution to the equation

(*) $p - C_y(y^*(p, w_1, w_2), w_1, w_2) = 0.$

This first-order condition must hold if we move p (or either of w_1, w_2) . Therefore, we can differentiate equation (*) w.r.t. p:

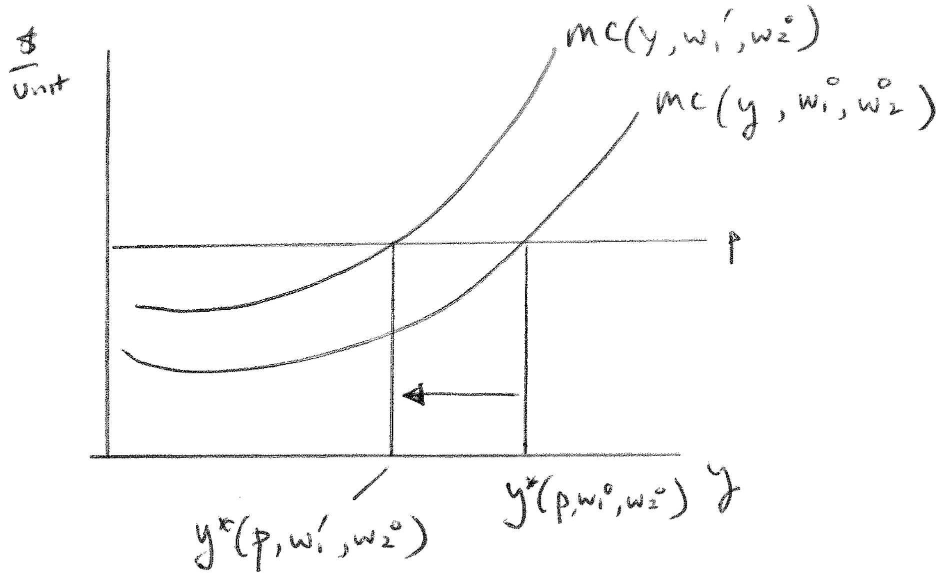
$$dp - C_{yy}(y^*(p, w_1, w_2), w_1, w_2) \times dy^* = 0 \qquad \text{(using the chain rule).}$$

Hence $\dfrac{dy^*}{dp} = \dfrac{1}{C_{yy}(y^*, w_1, w_2)}$,

and, from the second-order conditions, $C_{yy}(y^*(p, w_1, w_2), w_1, w_2) > 0$, implying that $\dfrac{dy^*}{dp} > 0$!

Changes in Input Prices

What is the effect of an increase in input prices on the firm's output decisions? Graphically, a rise in w_1 causes the MC curve to shift. Intuitively, an increase in input prices (say w_1) is associated with a shift in MC. So we have the following diagram:



In the case where MC rises as w_1 rises, we have $\dfrac{dy^*}{dw_1} < 0$. Is that always true? See the next lecture!

Economics 101A

Input Demand for a Competitive Firm

In this lecture we describe the determination of input demands for a competitive firm that sells output y at a price p. Its production function is $y=f(x_1, x_2)$. Inputs 1 and 2 have prices w_1 and w_2.

The firm's optimal choices of x_1, and x_2 are determined in two steps. First, the firm constructs its cost function $C(w_1, w_2, y)$. This implicitly defines the optimal input demands x_1, and x_2 for each level of y, and given input prices.

$$C(w_1, w_2, y) = w_1 x_1 + w_2 x_2 \quad \text{s.t.} \quad y=f(x_1, x_2).$$

$$= w_1 x_1^c(w_1, w_2, y) + w_2 x_2^c(w_1, w_2, y)$$

where $x_1^c(w_1, w_2, y)$ and $x_2^c(w_1, w_2, y)$ are the "conditional factor demands". The word "conditional" denotes that fact that these input demands are conditioned on the output choice. Note that $x_1^c(w_1, w_2, y)$ and $x_2^c(w_1, w_2, y)$ are very much like the compensated demands for the consumer. In particular, setting up the Lagrangean for the cost-min problem:

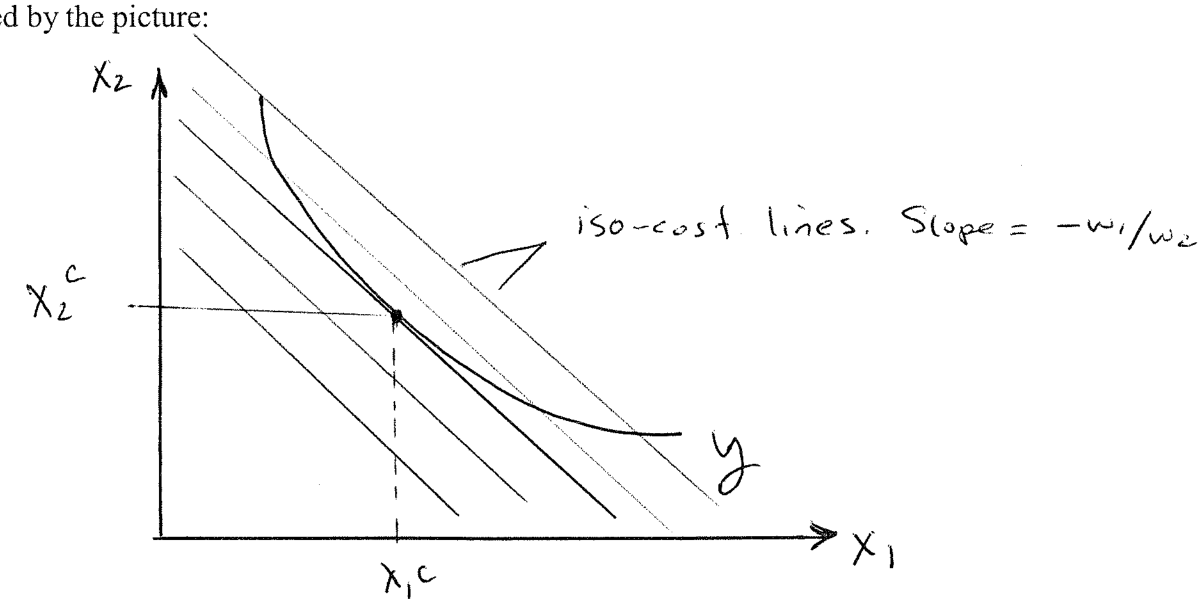
$$L(x_1, x_2, \mu) = w_1 x_1 + w_2 x_2 - \mu(y - f(x_1, x_2))$$

$$\partial L/\partial x_1 = w_1 - \mu f_1(x_1, x_2) = 0$$

$$\partial L/\partial x_2 = w_2 - \mu f_2(x_1, x_2) = 0$$

$$\partial L/\partial \mu = -y + f(x_1, x_2) = 0$$

The ratio of the first two implies that $w_1 / w_2 = f_1(x_1, x_2) / f_2(x_1, x_2)$. Recall that $f_1(x_1, x_2)$ is the "marginal product" of input 1. The ratio $f_1(x_1, x_2) / f_2(x_1, x_2)$ is called the "marginal rate of technical substitution (MRTS). This is the firm's equivalent of the consumer's MRS. It gives the slope of an isoquant at a point (w_1, w_2). So the first order conditions for cost-min are described by the picture:



For future reference, recall from earlier lectures that

$x_1^c(w_1, w_2, y) = \partial C(w_1, w_2, y)/\partial w_1$ and $x_2^c(w_1, w_2, y) = \partial C(w_1, w_2, y)/\partial w_2$.

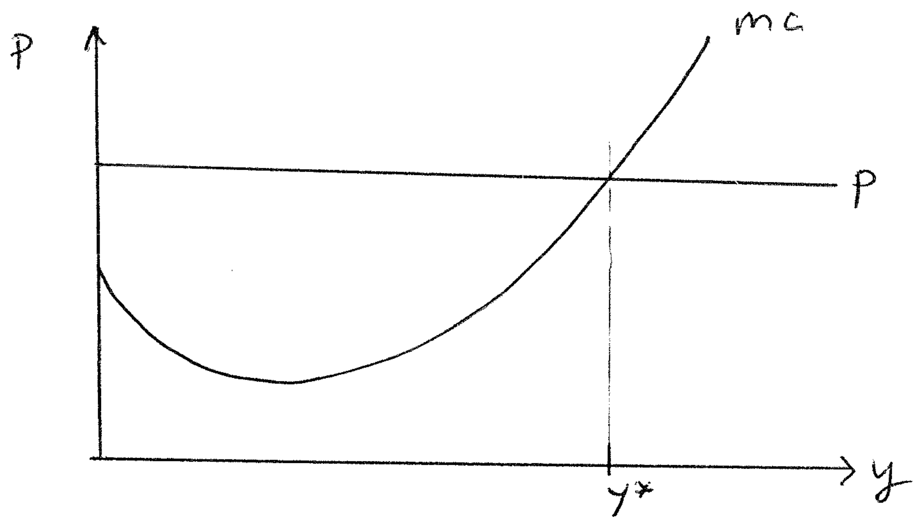
The second step is for the firm to decide what level of output to choose.
Having found C( ), the firm's decision is to maximize profit by choosing y:

$\max \; p \cdot y - C(w_1, w_2, y)$

The f.o.c. is $p - \partial C(w_1, w_2, y)/\partial y = 0$ or $p = MC(w_1, w_2, y)$ = "marginal cost"

the s.o.c is $-\partial^2 C(w_1, w_2, y)/\partial y^2 < 0$ or $\partial MC(w_1, w_2, y)/\partial y > 0$ "rising marginal cost"

This gives us the picture:



The optimal choice of y for given (p, w_1, w_2) is $y^*(p, w_1, w_2)$ which is implicitly defined by:

$$p = \partial C(w_1, w_2, y^*(p, w_1, w_2))/\partial y$$

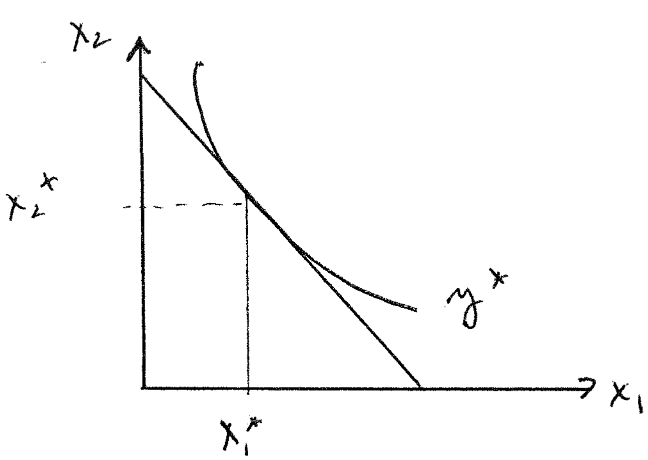
$$= MC(w_1, w_2, y^*(p, w_1, w_2)) .$$

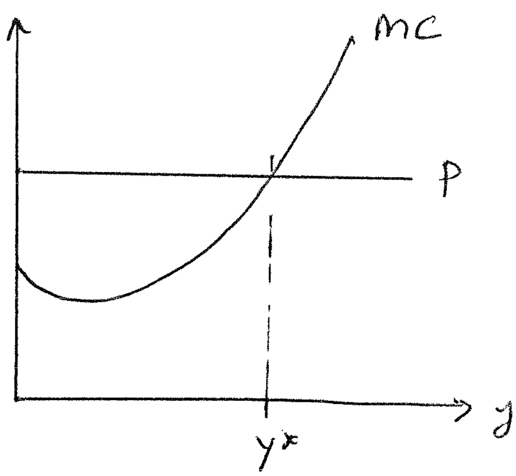
In other words, y^* is the value of y that equates MC to p.

Now we are ready to define the firm's **unconditional** input choices. The firms unconditional input demands are simply:

$x_1(p, w_1, w_2) = x_1^c(w_1, w_2, y^*(p, w_1, w_2))$
$x_2(p, w_1, w_2) = x_2^c(w_1, w_2, y^*(p, w_1, w_2))$.

These equations say that the unconditional input demands are the conditional demands for the "optimized" choice of y. We can think of the problem of finding optimal input demand choices as one of solving two problems simultaneously: cost-min and p=MC



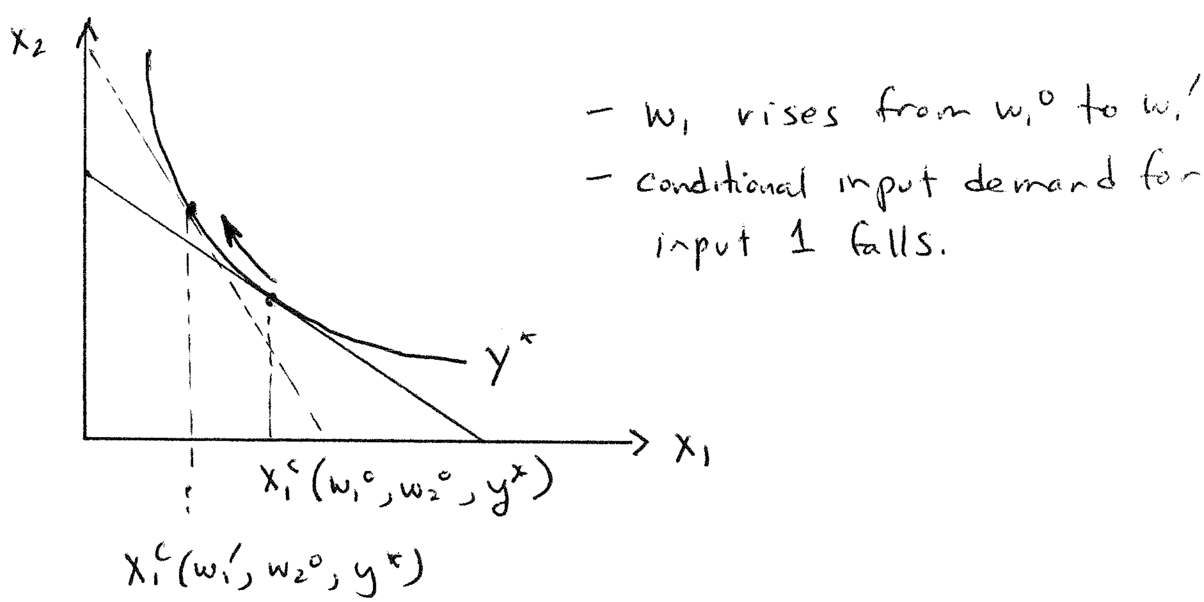


What happens when w_1 rises? Since

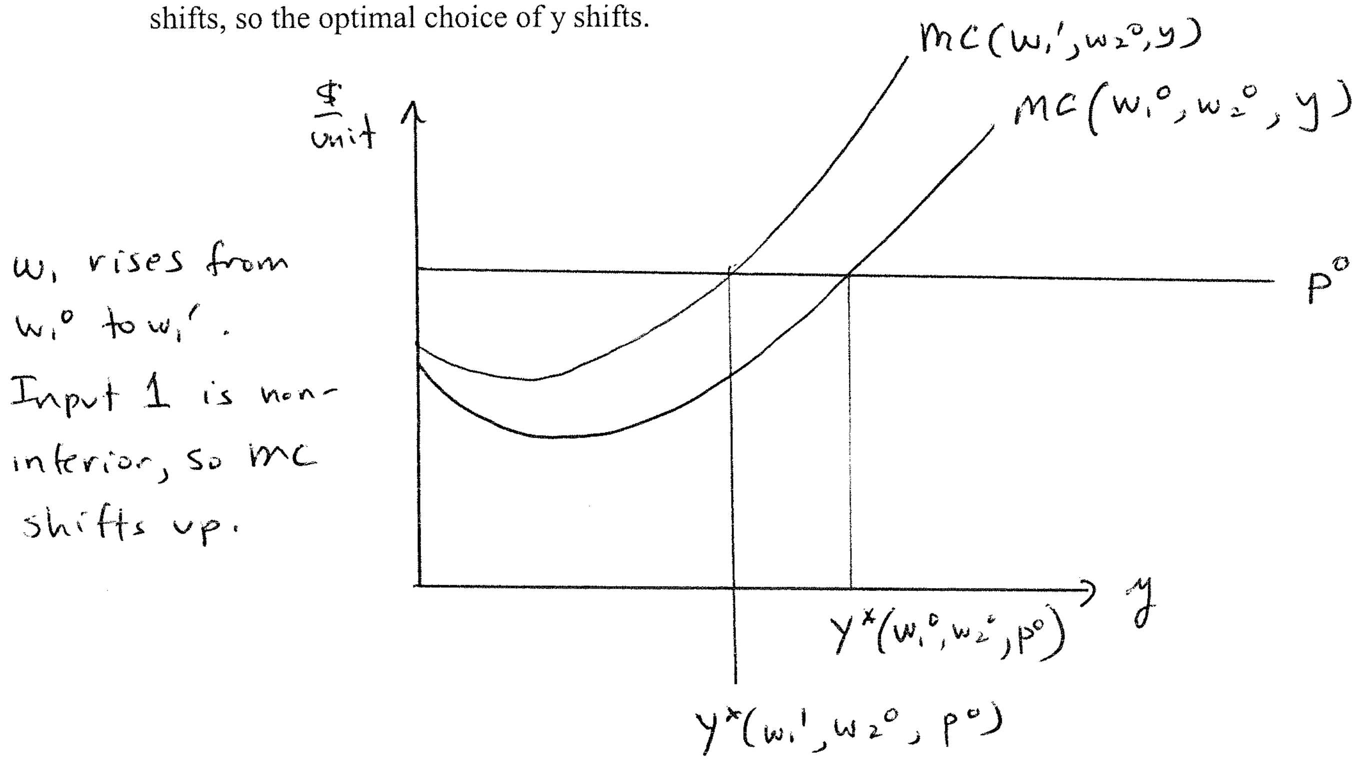
$$x_1\,(p, w_1, w_2) \;=\; x_1^c(\,w_1, w_2, y^*(\,p, w_1, w_2\,)\,)$$

$$\partial x_1\,(p, w_1, w_2)/\partial w_1 \;=\; \partial x_1^c/\partial w_1 \;+\; \partial x_1^c/\partial y \times \partial y^*(\,p, w_1, w_2\,)/\partial w_1$$

The first term is the response of optimal input demand, **holding constant y**. This is called the substitution effect. It is just like the consumer's substitution effect, which is defined as the change in demand holding utility constant. For the firm, however, the substitution effect holds y constant, giving a movement along an isoquant.



The second term is called the "scale effect". It has some similarity to the consumer's income effect, but the analogy can be misleading. It reflects the fact when w_1 rises, the firm's MC curve shifts, so the optimal choice of y shifts.



Recall from earlier lectures that if input 1 is a "non-inferior" input, when w_1 rises MC shifts up. Why?

$$\partial MC(w_1, w_2, y)/\partial w_1 = \partial/\partial w_1 \{ \partial C(w_1, w_2, y)/\partial y \}$$

$$= \partial^2 C(w_1, w_2, y)/\partial w_1 \partial y$$

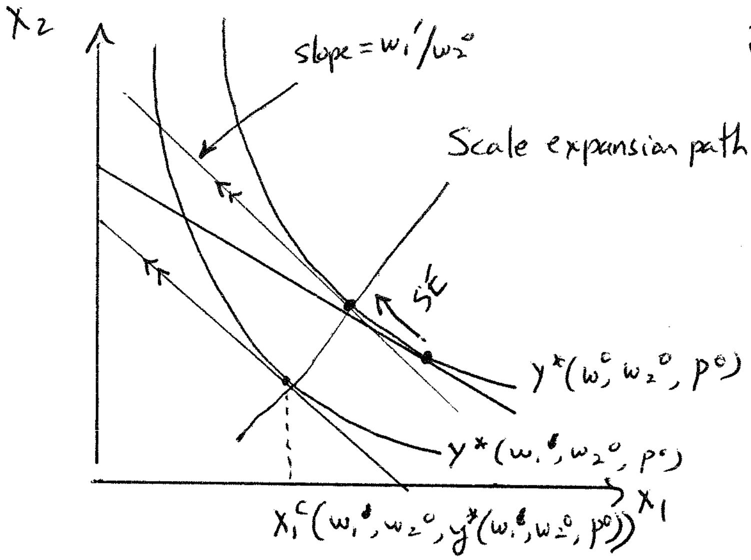
$$= \partial^2 C(w_1, w_2, y)/\partial y \partial w_1 \quad \text{since we can interchange order}$$

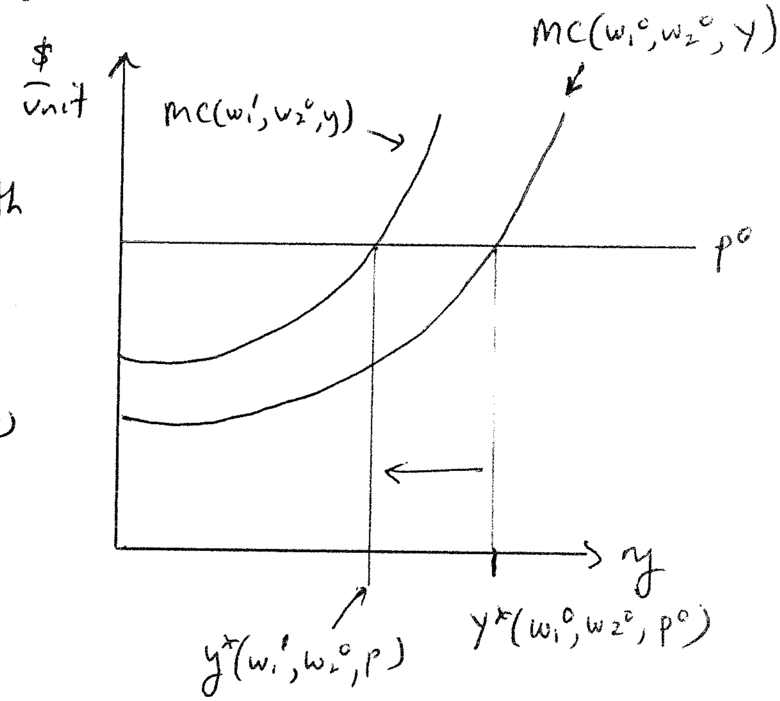
$$= \partial/\partial y \{ \partial C(w_1, w_2, y)/\partial w_1 \}$$

$$= \partial x_1^c(w_1, w_2, y)/\partial y$$

Thus, the derivative of MC w.r.t. w_1 **is the same thing** as the derivative of conditional input demand w.r.t. y. If input 1 is non-inferior, $\partial x_1^c/\partial y > 0$, so when w_1 rises MC shifts up.

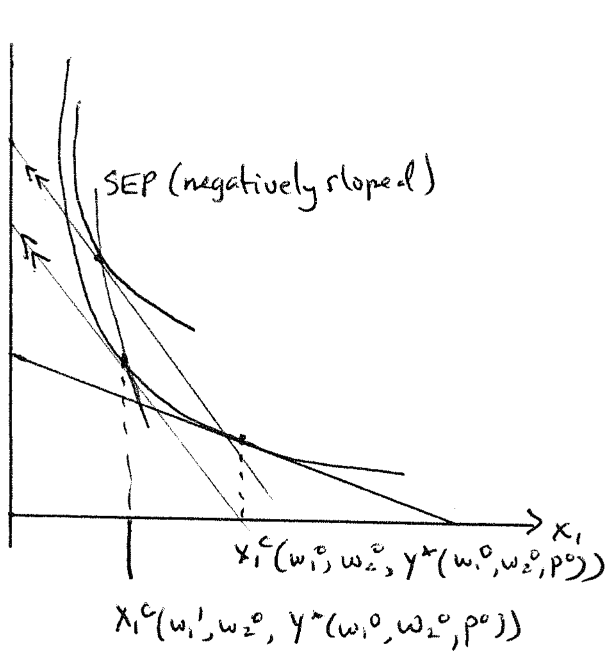
In this case we have the following pair of pictures. When w_1 rises, the substitution effect and scale effect both cause a reduction in demand for input 1.

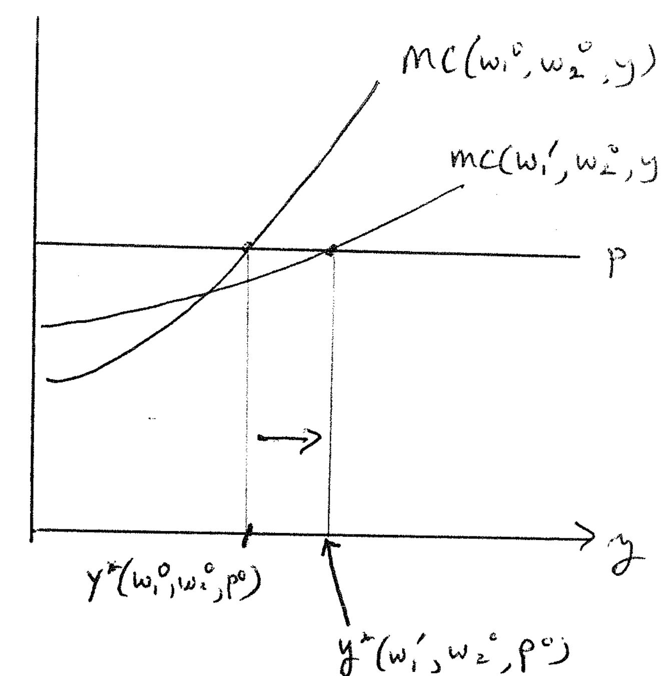




- substitution effect (SE) causes x_1 ↓
- Scale effect (movement along scale expansion path) also causes x_1 ↓

If we have an inferior input, when w_1 rises MC shifts down. (When shovels rise in prices the marginal cost of holes goes down). Then the scale effect is also negative, because although the rise in w_1 makes the firm want to increase output, input 1 is inferior so the expansion in output lowers demand!





For inferior input, $w_1 \uparrow \Rightarrow y^*$ rises!

There is another way to look at the problem of input demands – a so-called "direct approach". Suppose the firm simply chose x_1 and x_2 to maximize:

$$\pi = p \cdot f(x_1, x_2) - w_1 x_1 - w_2 x_2 \quad .$$

This is an unconstrained problem so the f.o.c. are:

a) $\quad p \cdot f_1(x_1, x_2) - w_1 = 0$

b) $\quad p \cdot f_2(x_1, x_2) - w_2 = 0$

Note that the ratio of a) to b) gives the "tangency condition" $w_1 / w_2 = f_1(x_1, x_2) / f_2(x_1, x_2)$.

Also, the firm sets $p = w_1 / f_1(x_1, x_2) = w_2 / f_2(x_1, x_2)$.

What do these mean? If the firm had to increase output by 1 unit, it could do it by increasing input 1 or input 2. If it used input 1, it would need $1/f_1(x_1, x_2)$ units of input 1 to add 1 unit of output. The marginal cost of this would be $w_1 / f_1(x_1, x_2)$. Similarly if it used input 2 the marginal cost would be $w_2 / f_2(x_1, x_2)$. The tangency condition implies that these are equal. So we can interpret the optimum conditions as p=MC plus the tangency condition.

Looking back at the Lagrangean for the cost min problem, notice that the f.o.c. are

$$w_1 = \mu \ f_1(x_1, x_2) \ \Rightarrow \ \mu = w_1 / f_1(x_1, x_2)$$

and

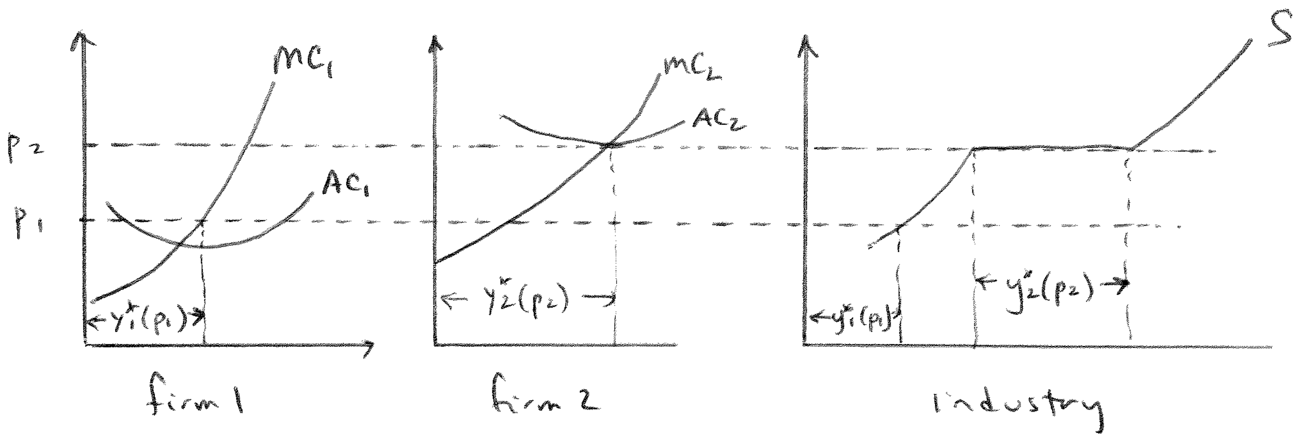
$$w_2 = \mu \ f_2(x_1, x_2) \ \Rightarrow \ \mu = w_2 / f_2(x_1, x_2)$$

Also, recall that μ is marginal cost. So when the firm solves the cost min **and** sets p=MC=μ it gets the same answer as the "direct approach".

Sometimes it is more convenient to work with the "cost-min & p=MC" approach. Other times it is easier to work with the "direct approach". They give the same answers.

Industry Supply

The supply curve for an industry consists of the "horizontal sum" of the supply curves of all individual firms:



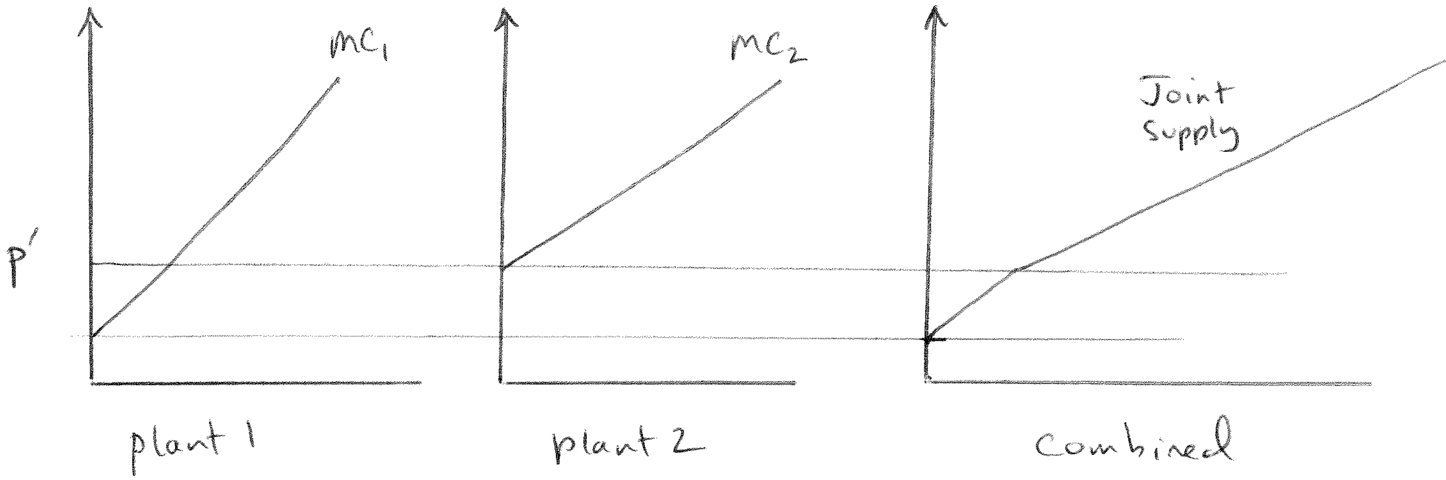
Notice that if firms vary in their costs, at any market price some firms are making profit, some are just on the margin of profitability, and others are "out of the market". A good example of this is the case of oil wells. Some wells have low production costs and are always profitable to operate. Others are high-cost, and only come "on-line" when crude prices are high. We usually call the profits earned by the infra-marginal suppliers "rents". Presumably, the lower costs of these firms arise from their control of a scarce resource.

A competitive market is in equilibrium if

(1) each existing firm has p = MC and $\pi \geq 0$

(2) no other firms can enter and earn profits

These ideas give a nice way to think about a mutli-plant firm (as in problem set 4). If a firm runs 2 plants, with MC schedules $MC_1(y_1)$ and $MC_2(y_2)$, then the firm can operate efficiently by thinking of the plants as independent suppliers. For example consider the following 2-plant firm:



At prices below p′, plant 2 does not operate. At prices above p′, both plants operate where

$$p = MC_1(y_1^*) = MC_2(y_2^*) .$$

Notice that the firm can run "as if" each plant was a separate entity. This is the so-called "principle of decentralization".

Lecture 12

MONOPOLY

A monopolist is the sole supplier to a given market. The critical feature of monopolistic behavior is the fact that a monopolist "sets price" (or sets quantity). Monopolies arise

(a) through exclusive control of inputs or resources: e.g. DeBeers monopoly of diamond marketing

(b) through exclusive legal rights: e.g., public utilities; drug companies with patents

Suppose the demand for output is represented by the function y = D(p). Then we can invert this to p = p(y), which is usually referred to as the "inverse demand function. A monopolist's profit is

$$\pi (y, w_1, w_2) = y\, p(y) - C(y, w_1, w_2)$$

The first order condition for profit maximization is

$$p(y) + y\, p'(y) - C_y(y , w_1 , w_2) = 0$$

$$p(y) + y\, p'(y) = C_y(y , w_1, w_2) .$$

The expression on the left hand side represents "marginal revenue", MR(y) = p(y) + yp'(y). If demand is downward sloping p'(y) < 0 , so **MR(y) is less than p**. This is the key point about a monopoly. Since a monopoly controls the market, it cannot treat price as exogenous. Rather, it has to take account of the fact that a rise in sales will necessarily come at the expense of a reduction in price. Note that there may be close substitutes for a product. But as long as a firm is the sole supplier of the product, it has monopoly power.

Define the elasticity of demand:

$$\eta = \frac{\partial y}{\partial p} \cdot \frac{p}{y}$$

$$= \frac{1}{p'(y)} \cdot \frac{P}{y} \Rightarrow p'(y) = \frac{1}{\eta} \frac{P}{y} .$$

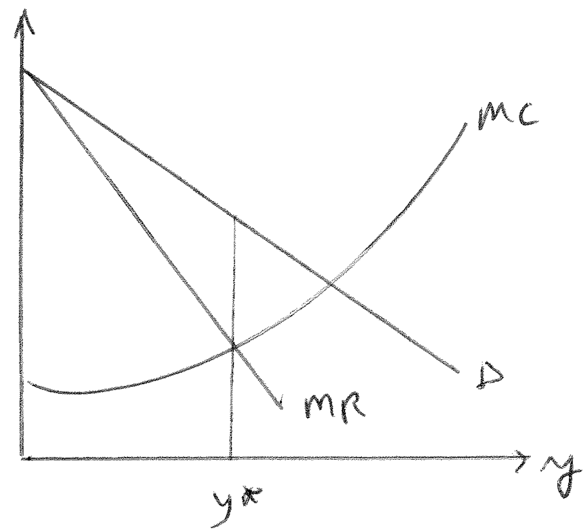
$$MR(y) = p(y) + yp'(y) = p(y) + y\left[\frac{1}{\eta}\frac{P}{y}\right] = p(y)\left[1 + \frac{1}{\eta}\right] ,$$

For a monopolist, then,

$$p(y)\left[1 + \frac{1}{\eta}\right] = MC .$$

As the market demand becomes closer and closer to a horizontal line, η goes to minus infinity,demand becomes perfectly elastic, and p = MC. So in the limiting case monopoly becomes perfect competition.

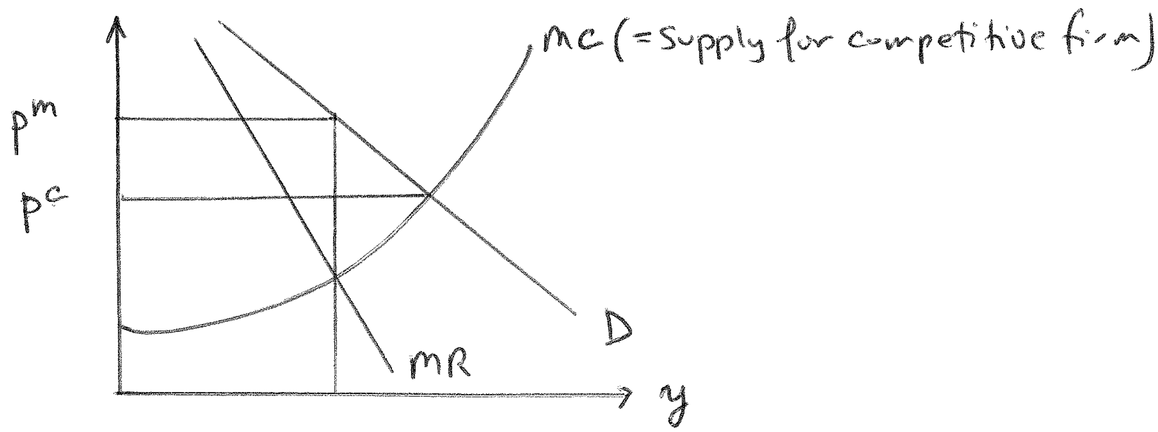
The picture associated with monopoly is the following:



NOTES

(1) A monopolist always sets MR=MC . Since MR = p (1 + 1/η) and $\eta < 0$, MR < p . If $|\eta| < 1$ then $1/\eta < -1$ and MR is <u>negative</u>. It follows that a monopolist never operates in a region where demand is inelastic. The intuition is that if demand is <u>inelastic</u>, you can increase revenue by raising price! This is a very powerful result. It says that some markets cannot be considered "monopoly" markets: namely those with measured elasticities of demand less than 1 in absolute value.

(2) If the monopolist's MC schedule was the MC schedule for a price taker, or a set of price takers (i.e. a competitive industry supply function) then equilibrium would occur where p = MC . This point would entail higher output and lower price, but lower profit to the industry as a whole.



(3) A monopolist does not have a supply schedule, per se. First, the monopolist looks at the demand function. <u>Then</u> she establishes a price. There is no schedule of price/quantity combinations.

4. In a second-price sealed bid auction, the best strategy is to bid your valuation.

Suppose your true value is v and you bid v-x (for $x \geq 0$). Suppose the highest bid among all the other bidders is w.

- if $v-x > w$, you win and pay w
- if $v-x < w$ you lose and get nothing.

Your profit/surplus from the bidding choice v-x is

$$S = P(v-x > w) \times [\, v-w] \quad = \quad \text{prob. you win} \times \text{surplus if you win}$$

this is maximized by setting x=0!

Now supppose you bid v+x (for $x \geq 0$), and the highest bid among the others is w.

- if $v+x < w$ you lose and get nothing
- if $v+x > w$ you win and pay w, and your surplus is v-w
  there are then 2 cases:
  a) $v>w$ –> you want to win
  b) $v<w$ –> you don't want to win

if you set x=0, you will win if and only if $v>w$. Best to choose x=0

5. Based on result 4, in a second price sealed bid auction the auction is won by the bidder with the highest valuation, who pays a price equal to the second highest valuation

6. Results 3 and 5 imply that **with private values** English auction == second price sealed bid

7. Without private values English auctions may be different because you can learn from the identity/size of the remaining pool of bidders.

8. How to bid in a first price auction

Assumptions – v_i independently distributed with
$P(\text{value} \le x) = F(x)$
i.e., distribution function F (cumulative d.f.)

– each bidder follows the same strategy and bids $b = B(v)$ if she gets valuation v.

Question: what does the bid function $B(v)$ look like?

– $B(v)$ will be increasing $\Rightarrow$ in an auction with $v_1 \ldots v_n$ the highest value will win, and will pay his/her bid.

– let $b = B(v) \Rightarrow v = g(b)$, g is the <u>inverse bidding fn</u>

– if you bid b, prob. you win is prob that all the other values are less than $g(b)$

$\Rightarrow$ prob win with bid b, assuming all others use inverse bid function $g(b)$ is $[F(g(b))]^{n-1}$

$S = S(b; v) =$ surplus when you bid b and value is v

$$= \underbrace{(v - b)}_{\text{surplus if win}} \cdot \underbrace{(F[g(b)])^{n-1}}_{\text{prob of winning}}$$

what is F.O.C for optimal b ?

$$\frac{\partial S}{\partial b} = -\left(F[g(b)]\right)^{n-1} + (v-b)\,F[g(b)]^{n-2} \times (n-1) \times \frac{d}{db} F(g(b))$$

by chain rule $\frac{d}{db} F(g(b)) = \frac{dF(v)}{dv} \times g'(b)$

and the derivative of the cumulative $F(v)$ is the <u>density</u> $f(v)$

setting $\frac{\partial S}{\partial b} = 0$

$$(v-b)\,F[g(b)]^{n-2}(n-1)\,f(g(b))\,g'(b) = F[g(b)]^{n-1}$$

$$\Rightarrow (v-b) = \frac{F(g(b))}{f(g(b))} \times \frac{1}{g'(b)} \times \frac{1}{(n-1)} \qquad (*)$$

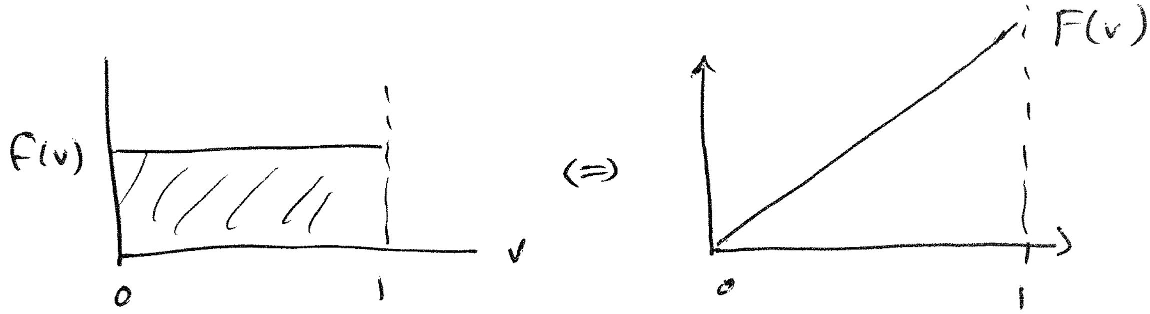
note that $g'(b) > 0$ so r.h.s > 0

$\Rightarrow$ $v - b > 0$: Always "shade your bid"

why: bid higher $\Rightarrow$ win more often but pay more

↳ like a monopolist you want to take this into account.

Example v_i distributed on $[0,1]$



$$F(v) = v \ , \quad f(v) = 1$$

Equation (*) says

$$g(b) - b = \frac{F(g(b))}{f(g(b))} \cdot \frac{1}{g'(b)} \cdot \frac{1}{n-1}$$

$$\Rightarrow \quad g(b) - b = \frac{g(b)}{g'(b)} \cdot \frac{1}{n-1} \quad \Big\} \text{ p.d. equation}$$

Solution: $g(b) = \frac{n}{n-1} \cdot b$

Check $g(b) - b = \frac{n}{n-1} \cdot b - b = \frac{n-(n-1)}{n-1} \cdot b = \frac{b}{n-1}$

$$\frac{g(b)}{g'(b)} \cdot \frac{1}{n-1} = \frac{\frac{n}{n-1} \cdot b}{\frac{n}{n-1}} \times \frac{1}{n-1} = \frac{b}{n-1} \quad \checkmark$$

$$\Rightarrow \boxed{B(v) = \frac{n-1}{n} \cdot v}$$

Economics 101A
Auctions - Lecture 2

Introduction

In this lecture we analyze the "winner's curse". The winner's curse arises in common value or affiliated value auctions, in which each bidder has a guess of the value of the item at auction. Bidders with higher guesses, **on average** have made a positive error. So bidders have to shade their bids to take account of the fact that when they win the auction they are on average overly optimistic.

A very simple example is the following. A used police car is being sold at auction, using sealed bid second price auction. Each potential buyer inspects the car, then the bidding starts.

Mathematically, we make the following assumptions:

1. the true value of the car is v, which is a random variable with mean μ, and variance σ^2_v. This expresses the idea that over many auctions, the average value of an old police car is μ. But there is variability from car to car, reflected in σ^2_v.

2. Each bidder hires a mechanic to estimate the value of the car. The mechanic reports back his/her estimated value $t = v + \epsilon$, where ϵ is the error the mechanic makes in assessing the car. Bidders don't know the values reported back to their competition.

3. We assume ϵ is a random variable with mean 0 and variance σ^2_ϵ. Note that the larger is σ^2_ϵ relative to σ^2_v the "noisier" are the mechanics' reports.

4. Based on the report of t, the ith bidder makes an estimate of the true value of the car. In particular, a buyer forms an estimate

$$y = \lambda t + (1-\lambda)\mu$$

The idea of this is that if t is really noisy, you should "downweight" the mechanic's report and assume instead that the true value is some weighted average of μ and t.

Note that $E(y) = \lambda E(t) + (1-\lambda)\mu = \lambda E(v+e) + (1-\lambda)\mu = \lambda\mu + (1-\lambda) = \mu$.

What λ should you use? The "forecast error" if you use a given value of λ is

$$\text{error} = v - [\lambda t + (1-\lambda)\mu] = v - \lambda(v+e) - (1-\lambda)\mu$$

$$= (1-\lambda)(v-\mu) - e.$$

The variance of the forecast error is

$$\text{Var[error]} = \text{Var} [(1-\lambda) (v - \mu) - e]$$

$$= (1-\lambda)^2 \sigma^2_v + \lambda^2 \sigma^2_\epsilon$$

Choosing a λ to minimize the variance of the forecast error, we differentiate w.r.t. λ and set the f.o.c. to 0:

$$\partial \text{Var}/\partial\lambda = -2 (1-\lambda) \sigma^2_v + 2\lambda \sigma^2_\epsilon = 0$$

$$=> (1-\lambda) \sigma^2_v = \lambda \sigma^2_\epsilon$$

or $\lambda^* = \sigma^2_v / [\sigma^2_v + \sigma^2_\epsilon]$.

You may have seen this before: λ^* is the "signal to total variance ratio". If σ^2_ϵ is small, λ^* is close to 1, and you use a weighted average that puts most weight on the mechanic's report.

4. Based on the report from the mechanic, and an optimal choice of $\lambda=\lambda^*$, each bidder now has a good idea of what the value of the car is in the current auction. Since its a second price auction, you might think each bidder should simply bid

$$y = \lambda^* t + (1-\lambda^*) \mu$$

But this will create a winner's curse! The highest bidder will win the auction - so that is the bidder whose mechanic made the biggest **positive error**. The bidder will pay the second highest price, which is the the **second biggest postive error**. So even with a second price auction, you have to take account of the fact that the second highest price contains the second-highest value of the error e.

5. In an auction with 10 bidders, if e is normally distributed with mean 0 and variance 1, the expected second highest value of e is 1.003. Here are some other cases:

Number of Bidders	Expected Second-highest error
10	1.003
25	1.53
35	1.692
100	2.146

You can see that the expected second highest error grows with the number of bidders.

This might make you worry a little about ebay. There are potentially thousands of bidders out there. If you don't know what the object is worth to you, you need to shade your bid quite a bit.

6. So what should you bid?

Suppose each bidder shades his/her estimated value by the amount k:

$$y = \lambda^* t + (1-\lambda^*)\mu - k$$

$$= \lambda^* v + (1-\lambda^*)\mu + \lambda^* e - k$$

Let e_2 = the expected value of the second highest draw from the distribution of e, when there are N bidders in the game. (Normally you estimate this by computer simulation).

The expected bid of the second-highest bidder is

$$\lambda^* v + (1-\lambda^*)\mu + \lambda^* e_2 - k$$

So if everyone sets $k = \lambda^* e_2$, each bidder will expect to pay

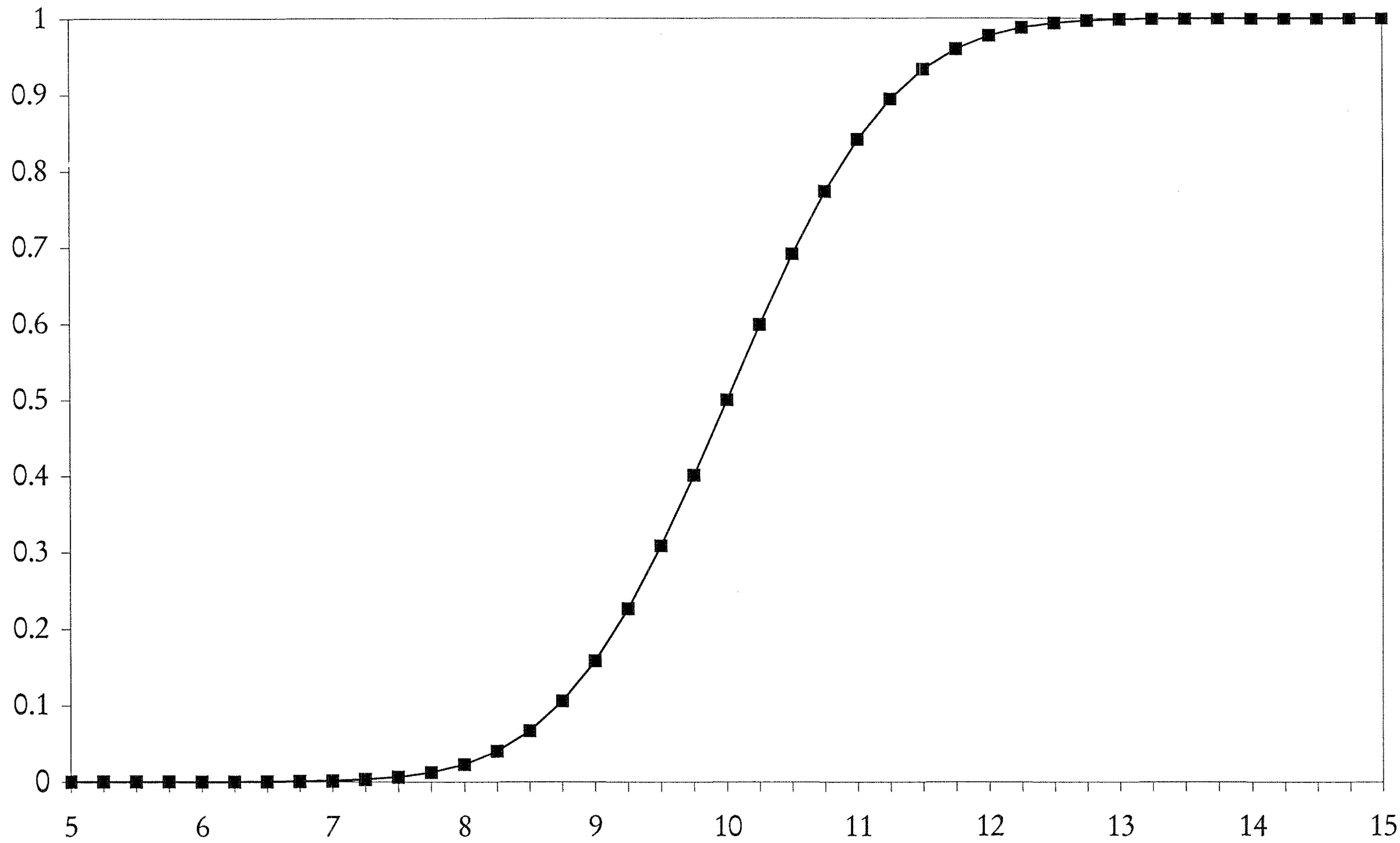
$$E[\lambda^* v + (1-\lambda^*)\mu] = \mu$$

if they win, which means **on average** they pay what the item is worth.

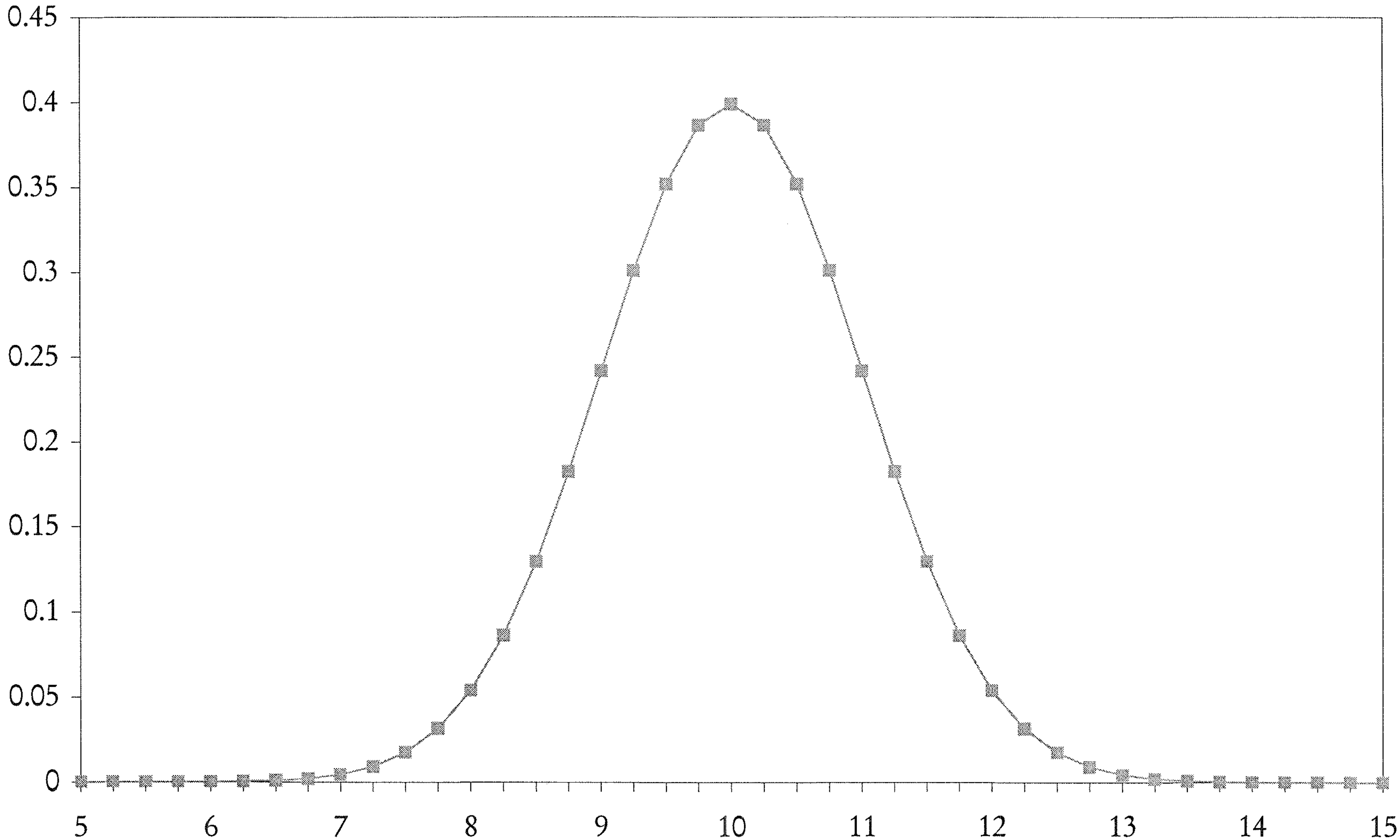
Bottom line: in the second price auction we have set up, each bidder bids that they think the item is worth, given their information, *minus a discount that is equal to the expected second-highest error in the observed signal.*

7. In real situations, people hire a consultant to simulate the auction (by making guesses about σ^2_ϵ and σ^2_v).

Normal Cumulative Distribution - F(x) - with Mean 10 and Standard Deviation 1



Normal density - f(x) - with Mean 10 and standard deviation 1



Economics 101A
Introduction to Finance - The CAPM Model

Overview

In this lecture we consider the implications of the simple assumption that people only hold portfolios of assets that are "mean-variance efficient". This turns out to have the surprising implication that the price of a stock (i.e. a the price of a share in a publicly traded company) depends on the *covariance* between the return yielded by that stock versus the return on market as a whole. This result was discovered (in the early 1960s) by William Sharpe, who shared the Nobel Prize for economics on the basis of his work. The theoretical model is called the Capital Asset Pricing Model (CAPM).

Introduction

1. We will assume that people can invest in a set of assets, numbered 1, 2, ...n. An investor with \$X selects a *portfolio*, which is a list of the amounts invested in each of the possible assets. Call the **share** invested in asset i α_i. Note that $\alpha_i \geq 0$, and $\Sigma_i\ \alpha_i = 1$. You can think of any n-dimensional vector with positive elements that sum to 1 as a "portfolio".

2. We make the simplifying assumption that an investor has a 1 period horizon (or "holding period"). An investment of \$1 in asset i will be worth $(1+R_i)$ at the end of the holding period. Thus R_i is the proportional "return" on asset i over the holding period. Note that $R_i \geq -1$ if assets have "limited liability", since the worst thing that can happen is you lose your entire investment. Each of the R_i's is a random variable.

3. The mean of R_i is $E[R_i]$ and the variance is σ^2_i. A "riskier" asset has a higher variance. We **do not** assume that the asset returns are independent of each other. Instead, we assume that there are potential covariances:

$$Cov[R_i, R_j] = E[\ (R_i - E[R_i]) \times (R_j - E[R_j])\ ] = \sigma^2_{ij}.$$

Note that the covariance of asset i with itself is

$$Cov[R_i, R_i] = E[\ (R_i - E[R_i])^2\ ] = Var[R_i] = \sigma^2_i.$$

To save notation we will sometimes use σ^2_{ii} to mean σ^2_i.

4. (For future reference). If you have taken econometrics, you may recall that if you have two variables, x and y, and you are interested in the "best-fitting line" explaining the relation between y and x, you fit a "linear regression" of y on x. This gives the "best fitting" coefficients *a* and *b*

for the linear model:

$$y = a + bx + e,$$

where e is the "residual". The best fit is defined to minimize the variance of the prediction error or residual e. Solving this (using the same methods we used in the winner's curse lecture) leads to the formula

$$b = \text{Cov}[x, y] / \text{Var}[x] = \sigma^2_{xy}/\sigma^2_x.$$

5. (Back to the portfolio problem).

If an investor with \$X to invest selects a set of shares $(\alpha_1, \alpha_2, ..., \alpha_n)$ what is his/her ending wealth?

- the total investment in asset i is $\$\alpha_i X$
- at the end of the holding period this is worth $\alpha_i X (1+R_i)$
- total wealth is $\$X \Sigma_i \alpha_i (1+R_i) = \$X \{ 1 + \Sigma_i \alpha_i R_i \}$ using the fact that $\Sigma_i \alpha_i = 1$.

- the realized proportional return on the portfolio is

$$\text{return} = [X \{ 1 + \Sigma_i \alpha_i R_i \} - X] / X$$

$$= \Sigma_i \alpha_i R_i$$

which is just a weighted average of the returns on the assets, with weights equal to the portfolio shares. The expected return is

$$E[\Sigma_i \alpha_i R_i] = \Sigma_i \alpha_i E[R_i]$$

What is the variance of the return on the portfolio? This is

$$E[(\Sigma_i \alpha_i R_i - \Sigma_i \alpha_i E[R_i])^2]$$

$$= E[(\Sigma_i \alpha_i (R_i - E[R_i]))^2]$$

$$= E[\Sigma_i \alpha_i (R_i - E[R_i]) \times \Sigma_j \alpha_j (R_j - E[R_j])$$

$$= E[\Sigma_i \Sigma_j \alpha_i \alpha_j (R_i - E[R_i]) (R_j - E[R_j])]$$

$$= \Sigma_i \Sigma_j \alpha_i \alpha_j \sigma^2_{ij} = \Sigma_i \alpha_i^2 \sigma^2_{ii} + 2 \Sigma_i \Sigma_{j \neq i} \alpha_i \alpha_j \sigma^2_{ij} .$$

6. The CAPM makes two assumptions:

a) there is a "risk free" asset – call this asset 1.

b) the market portfolio – the portfolio which has an equal distribution of all shares – is "mean variance efficient", in the sense that no other portfolio gets the same return and has a lower variance in its return

7. Consider an efficient portfolio $\alpha^p = (\alpha^p_1, \alpha^p_2, \ldots \alpha^p_n)$ that has the lowest variance subject to yielding an expected return r_p. To find the properties of such an "efficient" portfolio, note that it must solve the problem

$$\min_{\alpha} \quad \Sigma_i \Sigma_j \alpha_i \alpha_j \sigma^2_{ij} \quad \text{s.t.} \quad \Sigma_j \alpha_j E[R_j] = r_p \quad \text{AND} \quad \Sigma_j \alpha_j = 1.$$

Set up a Langrangean, with a multiplier λ for the first constraint, and another multiplier μ for the second:

$$L(\alpha_1, \alpha_2, \ldots\alpha_n, \lambda, \mu) = \Sigma_i \Sigma_j \alpha_i \alpha_j \sigma^2_{ij} - \lambda (\Sigma_j \alpha_j E[R_j] - r_p) - \mu (\Sigma_j \alpha_j - 1)$$

Differentiating w.r.t. α_j we get

$$(*) \quad 2 \Sigma_i \alpha_i \sigma^2_{ij} - \lambda E[R_j] - \mu = 0 \quad \text{for } j=1,,,n.$$

the first term in (*) follows from the alternative way of writing the variance of the portfolio in the last line of the previous page.

Note that (*) has to hold for asset #1, which by assumption is risk-free. Hence it has $\sigma^2_{i1} = 0$. Thus, for asset #1, equation (*) implies

$$- \lambda E[R_j] - \mu = 0$$

or $$\mu = - \lambda E[R_1] = - \lambda R_F =$$

where we are denoting the risk-free return by R_F.

Hence we can re-write (*) as

$$(*) \quad 2 \Sigma_i \alpha_i \sigma^2_{ij} = \lambda (E[R_j] - R_F), \quad \text{which holds for } j=1,2,,,n$$

Multiply (*) by α_j :

(**) $2\,\alpha_j \;\; \Sigma_i\, \alpha_i\, \sigma^2_{ij} \;\; = \;\; \alpha_j \;\; \lambda\,(\,E[R_j] \; - \; R_F\,)$, true for j=1,2,....n.

Now sum equations (**) across all n:

$$\Sigma_j \; 2\,\alpha_j \;\; \Sigma_i\, \alpha_i\, \sigma^2_{ij} \;\; = \; \Sigma_j \;\; \alpha_j \;\; \lambda\,(\,E[R_j] \; - \; R_F\,)$$

=> $$2 \;\; \Sigma_j \;\; \Sigma_i\, \alpha_j\, \alpha_i\, \sigma^2_{ij} \;\; = \; \lambda\,(\;\; \Sigma_j\, \alpha_j\, E[R_j] \;\; - \; R_F\,)\,.$$

or

(***) $2 \;\; \Sigma_j \;\; \Sigma_i\, \alpha_j\, \alpha_i\, \sigma^2_{ij} \;\; = \;\; \lambda\,(\,r_p \; - \; R_F\,)$, using the fact that $\Sigma_j\, \alpha_j\, E[R_j] = \; r_p$.

Call $\Sigma_j \;\; \Sigma_i\, \alpha_j\, \alpha_i\, \sigma^2_{ij} = \sigma^2_p$, the minimized variance of the portfolio with expected return r_p.

Then (***) says $2\,\sigma^2_p = \lambda\,(\,r_p \; - \; R_F\,)$ or

$$\lambda = \;\; 2\,\sigma^2_p\,/(\,r_p \; - \; R_F\,)\,.$$

Now equation (*) says:

$$2\,\Sigma_i\, \alpha_i\, \sigma^2_{ij} \;\;\; = \lambda\,(\,E[R_j] \; - \; R_F\,)$$

and substituting for λ we get

(****) $(\,E[R_j] \; - \; R_F\,) = \,(\,r_p \; - \; R_F\,)\,\Sigma_i\, \alpha_i\, \sigma^2_{ij}\,/\;\sigma^2_p$

Finally notice that

$$\Sigma_i\, \alpha_i\, \sigma^2_{ij} \;\; = \;\; E\,[\;\;(R_j - E[R_j]\,) \times \; \Sigma_i\, \alpha_i\,(R_i - E[R_i]\,)\,]$$

$$= Cov\,[\,R_j\,, \;\; \Sigma_i\, \alpha_i\, R_i\,]$$

In other words, $\Sigma_i\, \alpha_i\, \sigma^2_{ij}$ is the covariance of asset j's return with the return on the "efficient portfolio" that has expected return r_p. Let's call the return on that portfolio R^p. So $E[R^p]=r_p$ and $Var[R^p]$ is the lowest among all portfolios with that expected return.

We can therefore write (****) as

(+) $E[R_j] \; - \; R_F \; = \; (\,r_p \; - \; R_F\,)\,Cov[R^p, R_j]\,/\,Var[R_j]$

Suppose now that (+) is true of the "market portfolio" R^M, which is the portfolio consisting of an equal share of all assets that are out there (also called "holding the market"). Let r_M denote the expected return on the market portfolio, and let $\sigma^2_M = Var[R^M]$, denote the variance of the market portfolio. Then (+) implies

$$(++) \qquad E[R_j] = R_F + (r_M - R_F) \, Cov[R^M, R_j] \, / \, Var[R^M]$$

Finally, note that

$$Cov[R^M, R_j] \, / \, Var[R^M] = \beta_j ,$$

the regression coefficient you would get if you did a linear regression of returns on asset j on the returns on the market portfolio. This regression coefficient is called the asset's "beta".

Economics 101A
Introduction to Finance II
Efficient Market Hypothesis

1. In the previous lecture we considered the implications of the "CAPM assumptions" for the expected rate of return on a risky asset i. In particular, if asset i has random return R_i, and if the "beta" of asset i is β_i (i.e. if on average when the market return is above/below average by 1%, asset i's return is β_i % higher/lower), then the expected return on asset i has to satisfy the CAPM equation

(*) $E[R_i] = \pi_0 + \beta_i \pi_1$

where (in the strict CAPM), $\pi_0 = R_F$, the "risk free" rate of return, and $\pi_1 = r_M - R_F$, the gap between the expected return on the "market portfolio" and the risk free rate.

2. Researchers in the 1970s and 1980s tried to test the CAPM by estimating β's for sets of assets, and then seeing if on average assets with bigger β's had higher expected returns. This was not always successful. These days, people interpret the model a little less literally. Often they augment the model with additional "factors". The strict CAPM says that all that is relevent about an asset is its covariance with the market. A more agnostic view is that β is one of the things that matter, but there may be others. So sometimes people look at models like

$$E[R_i] = \pi_0 + \beta_i \pi_1 + \lambda_i \pi_2$$

where λ_i is some other "factor" (often there are 2 or more of these). A typical additional factor is included that reflects how a given asset covaries with a portfolio of small stocks, or with a portfolio that is mainly bonds (versus mainly stocks).

3. One key use of equation (*) is to think about how to discount the expected returns from different assets. For example, if we are thinking about an asset that is priced at P_0 today (period 0) and if the selling price 1 period in the future is P_1, then the expected return from holding the asset is

$$\{ E[P_1] - P_0 \} / P_0 .$$

If the asset has a given β, then the CAPM says that this expected return should equal $\pi_0 + \beta \pi_1$. Thus, we get

$$\{ E[P_1] - P_0 \} / P_0 = E[P_1] / P_0 - 1 = \pi_0 + \beta \pi_1$$

$$\Rightarrow P_0 = E[P_1] / (1 + \pi_0 + \beta \pi_1)$$

This very simple equation has many immediate implications.

4. Consider a very short holding period (e.g. 1 week). We can almost ignore the discounting and we get the implication

(**) $P_0 = E[P_1]$.

This says that the price today has to be the expected value of the price next week. A "stochastic process" is a sequence of random variables $\{ y_1, y_2, y_3, ... \}$. (An example is the height of the Nile River at a given location on June 1. Another example is the closing price of the S&P 500 on Friday). You may have heard the term "random walk". Roughly speaking, a random walk is a stochastic process with

$$E[y_{t+1} \mid t] = y_t$$

i.e., the best forcast of the value one period ahead is the value today. So sometimes people say that asset prices should follow a "random walk".

5. Equation (**) is sometimes called the 'efficient market hypothesis'. The key insight of this equation is that all the information we have to forecast the value of the asset tomorrow **is built in to the price today**. A very practicle example is the following. Suppose you think that Google stock will be worth $X at the end of June. Then you should be willing to pay very close to $X for the stock now (subject only to the discount factor).

6. Suppose (**) is true. Then the realized gain on buying the stock today and selling at the end of the holding period is

$$P_1 - P_0 = P_1 - E[P_1]$$

But the deviation of a random variable from its expectation **has to be "unpredictable"**. This implies that techniques like drawing charts (so called "technical analysis") cannot work!

7. Suppose new information comes out about an asset. A nice example would be news concerning a drug company (e.g. Merck): the news could be about a prospective drug that is being evaluated in a randomized clinical trial, or about the discovery of side effects, or about a decision of the FDA. Equation (**) says the "news" should cause an instantaneous adjustment of the stock price, either up or down. Likewise, economy-wide news, like the results of a Federal Reserve "Open Market" meeting, should cause the market as a whole to adjust up or down instantaneously, as people adjust their expectations.

8. This leads to the idea of an "event study". If you are trying to evaluate the effect of "news" on the value of a firm, you look at the so-called "excess returns" on the firm's shares:

$$XR_{it} = (P_{it} - P_{it-1})/P_{it-1} - \beta_i (M_{it} - M_{it-1})/M_{it-1}$$

where P_{it} is the value of a share in the firm under study at the close of the market on day t, and M_{it} is the value of the market index at the close of the market on day t.

The "cumulative excess return" is defined as the cumulative return over some horizon

$$CXR(t) = XR_{i1} + XR_{i2} + XR_{i3} + \ldots + XR_{it}$$

starting from a period "0" some time before the news, e.g. 7-14 days before.

One then plots CXR(t) from the starting period to several days after the release of "news". Ideally, there should be no systematic pattern before the news comes out; followed by a "jump" one the day the news is released, followed by unsystematic movements. A good reference for this technique is

W. Craig McKinley, Event Studies in Finance and Economics, Journal of Economic Literature, March 1997.

This is a plot from McKinley showing a typical pattern of cumulative excess returns for averages of events that are "good news", "bad news", and no news.

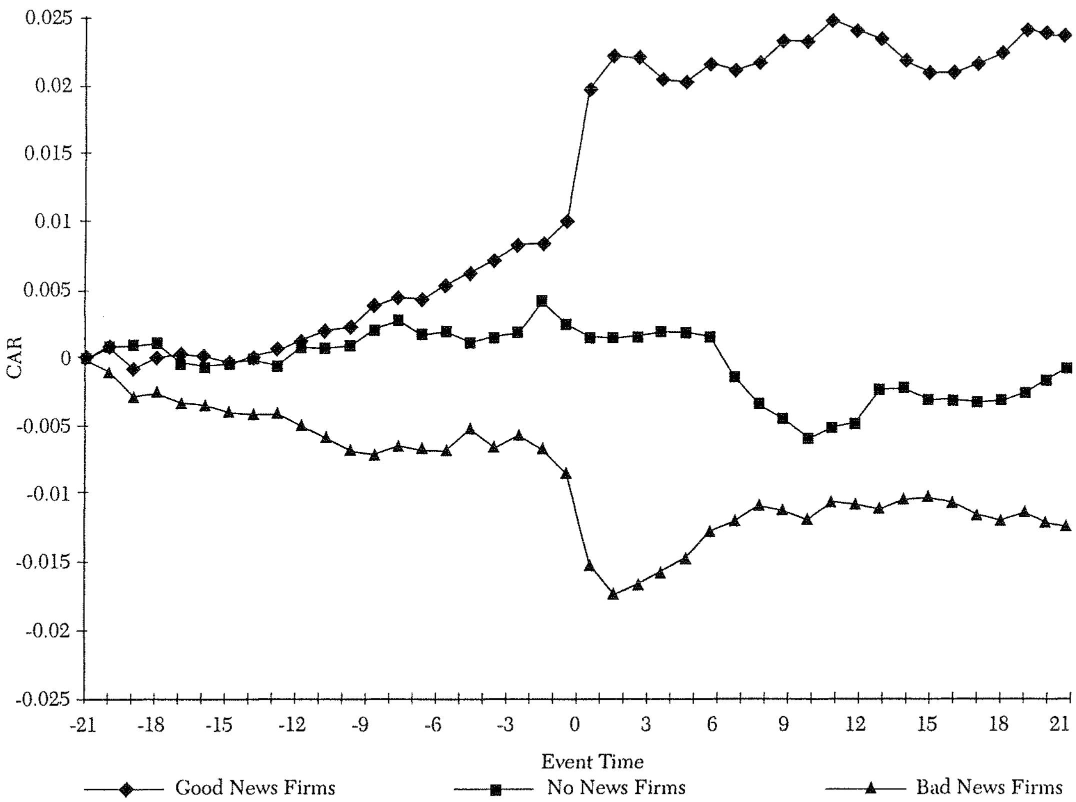


Figure 2a. Plot of cumulative abnormal return for earning announcements from event day -20 to event day 20. The abnormal return is calculated using the market model as the normal return measure.

9. An implication of the efficient markets hypothesis is that on average there are no advantages to following the advice of advisors (at least, adjusting for the excess risk of the portfolios they recommend). Many people believe that the high returns earned by some funds in some periods are just strings of "luck".

Here are some tables showing this idea from a paper Burton Malkiel, "The Efficient Market Hypothesis and Its Critics", Journal of Economic Perspectives, Winter 2003.

Exhibit 5

Percentage of Large Capitalization Equity Funds Outperformed by Index Ending 6/30/2002

	1 year	3 years	5 years	10 years
S&P 500 vs. Large Cap Equity Funds	63%	56%	70%	79%
Wilshire 5000 vs. Large Cap Equity Funds	72%	64%	69%	74%

Source: Lipper Analytic Services.
Note: All large capitalization mutual funds in existence are covered with the exception of "sector" funds and funds investing in foreign securities.

Exhibit 6

Median Total Returns Ending 12/31/2001

	10 years	15 years	20 years
Large Cap Equity Funds	10.98%	11.95%	13.42%
S&P 500 Index	12.94%	13.74%	15.24%

Source: Lipper Analytic Services, Wilshire Associates, Standard & Poor's and The Vanguard Group.

Exhibit 8

Getting Burned by Hot Funds

	1998–1999		2000–2001	
Fund Name	Rank	Average Annual Return	Rank	Average Annual Return
Van Wagoner:Emrg Growth	1	105.52	1106	−43.54
Rydex:OTC Fund;Inv	2	93.43	1103	−36.31
TCW Galileo:AGr Eq;Instl	3	92.78	1098	−34.00
RS Inv:Emerg Growth	4	90.19	1055	−26.17
PBHG:Large Cap 20	5	84.56	1078	−29.03
Janus Olympus Fund	6	77.24	1061	−27.03
Van Kampen Aggr Gro;A	7	76.70	1067	−28.04
Janus Mercury	8	76.31	1057	−26.35
PBHG:Sel Equity	9	76.21	1097	−33.19
WM:Growth;A	10	74.77	1046	−25.82
Berger New Generatn;Inv	11	73.31	1107	−45.96
Janus Enterprise	12	72.28	1101	−35.40
Janus Venture	13	72.22	1091	−30.89
Fidelity Aggr Growth	14	70.56	1105	−38.02
Janus Twenty	15	69.09	1090	−30.83
Amer Cent:New Oppty.	16	67.64	1033	−24.11
Morg Stan Sm Cap Gro;B	17	66.59	1102	−35.96
Van Kampen Emerg Gro;A	18	65.67	1021	−22.70
TCW Galileo:SC Gro;Instl.	19	64.87	1099	−34.77
BlackRock:MdCp Gro;Instl.	20	64.44	1009	−22.18
Average Fund Return		76.72		−31.52
S&P 500 Return		24.75		−10.50

Source: Lipper Analytic Services and Bogle Research Institute, Valley Forge, PA.

Economics 101A
Public and "Near Public" Goods

A "pure" public good is a good like public radio that has 2 properties:

1) the amount of the good consumed by one person has no effect on the availability of the good for others. This is called the"no rivalry" or "no congestion" condition.

2) people cannot be prevented from consuming the good. This is called the "non-exclusionary" condition.

Sometimes condition 1 is true but 2 is not. This is arguably the case for "intellectual property" distributed by the internet. If I download a piece of music, or software, my use does not affect your use.

Other examples of things that are "close" to public goods:
- parks and wildlife reserves (in some cases these can become congested)
- national defense

There are many things that people call "public goods" but really aren't: for example schools, which are subject to congestion and are also excludeable.

<u>Optimal Provision of Goods with No-rivalry Characteristics</u>

Consider a public good which can be varied in quantity. Let x be the quantity provided, and assume that the cost per unit of x is \$p.

An economy has n consumers, i=1,...n. Each has income y_i. Consumer i pays a tax t_i to pay for the public good. The utility of i is $U^i(c_i , x) = U^i(y_i - t_i , x)$.

<u>Case 1:</u> 1 consumer. $x = t_1/p$

The problem is: $\max_{\{t_1\}} U^1(y_1 - t_1 , t_1/p)$

foc: $- U^1_c(y_1 - t_1 , t_1/p) + 1/p\, U^1_x(y_1 - t_1 , t_1/p) = 0$

=> $U^1_x(y_1 - t_1 , t_1/p) / U^1_c(y_1 - t_1 , t_1/p) = p$

i.e. $MRS^1 (y_1 - t_1 , t_1/p) = p$

Recall that MRS^1 is 1's "willingness to pay" for the last unit of x, in units of good "c".

Case 2: 2 consumers. $x = (t_1+t_2)/p$

The problem is: $\max_{\{t_1, t_2\}} U^1(y_1 - t_1, (t_1+t_2)/p)$ s.t. $U^2(y_2 - t_2, (t_1+t_2)/p) \geq k_2$.

Why? Any social optimum must maximize 1's utility subject to whatever utility 2 is getting. This is called the condition of "pareto optimality". (If its not true, we could re-allocate resources and make both better off). Varying the minimum level of k_2 traces out a full range of potential social optima.

Set up the lagrangean

$$L(t_1, t_2, \lambda; k_2) = U^1(y_1 - t_1, (t_1+t_2)/p) + \lambda\{ U^2(y_2 - t_2, (t_1+t_2)/p) - k_2 \}$$

From the 'envelope theorem' we know that the rate of change of the maximized value of U^1 with respect to k_2 is $\partial L(t_1, t_2, \lambda; k_2)/\partial k_2 = -\lambda$. So λ will be positive. A higher value of λ puts more "weight" on 2's outcome.

FOC:

$$\partial L/\partial t_1 = -U^1_c(y_1 - t_1, (t_1+t_2)/p) + 1/p\, U^1_x(y_1 - t_1, (t_1+t_2)/p) + \lambda/p\, U^2_x(y_2 - t_2, (t_1+t_2)/p) = 0$$

$$\partial L/\partial t_2 = -\lambda U^2_c(y_2 - t_2, (t_1+t_2)/p) + 1/p\, U^1_x(y_1 - t_1, (t_1+t_2)/p) + \lambda/p\, U^2_x(y_2 - t_2, (t_1+t_2)/p) = 0$$

Note that the second and third terms are the same in each of these. So we get

$$U^1_c(y_1 - t_1, (t_1+t_2)/p) = \lambda U^2_c(y_2 - t_2, (t_1+t_2)/p) \quad \text{or} \quad \lambda = U^1_c/U^2_c.$$

The intuition for this is that the social planner can re-arrange taxes on 1 and 2 to keep x constant. If 1 pays a dollar less in taxes the gain in utility is U^1_c. If 2 pays a dollar less in taxes the gain to 2 is U^2_c. At the optimum, each additional gain in utility of 2 is evaluated as a λ unit increase in utility of 1.

The first of the FOC can be written (simplifying notation) as:

$$U^1_c = 1/p\, U^1_x + \lambda/p\, U^2_x = 1/p\, U^1_x + \{ U^1_c/U^2_c \}\, 1/p\, U^2_x$$

$$= 1/p\, \{ U^1_x + U^1_c\, U^2_x/U^2_c \}.$$ Divide both sides by U^1_c and multiply by p

$$p = U^1_x / U^1_c + U^2_x/U^2_c = MRS^1 + MRS^2.$$

This says that the optimum choice of x has the property that p equals the sum of the willingness to pay of 1 and 2!

Case 3: n consumers. $x= (t_1+t_2+t_3+....+t_n)/p$

Problem is

$$\max_{\{t_1, ...t_n\}} U^1(y_1 - t_1 , (t_1+t_2+t_3+....+t_n)/p) \quad \text{s.t.} \quad U^2(y_2 - t_2 , (t_1+t_2+t_3+....+t_n)/p) \geq k_2,$$

$$U^3(y_3 - t_3 , (t_1+t_2+t_3+....+t_n)/p) \geq k_3,$$

$$.... \quad U^n(y_n - t_n , (t_1+t_2+t_3+....+t_n)/p) \geq k_n .$$

This is the n-person version of pareto optimality. Given mimumum utility values for the other n-1 people, the choices of taxes must maximize 1's utility.

$$L = U^1(y_1 - t_1 , (t_1+t_2+t_3+....+t_n)/p) + \lambda_2\{ U^2(y_2 - t_2 , (t_1+t_2+t_3+....+t_n)/p - k_2 \} +$$

$$+ \lambda_3\{ U^3(y_3 - t_3 , (t_1+t_2+t_3+....+t_n)/p - k_3 \} + ...$$

$$..... + \lambda_n\{ U^n(y_n - t_n , (t_1+t_2+t_3+....+t_n)/p - k_n \} .$$

FOC:

$$\partial L/\partial t_1 = - U^1_c + 1/p\, U^1_x + \lambda_2/p\, U^2_x + \lambda_3/p\, U^3_x .... + \lambda_n/p\, U^n_x = 0$$

$$\partial L/\partial t_2 = - \lambda_2 U^2_c + 1/p\, U^1_x + \lambda_2/p\, U^2_x + \lambda_3/p\, U^3_x .... + \lambda_n/p\, U^n_x = 0$$

$$\partial L/\partial t_3 = - \lambda_3 U^3_c + 1/p\, U^1_x + \lambda_2/p\, U^2_x + \lambda_3/p\, U^3_x .... + \lambda_n/p\, U^n_x = 0$$

....

$$\partial L/\partial t_n = - \lambda_n U^n_c + 1/p\, U^1_x + \lambda_2/p\, U^2_x + \lambda_3/p\, U^3_x .... + \lambda_n/p\, U^n_x = 0$$

Note that in each FOC, all the terms involving the U^j_x 's are the same.

So we must have

$$(*) \qquad U^1_c = \lambda_2 U^2_c = \lambda_3 U^3_c = ... = \lambda_n U^n_c .$$

Thus: $\lambda_2 = U^1_c / U^2_c$, $\lambda_3 = U^1_c / U^3_c$, ... $\lambda_n = U^1_c / U^n_c$.

Substitute for the λ's into the first of the FOC's and we get

$$U^1_c = 1/p \{ U^1_x + \{U^1_c /U^2_c\}U^2_x + \{U^1_c /U^3_c\}U^3_x + ... + \{U^1_c /U^n_c\}U^n_x \}.$$

Dividing by U^1_c and multiplying by p, we get

$$p = U^1_x / U^1_c + U^2_x / U^2_c + U^3_x / U^3_c + ... + U^n_x / U^n_c .$$

$$= MRS^1 + MRS^2 + MRS^3 + ... MRS^n .$$

Again, as in the 2-person problem, p = the sum of the willingnesses to pay.

Implications

1) For a good that has no rivalry in consumption, the optimal provision of the good has the property that the marginal cost (p) equals the SUM of the willingnesses to pay. This is called the "Samuelson condition" because it was derived by the great American economist Paul Samuelson in 1954.

2) A simple market mechanism won't necessarily achieve the optimality condition. In the case of a non-excludeability condition, in fact, it is hard to model why anyone is willing to contribute voluntarily (though people do). Thus, provision of pure public goods is usually left to political mechanisms.

3) In the case of an excludeable good, like a piece of software, a "per user" use fee may be reasonable. Note that the producer gets the SUM of the user fees.

4) For issues like deciding how much to devote to wilderness areas, people sometimes suggest polling the public and asking how much they would be willing to pay to expand the wilderness, or sell it off. This is controversial because its not clear whether people really understand the questions, or tell the truth. Moreover, goods like wilderness areas are valued in a passive way, since most people will never experience them first hand. Unlike ordinary consumer goods, there is no observable behavior that we can actually trace back to a person's willness to pay. Despite these issues, this method, known as "Contingent Valuation" was used to value the environmental damage (or lost passive use) attributed to the Exxon Valdez oil spill.

Aside on Social Optimum with "Ordinary" Goods

You may be wondering how the idea of a social optimum works with "regular" goods. Lets consider the case of trying to decide how to allocate an ordinary good, x. The government collects a tax t_j on the jth consumer, and allocates him/her an amount x_j of the good. The budget constraint for the government in this case is $t_1+t_2+t_3+....+t_n = p\,(x_1+x_2+x_3+....+x_n)$, where as before p is the price of the good x. Assume as before that consumer j has income y_j and uses his/her after-tax income to buy an amount $c_j = y_j - t_j$ of the base good.

For this problem the social allocation problem is

$$\max_{\{t_1, \ldots t_n,\, x_1, \ldots x_n\}} \quad U^1(y_1 - t_1, x_1) \quad \text{s.t.} \quad U^2(y_2 - t_2, x_2) \geq k_2, \ldots, \; U^n(y_n - t_n, x_n) \geq k_n$$

$$\text{and} \quad t_1+t_2+t_3+....+t_n = p\,(x_1+x_2+x_3+....+x_n)\,.$$

Lagrangean is:

$$L = U^1(y_1 - t_1, x_1) + \lambda_2\{U^2(y_2 - t_2, x_2) - k_2\} +$$

$$+ \lambda_3\{U^3(y_3 - t_3, x_3) - k_3\} + ... + \lambda_n\{U^n(y_n - t_n, x_n) - k_n\}$$

$$+ \mu\,\{t_1+t_2+t_3+....+t_n - p(x_1+x_2+x_3+....+x_n)\}\,.$$

FOC:

$$\partial L/\partial t_1 = -U^1_c + \mu = 0$$

$$\partial L/\partial t_2 = -\lambda_2 U^2_c + \mu = 0$$

.....

$$\partial L/\partial t_n = -\lambda_n U^n_c + \mu = 0$$

$$\partial L/\partial x_1 = U^1_x - \mu p = 0$$

$$\partial L/\partial x_2 = \lambda_2 U^2_x - \mu p = 0$$

....

$$\partial L/\partial x_n = \lambda_n U^n_x - \mu p = 0$$

From the first condition we get that $\mu = U^1_c$.
Combining the first with the FOC for t_j we get $\lambda_j U^j_c = U^1_c \Rightarrow \lambda_j = U^1_c/U^j_c$.

Combining the FOC's for t_1 and x_1, we get

$U^1_x = \mu p \Rightarrow U^1_x / U^1_c = p$ i.e. $MRS^1 = p$

Finally, combining the FOC's for x_j with the expressions for λ_j, and using $\mu = U^1_c$, we get

$\lambda_j U^j_x = \mu p \Rightarrow \{ U^1_c/U^j_c \} U^j_x = U^1_c p \Rightarrow U^j_x/U^j_c = p$ i.e. $MRS^j = p$ for all j

Thus, at a social optimum, we have

(***) $MRS^1 = MRS^2 = ... = MRS^n = p$.

Notice that this is the same condition we would get if we opened up a market in good x, and charged a price equal to p. However, to get to a particular social optimum we would have to redistribute incomes (via choosing the t's).

It can be shown that

a) any particular social optimum can be achieved by having a "free market" in good x, and using taxes to redistribute incomes

b) for any given distribution of incomes, setting all taxes to 0 achieves one possible pareto optimum. This may not be one that people particularly like (it will give highest utility to the person with highest income) but it is "efficient" in the sense that it satisfies the optimality condition (***).

Economics 101A
Externalities

Externalities arise when the consumption or production of a good by one economic agent causes a "side-effect" on others. Examples include the air pollution caused by the burning of fossil fuels, the playing of loud music, etc. Externalities can be positive: a classic example is bees, which are needed to pollinate fruit trees.

We will focus this lecture on air pollution, which has a "public good" character: air quality affects the entire population of an area.

Externalities in Consumption

We will use an extended version of the model used in the last lecture. Suppose that consumers care about three things: consumption of a basic good (c), consumption of a good with an externality (x), and the level of the externality, z. Think of x as gasoline, and z as the amount of smog in the air. The economy has n consumers, i=1,...n. Each has income y_i, and chooses c_i and x_i. The level of z is determined by the sum of all the x's:

$$z = \alpha\,(x_1 + x_2 + \ldots + x_n) = \alpha\,\Sigma_i x_i$$

where α is the amount of smog produced per gallon of gas consumed. Let p represent the price (and the marginal cost) of x_i. The utility of consumer i is

$$U^i(c_i\,,\,x_i,\,z) = U^i(y_i - px_i\,,\,x_i\,,\,\alpha\,\Sigma_i x_i\,)$$

We are assuming that $U^i_c > 0$, $U^i_x > 0$, and $U^i_z < 0$ so z is "bad". Notice that z acts like a public good: i's "consumption" of z has no effect on the amount of z available for other consumers.

Market Equilibrium: Consumer 1 takes price p as given. Realizes $z = \alpha\,(x_1 + x_2 + \ldots + x_n)$ but takes x_2, x_3, x_n (gas consumption of others) as given.

The problem is: $\max_{\{x_1\}} \; U^1(y_1 - px_1\,,\,x_1,\,\alpha\,\Sigma_i x_i\,)$

FOC: $-pU^1_c \;+ 1/p\,U^1_x \;+ \alpha\,U^1_z \;= 0$

$$\Rightarrow \quad U^1_x / U^1_c \;=\; p \;-\; \alpha\,U^1_z/U^1_c$$

i.e. $MRS^1\,(x,c) = \;p \;-\; \alpha\,U^1_z/U^1_c$

The consumer should set her/his MRS for x relative to c equal to $p - \alpha\,U^1_z/U^1_c$. If $U^i_z < 0$, the second term is positive, so the consumer should act "as if" the price of x is actually higher than p. The price gap is $\alpha\,U^1_z/U^1_c$, which is the product of α (the rate of production of z per unit of x) and the marginal willingness to pay for clean air, U^1_z/U^1_c.

Social Optimum

A social planner has to allocate x_i and collect a tax t_i (for i=1,...n) that balances the government's costs: $t_1+t_2+t_3+....+t_n = p\,(x_1+x_2+x_3+....+x_n)$. As in the last lecture we look for Pareto optima. The social planner tries to solve the problem:

$$\max_{\{t_1, \dots t_n, x_1, \dots x_n\}} U^1(y_1 - t_1, x_1, \alpha \Sigma_i x_i) \text{ s.t. } U^2(y_2 - t_2, x_2, \alpha \Sigma_i x_i) \geq k_2, \dots$$

$$U^n(y_n - t_n, x_n, \alpha \Sigma_i x_i) \geq k_n$$

$$\text{and} \quad t_1+t_2+t_3+....+t_n = p\,(x_1+x_2+x_3+....+x_n) .$$

Lagrangean is:

$$L = U^1(y_1 - t_1, x_1, \alpha \Sigma_i x_i) + \lambda_2\{ U^2(y_2 - t_2, x_2, \alpha \Sigma_i x_i) - k_2 \} + ...$$

$$+ \lambda_n\{ U^n(y_n - t_n, x_n, \alpha \Sigma_i x_i) - k_n \} +$$

$$+ \mu \{ t_1+t_2+t_3+....+t_n - p(x_1+x_2+x_3+....+x_n) \} .$$

FOC:

$$\partial L/\partial t_1 = - U^1_c + \mu = 0$$

$$\partial L/\partial t_j = -\lambda_j U^j_c + \mu = 0 \quad \text{(for j=2,3,...n)}$$

$$\partial L/\partial x_1 = U^1_x + \alpha U^1_z + \alpha \lambda_2 U^2_z + ... + \alpha \lambda_n U^n_z - \mu p = 0$$

$$\partial L/\partial x_j = \lambda_j U^j_x + \alpha U^1_z + \alpha \lambda_2 U^2_z + ... + \alpha \lambda_n U^n_z U^n_x - \mu p = 0 \quad \text{(for j=2,3,...n)}$$

From the first condition we get that $\mu = U^1_c$.
Combining this with the FOC for t_j we get $\lambda_j U^j_c = U^1_c \Rightarrow \lambda_j = U^1_c/U^j_c$ (*)

From $\partial L/\partial x_1 = 0$, we get

$U^1_x = \mu p - \alpha \{ U^1_z + \lambda_2 U^2_z + ... + \lambda_n U^n_z \}$. Divide through by $\mu = U^1_c$ and use (*)

$$\Rightarrow \quad U^1_x/U^1_c = p - \alpha \{ U^1_z /U^1_c + U^2_z/U^2_c + ... + U^n_z/U^n_c \} \quad (**).$$

Finally, from $\partial L/\partial x_j = 0$ we get

$\lambda_j U^j_x = \mu p - \alpha \{ U^1_z + \lambda_2 U^2_z + ... + \lambda_n U^n_z \}$. Divide through by $\mu = U^1_c$ and use (*)

$$\Rightarrow \quad U^j_x/U^j_c = p - \alpha \{ U^1_z /U^1_c + U^2_z/U^2_c + ... + U^n_z/U^n_c \} \quad (***)$$

Notice that (**) and (***) imply that everyone has to set

$$MRS^j = U^j_x/U^j_c = p + \tau$$

where $\tau = -\alpha v$, where $v = U^1_z/U^1_c + U^2_z/U^2_c + ... + U^n_z/U^n_c$ is the SUM of the marginal willingnesses to pay for improved air quality.

Comparing the Market and the Social Optimum

Market: $MRS^j(x,c) = p - \alpha U^j_z/U^j_c$

Social Opt: $MRS^j(x,c) = p - \alpha \{ U^1_z/U^1_c + U^2_z/U^2_c + ... + U^n_z/U^n_c \} = p + \tau$

So in the social optimum, person j has to take account of the effect of his/her gas consumption on everyone else whereas in the market he/she only cares about him/herself.

The sum $p + \tau$ is the "social marginal cost" of consuming gas. It exceeds the "private cost" p if α is non-zero, and if there is some value to clean air (i.e., $U^j_z/U^j_c < 0$ for all j's.) Realistically, α is very small but n is very large, so $\alpha U^j_z/U^j_c$ is negligible but τ could be important.

In the 1920's the English economist Arthur C. Pigou figured out that you could "correct" an externality by taxing the activity that creates it with a tax τ. We have shown that the optimal Pigouvian tax for a consumption externality that affects the entire population is

$$\tau = \alpha \, \Sigma_i \text{ (i's willingness to pay for a marginal reduction in the externality) .}$$

Other examples

1. Taxes for "wear and tear" on the road. The usual justification for a gas tax (apart from the air pollution effect) is that driving causes wear and tear. If the wear and tear of different cars and trucks per mile of driving is proportional to the amount of gas they use per mile, a Pigouvian tax on gas is sensible.

2. Taxes on cigarettes are sometimes justified because they tax second hand smoke.

3. Some people have proposed taxing foods that cause obesity. This is a more complicated case, but the jist of the argument is that health care costs for people over 65 (which is when most costs are incurred) are heavily subsidized through the Medicare program. Thus, if someone eats too much, and gets diabetes later in life, they contribute to the Medicare bill, which we all pay.

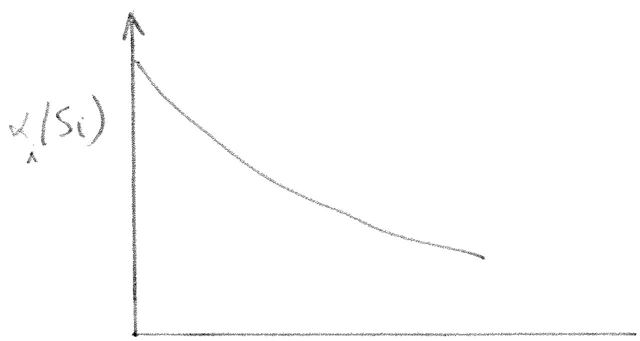
Production Externalities

We will consider a very simple example of a production externality. The example is motivated by the electric power generation industry, which (in most places) uses coal to create electricity.

We will assume there are n plants, i=1,2,...n. Plant i has the cost function $C_i(s_i, y_i)$, where y_i is a measure of kilowatt hours produced by the plant, and s_i is a choice variable representing the choice of things that affect the amount of SO2 (sulphur dioxide) produced. For example, s_i could represent the choice of what kind of coal to use (more expensive coal from the Western US, which burns cleaner, or cheaper eastern coal), or the choice of what kind of scrubber to install. The total SO2 emitted by the plant is

$$z_i = y_i \, \alpha_i(s_i).$$

We assume that $\alpha_i(s_i)$ looks like this:



$\alpha_i'(s_i) < 0$

$\alpha_i''(s_i) > 0$

Suppose the sum of the population's willingness to pay to avoid SO2 is v, and that the value of a kilowatt hour of electricity is p. We will think of an industry regulator as trying to maximize:

$$p \, \Sigma_i \, y_i \; - \; \Sigma_i \, C_i(s_i, y_i) \; - \; v \, \Sigma_i \, y_i \, \alpha_i(s_i)$$

Notice this is the total surplus created by the industy, valuing SO2 at -v per unit. (We could alternatively set up the problem by having U-functions for all the local people, who each use electricity, consume another good c, and dislike the total SO2 in the air. As an exercise try to set it up that way).

The FOC for maximizing w.r.t. y_i is

(1) $\quad p - \partial C_i(s_i, y_i)/\partial y_i - v \, \alpha_i(s_i) = 0 .$

This says we should choose output at plant i so that

$$\partial C_i(s_i, y_i)/\partial y_i + v\,\alpha_i(s_i) = p$$

The two terms on the left are the "marginal social cost" of production at plant i. We want to set this equal to p, the social value of a kilowatt hour.

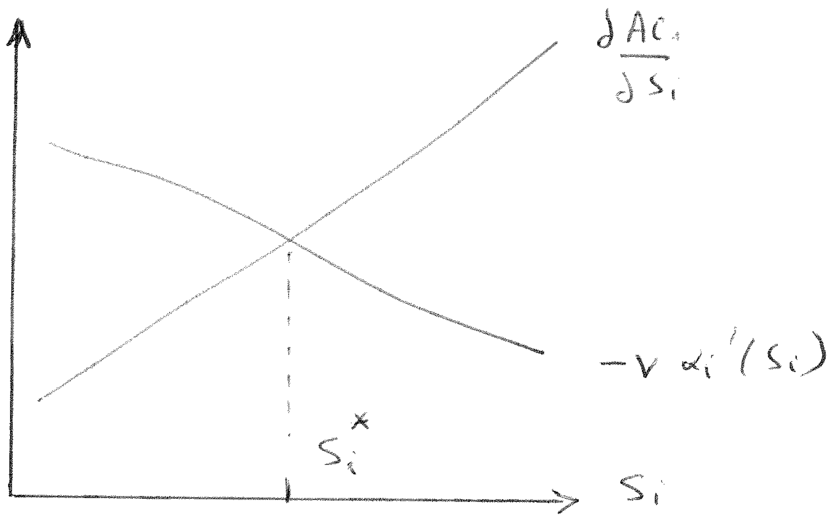
The FOC for maximizing with repect to α_i is

(2) $$-\partial C_i(s_i, y_i)/\partial s_i - v\, y_i\, \alpha_i'(s_i) = 0.$$

This says that

$$1/y_i\ \partial C_i(s_i, y_i)/\partial s_i = \partial AC_i(s_i, y_i)/\partial s_i = -v\,\alpha_i'(s_i).$$

The optimal choice sets the increase in average cost equal to the marginal value of the reduced pollution per unit of output. Assuming that $AC_i(s_i, y_i)$ is convex in s_i (so that at a higher level of s, an additional increase in s has a bigger effect on AC) and that α_i looks like we showed above, we have:



How can the regulator get plants to "do the right thing"?

Method 1:

Tax each plant an amount v per ton of SO2 produced, and pay them p per kwh. The manager of plant i will then try to solve the problem

$$p\,y_i - C_i(s_i, y_i) - v\, y_i\, \alpha_i(s_i)$$

which has the same FOC as (1) and (2) above. This is the Pigouvian tax idea.

Method 2.

Allocate the industry a series of $SO2$ emission rights. Each emission right allows the plant that holds it to produce a ton of $SO2$. Let the plants trade the emission rights between themselves. Buy electricity at price p.

Suppose that an emission right has value $q>0$. A plant manager who owns K tons of emission rights will then try to maximize

$$p\, y_i \quad - C_i\,(s_i, y_i\,) \;+\; \{\, Kq \;\; - q\;\, y_i\;\, \alpha_i(s_i\,)\,\}$$

The term in { } is the value of the emission rights he/she can sell to the market (or will have to buy). Notice that if $q=v$, the FOC for maximizing this are the same as (1) and (2). This is how $SO2$ is really regulated.

Why use method 2?

- in reality, no one knows that v to charge. So instead the regulator looks at total output of $SO2$ at some reference point, then issues a somewhat smaller number of permits (e.g, 80%). This method will ensure that $SO2$ is reduced by 20% "efficiently".

- firms like it better because they get the emission rights "for free". (Plants were distributed rights in the early 1990s, and allowed to trade them, but the rules requiring that they only emit $SO2$ up to the limit only took hold in 1995.

- it is claimed that enforcement is easier.

Economics 101A
Empirical Methods in Microeconomics

This lecture will provide an overview of how micro economists use real data to test alternative theories and (in some cases) estimate the relevent parameters of a particular model. The examples will be drawn from my own work in labor economics.

1. Experiments and Counterfactuals

Suppose you are interested in testing a prediction from microeconomic theory. To be concrete, we will consider four examples:

1) If single mothers who are currently on welfare are offered an *earnings subsidy*, will they work more?
2) If the supply of low-skilled workers in a local labor market is increased by an inflow of *immigrants*, will wages of low-wage native workers fall?
3) If the *minimum wage* is increased, will low-wage employers hire fewer workers?
4) If people without health insurance are provided *insurance*, will they use more health care services? Will they become healthier?

The "classical" scientific approach to such questions would be to conduct a randomized experiment. In such an experiment, a population whose behavior is to be studied would be randomly divided into two groups: the **treatment group,** who receive the treatment, and a **control group**, who do not. For the welfare question, the population would be single mothers currently on welfare. For the minimum wage question, the population would be employers. For the immigration question, the population would be cities (or other geographic units, like counties). For the final question, the population would be uninsured people. Note that it seems harder to imagine some of these experiments actually taking place than others.

Let's assume that you could conduct a randomized experiment on welfare mothers. (In reality, such an experiment was conducted in two Canadian provinces in the mid-1990s. We will look at the data shortly). What would you do? Presumably, you would tabulate the employment rates of the treatment group (Y_T) and the control group (Y_C) in some period after the subsidy was in place. You would then calculate the "treatment effect":

$$\Delta = Y_T - Y_C$$

The idea of a randomized experiment is that the behavior of the control group tells you what the treatment group would look like *in the absence of the treatment*. Randomization is key: if "treatment status" is really randomly assigned to the overall population, then except by accident, you would expect the two groups to exhibit the same behavior in the absence of treatment. The impact of "accidents" is minimized by having large numbers in the two groups. The behavior of the control group represents a **counterfactual** for assessing the effect of treatment. If a theory predicts that an subsidy will increase work effort, for example, then we want to test the null hypothesis $\Delta=0$ against the alternative that $\Delta>0$.

A randomized experiment is considered the "gold standard" for scientific evidence. The FDA, for example, requires drug companies to evaluate the efficacy of a new drug by a randomized experiment. The high status of randmized experiments arises from several features:

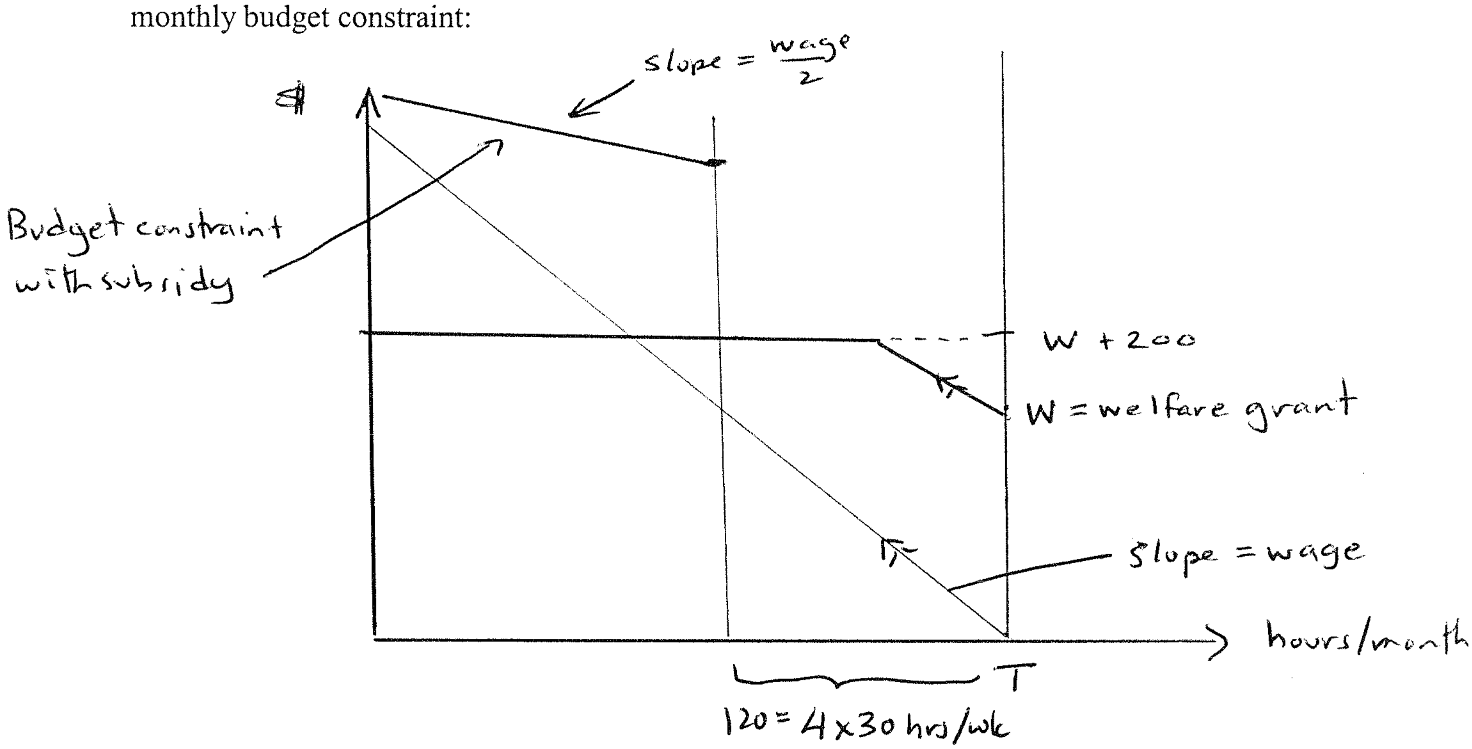
a) randomization ensures that Y_C is a *valid counterfactual*. So except for chance errors, the difference Δ is truly attributable to the treatment, not to some inherent differences between the treatment and control groups.

b) once the "experimental design" is set, the researcher's hands are tied. There is no room for weazling. (The experimental design is a full description of the population, the sample size, the randomization procedure, the treatment, and the data collection process).

c) because of a) and b), randomized experiments are easy to understand, and have a lot of credibility.

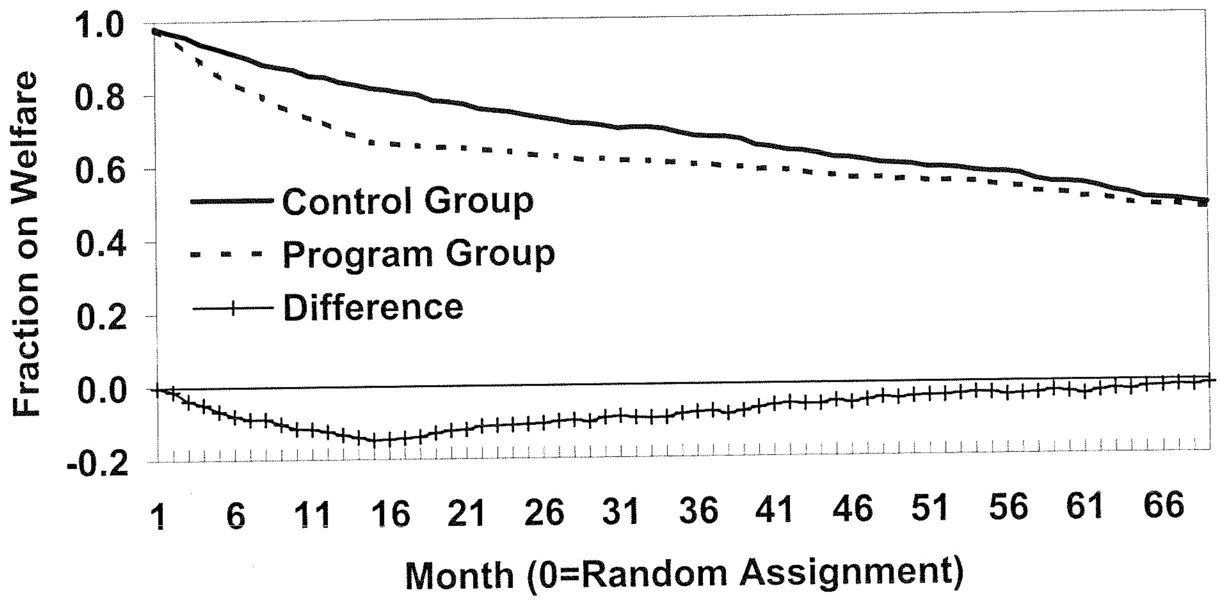
Example 1: The Self Sufficiency Project"

SSP was a randomized experiment conducted in Canada. One half of a sample of single parents who had been on welfare for at least a year was assigned to the treatment group. The other half were assigned to the control group. People in the control group were eligible to receive the regular welfare benefit, which was a fixed monthly sum (based on number of children and province, e.g. $712 per month for a mom with 1 child in New Brunswick). Welfare payments are reduced dollar for dollar if you have earnings over $200 per month. People in the treatment group could stay on welfare, but were offered an earnings subsidy S = ½ (M − E) where M is a monthly earnings target ($2500 per month) and E is actual earnings. So, if you earn $650 in a month, you receive a subsidy of $925. People could only get the subsidy if they worked at least 30 hours per week, for up to 3 years. They also had to receive their first subsidy payment within a year of random assignment or they forfeited all future eligibility. Here is a graph of the monthly budget constraint:

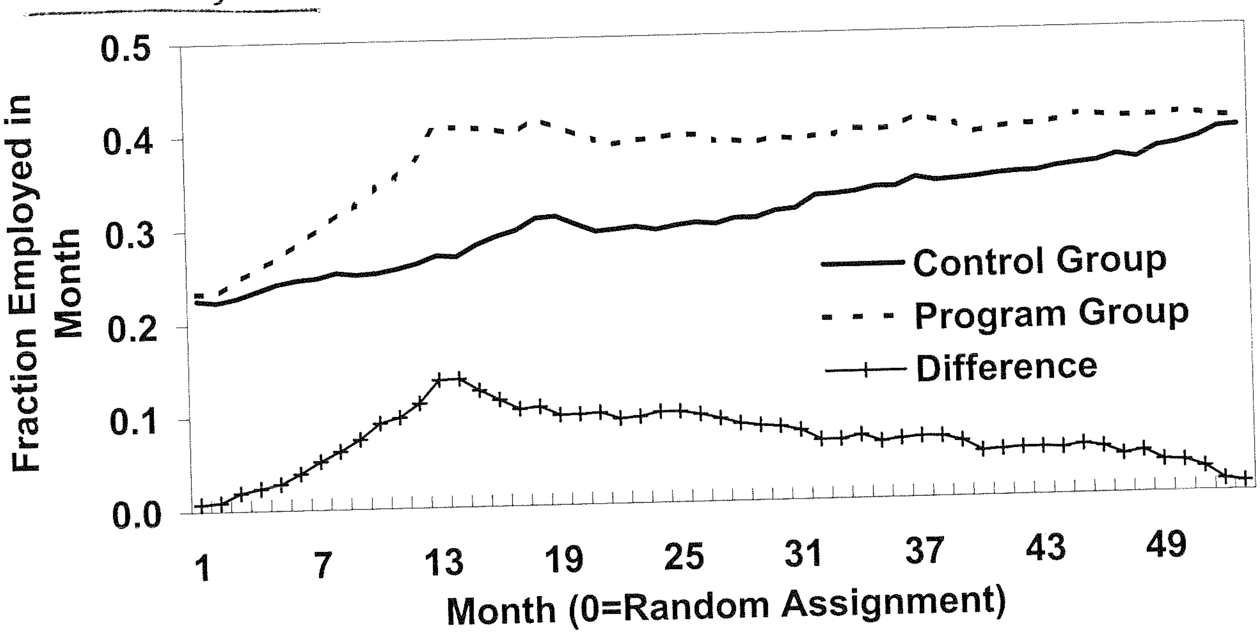


Here is a picture showing the fractions on welfare by month for the two groups, by months after random assignment, along with a picture average employment rates of the two groups.

(a) On welfare



(b) Employed



Source: D. Card and D. Hyslop, "Estimating the Effects of a Time-limited Earnings Subsidy for Welfare Leavers". NBER Working Paper 2004. Forthcoming in Econometrica.

2. Research Designs Based on "Natural" Experiments

Lots of times we cannot run an experiment, either because it would cost a lot, and be quite invasive (like SSP), or because it would be impractical. How can we proceed?

One approach is to think about events that happen, and consider whether an analysis of the event could be interpreted "as if" the event was a random experiment. A very simple example is a paper I wrote on the Mariel Boatlift. (Coincidentally, this spring is the 25th anniversary of the boatlift). In that paper, I looked at movements in wages, unemployment rates, and employment rates in Miami (where the Marielito's landed) and a "control group" based on 4 other cities: Tampa, Houston, Atlanta, and Los Angeles. A key difference between a true randomized experiment and a "natural experiment" is that treatment is not randomly assigned. So it is debatable whether the control group cities provide a valid counterfactual. To test that, I looked at trends in employment in Miami versus the average of the four cities in the period from 1970 to 1980: the two moved in close parallel. (Ironically, the editor made me cut this graph from the published paper!)

In a natural experiment setting, it may not happen that outcomes are exactly the same in the treatment and control cities, even before treatment. Let

$$\Delta(0) = Y_T(0) - Y_C(0)$$

represent the gap in some outcome (e.g. average wages) between the treatment and control group *before the treatment*, and let

$$\Delta(1) = Y_T(1) - Y_C(1)$$

represent the gap in some period after. Then we might want to look at the "difference-in-differences"

$$\Delta(1) - \Delta(0) = \{ Y_T(1) - Y_T(0) \} - \{ Y_C(1) - Y_C(0) \} .$$

This is the change in the treatment group *relative to the change in the control group*. The implicit assumption is that in the absence of treatment, the pre-existing gap $\Delta(0)$ would have held constant.

Example 2: The Mariel Boatlift

In the Boatlift, about 125,000 Cuban immigrants were transported on a flotilla of small boats to Miami, over the period from April to July of 1980. This represented an increase of about 7% in the Miami labor force – mainly in the ranks of the unskilled. One simple hypothesis is that such an inflow would lower wages of less-skilled natives in Miami. Here is a table showing data on outcomes of blacks in Miami relative to the comparison cities.

Table 3. Logarithms of Real Hourly Earnings of Workers Age 16–61 in Miami and Four Comparison Cities, 1979–85.

Group	*1979*	*1980*	*1981*	*1982*	*1983*	*1984*	*1985*
Miami:							
Whites	1.85 (.03)	1.83 (.03)	1.85 (.03)	1.82 (.03)	1.82 (.03)	1.82 (.03)	1.82 (.05)
Blacks	1.59 (.03)	1.55 (.02)	1.61 (.03)	1.48 (.03)	1.48 (.03)	1.57 (.03)	1.60 (.04)
Cubans	1.58 (.02)	1.54 (.02)	1.51 (.02)	1.49 (.02)	1.49 (.02)	1.53 (.03)	1.49 (.04)
Hispanics	1.52 (.04)	1.54 (.04)	1.54 (.05)	1.53 (.05)	1.48 (.04)	1.59 (.04)	1.54 (.06)
Comparison Cities:							
Whites	1.93 (.01)	1.90 (.01)	1.91 (.01)	1.91 (.01)	1.90 (.01)	1.91 (.01)	1.92 (.01)
Blacks	1.74 (.01)	1.70 (.02)	1.72 (.02)	1.71 (.01)	1.69 (.02)	1.67 (.02)	1.65 (.03)
Hispanics	1.65 (.01)	1.63 (.01)	1.61 (.01)	1.61 (.01)	1.58 (.01)	1.60 (.01)	1.58 (.02)

Note: Entries represent means of log hourly earnings (deflated by the Consumer Price Index—1980 = 100) for workers age 16–61 in Miami and four comparison cities: Atlanta, Houston, Los Angeles, and Tampa-St. Petersburg. See note to Table 1 for definitions of groups.

Source: Based on samples of employed workers in the outgoing rotation groups of the Current Population Survey in 1979–85. Due to a change in SMSA coding procedures in 1985, the 1985 sample is based on individuals in outgoing rotation groups for January–June of 1985 only.

Source: D Card, "The Impact of the Mariel Boatlift on the Miami Labor Market" *Industrial and Labor Relations Review*, January 1990.

3. Natural Experiments with Several Control Groups

In a natural experiment, one can never be sure that the control group provides a valid counterfactual. Sometimes it is possible to do more checking by having two or more different control groups. Then you can construct

$$DD_1 = \{ Y_T(1) - Y_T(0) \} - \{ Y_{C1}(1) - Y_{C1}(0) \}$$

$$DD_2 = \{ Y_T(1) - Y_T(0) \} - \{ Y_{C2}(1) - Y_{C2}(0) \}$$

and

$$DD_3 = \{ Y_{C2}(1) - Y_{C2}(0) \} - \{ Y_{C1}(1) - Y_{C1}(0) \}$$

where "C1" refers to control group 1, and "C2" refers to control group 2. If things are "working", you should see: $DD_1 = DD_2$, or equivalently, $DD_3=0$.

Example 3: The New Jersey Minimum Wage

In April 1992, the minimum wage rose from \$4.25 to \$5.05 per hour in the state of New Jersey. Elsewhere, it remained at \$4.25. The statute that raised the min. wage had been passed in the preceeding fall, and in anticipation Alan Krueger and I developed a survey of fast food stores in New Jersey and Pennsylvania. We surveyed a set of about 400 stores first in February-March of 1992 (just before the increase) and again in late fall. We were extremely careful to track down all the stores that were surveyed in the first round. The treatment group consisted of stores in NJ that had "starting wages" less than \$5.05 in wave 1. There were two control groups: stores in PA, and stores in NJ that were already paying relatively high wages (\$5.00 or more per hour in wave 1). Here is a table showing the comparisons of employment growth.

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33	20.44	−2.89	19.56	20.08	22.25	−2.69	−2.17
	(1.35)	(0.51)	(1.44)	(0.77)	(0.84)	(1.14)	(1.37)	(1.41)
2. FTE employment after, all available observations	21.17	21.03	−0.14	20.88	20.96	20.21	0.67	0.75
	(0.94)	(0.52)	(1.07)	(1.01)	(0.76)	(1.03)	(1.44)	(1.27)
3. Change in mean FTE employment	−2.16	0.59	2.76	1.32	0.87	−2.04	3.36	2.91
	(1.25)	(0.54)	(1.36)	(0.95)	(0.84)	(1.14)	(1.48)	(1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	−2.28	0.47	2.75	1.21	0.71	−2.16	3.36	2.87
	(1.25)	(0.48)	(1.34)	(0.82)	(0.69)	(1.01)	(1.30)	(1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	−2.28	0.23	2.51	0.90	0.49	−2.39	3.29	2.88
	(1.25)	(0.49)	(1.35)	(0.87)	(0.69)	(1.02)	(1.34)	(1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

[a]Stores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

[b]Difference in employment between low-wage (\$4.25 per hour) and high-wage (≥ \$5.00 per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

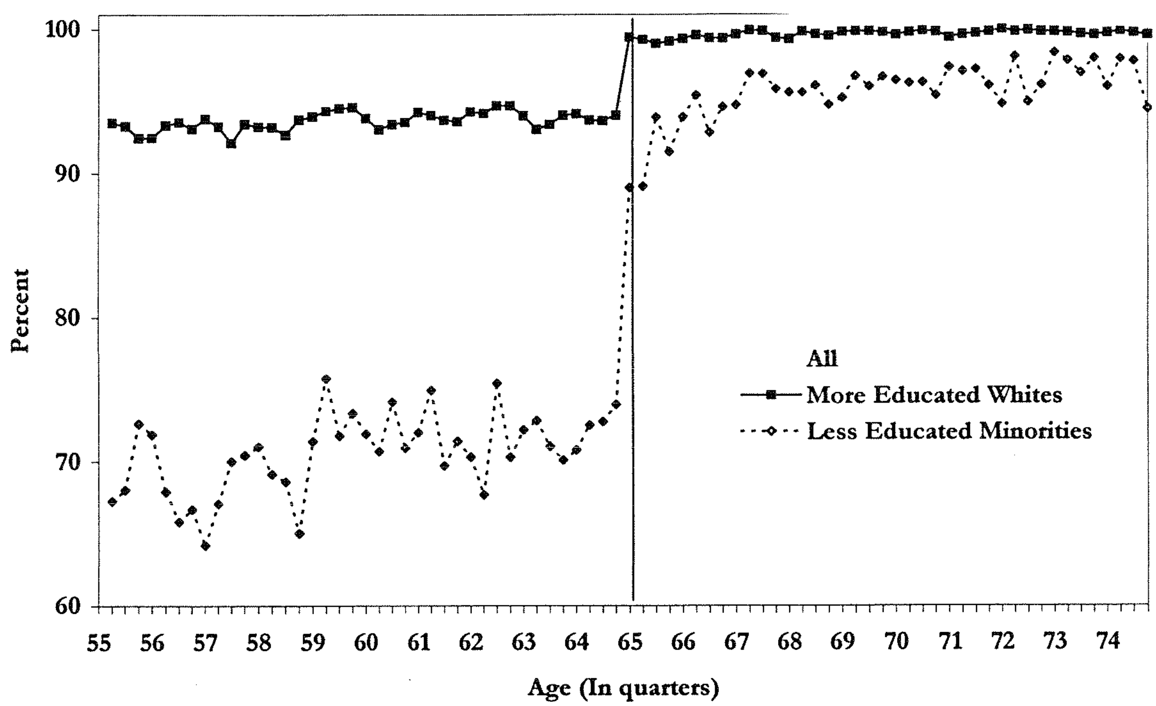
[c]Subset of stores with available employment data in wave 1 and wave 2.

[d]In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

Source: Card and Krueger, Myth and Measurement, Princeton Univ. Press, 1995.

4. The "Discontinuity" Reseach Design

Sometimes you can't find a good natural experiment, but it is still possible to think of a good counterfactual by looking at treatments that affect some groups, but not others who are "almost the same". A very good example is Medicare. When people who have worked for at least 10 years turn 65, they become eligible for "free" health insurance. (You also can get covered if your spouse worked 10 years). This age limit suggests that we compare people who are just a few months younger than 65, with those who are just a few months older. Here is graph showing the fractions of peple with health insurance, by age measured in quarters. The plots are for two groups: more educated whites (>12 years of education), and less educated minorities (blacks and Hispanics with <12 years of education).



The idea of the discontinuity design is that the rule that creates eligibility for those who reach their 65th birthday creates an "experiment": so we think about people just over 65 as the treatment group, and those just under 65 as the control group. There are some potential problems with this idea, depending on the application.

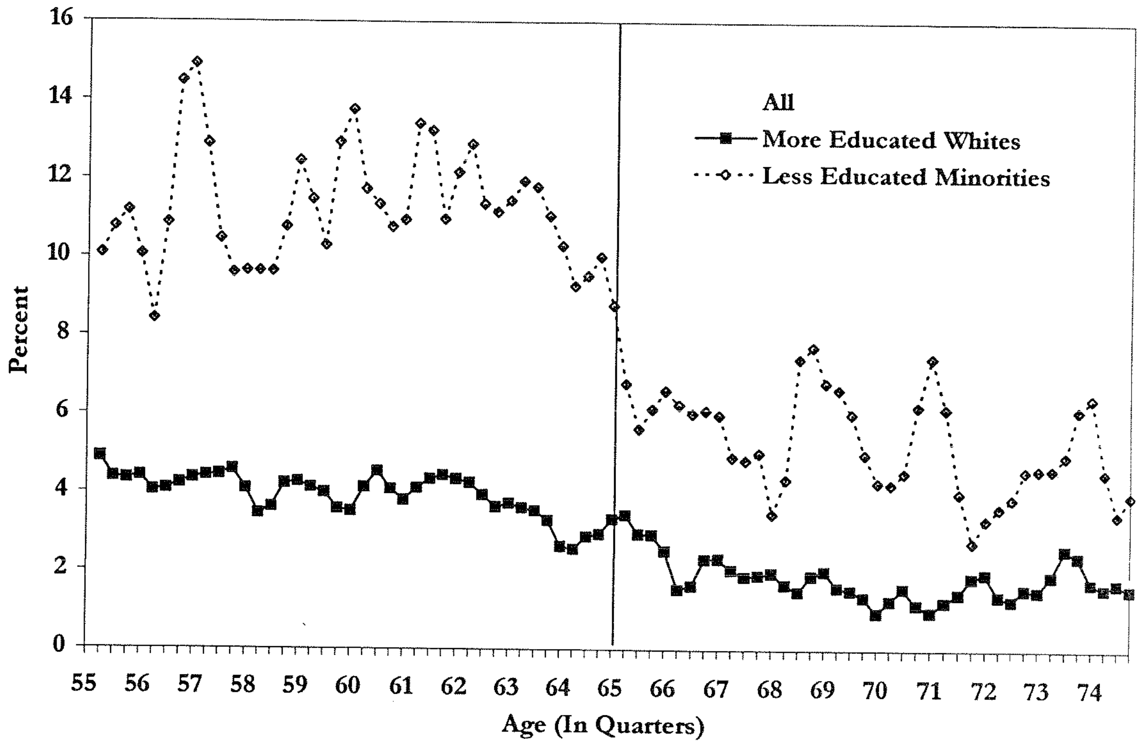
a) It may be that other things besides the primary treatment also shift at the same point. So normally you have to check very carefully that "other things" are very similar for people on either side of the boundary.

b) There may be an age trend in the outcome of interest, so even in the absence of treatment, people who are a little over 65 are a little different than those under 65. This can be checked by looking at the age profile of the outcome of interest.

c) If people know that Medicare is "coming" they may act differently when they are just under 65 than they would if there was no such rule.

Here are some graphs showing (a) the fractions of people who report that they did not get medical care last year because they could not afford it (b) the number of cataract surgeries by age of the patient in Florida. You can see some of the issues mentioned above in the age profile of cataracts.

Figure 5a: Percent Who Did Not Get Medical Care Last Year for Cost Reasons (NHIS)



Cataract Surgeries: Florida Outpatient Data, 1997-2002