DEMAND FOR AND SUPPLY OF HEALTH CARE

EE 474 Health Economics

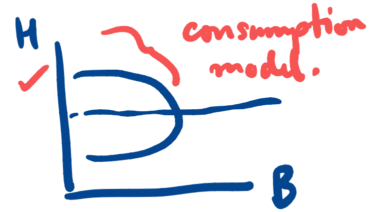
Semester 2/2021

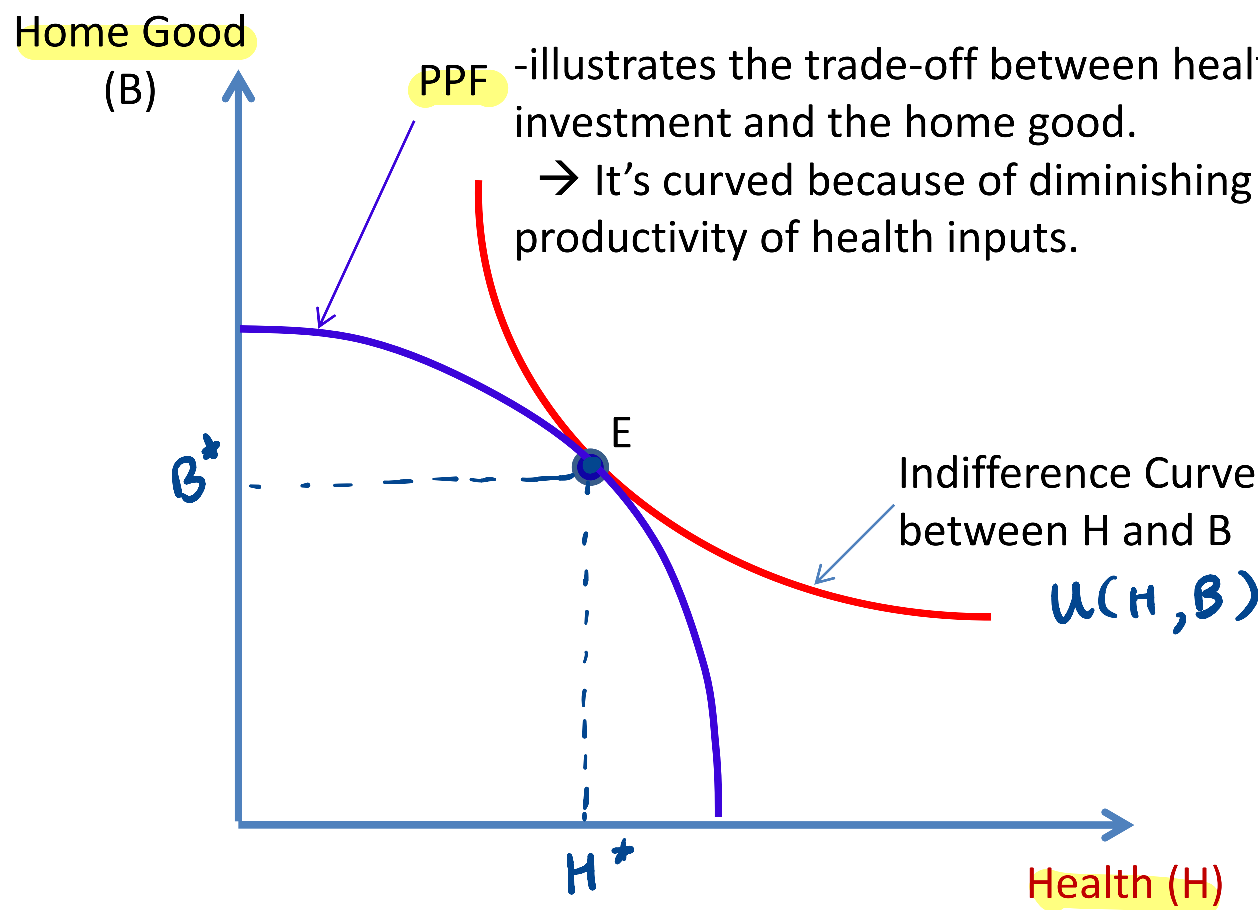
DEMAND FOR HEALTH CARE

Topics

- Demand for Health Capital (Revisited)
- Demand for Health and Health Care
- Demand for Health Care in the Standard Budget Constraint Model
- Comparative Statics - Factors affecting medical care. (health care).
- Empirical Studies
- Other Variables Affecting Demand

Recall: Demand for Health Capital





-illustrates the trade-off between health investment and the home good.

→ It's curved because of diminishing marginal productivity of health inputs.

Switching from Health Capital (H) to Medical Care (M)

health care : H = f(health care) (M)

- Now, we want to derive *the demand for medical care.*
  - We've shown that H=*f*(M), so the demand for M is derived from the demand for H.

    ↑ health capital
- Consider the utility as a function of other good (Y) and medical care (M), rather than health (H): U = U(Y, M)
- Assume that M is homogeneous and represents the number of units of medical care.
- The budget line is a straight line because each unit of M costs the same number of dollars and means the same reduction in Y.
  - The slope of the budget constraint is constant.

Consumer's Maximization Problem

- The consumer's problem now is:

Maximize $U = U(Y,M)$

subject to $I = P_M{*}M + P_Y{*}Y$,

where I = total income,

P_M = Medical care price per unit

P_Y = Price per unit of other goods

- Let $P_Y = 1$. So, the budget can be simplified to
  - $I = P_m^* M + Y^*$

$\Rightarrow Y^* = I - \underbrace{P_m^*}_{\text{slope}} M$

Standard Budget Constraint Model

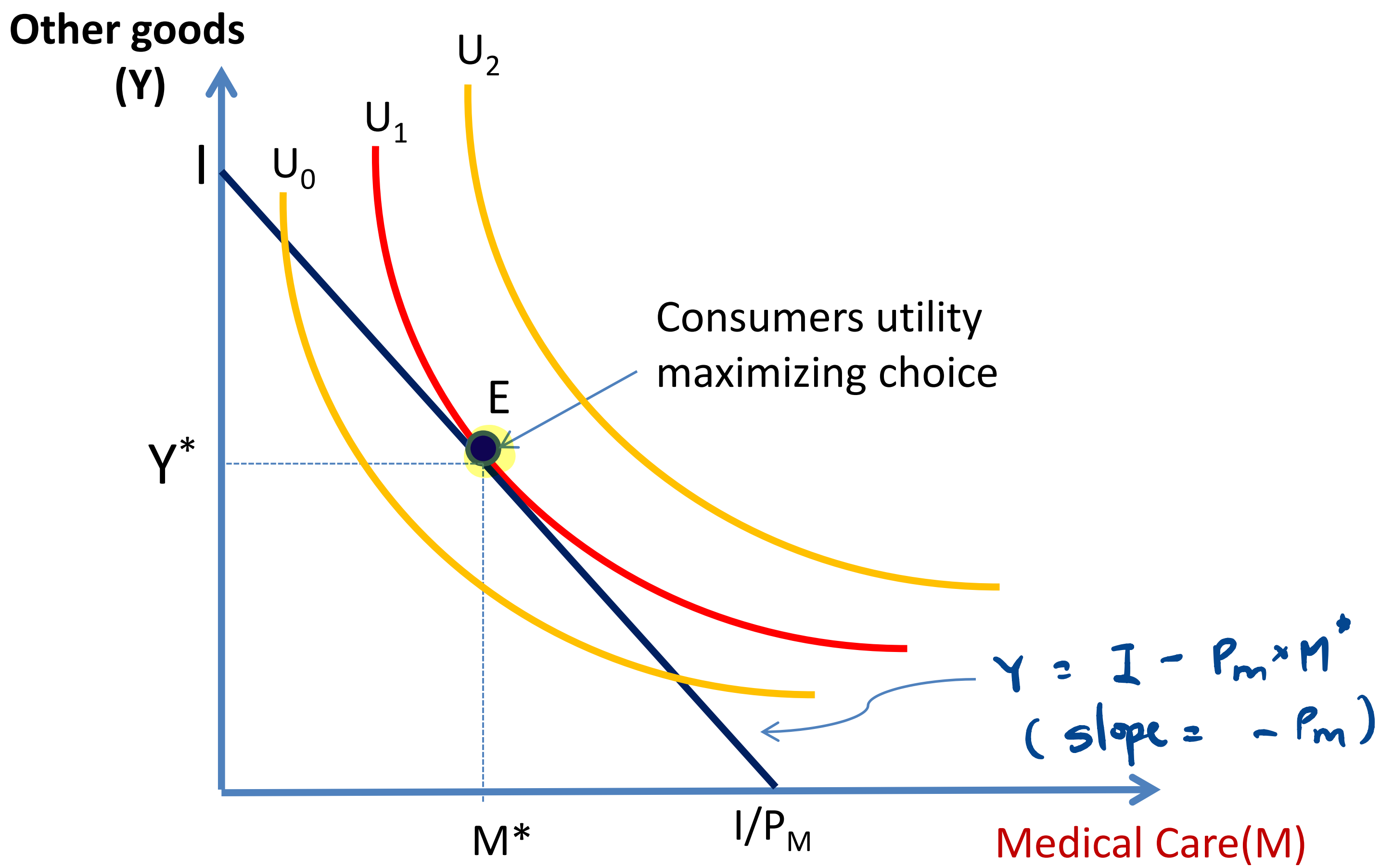
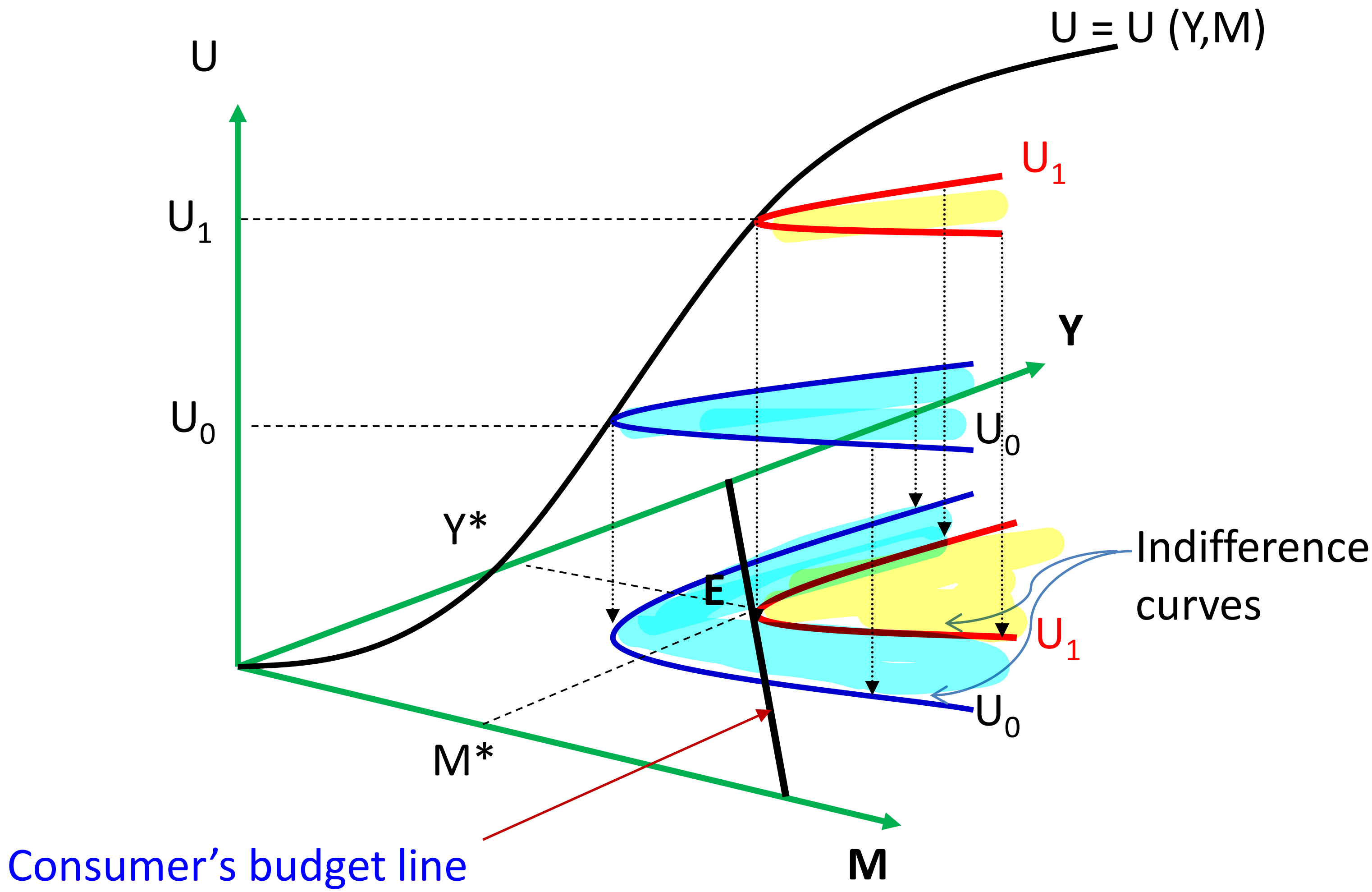


Illustration of U = U (Y, M) in a 3-space diagram.



Utility Maximization

Tangency of IC and BL.

- At point E the slope of indifference curve U_1 (marginal rate of substitution) is just equal to the price ratio:

$$MRS_{MY} = -\frac{MU_M}{MU_Y} = -P_M$$

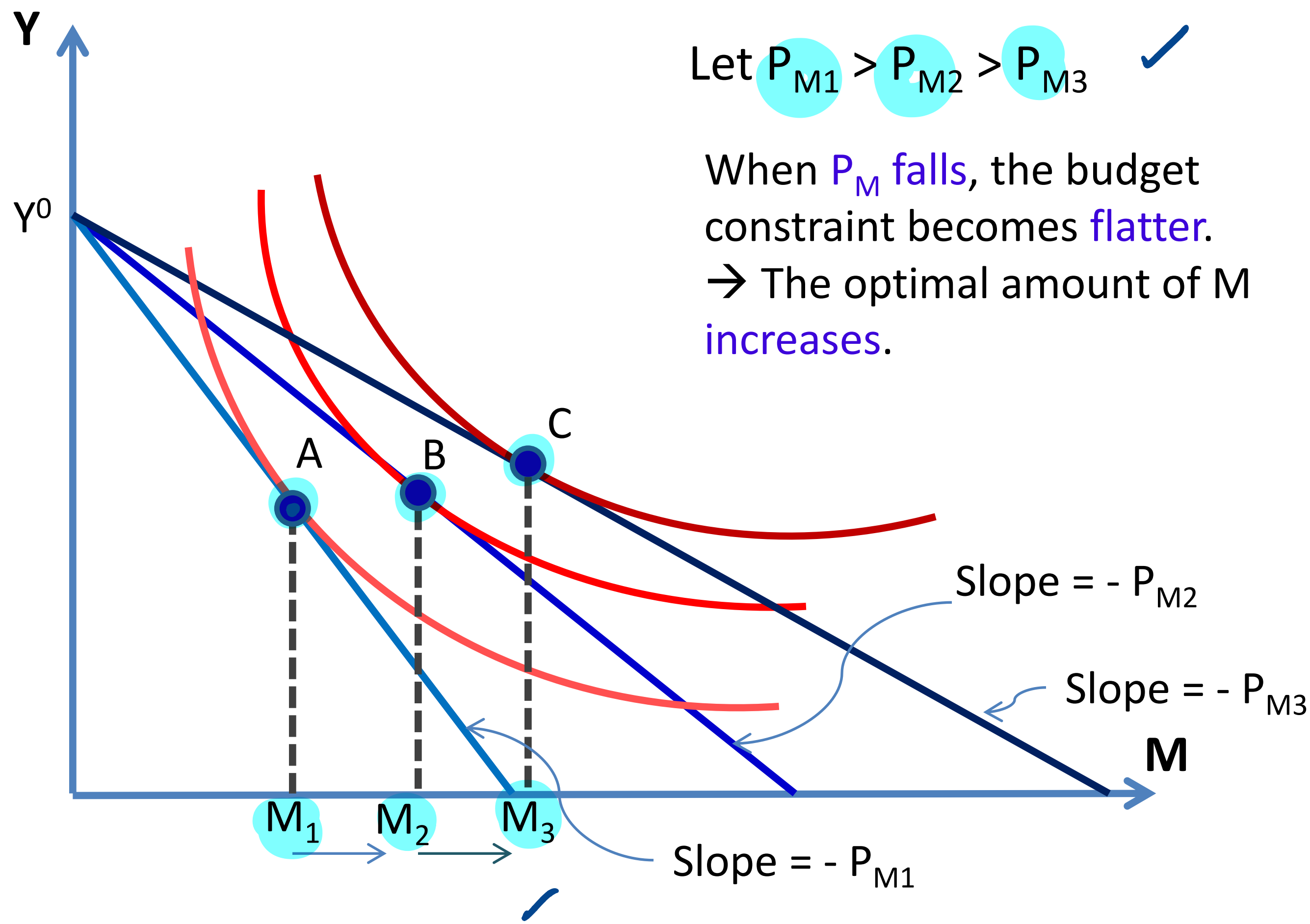
$\Rightarrow |MRS| = \left|\frac{MU_m}{MU_y}\right| = |P_m|$

- The marginal rate of substitution (MRS) is a measure of the rate at which a consumer is *willing* to trade other goods for medical care.
- The price ratio is a measure of the rate at which a consumer *can* trade other goods for medical care.

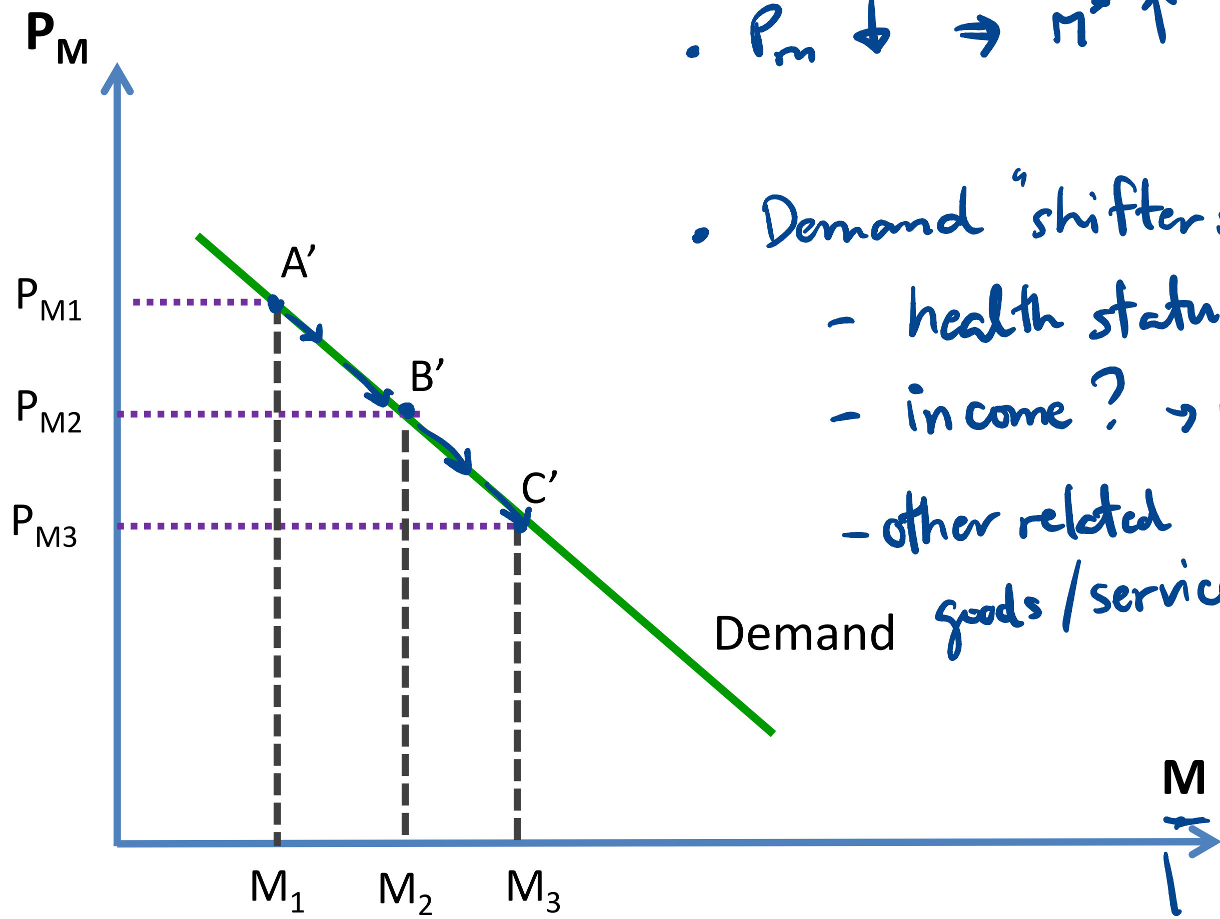
What Happens when P_M Falls?

Let $P_{M1} > P_{M2} > P_{M3}$

When P_M falls, the budget constraint becomes flatter.
→ The optimal amount of M increases.



The Demand for Medical Care



- $P_m \downarrow \Rightarrow M^* \uparrow$ (movement along D_m)
- Demand "shifters":
  - health status
  - income? → depends on health status
  - other related goods/services.

eg. surgical mask

Comparative Statics

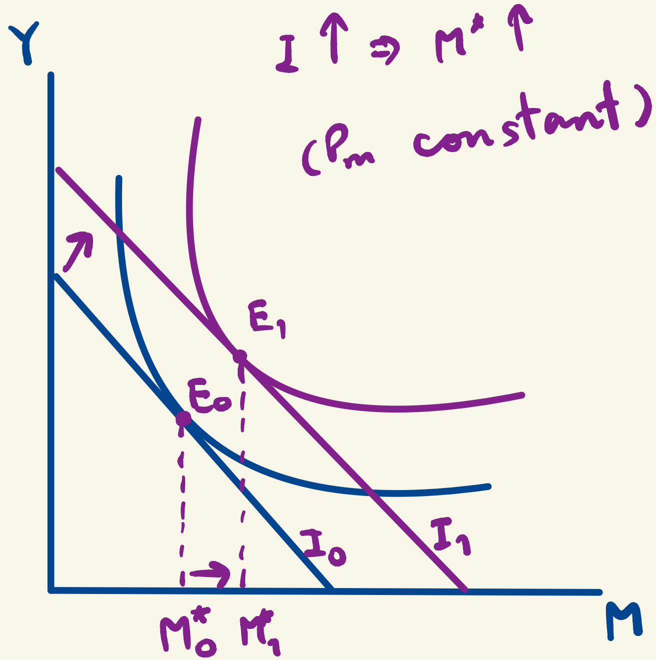
- Comparative statics is a type of analysis where the original equilibrium is identified, and *after a change occurs*, the new equilibrium is compared to the old one.

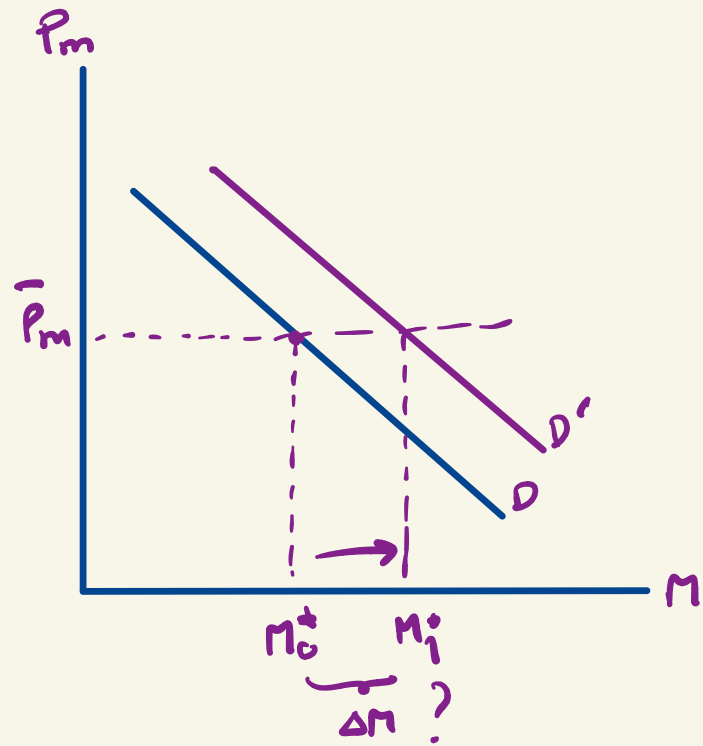
exogenous

- We will consider how the following changes affect the demand for medical care.
- ✓ Change in health status
- ✓ A decrease in medical care price
- ✓ An increase in income ✓
- ✓ An increase in the price of a substitute (or complement)
- * Two other factors that affect demand for medical care:
  - Time cost
  - Coinsurance → payment that consumers share a part of total health expenditure.

proportion / fixed payment.

Impact of income change on D_m

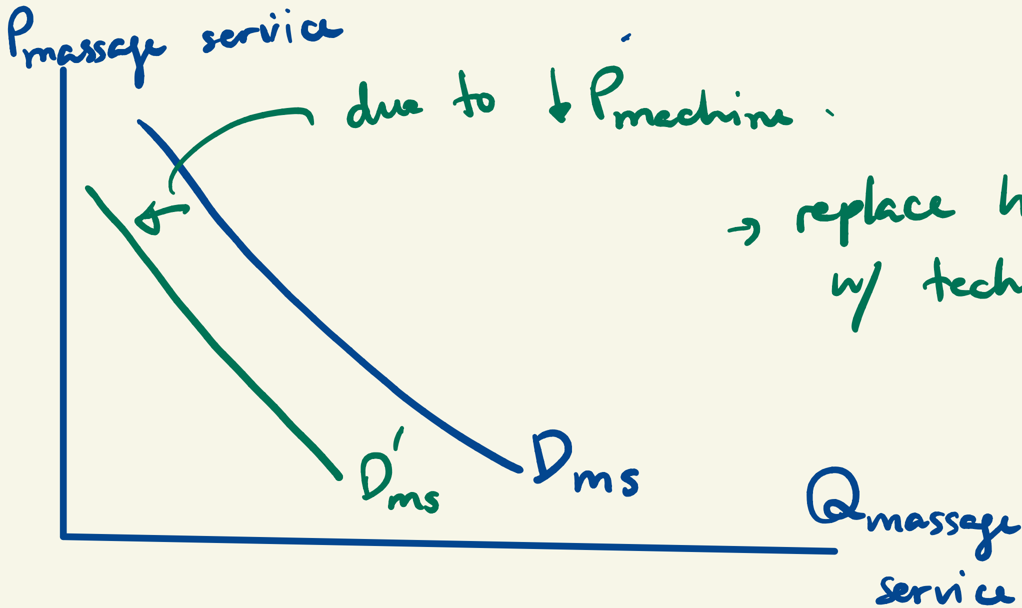




$$\varepsilon_I = \frac{\%\Delta M}{\%\Delta I} = \frac{\Delta M / M}{\Delta I / I} \; ; \; |\varepsilon_I| > 0$$

Change of Price of massage machine ✓ (P_{mach})

P_{mach} ↓ ✓ ⇒ $D_{massage}$? — Thai massage service ✓ substitute



→ replace human service w/ technology

Role of Time

"congestion"

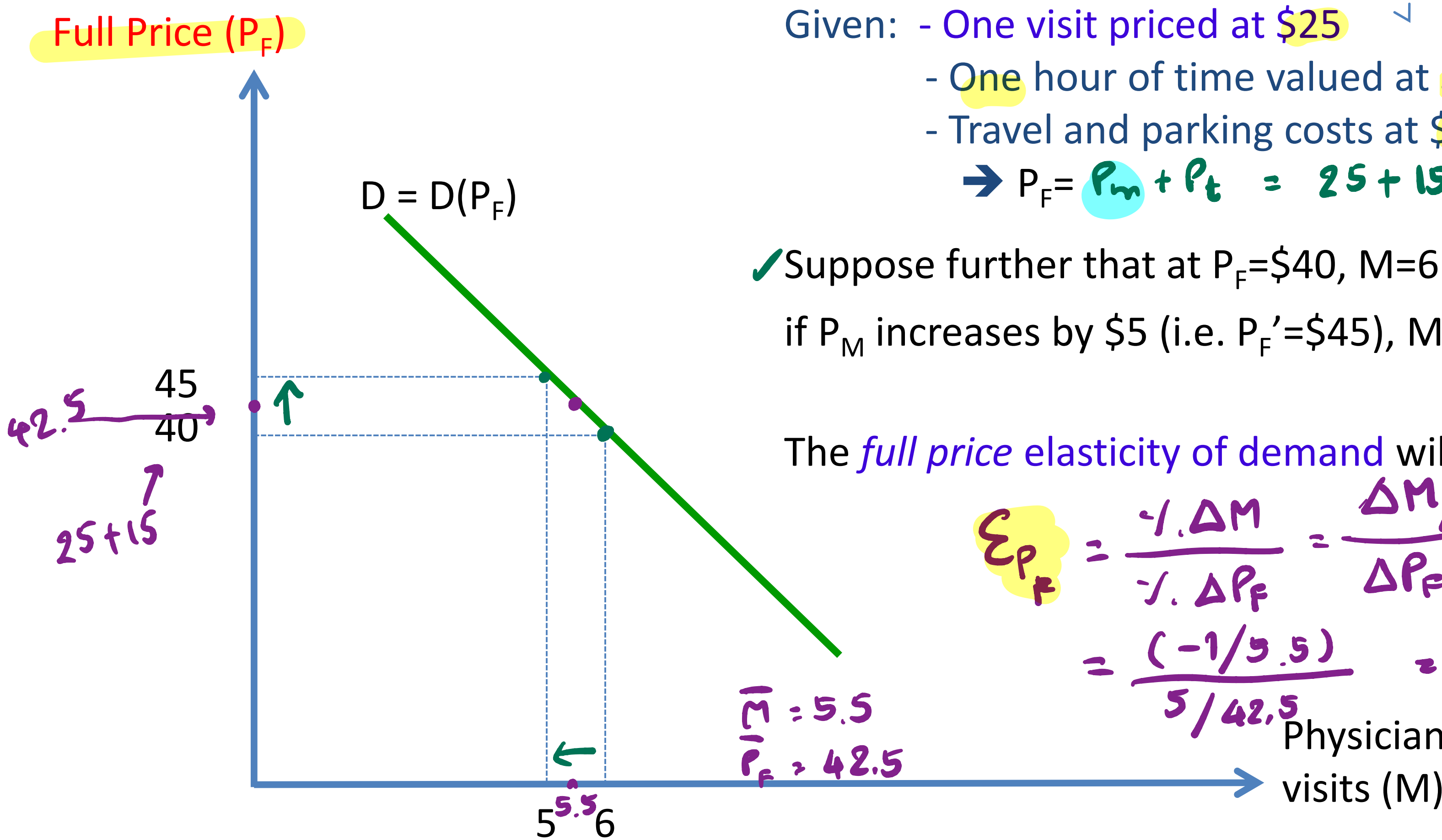
- Suppose the medical care is physician visits.
- The total cost of a physician visit might include both the money price of the visit itself and the cost of being away from work and not getting paid:
  - $P_F = P_M + P_T = P_M + wT$,

  / foregone earnings.

  Where P_F= Full price, P_M = Money price, P_T =Time price, w = wage rate, and T is time spent in obtaining care.
- So, the longer the time (T) involved in traveling, waiting, and receiving services, the higher the total price of medical care, and the smaller the quantity demanded.
- Similarly for higher wages (w) , higher the time price.

"What money can't buy"

Example: Demand and Time Price for Physician Visits



Given: - One visit priced at \$25 (P_m)
- One hour of time valued at \$10
- Travel and parking costs at \$5

➔ $P_F = P_m + P_t = 25 + 15 = \$40$

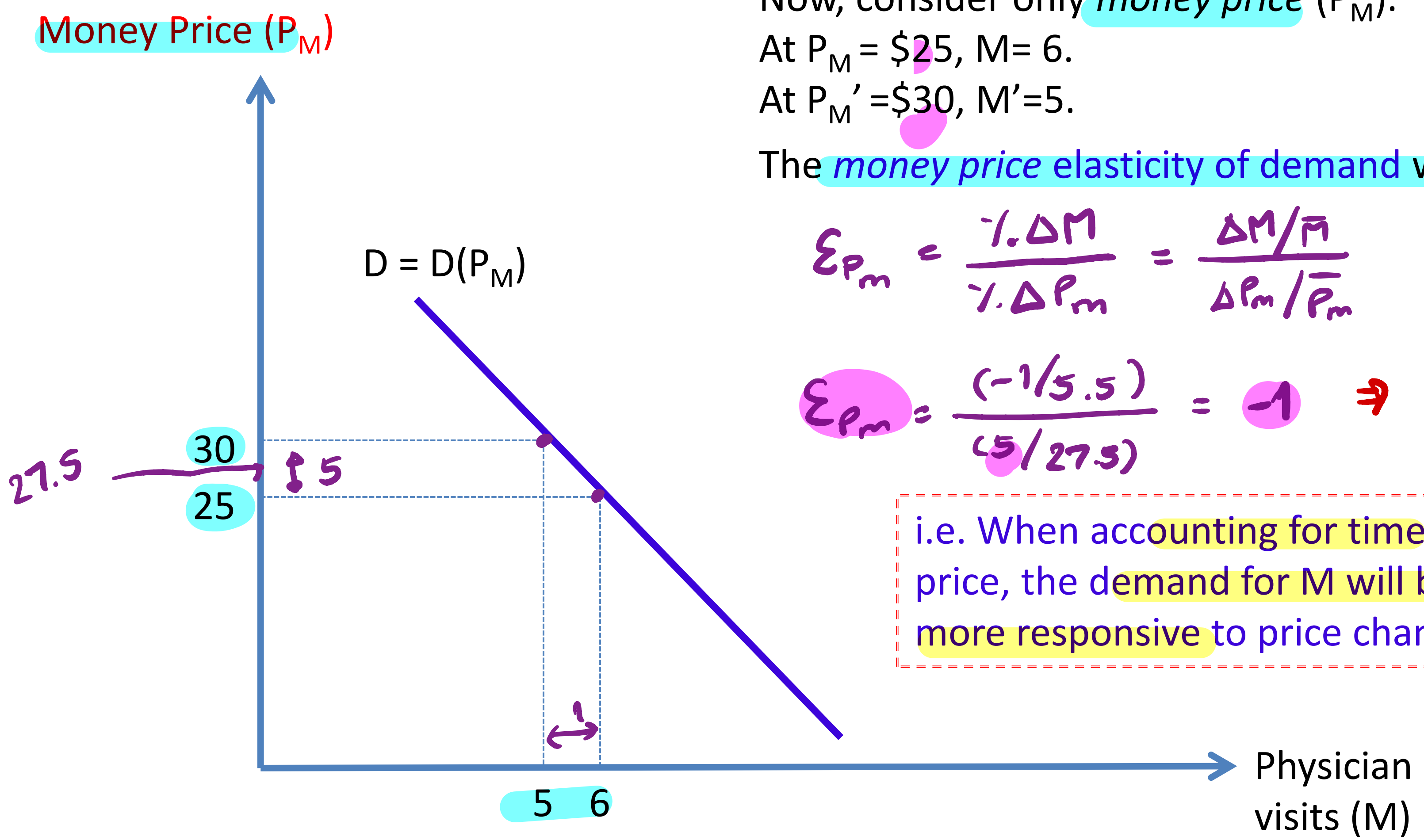
✓ Suppose further that at P_F=\$40, M=6 and if P_M increases by \$5 (i.e. P_F'=\$45), M'=5.

The *full price* elasticity of demand will be:

$$\varepsilon_{P_F} = \frac{\%\Delta M}{\%\Delta P_F} = \frac{\Delta M / M}{\Delta P_F / P_F} = \frac{(-1/5.5)}{5/42.5} = -1.545$$

$$\varepsilon_d = \frac{\%\Delta Q_D}{\%\Delta P} = \frac{\Delta Q_D / Q_D}{\Delta P / P}$$

Example: Demand and Time Price for Physician Visits (cont'd)



Now, consider only *money price* (P_M).
At P_M = \$25, M= 6.
At P_M' =\$30, M'=5.

The *money price* elasticity of demand will be:

$$\varepsilon_{P_m} = \frac{\%\Delta M}{\%\Delta P_m} = \frac{\Delta M/\bar{M}}{\Delta P_m/\bar{P}_m}$$

$$\varepsilon_{P_m} = \frac{(-1/5.5)}{(5/27.5)} = -1 \Rightarrow |\varepsilon_{P_m}| < |\varepsilon_{P_F}|$$

1, 1.5

i.e. When accounting for time price, the demand for M will be more responsive to price change.

Applications: Time Costs

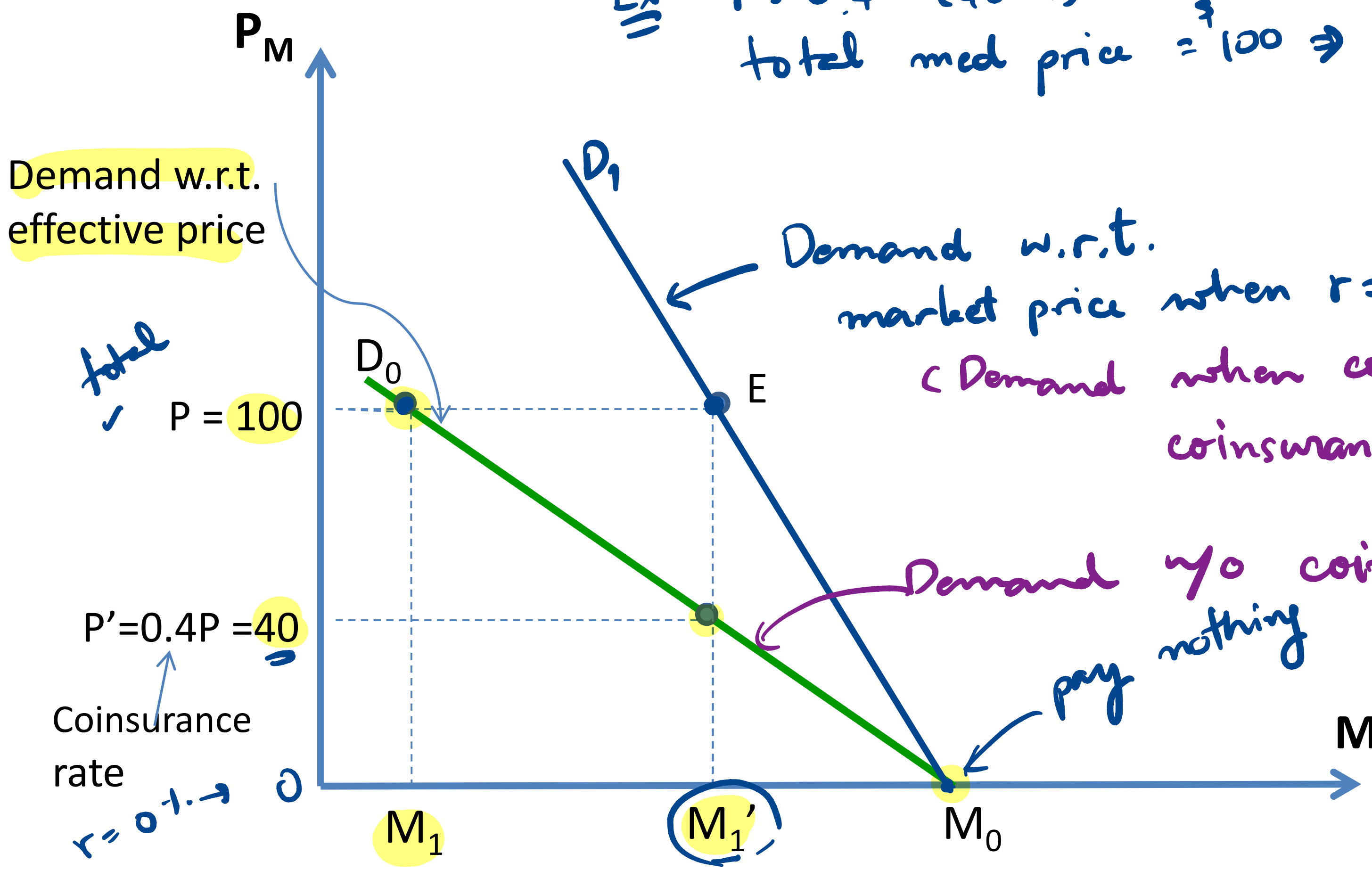
- People with lower opportunity costs of time would be more likely tolerate or endure long waiting times in clinics or physician offices. → Public hospital services
- Even for the poor whose medical care is subsidized, they still have to pay for the time costs.
  - Examples: program for the poor and disadvantage group.
    - Medicaid in the US,
    - UC Scheme in Thailand (long waiting time) — obstacle to seek health care.
  - Solutions:
    - Build health facilities near people's residence — District hospitals. small clinics
    - ??? — self-prescription. (self-medication).

Role of a Coinsurance Rate

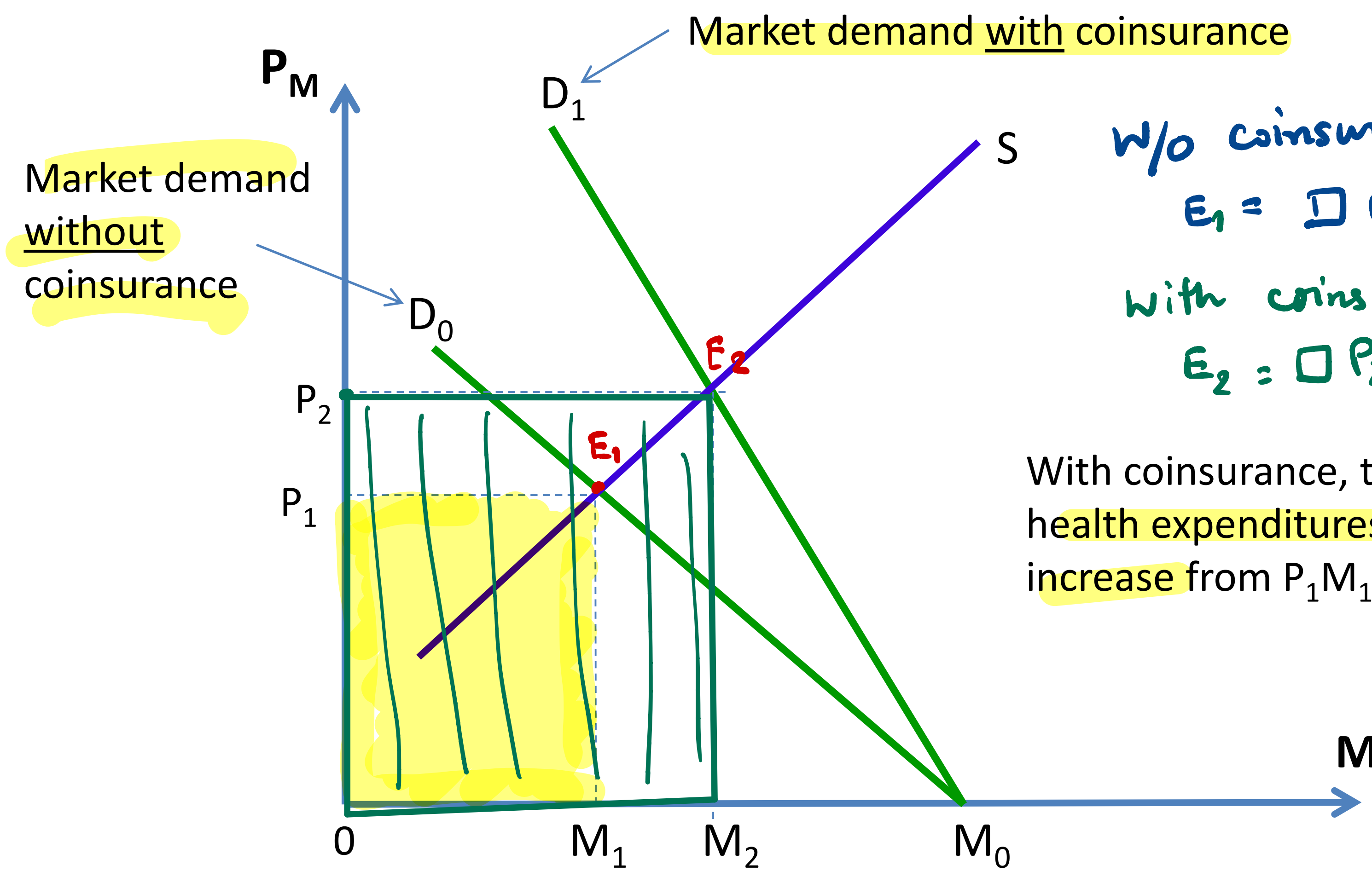
reduce demand (r)

rate at which patients pay out of total health expenditure.

Ex r = 0.4 (40 %)
total med price = $100 ⇒ patient pays $40.



Market Effects



W/o coinsurance:

$E_1 = \square P_1E_1M_1O = P_1M_1$

with coinsurance:

$E_2 = \square P_2E_2M_2O = P_2M_2$

With coinsurance, total health expenditures will increase from P_1M_1 to P_2M_2.

Empirical Studies

- Demand function for physician visits (V) can be written as:

Qm

$$V = f(P, r, t, P_0, Y, HS, AGE, ED, \ldots)$$

where P = price per visit,

r = patient's coinsurance rate,
t = time price,
P_0 = price of other goods,
Y = income,
HS = the patient's health status,
AGE = age, ED = education

$dv \approx \Delta V$
$dp \approx \Delta P$

- *'Usual'* econometric specification for the above demand function:

$$V = b_0 + b_1 P + b_2 r + b_3 t + b_4 P_0 + b_5 Y + b_6 HS + b_7 AGE + b_8 ED + \varepsilon$$

or $$\ln V = b_0 + b_1 \ln P + b_2 r + b_3 t + b_4 P_0 + b_5 Y + b_6 HS + b_7 AGE + b_8 ED + \varepsilon$$

$$\Rightarrow b_1 = \frac{d(\ln V)}{d(\ln P)} = \left(\frac{1}{V}\cdot dV\right) / \left(\frac{1}{P}\cdot dP\right) = \frac{dV}{dP}\cdot\frac{P}{V} = \varepsilon_P$$

Issues in Empirical Studies on the Demand for Medical Care

- Measurement
  - Measure "medical care/services" in monetary values
    - One problem is that expenditures reflect a complex combination of price of care, quantities of care, and qualities of care
  - Alternative measures :*quantity of visits, patient days, or cases treated*
    - Still have problems with the intensity of care.
- Definition of the price of services → affect demand estimation
  - Many patients do not pay the full price for their treatments.
- Differences in study populations

Example: Price Elasticities

- sensitivity to price change

Study	Dependent Variable	Price Elasticity
All Expenditures:		
Manning et al. (1987)	All expenditures	–0.17 to –0.22
Physician Services:		
Newhouse and Phelps (1976)	Physician office visits	–0.08
Cromwell and Mitchell (1986)	Surgical services	–0.14 to –0.18
Wedig (1988)		
Health perceived excellent/good	Physician visits	–0.35
Health perceived fair/poor	Physician visits	–0.16
Hospital Services:		
Newhouse and Phelps (1976)	Hospital length of stay	–0.06
Manning et al. (1987)	Hospital admissions	–0.14 to –0.17
Nursing Homes:		
Chiswick (1976)	Nursing home residents per elderly population	–0.69 to –2.40
Lamberton et al. (1986)	Nursing home patient days per capita elderly	–0.69 to –0.76

Source: Table 9.2 in Folland et al. (2013)

- sicker people are less sensitive to price change

Relatively more elastic

$|\varepsilon_p| < 1$

⇒ Demand is inelastic.

Example: Income Elasticities

$\varepsilon_I > 0$

Study	Dependent Variable	Income Elasticity
All Expenditures:		
Rosett and Huang (1973)	Expenditures	0.25 to 0.45
Hospital Services:		
Newhouse and Phelps (1976)	Admissions	0.02 to 0.04
Physician Services:		
Newhouse and Phelps (1976)	Visits	0.01 to 0.04
Nursing Homes:		
Chiswick (1976)	Residents per elderly population	0.60 to 0.90

Source: Table 9.4 in Folland et al. (2013)

$0 < \varepsilon_I < 1$

All income elasticities are positive, suggesting that medical care is normal goods,
And their values range between 0 and 1, suggesting that it is a necessity.

Other Factors Affecting Demand for Medical Care

- Ethnicity and Gender
  - In general, ‘blacks’ tend to consume less medical care.
  - Females, particularly at child-bearing ages, are heavy users of health care. Also, females generally live longer, so they are predominant among elder patients.
- Urban vs. Rural
  - People living in rural areas use less care. (Access problem? Or, socioeconomic status?)
- Education
  - Educated people are more informed and may demand more health care.
  - Often, education is correlated with income, and income actually leads to more demand for health care.

Other Factors Affecting Demand for Medical Care (Cont'd)

- Uncertainty
  - People demand health care for precautionary purposes, e.g. preventive care.
- Age
  - Grossman: As we age, the depreciate rate gets larger. So, more health inputs are required to restore health.
- Health status
  - The sicker tend to demand more health care.

THE PRODUCTION, COST AND TECHNOLOGY OF HEALTH CARE

Topics

- Production and the Possibilities for Substitution
- Costs in Theory and Practice
- Technical and Allocative Efficiency
- Productive Efficiency
- Frontier Analysis
- Health Care Technological Changes
    - Technological Changes and Costs
    - Price Adjustment When Technology Change Occurs
    - Diffusions of New Health Care Technologies

Production

- Short-run production
  - A production in which *at least* one input cannot be modified within a given period of time.
  - Fixed inputs – fixed at all levels of output, e.g. capital, salary
  - Variable inputs – vary by the amount of output, e.g. labor
- Long-run production
  - A production in which *all inputs* can be changed over time

Possibilities for Substitution

in Long-run productⁿ.

- Now, consider the production with *more than two variable inputs* (e.g. physicians and nurses).
- We are interested in the ability to substitute one input (e.g. physicians) for another input (e.g. nurses) while maintaining the same level of output.
- The degree to which one input can be replaced by another input, while output remains constant, is measured by the *slope of the isoquant*, which is equal to the marginal rate of technical substitution (MRTS).

Examples : . doctors (surgeons) and Robots

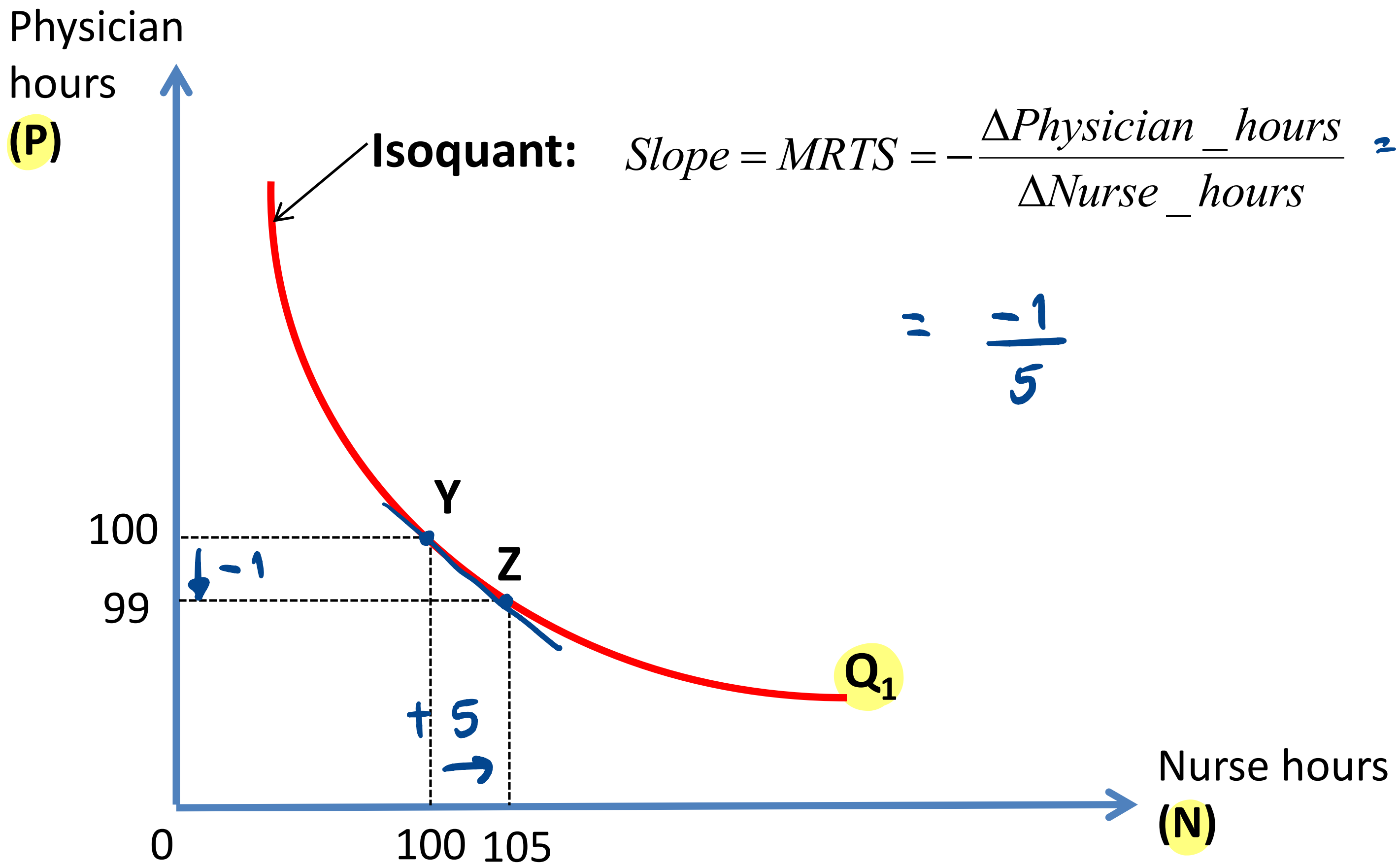
. lab technician and machine.

Substitution between Physicians and Nurses

y = f(P, N)

Isoquant: $Slope = MRTS = -\frac{\Delta Physician\_hours}{\Delta Nurse\_hours} = -\frac{\Delta P}{\Delta N}$

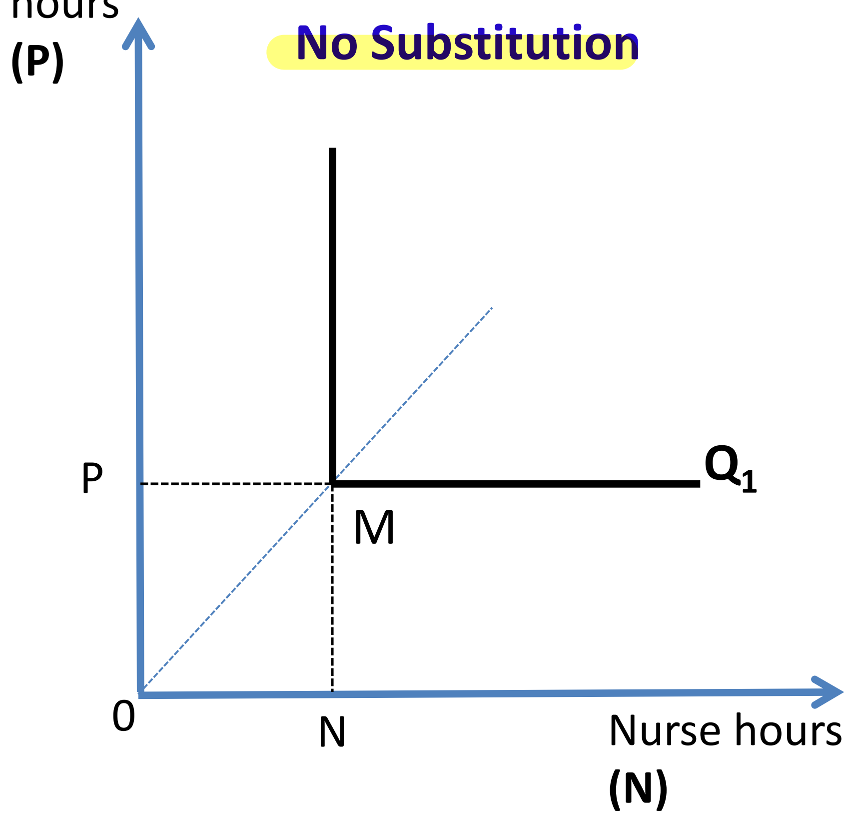
$= \frac{-1}{5}$



Substitution: Extreme Cases

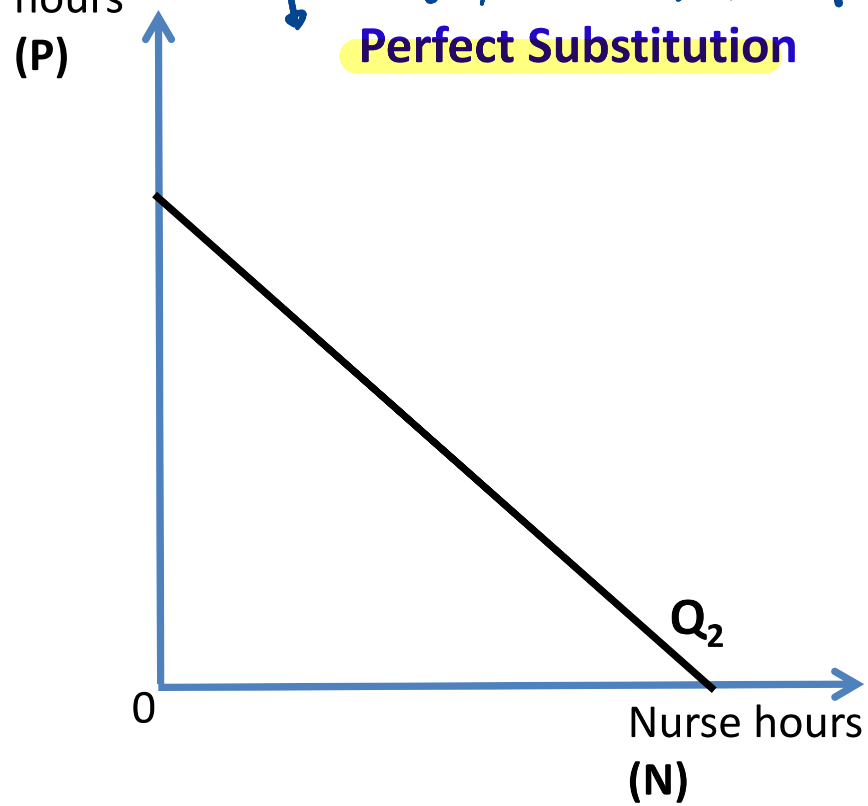
Perfect complement

No Substitution



Ex: • Thermometers
– many types
• ATK vs. RT-PCR

Perfect Substitution



Question: Would these two cases be possible in the real world?

Ex. Echo → doctor and machine : Perfect complement.

Elasticities of Input Substitution

- If an input price changes, a cost-minimizing firm would respond by shifting away from the *costlier* input to the *cheaper* input.
- The new input combination is determined by the elasticity of substitution of inputs.
- The elasticity of substitution (E_S) measures the responsiveness of a cost-minimizing firm to changes in relative input prices:

$$E_s = \frac{\text{Percentage change in factor input ratio}}{\text{Percentage change in factor price ratio}}$$

ratio

$$= \frac{\% \Delta R_Q}{\% \Delta R_P}$$

- Example: Suppose M = f(P, N), and factor prices are Wp and Wn.

$$E_S = \frac{\%\Delta(P/N)}{\%\Delta(W_p/W_n)} = \frac{\Delta(P/N)}{P/N} \times \frac{W_p/W_n}{\Delta(W_p/W_n)}$$

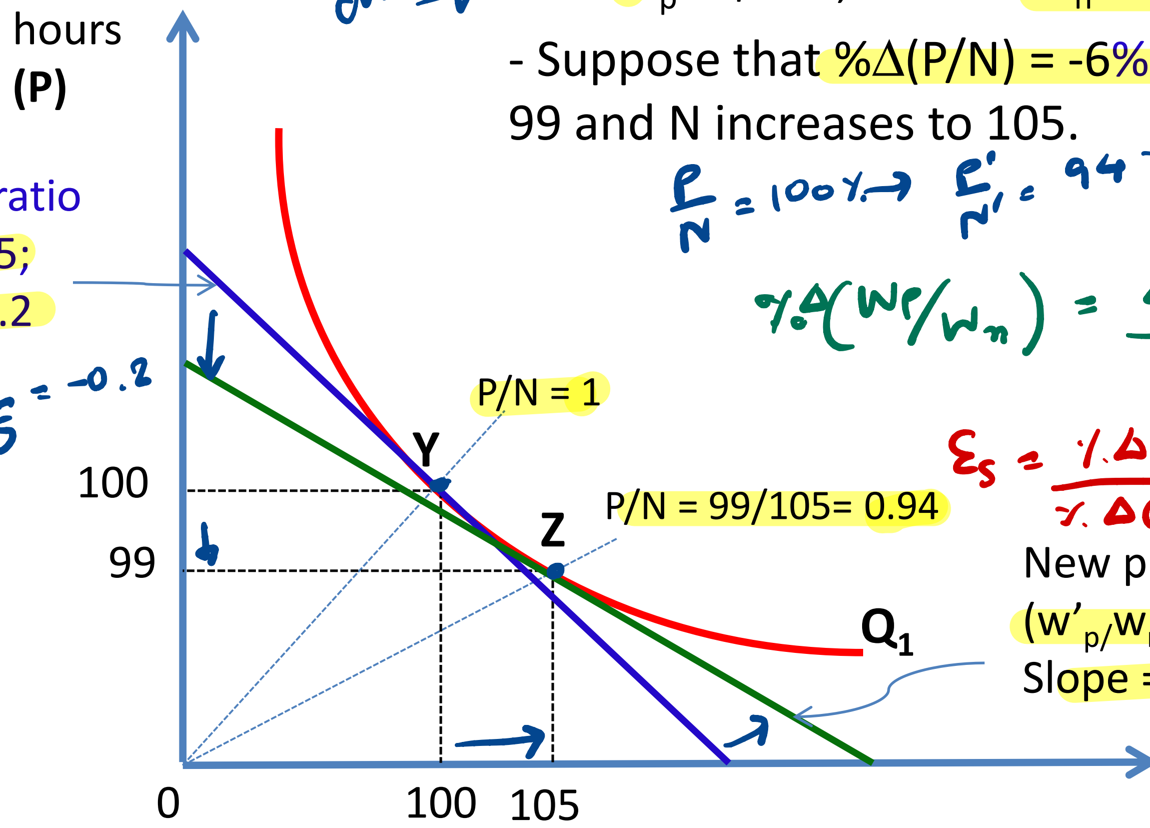
Substitution between Physicians and Nurses

- Suppose w_p = \$200,000 and w_n = \$40,000
And w'_p = \$220,000 and w_n is constant.

- Suppose that $\%\Delta(P/N) = -6\%$, so P reduces to 99 and N increases to 105.

Old price ratio
$(w_p/w_n) = 5$;
Slope = -0.2

New price ratio
$(w'_p/w_n) = 5.5$;
Slope = -0.18



$\frac{w_p}{w_n} = \frac{200,000}{40,000} = 5$ ✓

After change

$\frac{w'_p}{w_n} = \frac{220}{40} = 5.5$

$\frac{P}{N} = 100\% \rightarrow \frac{P'}{N'} = 94\% \Rightarrow \%\Delta\left(\frac{P}{N}\right) = -0.06$

$\%\Delta(w_p/w_n) = \frac{\Delta(w_p/w_n)}{w_p/w_n} = \frac{(5.5-5)}{5}$

$= \frac{w_n}{w_p} = -\frac{1}{5} = -0.2$

$= -\frac{1}{5.5}$

$\varepsilon_s = \frac{\%\Delta(P/N)}{\%\Delta(w_p/w_n)} = \frac{-0.06}{\frac{0.5}{5}} = -0.6$

$\varepsilon_s = -0.6$ → If input price ratio increases 1%, the input ratio would ↓ by 0.6%.

Input Substitution: Application

Can E_s be positive?

- E_s = -0.6 means that, as input price ratio increases by 1%, the input ratio would decrease by 0.6%.
- The absolute values of E_s range between 0 and ∞.
  - E_s =0 ➔ No substitutability ie. Perfect complement
  - Larger E_s ➔ Greater potential for substitutability — Nasal swap ATK & Saliva ATK
- The concept of input substitution can be applied particularly to the *long-run production*, where all inputs are flexible.
- Other examples of input substitution:
  - Substitution between health workers and new technology
  - Substitution between local physicians and foreign physicians

Deriving Cost Curve

- Hospital's cost minimization problem:

  Min TC = rK + wL

  Subject to Q*=Q(K, L)

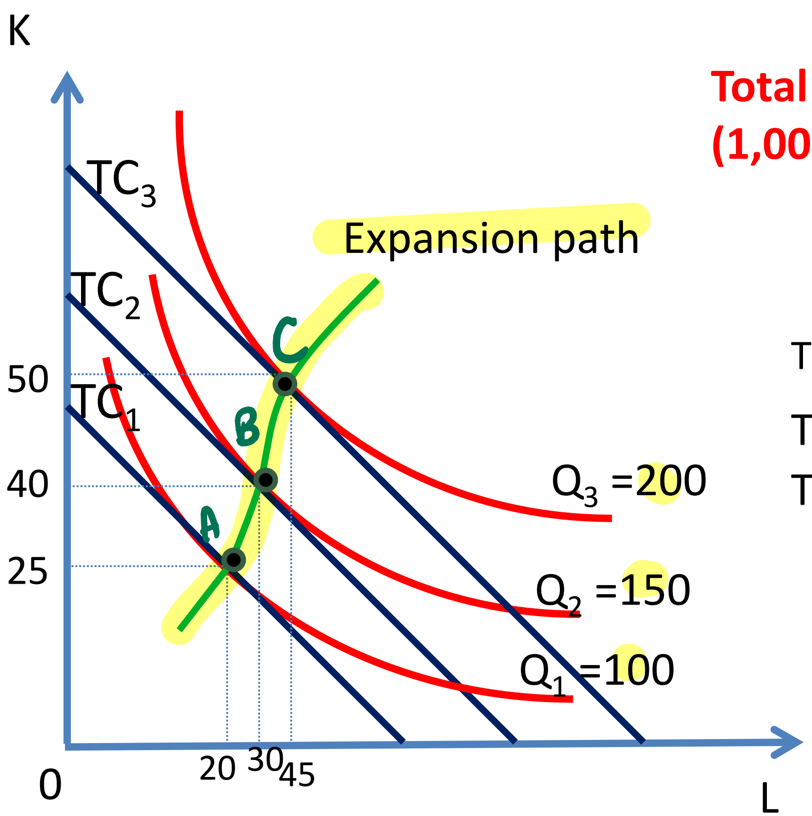
  ➔ (K*,L*) is the input combination that gives the least cost at a given output level.

- The set of all possible points of tangency between the isocost curves and the isoquants is called the expansion path.
  - This curve tells the relationship between a given output level and its minimum cost.

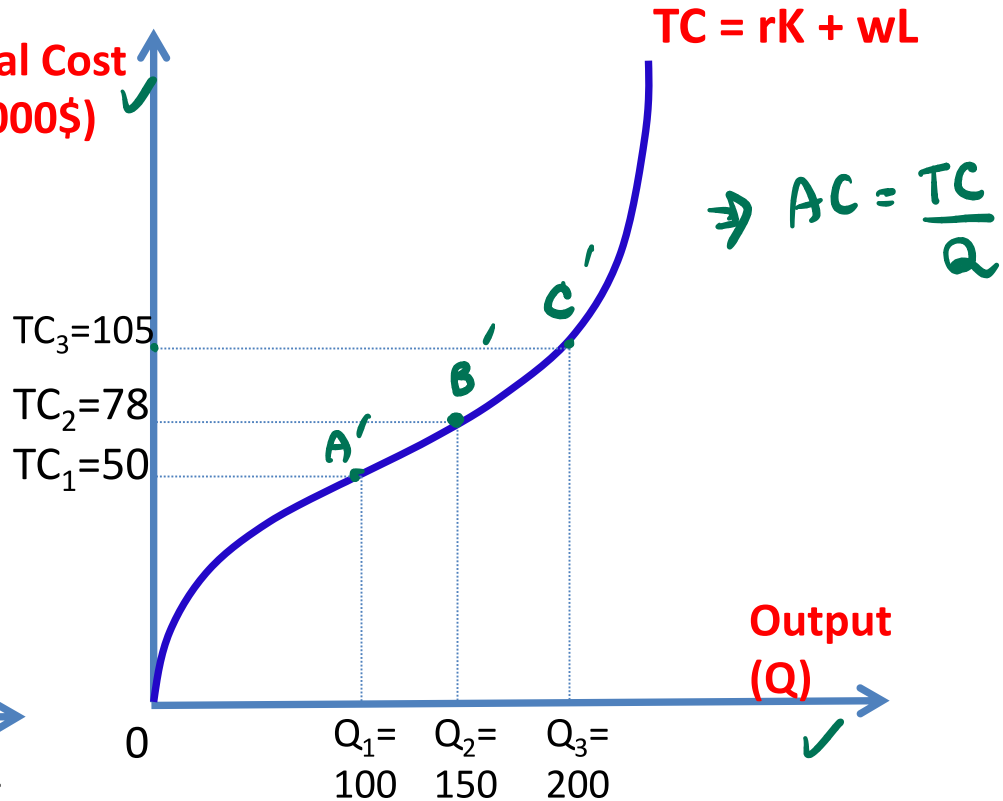
Cost Function

A': $TC_1 = (20 \times 1000) + (25 \times 1200)$
$= \$50,000$

Production Isoquants



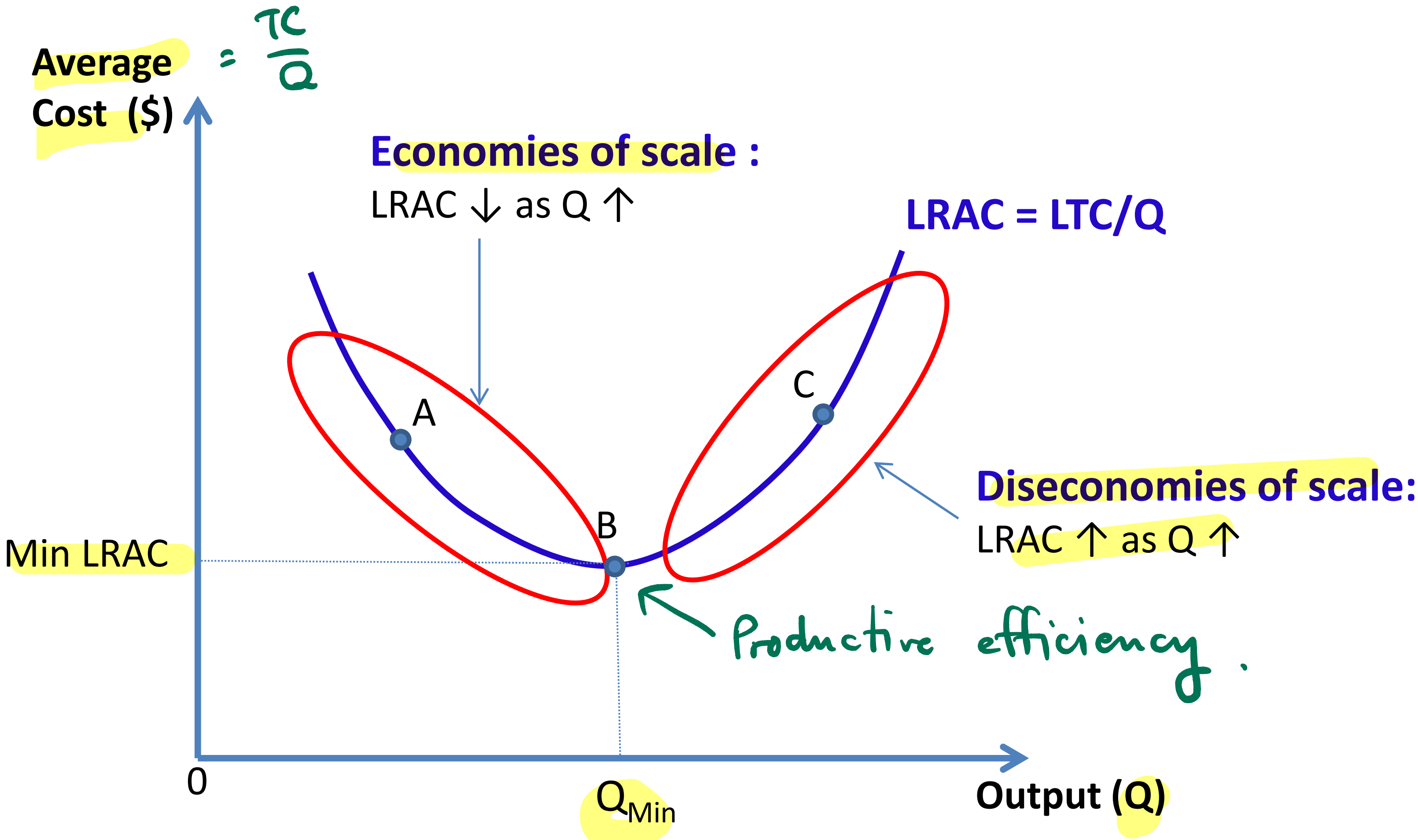
Total Cost Function



→ $AC = \frac{TC}{Q}$

Suppose input prices are r = $1,200 and w = $1,000.

Economies of Scale



Economies of Scope

- Definition:

  For a multiproduct firm, economies of scope occur whenever it is possible to produce two or more goods jointly more cheaply than they can be produced separately.

- Mathematically, suppose a firm has 2 outputs Q_1 and Q_2. Economies of scope exists if:

$$TC(Q_1, Q_2) < TC(Q_1, 0) + TC(0, Q_2)$$

where $TC(Q_1,Q_2)$ = the joint cost of producing both outputs together

$TC(Q_1,0)$ = the cost of producing output 1 only

$TC(0,Q_2)$ = the cost of producing output 2 only

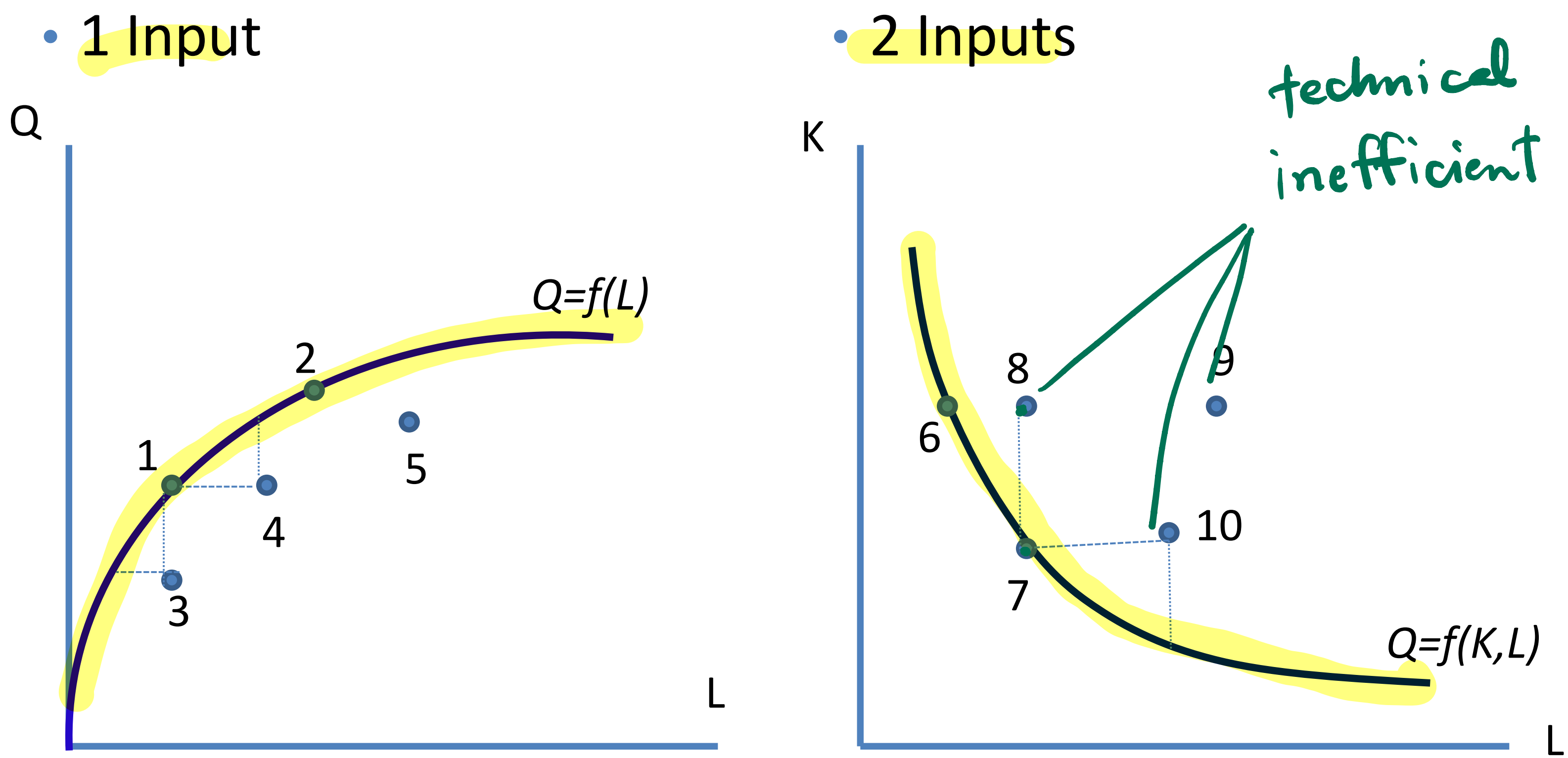
Significance of Economies of Scale and Economies of Scope

- Both concepts can provide implications to public policy and to managerial policy.
- Economies of scale:
  - Profit-maximizing firm
    - Prefers to produce where AC is still decreasing
  - Society's perspective - eg Public hospitals, NGOs.
    - Prefers lowest average costs (not necessary where it max π)
- Economies of scope:
  - Provision of different departments in the same hospital
  - Subsidization in teaching hospitals

Efficiency

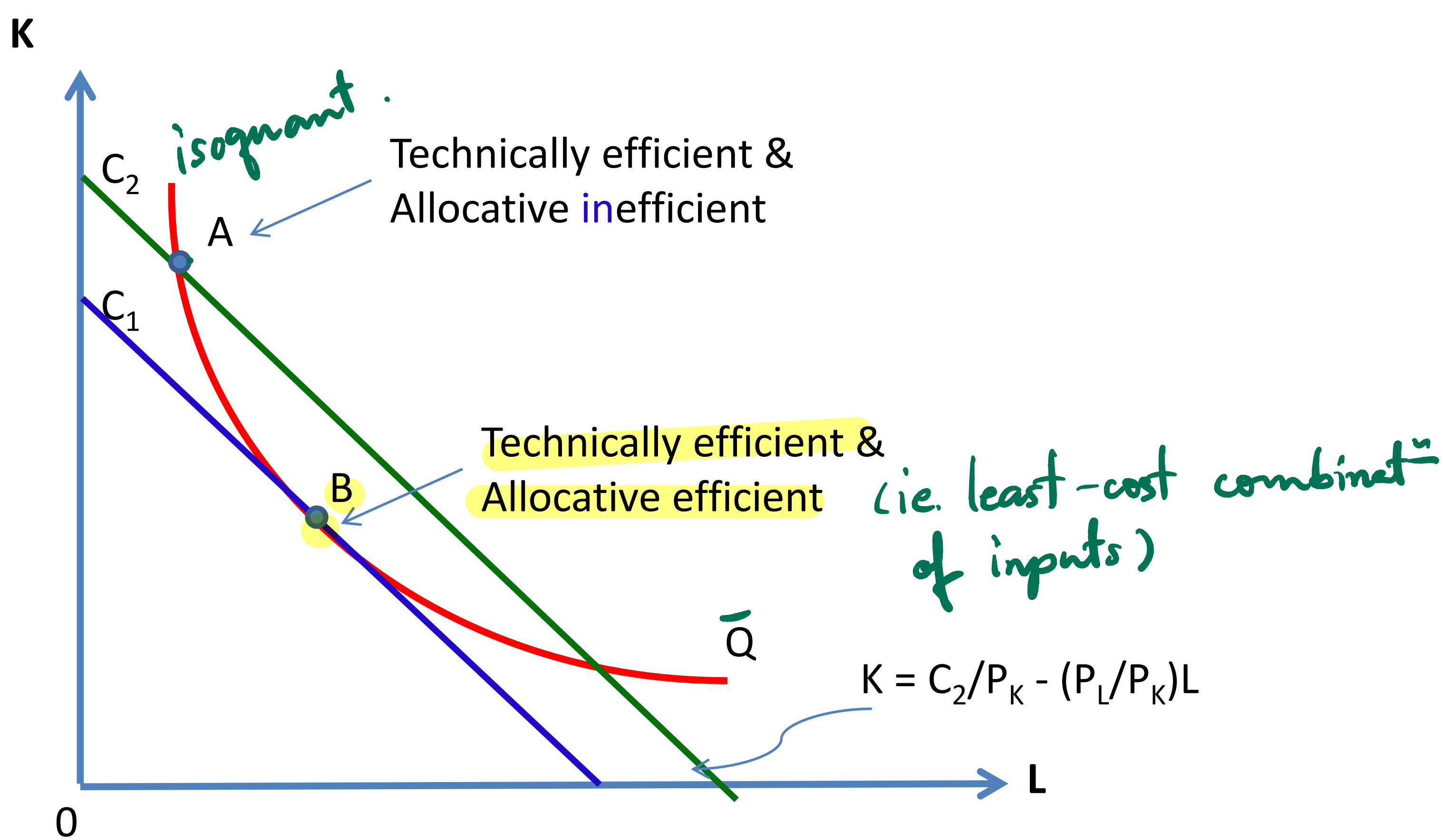
- Technical efficiency
  - Applies to production within a given firm
  - Technical efficiency is achieved when a maximum output is being produced from a given input combination.
- Allocative efficiency → consider input prices
  - Requires the efficient allocation of inputs between firms and between outputs.
  - Allocative efficiency is achieved when inputs are put into their best uses so that no further gains in output are possible.
- Productive efficiency
  - There could be many different levels of output that are allocative efficient (i.e. total costs are minimized).
  - Of all the least cost levels of output, the "most least cost" level of output is productively efficient. lowest AC.

1. Technical Efficiency



Technical inefficiency results when a firm uses more resources than necessary to produce a given level of output.

2. Allocative Efficiency



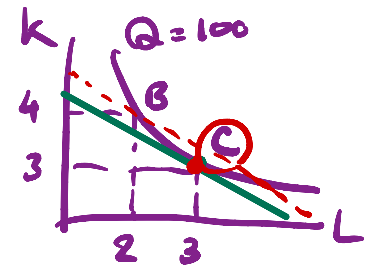
Technical and Allocative Efficiencies

$\bar{Q} = 100$

Technology	K	L	Cost (if P_K=10, P_L=5)
~~A~~	3	~~4~~ ←	~~$50~~
B	4	2	$50
C	3	3 ←	$45 ✓

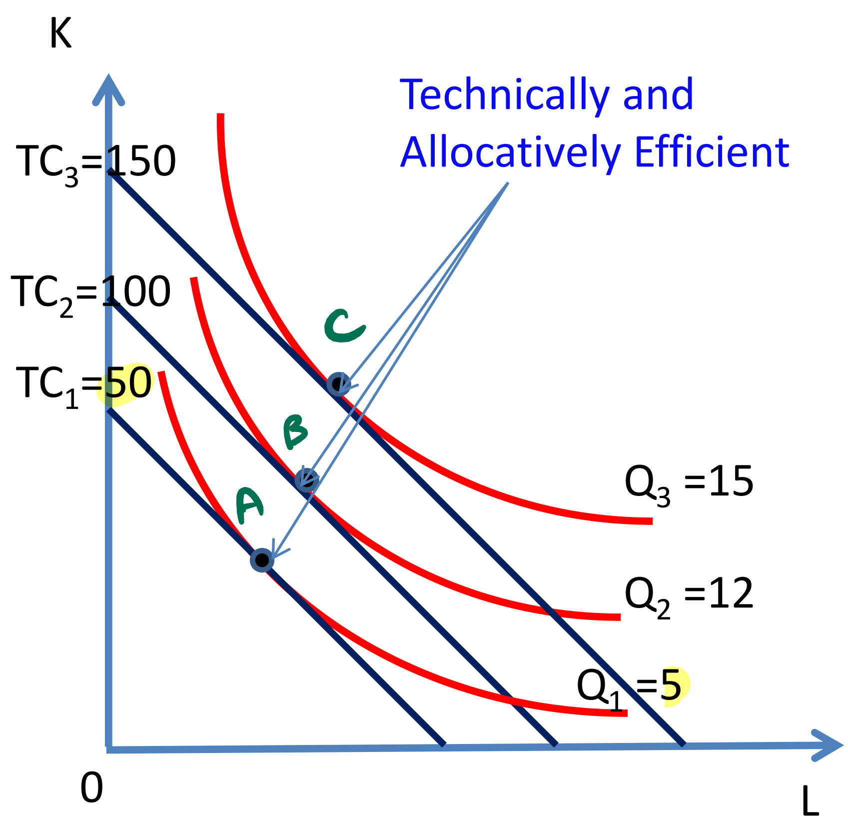
Technical efficient

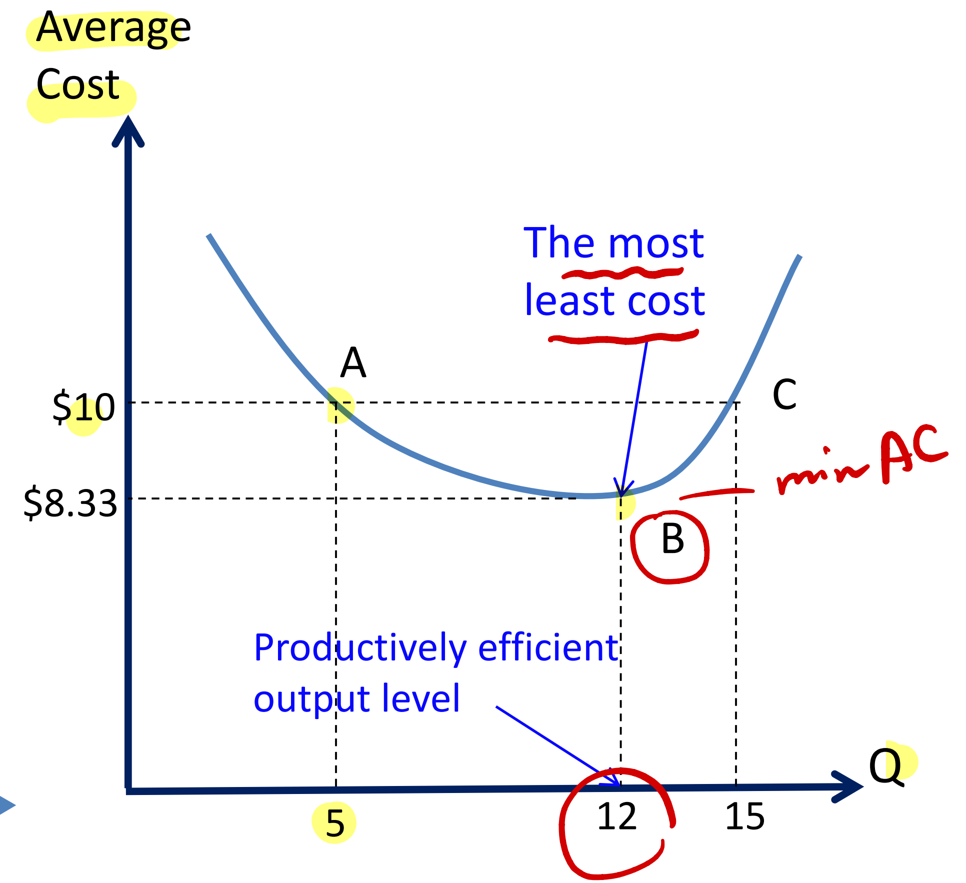
- Suppose P_K = 10 and P_L= 5, which technology is allocatively efficient? → C



- Which technology should not be used?
  ↳ A

3. Productive Efficiency





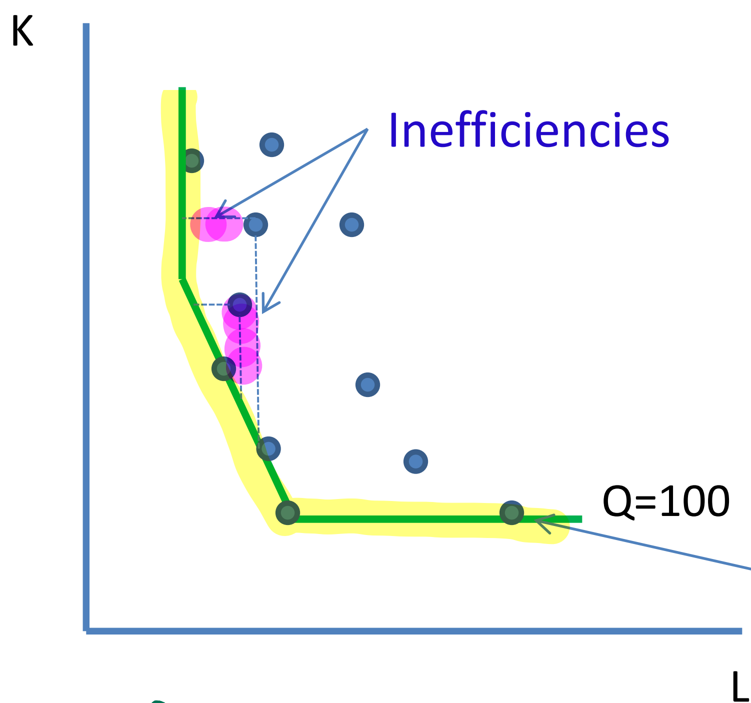
@ A, AC = 50 ÷ 5 = 10

@ B, AC = 100 ÷ 12 = 8.3

@ C, AC = 150 ÷ 15 = 10

Frontier Analysis: Data Envelope Analysis

(DEA)



The Data Envelope Analysis (DEA) identifies the efficient outer shell enveloping the data.

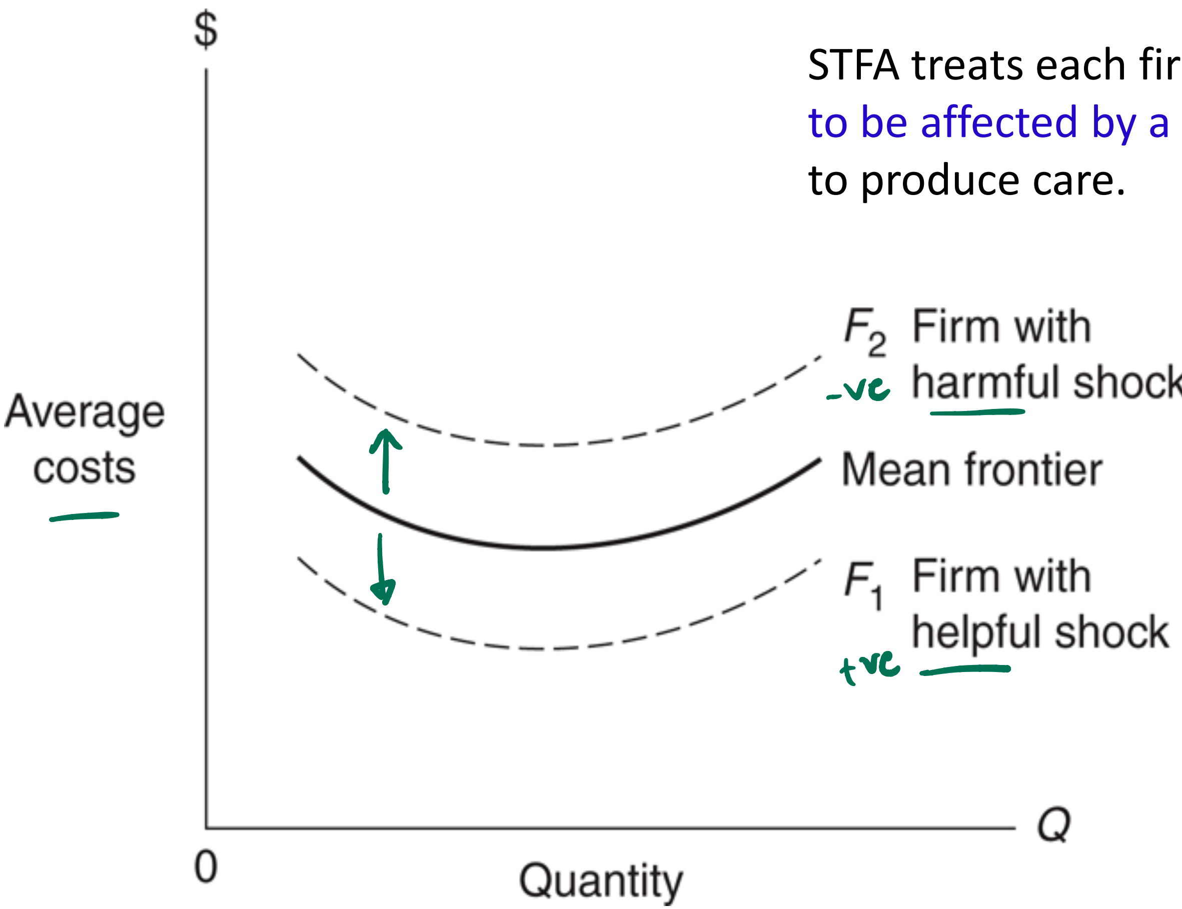
Deterministic frontier (found by using linear programming)

non-parametric method,

$Q = f(K, L)$

Assume CRS, $Q = K^a L^b$, $a + b = 1$

Frontier Analysis: Stochastic Frontier Analysis (STFA)

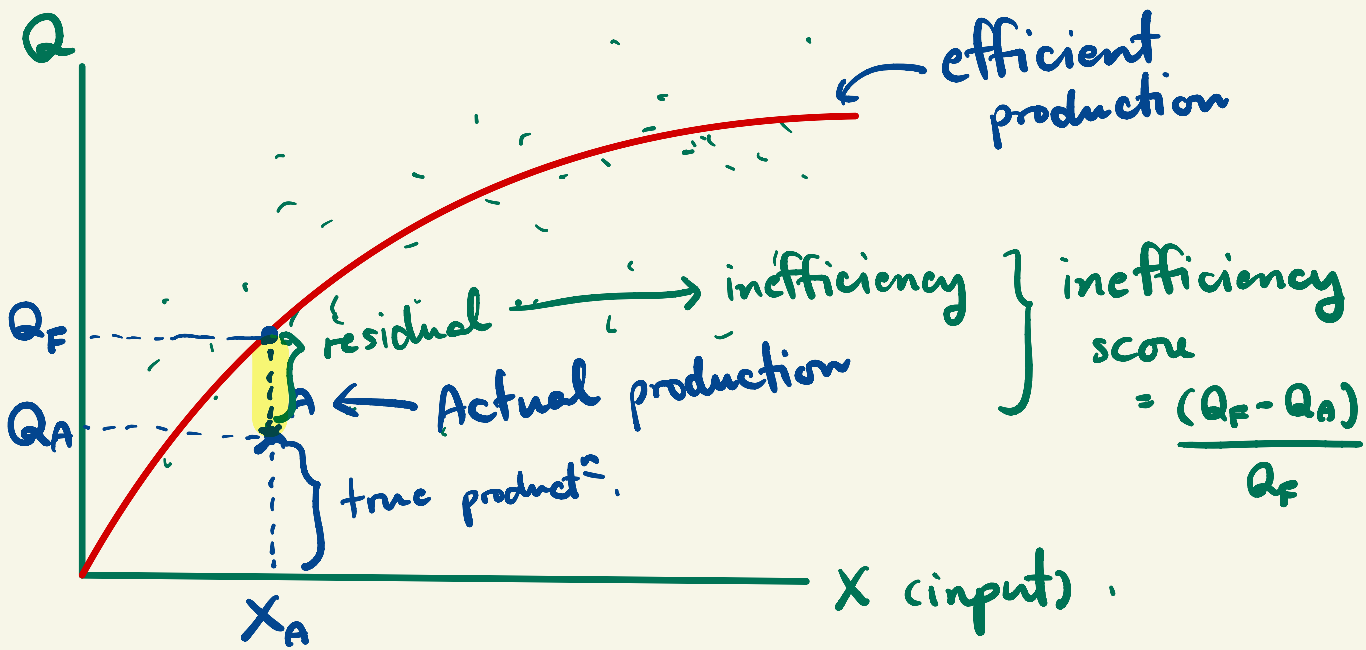


STFA treats each firm uniquely by assuming it to be affected by a potential shock to its ability to produce care.

Frontier is stochastic.

Inefficiency score is calculated from the residuals obtained from SCF model.

Ex 1 input-production : STFA



Case Study:
Efficiency of Public Hospitals in Thailand

(Source: Patamasiriwat, 2011)

- Examined cost efficiency of 3 different types of public hospitals in Thailand:
  - Regional hospitals (23)
  - Provincial hospitals (58)
  - Community hospitals (629)
- Focused on technical efficiency and looked at the cost side
  - Costs: Wage, personal compensations, operating expenses, material costs, utilities, others
- Estimated cost function: $C = f(OP, IP) = OP^a x IP^b$
  - Regional hospitals: constant returns to scale (i.e. a+b = 1)
  - Community & Provincial hospitals: decreasing cost (i.e. a+b <1)

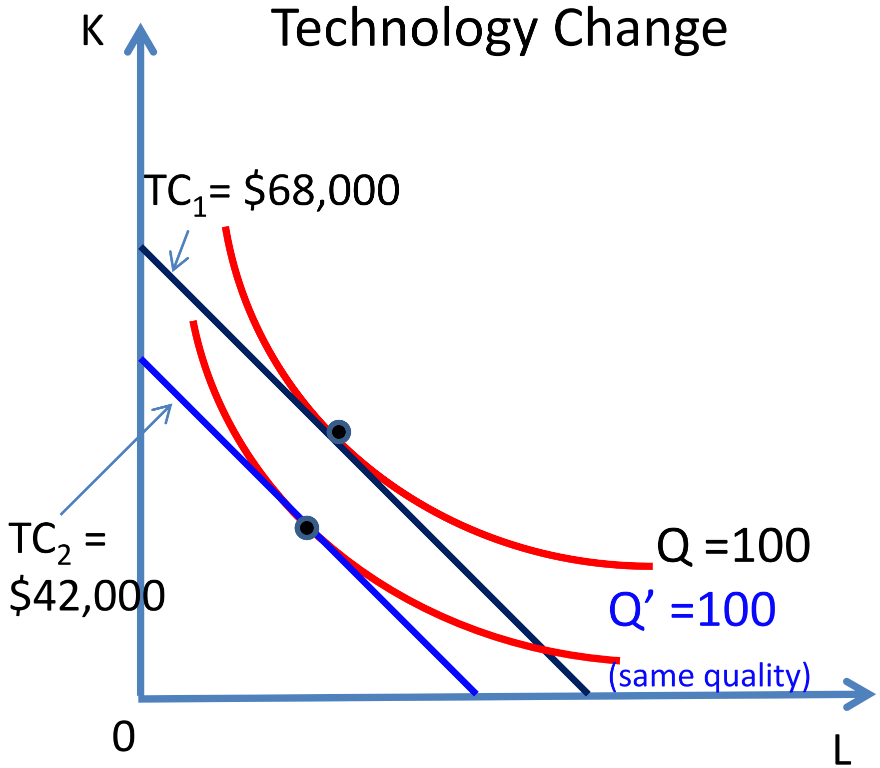
Case Study:
Efficiency of Public Hospitals in Thailand

(Source: Patamasiriwat, 2011)

- 2 Methods to determine efficiency: DEA and STFA
  - Regional hospitals: Efficiency score = 0.94 (i.e. "output slack" =6%)
  - Provincial hospitals: Efficiency score = 0.64
  - Community hospitals: Inefficiency score=0.186 → Efficiency score = 0.814
- Limitations:
  - Ignore other types of hospital outputs (teaching, research, health promotion)
  - Did not consider quality of care or severity of different health care cases
- Discussion:
  - Provincial and community hospitals might operate at less than full capacity
  - There were signs of resource underutilization.

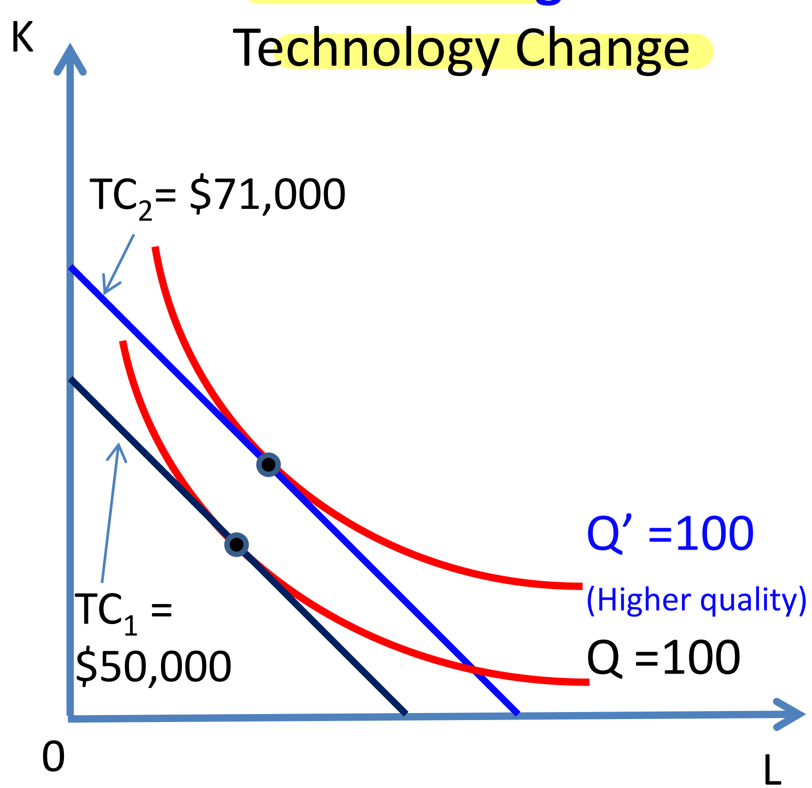
Technological Changes and Costs

Cost-Decreasing due to Technology Change



Technology change increases the productivity of health resources.

Cost-Increasing due to Technology Change



Technology change improves quality of care or introduce costlier products.

Health Care Prices with Technological Change Occurs

- Question: Whether technological changes really cause the inflation in the health care sector?

➢ Not necessarily!

  ➢ New treatments are *not* included in the fixed basket when calculating CPI.

  ➢ Improvement in the treatment effectiveness could be omitted.

    - Example: New treatments reduce the length of hospital stay. By putting more weight on the room charge, the price index is overestimated.

  ➢ Not taken into account the improvement in quality of life of the patients.

Adjustment in Health Care Prices

- <u>Example</u>: Cutler et al (1999) developed two quality adjusted price indexes of myocardial infarction treatment

Unadjusted Indexes	Avg. Price Changes	Quality-Adjusted Indexes	Avg. Price Changes
Official MCPI	3.4%	Quality (extra years of life)	-1.5%
Heart attack episode approach	2.8%	Quality (extra QALYs)	-1.7%

- Technological change has *improved the quality of treatment*.
- The quality adjustment can prove that the heart attack treatment *price* is actually *declining* (price deflation).