Production of Health

7MHPH010 – Health Economics and Health Policy

Department of Primary Care and Public Health Sciences
Kings College London

Farasat A.S. Bokhari
farasat.bokhari@kcl.ac.uk
http://phs.kcl.ac.uk/fbokhari/

Spring 2012

OUTLINE
FIRMS

- Production Functions
- Cost Functions

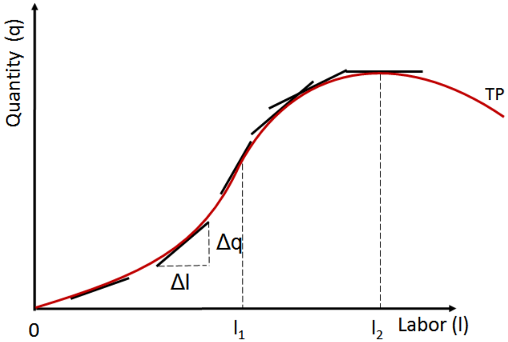
- Firm produces a particular good (q) using combinations of inputs
  - The most common used inputs are capital (k) and labor (l)
  - One could think in introducing different inputs: skilled labor, unskilled labor, raw materials, intermediate products, technology
- Firm faces an incentive to produce as efficiently as possible
- They possesses perfect information regarding the demand for their product
- A firm's **production function** for a particular good (q) shows the maximum amount of the good that can be produced using alternative combinations of inputs (for example, capital (k) and labor (l))

$$q = f(l, k)$$

- In the short run, quantity of capital is fixed at some amount (say $\bar{k}$), then

$$q = f(l, \bar{k})$$

- Total output produced by different levels of the variable input, holding all other inputs constant



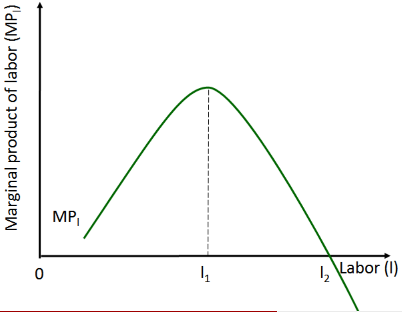
- **Diminishing marginal productivity**: Total output at first increases at an increasing rate – but after some point increases at a decreasing rate
- The total product curve shows that output initially increases at an increasing rate from 0 to l_1, then increases at a decreasing rate from l_1 to l_2, and finally declines after l_2 as the firm employs more units of labor

- Diminishing marginal productivity provides the reason why output fails to expand at an increasing rate after l_1 units of labor

- **Marginal Product:** Change in total output associated with a one-unit change in one input holding all other inputs constant

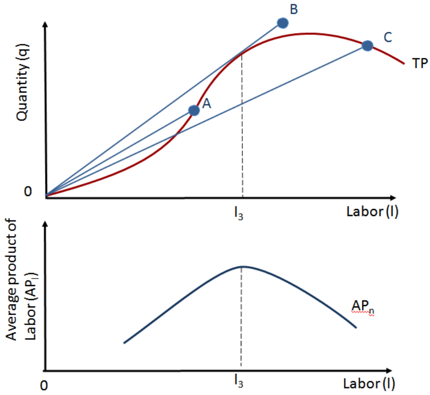
$$MP_l = \frac{\Delta q}{\Delta l} \qquad\qquad MP_k = \frac{\Delta q}{\Delta k}$$

- MP_l : Additional output produced by each additional unit of labor
- Slope of the total product curve



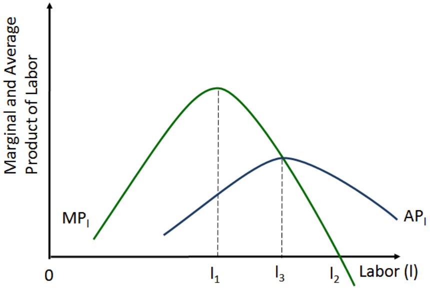
- Initially, increasing marginal productivity (MP_l is positive and increasing)
- Next, diminishing marginal productivity sets in (MP_l is positive but decreasing)
- Next, $MP_l = 0$ (when total product is maximum)
- Eventually, MP_l is negative

- **Average Product:** Total output divided by the number of units employed – Average Product of Labor $AP_l = \frac{q}{l} = \frac{f(l,k)}{l}$



- If labor is measured in hours, the AP_l provides a measure of quantity produced per hour
- Extend a ray from origin to each point on TP curve – then AP_l is the slope of the ray
- Average productivity first increases with labor and then declines because of increasing and then diminishing marginal productivity

- Average productivity rises when marginal productivity exceeds average productivity
- Average productivity falls when marginal productivity lies below average productivity
- Marginal productivity equals average productivity when average productivity is maximized



- MP curve cuts AP curve at its maximum point
- MP is above AP whenever AP is increasing
- MP is below AP whenever AP is declining

- Note that marginal productivity (of labor or capital) generally depends on all the inputs
- For instance, the marginal productivity of labor also depend on changes in other inputs such as capital
- An additional unit of labor will increase the production of the good more if there is more capital available
- Similarly, average product of labor (AP_l) also depends on the amount of capital employed
- However, in the short run we typically assume that capital cannot be changed

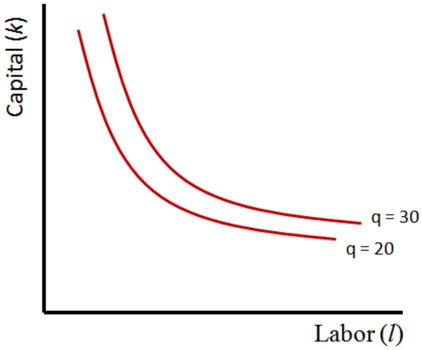
Production Functions

Isoquants

- An isoquant shows combinations of inputs (k,l) that can produce a given level of output (q_0) (rings a bell ... indifference curves)

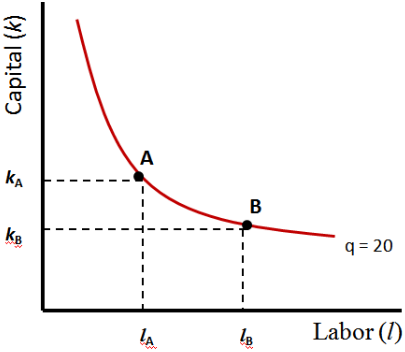
$$f(k,l) = q_0$$

- Isoquants illustrate the possible substitution of one input for another



- Isoquants are typically drawn on capital-labor graphs, showing the technological tradeoff between capital and labor in the production function, and the decreasing marginal returns of both inputs
- Each isoquant represents a different level of output
- Output rises as we move northeast
- As with indifference curves, two isoquants can never cross

- The slope of an isoquant shows the rate at which one input can be substituted for another (say l for k) while holding the output constant
- Marginal Rate of Technical Substitution (MRTS) = − slope of isoquant



- MRTS(l for k) or $\text{MRTS}_{lk} = -\frac{dk}{dl}\big|_{q=q_o}$
- MRTS(l for k) > 0 and is diminishing for increasing inputs of labor
- $\text{MRTS}_{lk} = \frac{MP_l}{MP_k}$
- An alternative measure, independent of units of measurement, is the elasticity of substitution σ

- Elasticity of substitution
  - Degree of substitutability depends on technical and legal considerations
  - Elasticity of substitution measures the curvature of an isoquant
  - Defined as percentage change in the input ratio divided by the percentage change in the ratio of the inputs's marginal productivities, holding constant the level of output
  - Equivalently, elasticity of substitution measures the percentage change in factor proportions due to a change in marginal rate of technical substitution
  - Elasticity of substitution between capital and labor (σ)

$$\begin{aligned}\sigma &= \frac{d\ln(k/l)}{d\ln(MP_l/MP_k)} \\ &= \frac{\Delta(k/l)}{(k/l)} \div \frac{\Delta(MP_l/MP_k)}{(MP_l/MP_k)} = \frac{\Delta(k/l)}{(k/l)} \div \frac{\Delta(MRTS_{lk})}{(MRTS_{lk})}\end{aligned}$$

  where ln( ) stands for the natural log
  - If $\sigma = 0$ then variable inputs cannot be substituted
  - If $\sigma = \infty$ then variable inputs are perfect substitutes

Production Functions

Returns to Scale

- So far, we have studied how the production changes when we change the amount of one input, leaving the rest constant (marginal productivity)
- But how does output respond to increases in all inputs together?
  - Suppose that all inputs are doubled, would output double?
- Returns to scale – two forces that come into operation as inputs are doubled
  - Greater division of labor and specialization of function
  - Loss in efficiency because management may become more difficult given the larger scale of the firm
- If the production function is given by $q = f(k,l)$ and all inputs are multiplied by the same positive constant $(t > 1)$, then
  - **returns to scale are constant** if $f(tk,tl) = tf(k,l)$
  - **returns to scale are decreasing** if $f(tk,tl) < tf(k,l)$
  - **returns to scale are increasing** if $f(tk,tl) > tf(k,l)$
- Whether there are constant, decreasing or increasing returns to scale depend on the production function ...

PRODUCTION FUNCTIONS

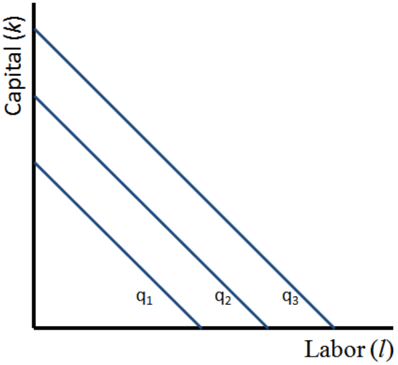
RETURNS TO SCALE

- It is possible for a production function to exhibit constant returns to scale for some levels of input usage and increasing or decreasing returns for other levels
- Note that while isoquants shows the extent to which the firm in question has the ability to substitute between the two different inputs, an isoquant map can also indicate decreasing or increasing returns to scale based on increasing or decreasing distances between the isoquant pairs of fixed output increment, as output increases
  - If the distance between those isoquants increases as output increases, the firm's production function is exhibiting decreasing returns to scale
  - Doubling both inputs will result in placement on an isoquant with less than double the output of the previous isoquant
  - Conversely, if the distance is decreasing as output increases, the firm is experiencing increasing returns to scale
  - Doubling both inputs results in placement on an isoquant with more than twice the output of the original isoquant

- Linear
  $q = f(k,l) = ak - bl$
- Fixed Proportions
  $q = \min(ak, bl)$ and $a, b > 0$
- Cobb-Douglas
  $q = f(k,l) = Ak^a l^b$ and $A, a, b > 0$
- Constant Elasticity of Substitution (CES)
  $q = A\big(\theta(aK)^\gamma + (1-\theta)(bL)^\gamma\big)^{1/\gamma}$

Production Functions

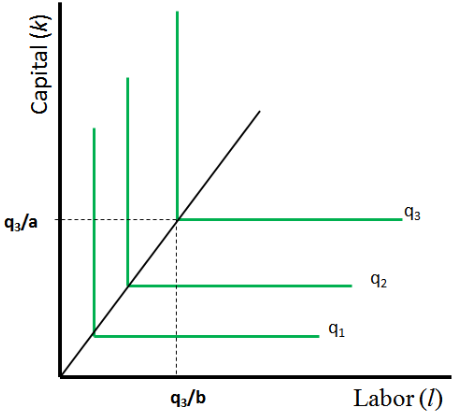
Linear Production Function



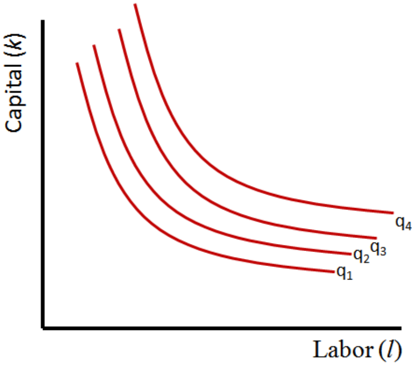
- Linear: $q = f(k, l) = ak - bl$
- Inputs are perfect substitutes
- $MP_k = a$
- $MP_l = b$

Production Functions

Fixed Proportions



- Fixed Proportions:
  $q = \min(ak, bl)$ and $a, b > 0$
- No substitution between labor and capital is possible
- They must always be used in a fixed ratio
- k/l is fixed at b/a
- Elasticity of substitution $\sigma = 0$



- Cobb-Douglas:
  $q = f(k,l) = Ak^a l^b$ and $A, a, b > 0$
- This production function can exhibit any returns to scale

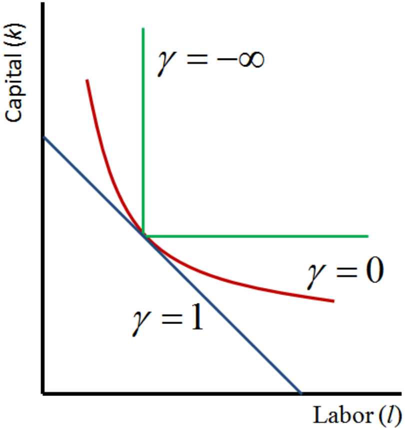
$$f(tk,tl) = A(tk)^a (tl)^b$$
$$= t^{a+b} f(k,l)$$

- Returns to scale
  - constant if $a + b = 1$
  - increasing if $a + b > 1$
  - decreasing if $a + b < 1$
- Elasticity of substitution $\sigma = 1$

- The Cobb-Douglas production function is linear in logarithms

$$\ln(q) = \ln(A) + a\ln(k) + b\ln(l)$$

where a and b are elasticity of output wrt k and l



- CES: $q = A\left(\theta(ak)^{\gamma} + (1-\theta)(bl)^{\gamma}\right)^{1/\gamma}$
- $0 < \theta < 1$ is the share parameter
- Parameters A, a, b depend upon the units in which the output and inputs are measured and play no important role
- $\gamma \leq 1$ (can be $-\infty$) and determines the degree of substitutability of the inputs
  - $\gamma = 1$: Inputs are perfect substitutes
  - $\gamma = -\infty$: Inputs are in fixed proportions
  - $\gamma = 0$: Elasticity of substitution is 1 (Cobb-Douglas)

PRODUCTION FUNCTIONS

A PRODUCTION FUNCTION FOR HOSPITAL ADMISSIONS

- Output
  - Heterogenous nature of hospital services
  - Hospital admissions serve as a measure of output
  - Adjust for case-mix: weighted sum of proportion of hospitals' patients in different diagnostic categories, where weights reflect average cost per case in each diagnostic category
  - Case-mix-adjusted hospital admissions
- Inputs
  - Physicians (labor)
  - Nurses (labor)
  - Other non-physician staff (labor)
  - Beds (rough measure of capital)
  - X (other production factors)
- Admissions $= f(\text{physicians, nurses, other labor, beds, X})$

- Jensen and Morrisey (1986)
  - Translog production function
  - Admissions $= f($physicians, nurses, other labor, beds, X$)$
  - Estimated production function for admissions using 3,500 nonteaching hospitals in the U.S. (1983)
- Marginal product for annual admissions wrt
  - Physicians 6.05
  - Nurses 20.3
  - Other Staff 6.97
  - Beds 3.04
- Marginal product for each input positive
  - Declined in magnitude with greater usage
  - Law of diminishing marginal product
  - Nurse input - by far the most productive input

TABLE 4.—ANNUAL MARGINAL PRODUCTS FOR ADMISSIONS EVALUATED AT THE MEAN

	Nonteaching		Teaching	
Input	Admissions	$CM \times$ Admissions	Admissions	$CM \times$ Admissions
1. Medical Staff	4.529[a]	6.049[a]	3.867[a]	2.209[c]
	(53.749)	(90.939)	(12.851)	(3.603)
2. Nursing Staff	19.117[a]	20.295[a]	11.669[a]	14.069[a]
	(942.246)	(1020.460)	(95.251)	(111.350)
3. Medical Residents	—	—	0.107	1.328
			(0.002)	(0.241)
4. Other Staff	5.263[a]	6.969[a]	1.324[b]	2.511[a]
	(222.557)	(350.424)	(5.885)	(17.244)
5. Beds	5.203[a]	3.036[a]	14.723[a]	15.935[a]
	(56.151)	(17.332)	(103.252)	(79.819)

Note: The number in parentheses is the F-statistic associated with the test that the marginal product is zero.
[a] Significant at the $\alpha = 0.01$ confidence level.
[b] Significant at the $\alpha = 0.05$ confidence level.
[c] Significant at the $\alpha = 0.10$ confidence level.

Production Functions

A Production Function for Hospital Admissions

- Elasticity of Substitution

Input Pair	Nonteaching Case-Mix Adjusted Admissions	Teaching Case-Mix Adjusted Admissions
1. Medical Staff with Nurses	0.547	0.159
2. Medical Staff with Beds	0.175	0.155
3. Nurses with Beds	0.124	0.211
4. Nurses with Residents	—	2.127
5. Medical Staff with Residents	—	0.292

Source: Jensen and Morrisey (1986).

- Substitution elasticities between Physicians and nurses, 0.547
  - A 10% increase in the marginal productivity of a doctor causes a 5.47% increase in the ratio of nurses to doctors
- Physicians and beds, 0.175
- Nurses and beds, 0.124
- Hospital policy makers can avoid some of the price (wage) increase in any one input by substituting with the others

Cost Functions

Introduction

- Just as production function describe input/output relationship, cost function describes the cost/output relationship
- The two are closely related and under correct conditions, one can be derived from the other
  - Two Simplifying Assumptions
  - There are only two inputs
    - homogeneous labor (l), measured in labor-hours
    - homogeneous capital (k), measured in machine-hours (entrepreneurial costs are included in capital costs)
  - Firms cannot influence the input prices, they are given . . . (they do not depend on firms decisions on the inputs to be used)

Cost Functions

Economic Costs

Differentiate between accounting cost and economic cost

- Accountants view of cost stresses out-of-pocket expenses, historical costs, depreciation, and other bookkeeping entries
- Economists focus on opportunity cost – opportunity costs are what could be obtained by using the input in its best alternative use
- The economic cost of any input is the payment required to keep that input in its present employment the remuneration the input would receive in its best alternative employment
    - **Labor Costs:** To both economist and accountants, labor costs are very much the same thing: labor costs of production (hourly wage)
    - **Capital Costs:** Accountants use the historical price of the capital and apply some depreciation rule to determine current costs – For economists, cost of the capital is what someone else would be willing to pay for its use (and this is what the firm is forgoing by using the machine)
    - **Entrepreneurial Costs:** Accountants use revenues or losses left over after paying all input costs – Economists consider the opportunity costs of time and funds that owners devote to the operation of their firms

- Let the production function be given by

$$q = f(l,k)$$

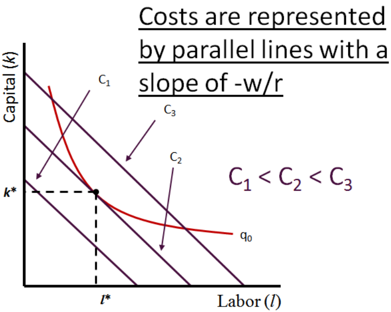
  - In the short run, assume k is fixed at $\bar{k}$
  - Derive the cost function from the production function
    - To achieve a given amount of output q, find the lowest cost input combination to do that
    - Then determine the total cost of those required inputs

- The cost function

$$\begin{aligned} TC(q) &= r\bar{k} + wl \\ &= FC + VC \end{aligned}$$

where q is the output, l and k are capital (in the short run fixed at $\bar{k}$), w and r are the wages and rental price, and FC and VC stand for fixed and variable costs

- Say the firm has already chosen the level of output (q_0) and wants to minimize its costs
- Given the input prices (w, r), how should the firm choose inputs so as to minimize costs at q_0



- The minimum cost of producing q_0 is C_2
- This occurs at the tangency between the isoquant and the iso-cost curve – at (l^*, k^*)
- Mathematically, we seek to solve for values of (l, k) that minimize total cost subject to $q = f(k, l) = q_0$
- This gives the solution

$$\frac{w}{r} = \frac{MP_l}{MP_k} = MRTS_{lk}$$

(or equivalently $\frac{MP_k}{r} = \frac{MP_l}{w}$)

- For costs to be minimized, the marginal productivity per dollar spent should be the same for all inputs
- The solution to this minimization problem for an arbitrary level of output q_0 give us the contingent demand functions for inputs

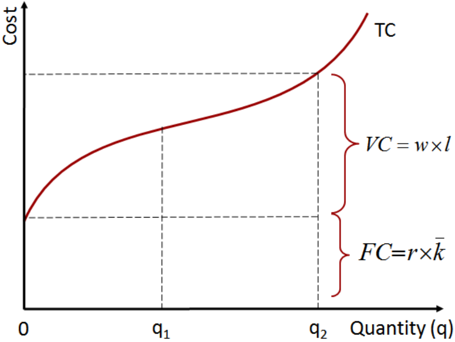
$$l^* = l^*(w,v,q_0) \qquad\qquad k^* = k^*(w,v,q_0)$$

- Thus, the **total cost function** gives the minimum cost incurred by the firm to produce any output level with given input prices – $TC = TC(r,w,q)$
- We compute it as $TC = C(r,w,q) = w \times l^*(w,v,q) + r \times k^*(w,v,q)$
- The **average cost function (AC)** is found by computing total costs per unit of output $AC(r,w,q) = \frac{C(r,w,q)}{q}$
- The **marginal cost function (MC)** is found by computing the change in total costs for a change in output produced $MC(r,w,q) = \frac{\Delta C(r,w,q)}{\Delta q}$

Cost Functions

Short-Run Total Cost Function

- Relationship between TP and (short-run) TC curve



- When TP is increasing at an increasing rate, TC is increasing at a decreasing rate
- When TP increasing at a decreasing rate TC is increasing at an increasing rate
- TC first increases at a decreasing rate up to point q_1 and then increases at an increasing rate with respect to producing more output
- TC increases at an increasing rate after q_1 because of diminishing marginal productivity

- **Fixed costs:** Occur in the short-run (operating period) – levels of some inputs are fixed
- **Variable costs:** All inputs are variable during the long run or planning period

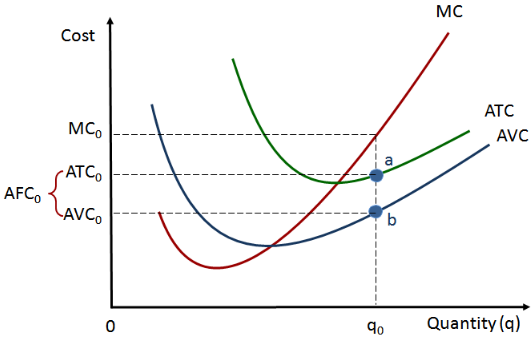
- Marginal Cost – Change in total cost with respect to output
- Marginal Cost depends on the marginal productivity of the variable factor

$$\begin{aligned} \text{MC} &= \frac{\Delta(TC)}{\Delta q} = \frac{\Delta(r\bar{k} + wl)}{\Delta q} \\ &= w\frac{\Delta l}{\Delta q} = \frac{w}{MP_l} \end{aligned}$$

- Average Cost – Total cost per unit of output
- Average Variable Cost depends on the average productivity of the variable factor

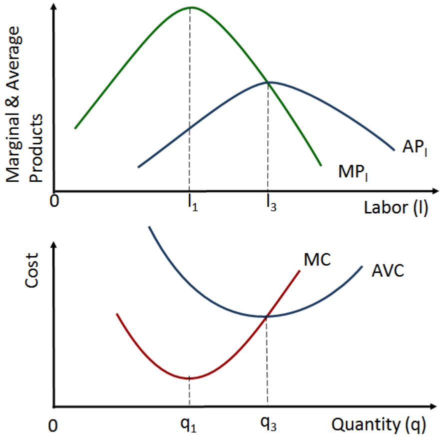
$$\text{AVC} = \frac{TVC}{q} = \frac{wl}{q} = \frac{w}{AP_l}$$

Relationships among Marginal, Average Variable, and Average Total Costs



- Average total cost, ATC, equals the sum of average variable cost, AVC, and average fixed cost, AFC
- AFC is reflected in the vertical distance between the ATC and AVC
- MC cuts both of the average cost curves at their minimum points
- MC lies above the AVC and ATC curves when they are rising and below them when they are falling

Relationships among Productivity and Cost Curves



- MC inversely related to marginal product of labor
- AVC inversely related to average product of labor
- Maximum point on MP curve corresponds to minimum point on the MC curve
- Maximum point on the AP curve corresponds to the minimum point on the AVC curve

Factors Affecting the Position of the Short-Run Cost Curves

- Prices of the variable inputs increase – cost curves shift upward
- Prices of the variable inputs decrease – cost curves shift downward
- Better quality of care – cost curves shift upward
- More severe patient case-mix – cost curves shift upward
- Excessive amounts of the fixed inputs – cost curves shift upward

- STVC = f(output, input prices, quality of care, patient case-mix, quantity of the fixed inputs)
- Cowing and Holtmann (1983)
  - $STVC = f(q_1,\ldots,q_5,w_1,\ldots,w_6,K,A)$
  - $q_i(i=1,5)$ – quantity of five different patient services (ER care, medical-surgical care, pediatric care, maternity care, and other inpatient care - measured in total patient days)
  - $w_j(j=1,6)$ – variable input prices (Nursing labor, auxiliary labor, professional labor, administrative labor, general labor, and material and supplies)
  - K – single measure of the capital stock (measured by the market value of a hospital)
  - A – fixed number of admitting physicians in the hospital

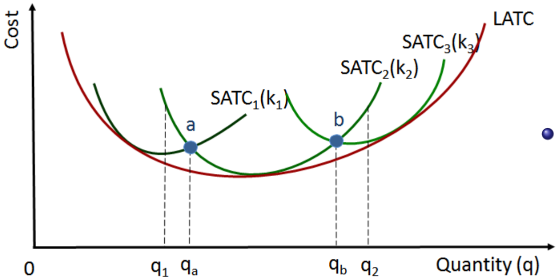
Cowing and Holtmann (1983), results

- Short-run economies of scale
- Limited evidence for economies of scope with respect to pediatric care and other services
- Limited evidence for diseconomies of scope with respect to emergency services and other services
- Short-run marginal cost of each output
  - Declined and then became constant over the levels of output observed in their study
- Estimate short-run elasticities of input substitution between all pairs of variable inputs
  - Substantial degree of substitutability between
    - Nursing and professional workers
    - Nursing and general workers
    - Nursing and administrative workers
    - Professional and administrative labor

Cost Functions

Long Run Costs of Production

- In the long run, all inputs are variable
- Capital (k) is no longer fixed



- Economies of scale
  - Average cost per unit of output falls as the firm increases output – LTAC has a downward slope
  - Due to specialization of labor and capital
  - Increasing returns to scale
- Diseconomies of scale
  - Average cost per unit of output rises as the firm increases output – LATC has an upward slope
  - Due to management issues
  - Decreasing returns to scale in production