Costs and Supply 1

Introduction

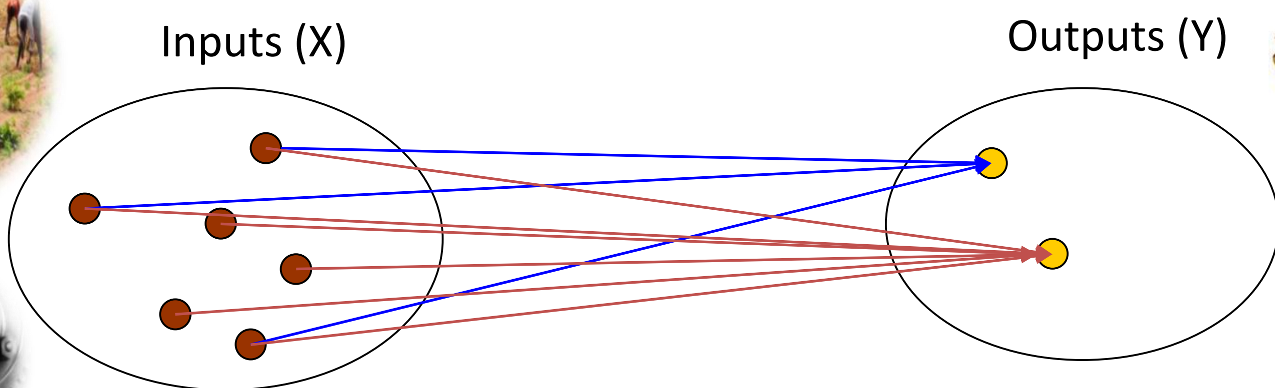
Production Function

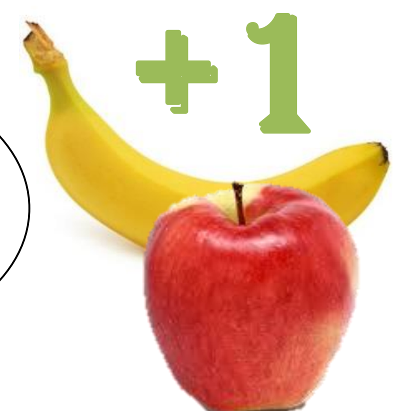
Economies of Scale

MICRO- AND MACROECONOMICS

The firm's technology: input and output

- We start by introducing the production function, which describes the firm's technology.
- An **input** (or **factor of production**) is a good or service used to produce output[s] (other goods or services).



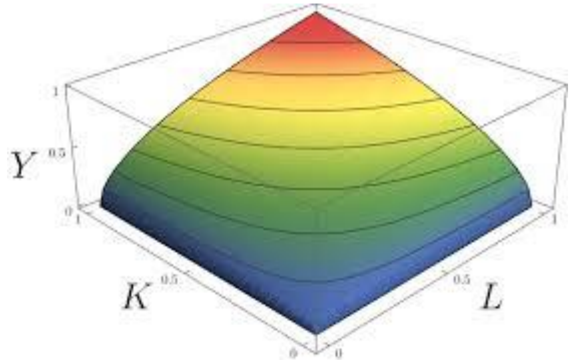


$$\underline{y} = f(\underline{x}) \quad : \underline{x} \in X;\ \underline{y} \in Y; \quad f(\underline{x}_i) \geq \underline{y}_i \ \forall i$$

$\underline{y}$: output(s); $\underline{x}$: input(s); f: production function; Y: set of outputs; X: set of inputs

Inputs (factors of production)

- Inputs include labour, machinery, buildings, raw materials, and energy.
- Suppose our firm uses inputs to make spoons. This is an engineering and management problem.
- Making spoons is largely a matter of technology and on-the-job experience.

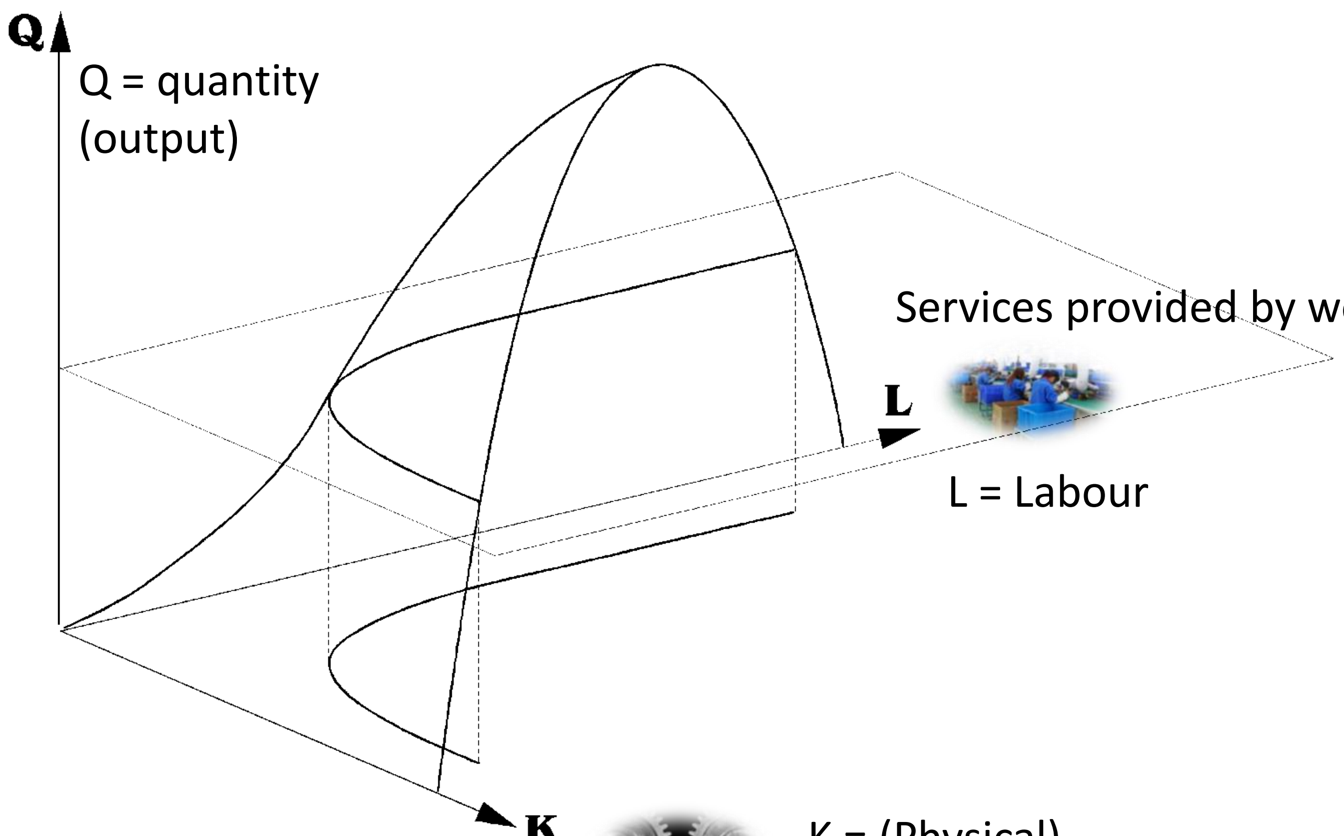


The production function

- The production function summarizes technically efficient ways to combine inputs to produce output.
- A production technique is **technically efficient** if there is no other way to make a given output using less of one input and no more of the other inputs.
- The **production function** is the set of all technically efficient techniques.

Production functions

- Production function (in theory): Q = f(L,K,A,E)
- Simplified version: Q = f(K,L)



Machines, tools, equipment, buildings

K = (Physical) capital

Wasteful production methods

- Since profit maximizing firms are not interested in wasteful production methods, we restrict our attention to those that are technically efficient.
- To make a spoon, method ***A*** needs 2 workers and 1 machine, but method ***B*** needs 2 workers and 2 machines. 

- Method ***B*** is less efficient than method ***A***. It uses more machines, but the same labour, to make the same output. Method ***B*** is not one of the efficient methods listed in the production function.

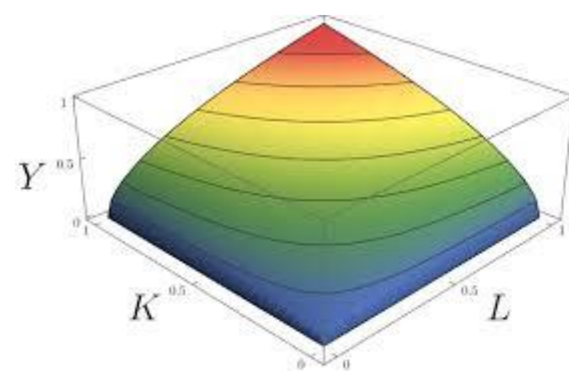
Example: A production function

Output 	Capital input	Labour input
100	4	4
100	2	6
106	2	7
200	4	12

- The table to the left shows some technically efficient methods in the production function. The first two rows show two ways to make 100 spoons: 4 machines and 4 workers, or 2 machines and 6 workers.
- Beginning from the latter, the third row shows the effect of adding an extra worker. Output rises by 6 spoons.
- The last row shows that doubling both inputs in the second row also doubles the output, though this need not be so: overcrowding a small factory can slow people down.

Why do firms need engineers?

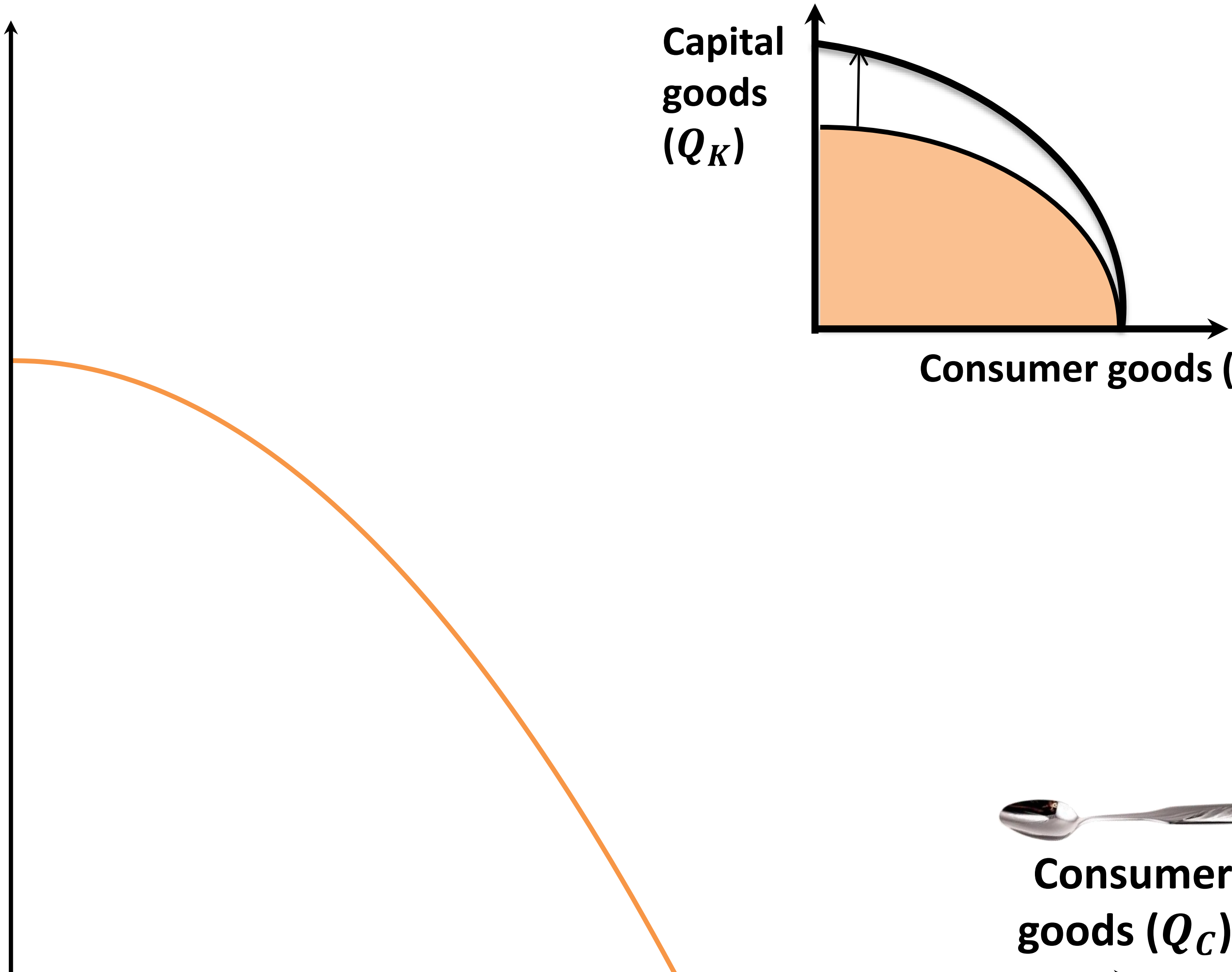
- The previous table could be enlarged to include other combinations of labour and capital that are also technically efficient.
- A firm discovers its production function, the complete set of technically efficient production techniques, by asking its engineers, designers, and time-and-motion experts; and by trial and error.



Technique, technology, and technical progress

- A **technique** is a particular way to combine inputs to make output.
- **Technology** is the list of all known techniques.
- **Technical progress** is a new technique allowing a given output to be made with fewer inputs than before.

Technical progress (PPF)



Technical progress

- A method previously technically efficient may become inefficient after a technical advance allows a better production technique.
- Technical progress alters the production function.
- For now, we assume a given technology and a given production function.

- The production function relates volumes of inputs to volume of output. To get costs, we also need to know input prices.
- Consider the lowest-cost way to make 100 spoons. Assume that there are two technically efficient techniques, the first two rows of our previous table, reproduced as the first two columns of the table below and labelled techniques ***A*** and ***B***.

Technique	Capital input	Labour input	Rental per machine	Wage per worker	Capital cost	Labour cost	Total cost
A	4	4	€320	€300	€1280	€1200	€2480
B	2	6	€320	€300	€640	€1800	€2440 Lower!

Economically efficient production methods

- It costs €320 to rent a machine, and €300 to hire a worker.
- To make 100 spoons, the table on the previous slide shows that the total cost is €2480 with technique ***A*** and €2440 with technique ***B***.
- The firm chooses ***B***. 100 spoons at a total cost of €2440 is one point on the total cost curve for spoons. It is the **economically efficient** (lowest-cost) production method at the going rental and wage rates.

Deriving the total cost curve

- To get the whole **total cost** curve, we repeat the calculation for each output. The production function tells us the inputs needed by each technique. Using input prices, we calculate the cost using each technique, and choose the lowest-cost production method.
- Joining up these points we get the total cost curve, which may switch from one production technique to another at different outputs. From the total cost curve we calculate the marginal cost curve – the rise in total cost at each output when output is increased by one more unit.

Factor intensity

- A technique using a lot of capital and little labour is ‘capital-intensive’.
- A technique using a lot of labour but relatively little capital is called ‘labour-intensive’.
- In our previous table, technique ***A*** was more capital-intensive and less labour-intensive than technique ***B***. (The ratio of capital input to labour is 1 in technique ***A*** but only one-third in technique ***B***.)

Factor prices and relative prices

- At the **factor prices** (prices per unit input) in our previous table, the more labour-intensive technique was cheaper.
- Suppose the wage rises from €300 to €340: labour is dearer but the rental on capital is unchanged.
- The **relative price** of labour has risen.

Effects of a price change

- We ask two questions: First, what happens to the total cost of making 100 spoons?
- Second, is there any change in the preferred technique?
- The table below recalculates production costs at the new factor prices.

Technique	Capital input	Labour input	Rental per machine	Wage per worker	Capital cost	Labour cost	Total cost
A	4	4	€320	€340	€1280	€1360	€2640 Lower!
B	2	6	€320	€340	€640	€2040	€2680

Effects of a price change (2)

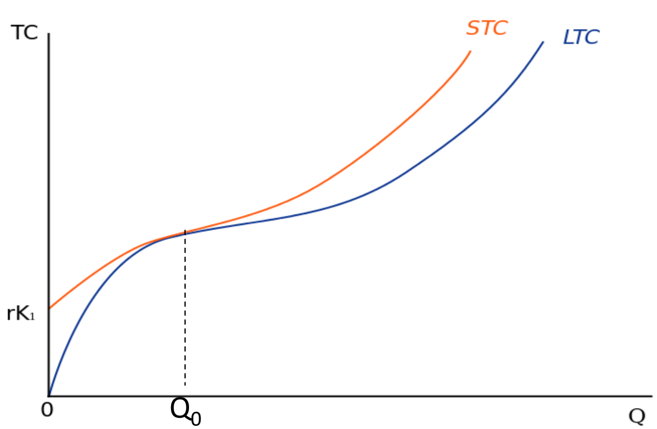
- As we could see, since both techniques use some labour, the total cost of making 100 spoons by each technique rises.
- Repeating this argument at all output, the total cost curve must shift upwards when the wage rate (or the price of any other input) rises.
- In this example, the rise in the relative price of labour leads the firm to switch techniques: it switches to the more capital-intensive technique ***A***.

Long-run analysis

- Faced with an upward shift in its demand and MR curves, a firm will expand output.
- However, adjustment takes time. Initially, the firm can get its existing workforce to do overtime. In the long run, the firm can vary its factory size, switch techniques of production, hire new workers and negotiate new contracts with suppliers of raw materials.

Long run and short run

- The **long run** is the period long enough for the firm to adjust **all its inputs** to a change in conditions.
- In the **short run** the firm can make only **partial** adjustment of its inputs to a change in conditions.



Partial adjustments are more costly: short-run total costs are at least as high, but usually higher than the long-run total costs if the firm increases or decreases its output.

Higher short-run costs

- The firm may be able to alter the **shift length** at once.
- **Hiring or firing** workers takes longer, and it might be years before a **new factory** is designed, built, and operational.
- For now, we will only analyse long-run cost curves, when the firm can make all the adjustments it desires.

Long-run total and marginal costs

- **Long-run total cost** is the minimum cost of producing each output level when the firm can adjust all inputs.
- **Long-run marginal cost** is the rise in long run total cost if output rises permanently by one unit.

Alternatively, it may be defined as the derivate of the total cost function LTC(q):

$$\mathrm{LMC} = \frac{\partial \mathrm{LTC(q)}}{\partial \mathrm{q}}$$

Long-run costs

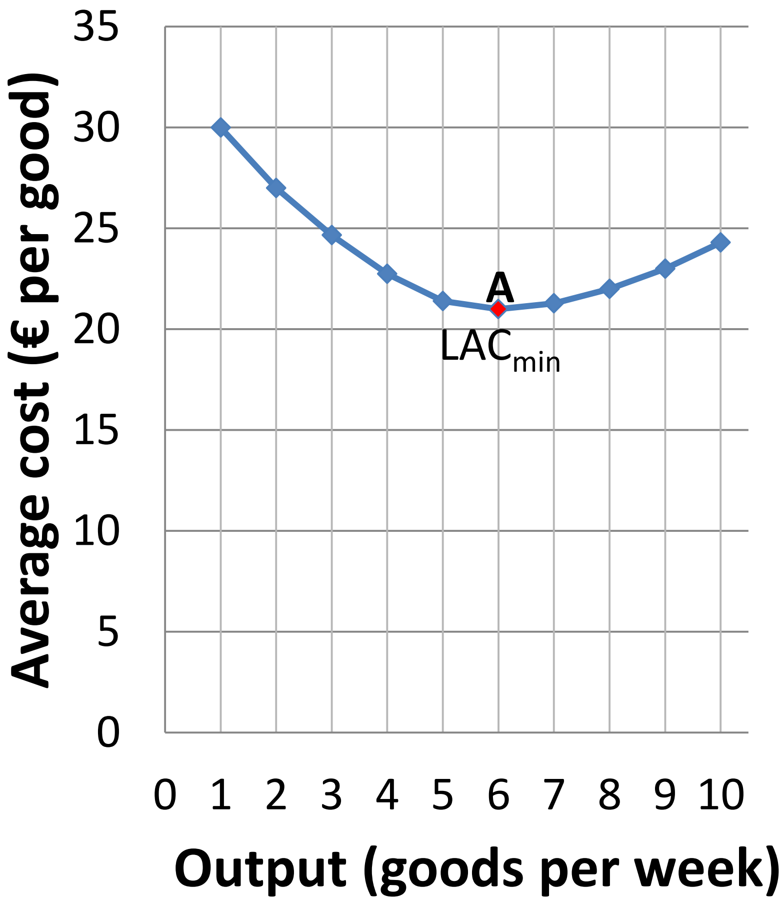
(1) Output	(2) Total cost (€)	(3) Marginal cost (€)	(4) Average cost (€)
0	0		
1	30	30	30
2	54	24	27
3	74	20	24.67
4	91	17	22.75
5	107	16	21.4
6	126	19	21
7	149	23	21.29
8	176	27	22
9	207	31	23
10	243	36	24.3

- The table to the left shows long-run total costs LTC and long-run marginal costs LMC of making each output.
- Since there is always an option to close down entirely, the LTC of producing zero output is zero. LTC describes the eventual costs after all adjustments have been made.
- LTC must rise with output: higher output always costs more to produce. LMC is always positive.

Long-run average cost

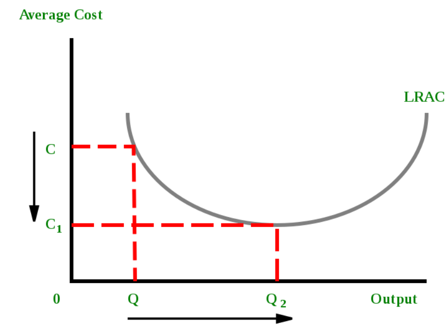
- Can large firms produce goods at a lower unit cost than small firms? Might it be a disadvantage to be (too) large?
- To answer these questions, we need to think about average cost per unit of output.
- **Long-run average cost** (LAC) is the total cost (LTC) divided by the level of output (Q).

The LAC curve (figure 1)



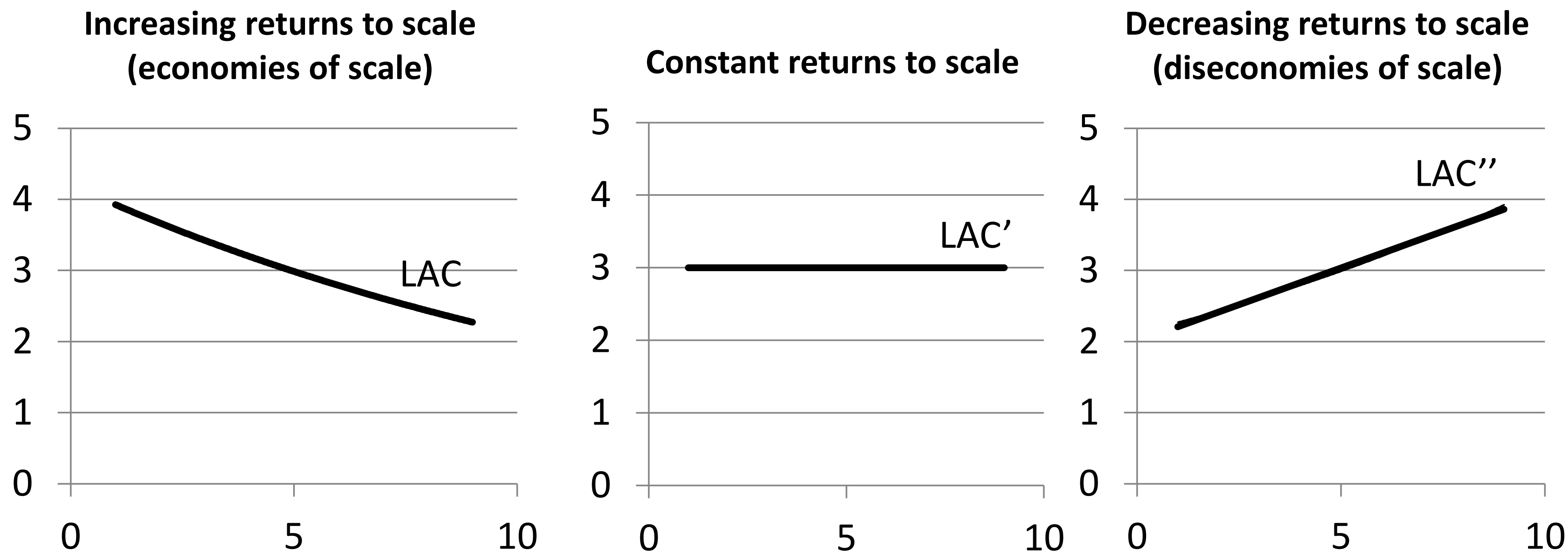
- Our previous table showed long-run average cost LAC (column 2 divided by column 1). These LAC data are plotted in the figure to the left (figure 1).
- Average cost starts out high, then falls, then rises again. This common pattern of average costs is called the U-shaped average cost curve. To see why U-shaped average cost curve is common in practice, we examine ‘returns to scale’.

Returns to scale

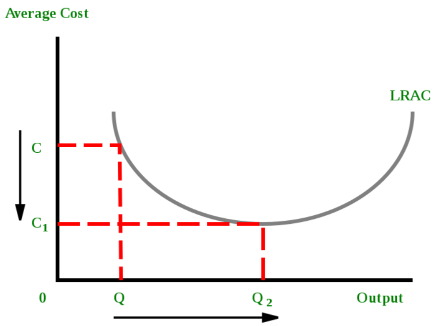


- **Economies of scale (or increasing returns to scale)** mean long-run average cost falls as output rises.
- **Diseconomies of scale (or decreasing returns to scale)** mean long-run average cost rises as output rises.
- **Constant returns to scale** mean long-run average costs are constant as output rises.

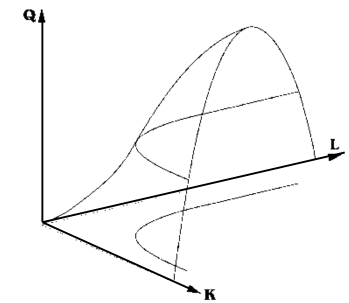
Returns to scale and long-run average cost (LAC) curves



- When LAC is declining, average costs of production fall as output increases and there are economies of scale.
- When LAC is increasing, average costs of production increase with higher output, and there are decreasing returns to scale.
- The intermediate case, where average costs are constant, has constant returns to scale.

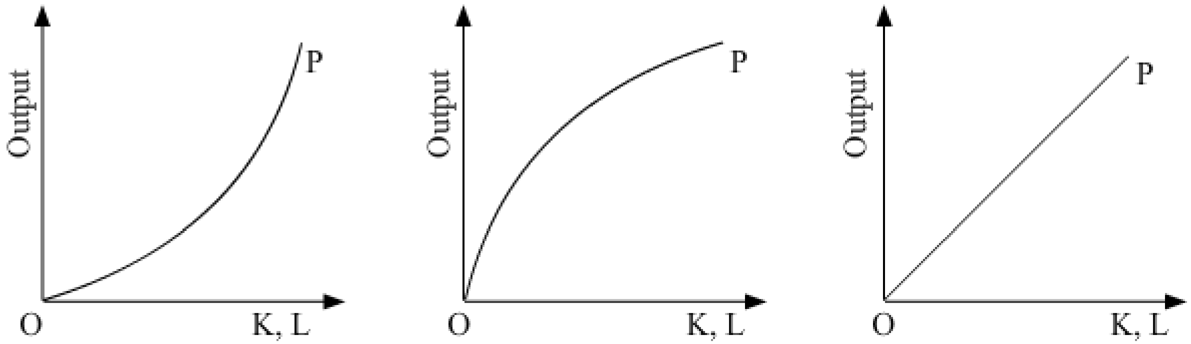


Returns to scale (2)



- In the figure above (and to the left), the U-shaped average cost curve has scale economies up to point Q_2, where average cost is lowest. At higher outputs there are diseconomies of scale.
- Why are there scale economies at low output levels but diseconomies of scale at high output levels? We draw a cost curve for given input prices. Changes in average costs as we move along the LAC curve cannot be explained by changes in factor prices. (Changes in factor prices ***shift*** cost curves.)
- The relationship between average costs and output as we move along the LAC curve depends on the technical relation between physical quantities of inputs and output, summarized in the production function.

Returns to scale (3)



(a) Increasing Returns to Scale | (b) Decreasing Returns to Scale | (c) Constant Returns to Scale

> Increasing returns to scale mean that one (big) company can produce more efficiently than more than one (smaller) firms utilizing the same factors of production → emergence of natural monopolies

Formally, „returns to scale", refers to changes in output (quantity, q) resulting from a proportional change (λ) in all inputs:

a) If output increases by more than that proportional change, there are increasing returns to scale (economies of scale):

e.g. $Q = KL \quad Q(\lambda K, \lambda L) = (\lambda K)(\lambda L) = \lambda^2 Q(K, L)$

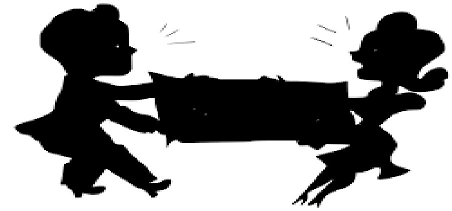
b) If output increases by less than that proportional change, there are decreasing returns to scale:

e.g. $Q = \sqrt[3]{KL} \quad Q(\lambda K, \lambda L) = \sqrt[3]{(\lambda K)(\lambda L)} = \lambda^{2/3} Q(K, L)$

c) If output increases by that same proportional change, there are constant returns to scale:

e.g. $Q = \sqrt{KL} \quad Q(\lambda K, \lambda L) = \sqrt{(\lambda K)(\lambda L)} = \lambda Q(K, L)$

Indivisibilities



- There are three reasons for economies of scale. The first is indivisibilities in the production process, a minimum quantity of inputs required by the firm to be in business at all whether or not output is produced.
- These are sometimes called quasi-fixed costs, because they do not vary with output level. To be in business a firm requires a manager, a telephone, an accountant, a market research survey. The firm cannot have half a manager and half a telephone merely because it wishes to operate at low output levels.

Quasi-fixed costs

- Beginning from small output levels, these costs do not initially increase with output.
- The manager can organize three workers as easily as two. As yet there is no need for a second telephone.
- There are economies of scale because these costs can be spread over more units of output as output is increased, reducing average cost per unit of output.
- However, as the firm expands further, it has to hire more managers and telephones and these economies of scale die away.

- The average cost curve stops falling.

Specialization

- The second reason for economies of scale is specialization.
- A sole trader must undertake all the different tasks of the business.
- As the firm expands and takes on more workers, each worker can concentrate on a single task and handle it more efficiently.

Machinery

- The third reason for scale economies is closely related. Large scale is often needed to take advantage of better machinery.
- No matter how productive a robot assembly line is, it is pointless to install one to make six cars a week. Average costs would be enormous.
- However, at high output levels the machinery cost can be spread over a large number of units of output and this production technique may produce so many cars that average costs are low.

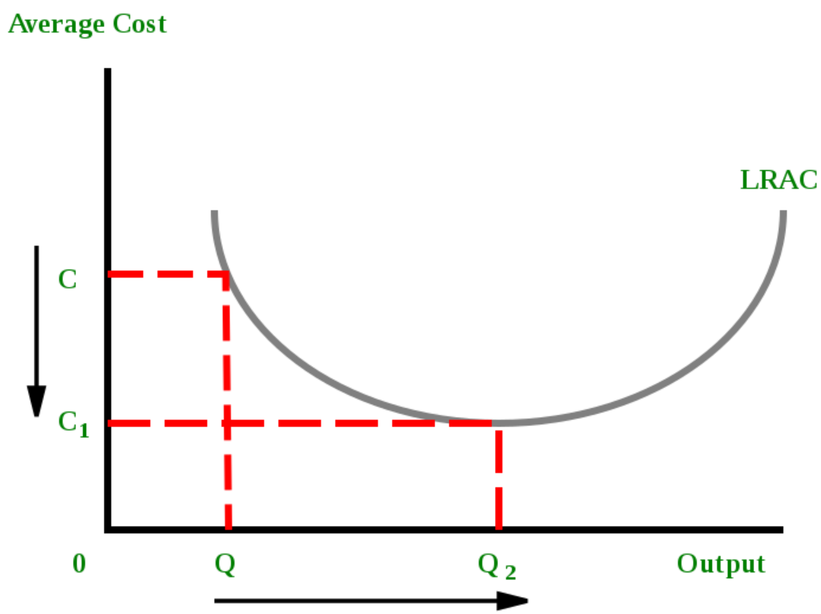
Managerial diseconomies of scale

- Beyond some output, the U-shaped average cost curve turns up again as diseconomies of scale begin.
- Management is harder as the firm gets larger: there are managerial diseconomies of scale.
- Large companies need many layers of management, themselves needing to be managed. The company becomes bureaucratic, co-ordination problems arise, and average costs begin to rise.

Other examples of diseconomies of scale

- Geography may also explain diseconomies of scale. If the first factory is located in the best site to minimize the cost of transporting goods to the market, the site of a second factory must be less advantageous.
- To take a different example, in extracting coal from a mine, a firm will extract the easiest coal first. To increase output, deeper coal seams have to be worked and these will be more expensive.

The shape of the average cost curve



- As output increases, the shape of the average cost curve thus depends on two things: how long economies of scale persist, and how quickly the diseconomies of scale set in.
- The balance of these two forces varies from industry to industry and firm to firm.

Returns to scale in practice

- To gather evidence on returns to scale we can talk to design engineers to see how production costs vary with output.
- Many studies of manufacturing industries confirm that scale economies continue over a wide range of output.
- The long-run average cost curve slopes down, albeit at an ever decreasing rate.

Manufacturing and service sector industries

- Scale economies in manufacturing industries are substantial.
- However, there are many industries, even in the manufacturing sector, where minimum efficient scale (MES) for a firm is small relative to the whole market and average costs are only a little higher if output is below minimum efficient scale.
- **Minimum efficient scale** is the lowest output at which the LAC curve reaches its minimum.
- For some industries, particularly personal services (hairdressers, doctors, etc.), economies of scale run out at quite low output levels.


Thank you