Unit 5. Producer theory: revenues and costs

Learning objectives

- to understand the concept of the short-run production function, describing the relationship between the quantity of inputs and the quantity of output;
- to understand within the context of the production function, average and marginal products as well as the law of diminishing marginal returns;
- to learn the link between productivity and costs;
- to examine the relationships among the short-run costs: total, average and marginal;
- to consider the process of cost minimization;
- to explain the properties of long-run costs;
- to examine economies and diseconomies of scale, as well as returns to scale;
- to make the distinction between accounting and economic profits;
- to establish the profit-maximizing rule, using marginal analysis.

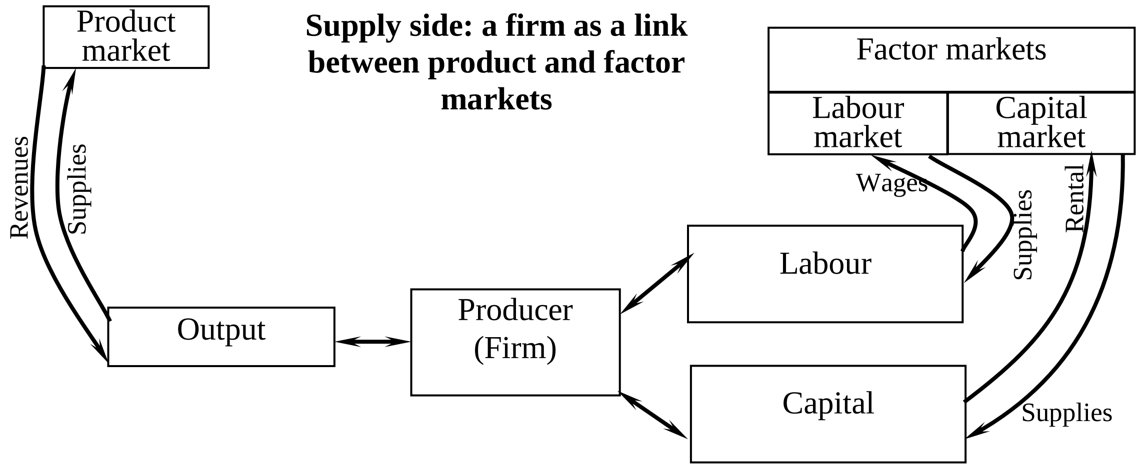
Questions for revision:

- ✓ Utility maximization: optimal consumption bundle;
- ✓ Opportunity cost.

5.1. Production functions: short and long run. Average and marginal product. Diminishing returns. Economies of scale

A firm is an institution that puts together factor and product markets. A factor of production is an aggregate input that plays a sufficient role in production of output. Main factors are capital (*K*) and labor (*L*). Other factors important factors are land and entrepreneurship.

A firm buys resourses at markets for factors and sells its output (*Q*) for the price (*P*) at product markets. Factor rewards – rental (*R*) and wage rate (*w*) – are production costs, and the product market via revenues is the source of financial resources for a firm (see the figure below).



Production function is a way to formalize transformation of inputs into output. Production function specifies the maximum output (Q) that can be produced from any given amount of inputs.

The long run is a sufficient period of time for a firm to adjust all its inputs to a change in economic environment. In the short run there are fixed and variable factors of production. The quantity of fixed factors can not be changed within some period of time. For instance, it takes rather long time to install a new plant. The quantity of a variable factor can be changed within rather short period of time. For instance, it may take a firm one day to hire and fire workers and thus to vary its staff. The short run is the period in which the firm can make only partial adjustment of its inputs to a change in economic environment.

A fixed factor of production is a factor whose level cannot be varied in the short-run. A variable factor can be adjusted, even in the short-run. Thus, the short run is a period of time when there is at least one fixed factor of production, and the long run is a period of time when all factors are veriable. We typically assume that capital is fixed and labour is the only variable factor in the short run.

Factor productivity can be treated either in average or in marginal sense. Average product of a factor (labour) is given by the output related to the quantity of the factor (labour):

$$AP_L = \frac{Q}{L}.$$

Average product of a variable factor (labour) is represented by the slope of an arrow coming from the origin and crossing total product curve in the point which corresponds to the given quantity of labour: $AP_L = tg\alpha$ (see the figure below).

Marginal product of a variable factor (labour) is a change in output obtained by a unitary change in the variable factor (labour), holding the other factors constant:

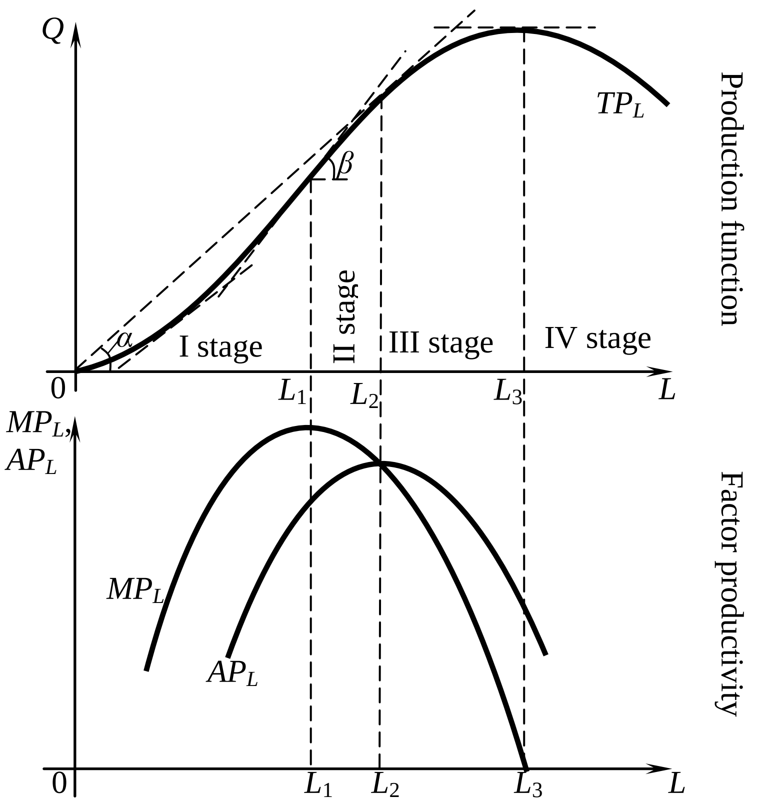
$$MP_L = \frac{\Delta Q}{\Delta L}.$$

In a limiting case when changes in labour are incremental marginal product of labour can be calculated as a derivative of production function with respect to the variable input:

$$MP_L = \lim_{\Delta L \to 0} \frac{\Delta Q}{\Delta L} = \frac{dQ}{dL}.$$

Marginal product of a variable factor (labour) is represented by the slope of a line tangent to total product curve: $MP_L = tg\beta$ (see the figure below).

Production technology in short run



In relation to changes in productivity one can point out four stages of production. If a firm hires is less than L_1 labour, output grows faster than employment due to gains in specialization and disposable production possibilities provided by installed capital. The I stage of production embraces the region of increasing marginal product of labour (see the figure below). When employment exceeds L_1, fixed factor starts to limit the growth of output. Each worker is equipped with less and less capital, and technology starts to exibit decreasing returns to a variable factor – marginal (stage II) and then average product of labour tends to decrease. One can imagine a situation when employment is so high (L_3) that workers begin hampering each other and total output goes down and marginal product of labour is negative (stage IV).

If marginal product of labour is less than average, i.e. the last hired worker produces less than the rest of the workers on the average, output with respect to total man-hours of labour, i.e. average product of labour, is going to decrease (stage III). And vice versa, if marginal product of labour exceeds average product, an additional worker hired will push average productivity up. So marginal product intersects average product of labour when the latter is at the maximum (see the figure below).

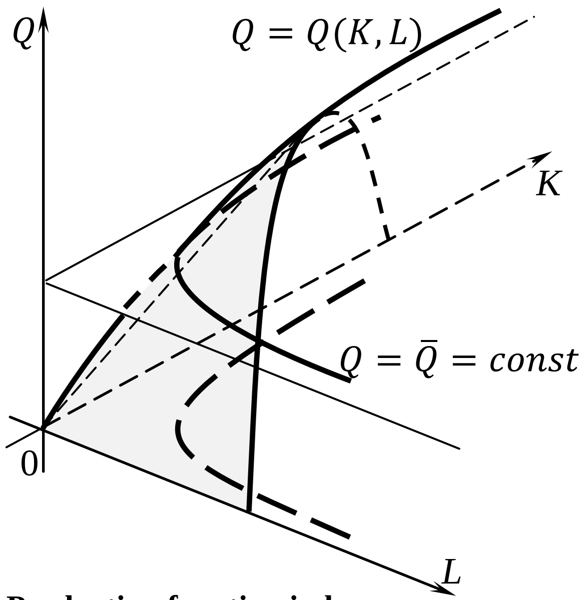
So the law of diminishing marginal returns says that beyond the Ist stage of production, continual increments in the variable input lead to the steadily decreasing marginal product.

Marginal and average product curves intersect at the point of maximum of average product of the variable factor. To prove it take a derivative of average product of labour:

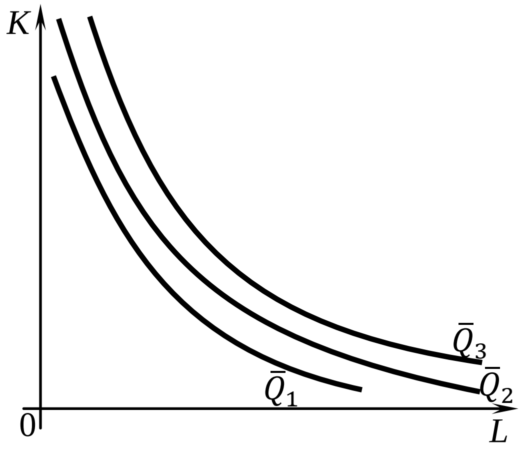
$$AP_L' = \frac{MP_L - AP_L}{L}.$$

At the point of maximum it is zero: $AP_L' = 0$. It follows that marginal product is equal to average product at this point: $MP_L = AP_L$.

Now let's consider production in long run when all the factors are variable. Production function can be represented by a set of isoquant lines (see the figure below). Each isoquant represents all the combinations of factors of production that correspond to a fixed level of output. The more distant from the origin isoquant curve corresponds to the higher output.



Production function in long run



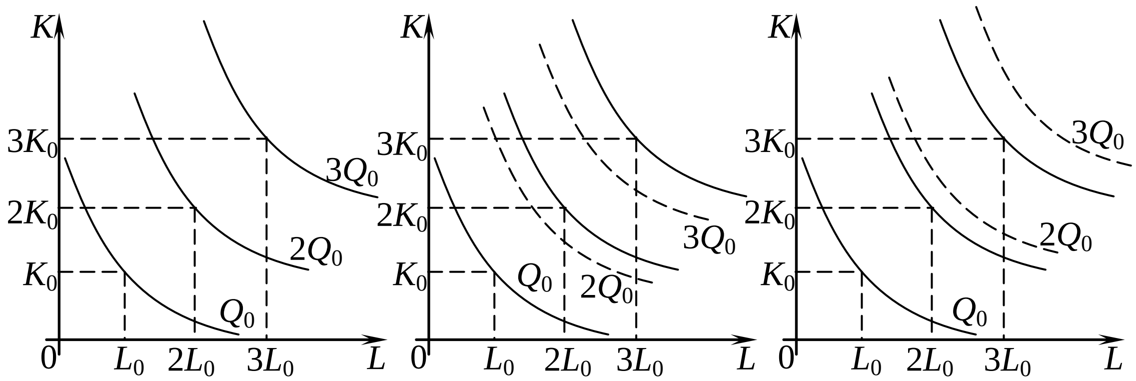
A set of isoquant curves

Production technology exhibits increasing returns to scale when with an increase in factors of production (in equal proportions) output grows in a greater proportion: $Q(\alpha K, \alpha L) > \alpha Q(K, L)$, where $\alpha > 1$ is a constant factor. Decreasing returns to scale are observed in the opposite situation: $Q(\alpha K, \alpha L) < \alpha Q(K, L)$. Constant returns to scale correspond to the boundary case, when $Q(\alpha K, \alpha L) = \alpha Q(K, L)$.

One should distinguish between diseconomies of scale and diminishing marginal returns. When we talk about (dis)economies of scale we mean that all factors can be varied. Economies of scale deal with relative disposition of isoquants that correspond to equal increments of output. Talking about diminishing marginal returns we suppose that one factor is varied and other are fixed. Marginal returns to a variable factor of production deal with relationship of quantities of output and the single variable input.

If technology exhibits constant returs to scale equiproportinal increments in labour and capital correspond to the equiproportional increments of output. If technology exhibits increasing returs to scale the higher output, achieved due to equiproportinal increment in labour and capital, corresponds to an isoquant which is situated closer to the origin as compared to the isoquant that corresponds to the equiproportional increment of output. If technology exhibits decreasing returs to scale the higher output, achieved due to equiproportinal increment in labour and capital, corresponds to an isoquant which is more distant from the origin as compared to the isoquant that corresponds to the equiproportional increment of output (see the figure below).

Constant returns to scale **Increasing returns to scale** **Decreasing returns to scale**



5.2. Opportunity cost and accounting costs. Total, average and marginal costs: short run and long run. Technology and costs. Cost minimizing input combination

Production costs (*TC*) are monetary expenditures for utilized resources. One can distinguish accounting and opportunity costs. Accounting costs are actual expenditures of the firm for acquired resources. Opportunity cost is the amount lost by not using the resource (labor or capital) in its best alternative use. Opportunity costs include actual as well as implicit costs that are foregone incomes of the owner of resoures in comparable business applications.

There are two huge types of production costs in short run (*STC*). The first class comprises fixed costs (*FC*), i.e. production expenditures that are one and the same whatever output is. These include installation costs; salaries of managers, other managerial costs; rent, insurance and adverticement costs; various information and communication costs etc.

The second class of short run production costs are variable costs (*VC*) – expenditures that depend on production output. These include expenditures for raw matherials, wages of workers, transportation costs, a fraction of power supply and so on.

Thus, $$STC = VC + FC.$$

Average costs (*AC*) are per unit total costs. In the short run these are the sum of average fixed (*AFC*) and variable (*AVC*) costs:

$$AC = \frac{TC}{Q} = AFC + AVC,$$

where $AFC = \frac{FC}{Q}$ is a unit fixed cost, and $AVC = \frac{VC}{Q}$ is a unit variable cost.

Marginal cost (*MC*) is the change in total cost due to an infinitesimal change in output. As fixed costs are independent of the volume of production, marginal cost is the change in variable cost due to an infinitesimal change in output:

$$MC = \frac{dTC}{dQ} = \frac{d(VC + FC)}{dQ} = \frac{dVC}{dQ}.$$

Short run total cost curve is the variable cost curve shifted upwards by the value of fixed cost. Fixed cost curve is a horizontal straight line. Average fixed cost curve is monotonically descending because AFC is a fixed value divided by (an increasing) output.

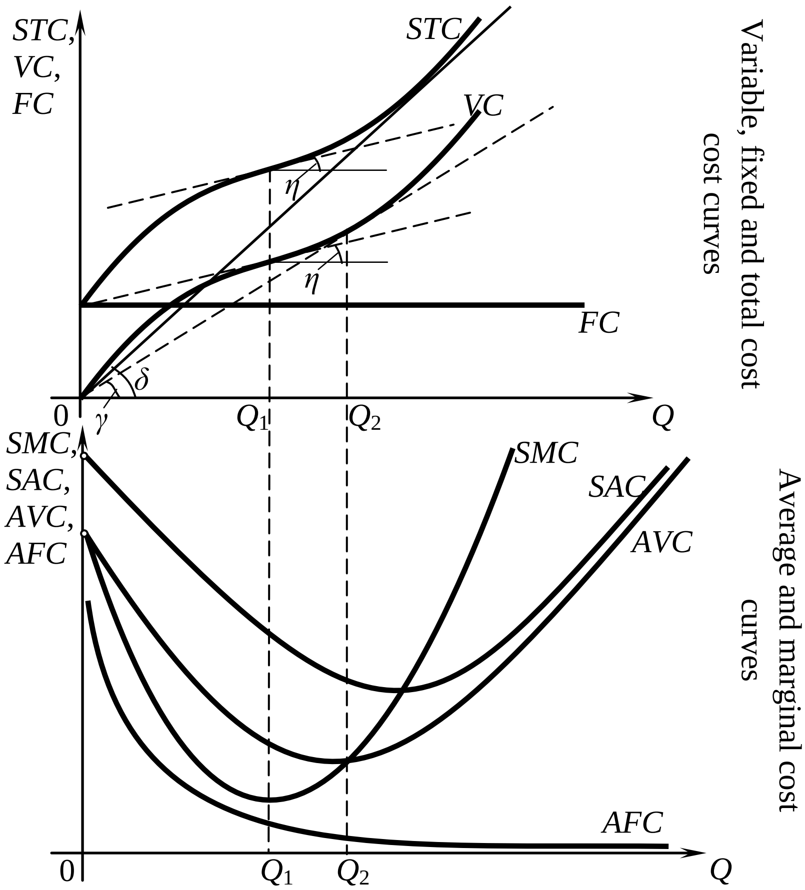
Suppose that marginal costs are less than average costs, i.e. the cost of producing the last, incremental unit of good is less than unit cost of production of the given amount of the good on the average. It means that average costs of producing total output, including the last unit, will go down, and average cost curve will be downward sloping. And vice versa, if marginal cost exceeds average cost, production of an additional unit of the good will push per-unit costs up, and average cost curve will be upward bending. It follows that marginal cost (*MC* and *SMC*) curve intersects average cost (*AC* and *SAC*) curve at the minimum of the latter.

To prove it take a derivative of average costs:

$$\frac{dAC}{dQ} = \frac{d}{dQ}\left(\frac{TC}{Q}\right) = \frac{Q \cdot MC - TC}{Q^2} = \frac{MC - AC}{Q}.$$

At the point of minimum of average costs $\frac{dAC}{dQ} = 0$, that yields $AC = MC.$

The same considerations can be applied to *MC* and *AVC* curves. *AVC* curve is declining whenever *MC* is below *AVC*, and rising whenever *MC* is above *AVC*. *MC* curve cuts the *AVC* curve at the minimum points of the average variable cost curve.



Short run cost curves

Suppose that in short run labour (L) is the only variable factor of production. So variable costs are expenditures for labour input ($VC=wL$), and AVC is inversely related to AP_L:

$$AVC = \frac{wL}{Q} = \frac{w}{Q/L} = \frac{w}{AP_L}.$$

Recall that in short run marginal costs show a change both in total and variable costs with respect to an incremental change in output. That's why there is an inverse relationship between MC and MP_L:

$$MC = \frac{d(wL(Q))}{dQ} = w\frac{dL}{dQ} = \frac{w}{dQ/dL} = \frac{w}{MP_L}.$$

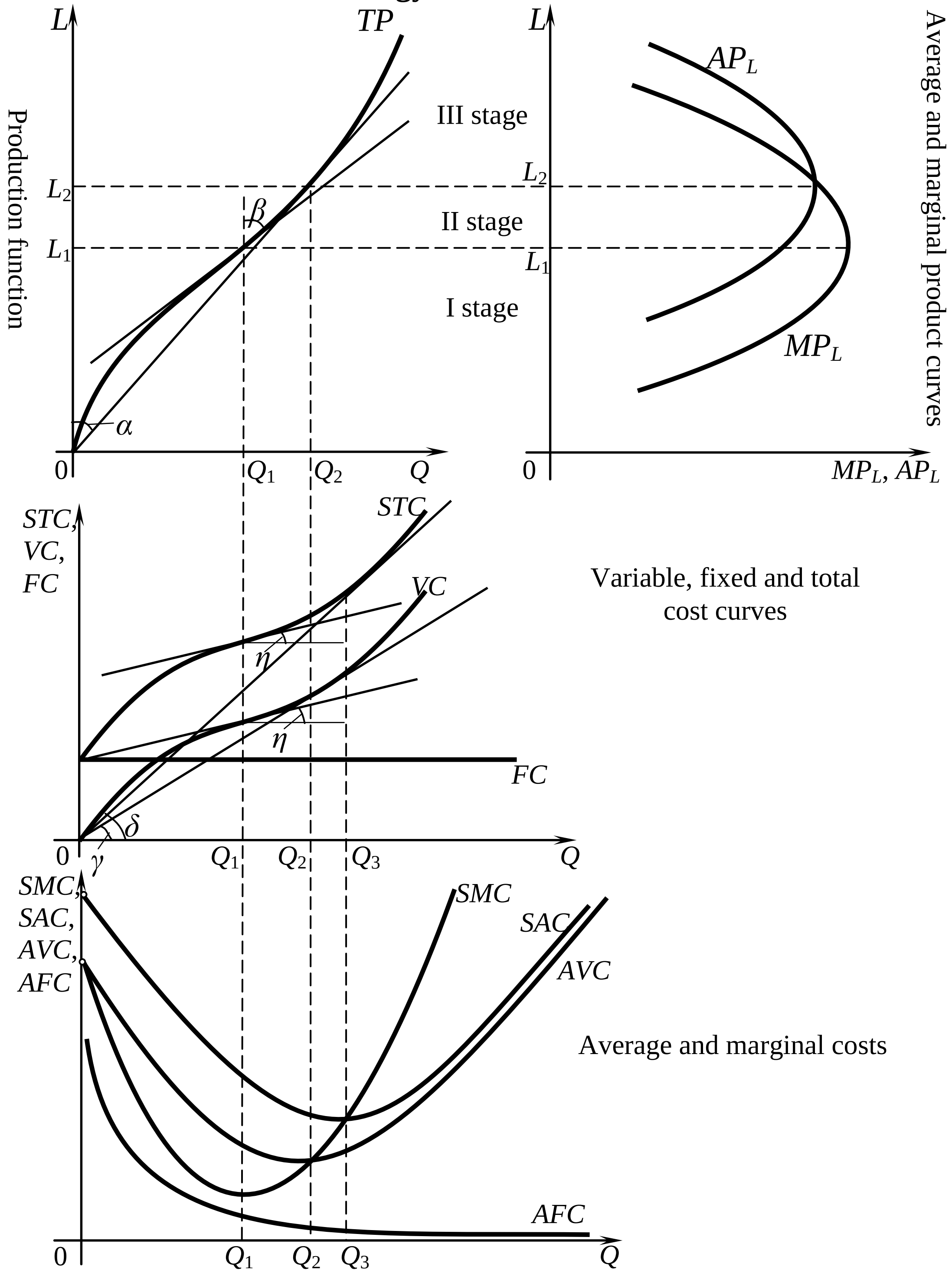
Consequently, with an increase in marginal product of labour marginal costs will go down. With an increase in average product average variable costs will be reduced. And vice versa: decreasing marginal productivity of labour implies increasing marginal costs (stages II and III), and decreasing average productivity of labour implies increasing average variable costs (stage III). When output is greater than Q_1, due to decreasing returns to a variable factor of production variable as well as total costs will be increasing faster than output, and TC and VC curves become steeper.

So marginal and average variable cost curves are first decreasing and then increasing. They are U-shaped opposite to MP_L and AP_L curves correspondingly. AVC curve is at the minimum when AP_L is at the maximum; and MC is at the minimum when MP_L is at the maximum. At the figure below $AP_L = tg\alpha$, $MP_L = tg\beta$, $AVC = tg\gamma$, $AC = tg\delta$, $MC = tg\eta$. Here the Ist stage of production corresponds to increasing marginal returns to a variable production factor (L), i.e. increasing marginal product of labour. The other stages (II to IV) correspond to decreasing marginal productivity of the factor.

L'Hopitale rule of exposing uncertainties of the kind 0/0 or ∞/∞ says that the limit of ratio of functions is equal to the limit of their derivatives. Using L'Hopitale rule we can see that marginal cost is equal to average variable cost for the first, extremely small unit of output:

$$\lim_{Q\to 0}\frac{VC(Q)}{Q} = \lim_{Q\to 0}\frac{\frac{dVC(Q)}{dQ}}{\frac{dQ}{dQ}} = \lim_{Q\to 0} MC(Q).$$

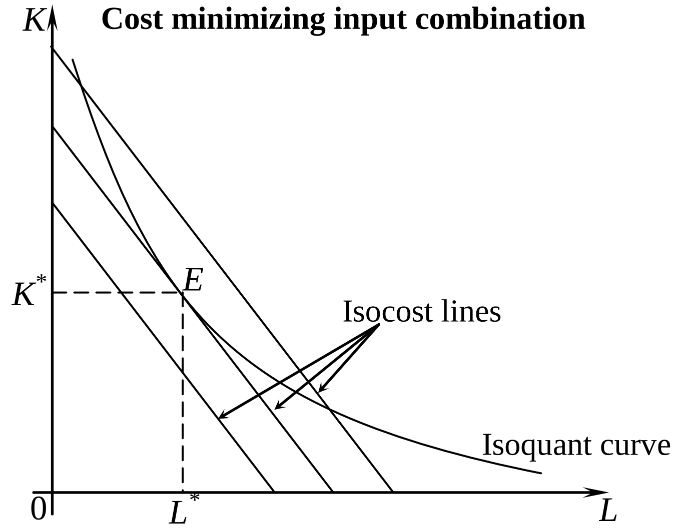
Production technology and cost curves in short run



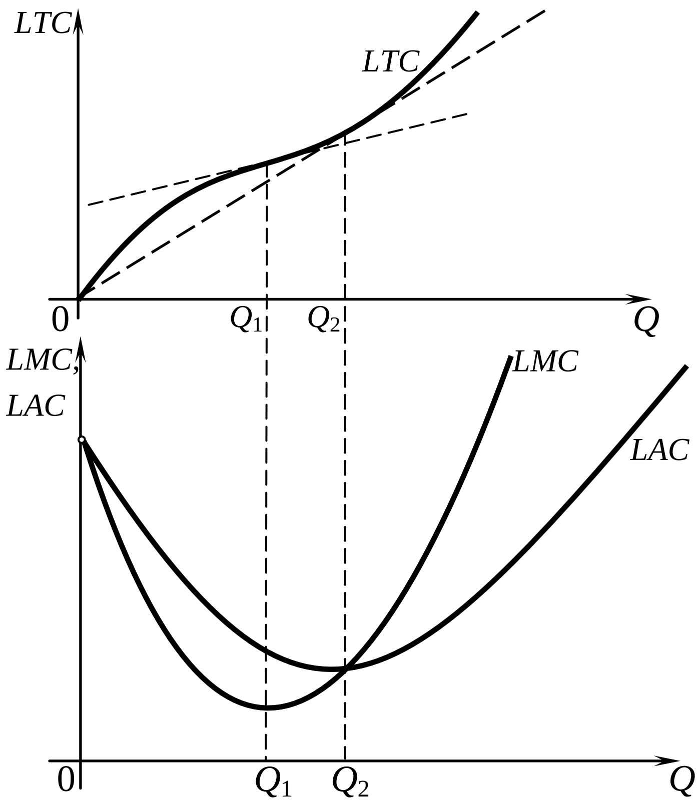
In the long run a firm is able to change its production capacities, and all factors are variable. In the long run a firm can produce a particular output using cost-minimizing input combination. To minimize production costs means to move along the set of isocost lines towards the origin up to the lowest isocost line which has at least one common point with the given isoquant line that corresponds to the particular output of production. So the cost-minimizing input combination is found at the tangency point of the given isoquant curve and the lowest possible isocost line (see the figure below).

The main factors that affect cost-minimizing input combination are prices for inputs and technology of production. Technology influences the shape and location of an isoquant curve and prices for inputs determine the slope of isocost lines.

Cost minimizing input combination



The correspondence between long run average and marginal costs is similar to that in the short run: long run marginal cost curve intersects lonf run average cost curve at the minimum of the latter.



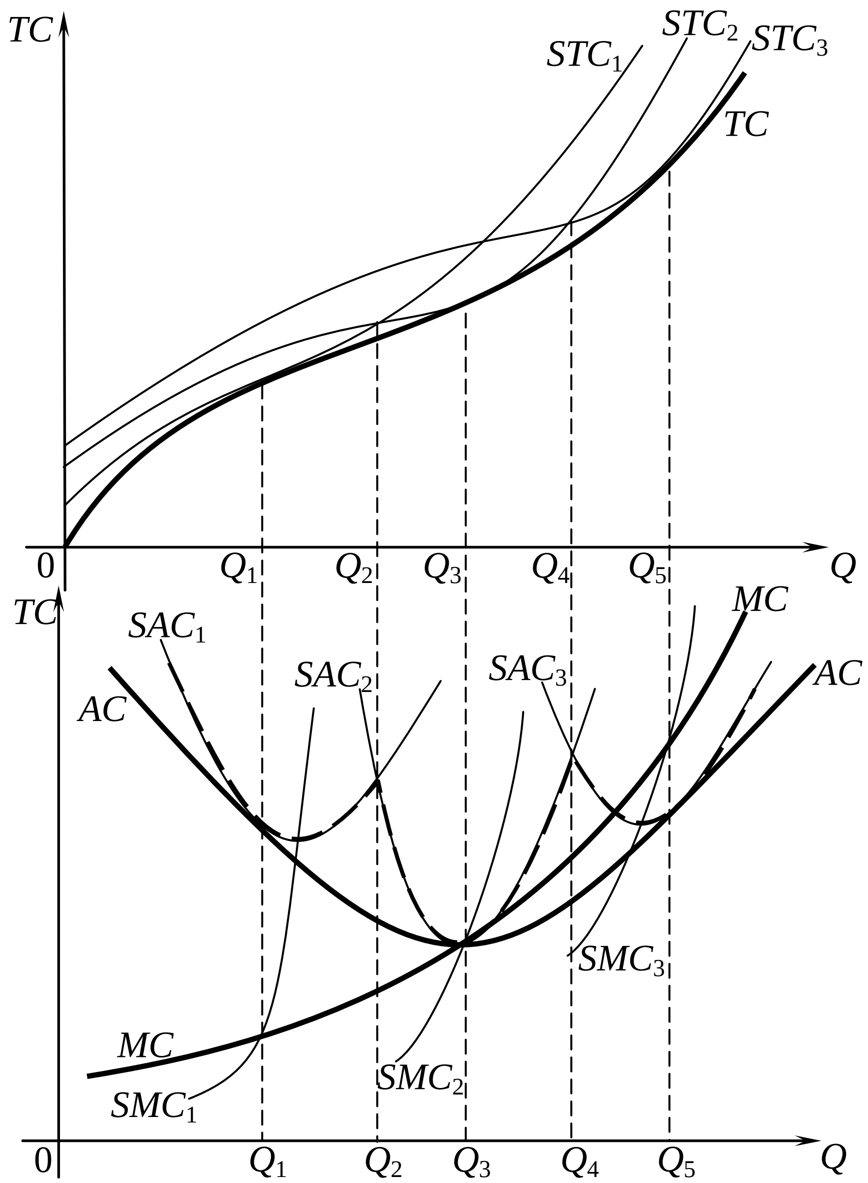
Long run total cost curve

Average and marginal cost curves

Recall that *AC* curve shows the minimum average cost to produce a given output when all factors can be varied. Different short-run cost curves correspond to different plant sizes.

Short run total and average costs exceed the corresponding long run costs because capital input cannot be varied and as usual is not optimal in the short run. This is true except the sole point of tangency of long run and short run curves, when capital input in short run coinsides with that in the long run. So long run total (*TC*) and average (*AC*) curves are envelopes from the bottom of the corresponding short run curves (*STC* and *SAC*).

Long run and short run cost curves



Production technology determines the shape of *AC* curve. Economies of scale deal with relationship between output and *LAC*. If the factor prices are constant the equiproportional increase in quantity of factors applied will result in the equiproportional growth of production cost. In case of increasing returns to scale, or economies of scale, the equiproportional increase in quantity of factors applied yields increase in output in greater proportion. So output grows faster than total cost, and *AC* curve is downward sloping. The opposite behavior of *AC* is observed in case of decreasing returns to scale, when the equiproportional increase in quantity of factors applied and production cost yields smaller increase in output. So diseconomies of scale correspond to upward sloping *LAC* curve. If there are constant returns to scale, *AC* is constant. Thus, output up to Q_3 at the figure above corresponds to positive economies of scale, and the output above Q_3 – to diseconomies of scale. In case of constant returns to scale there would have been a horizontal segment of *AC* curve.

5.3. Total, average and marginal revenue. Accounting versus economic profits. Profit maximization

Producer (a firm) cares about costs which depend on technology, inputs used, proportions and prices of inputs; and revenues which depend on demand.

A firm's revenue is the amount it earns by selling goods or services. Total revenue is the product of the quantity of output (Q) and the price of the good: $TR=PQ$. Average revenue is the total revenue with respect to the volume of output. Average, or per-unit, revenue is the price of the good produced by the firm: $AR = \frac{TR}{Q} = P$.

Marginal revenue is an extra revenue from selling last unit of output. Marginal revenue is the change in total revenue with respect to an infinitesimal change in output: $MR = \frac{dTR}{dQ}$. *MR* under imperfect competition declines with an increase in output, because a firm can sell more output only if the price is lower. To sell an additional unit of output means to loose revenue from the previous units, selling them at lower price (see the figure below).

A business enterprise is supposed to seek maximum of profits. Profit is the objective of every business enterprise. Profit (PR) is the difference between revenues and production costs: $PR(Q) = TR(Q) – TC(Q)$.

Similar to production costs, one can distinguish accounting and economic profits (see fig. below). Accounting profit is the difference between total revenue and accounting costs. Economic profit is the difference between total revenue and economic costs.

Economic and accounting profit and costs

<table>
<tr><td rowspan="3">Total revenue (TR)</td><td colspan="2">Economic profit (PR_{EC})</td><td rowspan="2">Accounting profit (PR_{AC})</td></tr>
<tr><td rowspan="2">Economic costs (TC_{EC})</td><td>Implicit costs (TC_{IMP})</td></tr>
<tr><td>Explicit costs (TC_{EXP})</td><td>Accounting costs (TC_{AC})</td></tr>
</table>

Thus,

$$TR - TC_{EC} = TR - (TC_{IMP} + TC_{EXP}) = PR_{EC},$$

$$TR - TC_{AC} = TR - TC_{EXP} = PR_{AC},$$

i.e. implicit costs are the difference between accounting and economic profit:

$$PR_{AC} - PR_{EC} = TC_{IMP}.$$

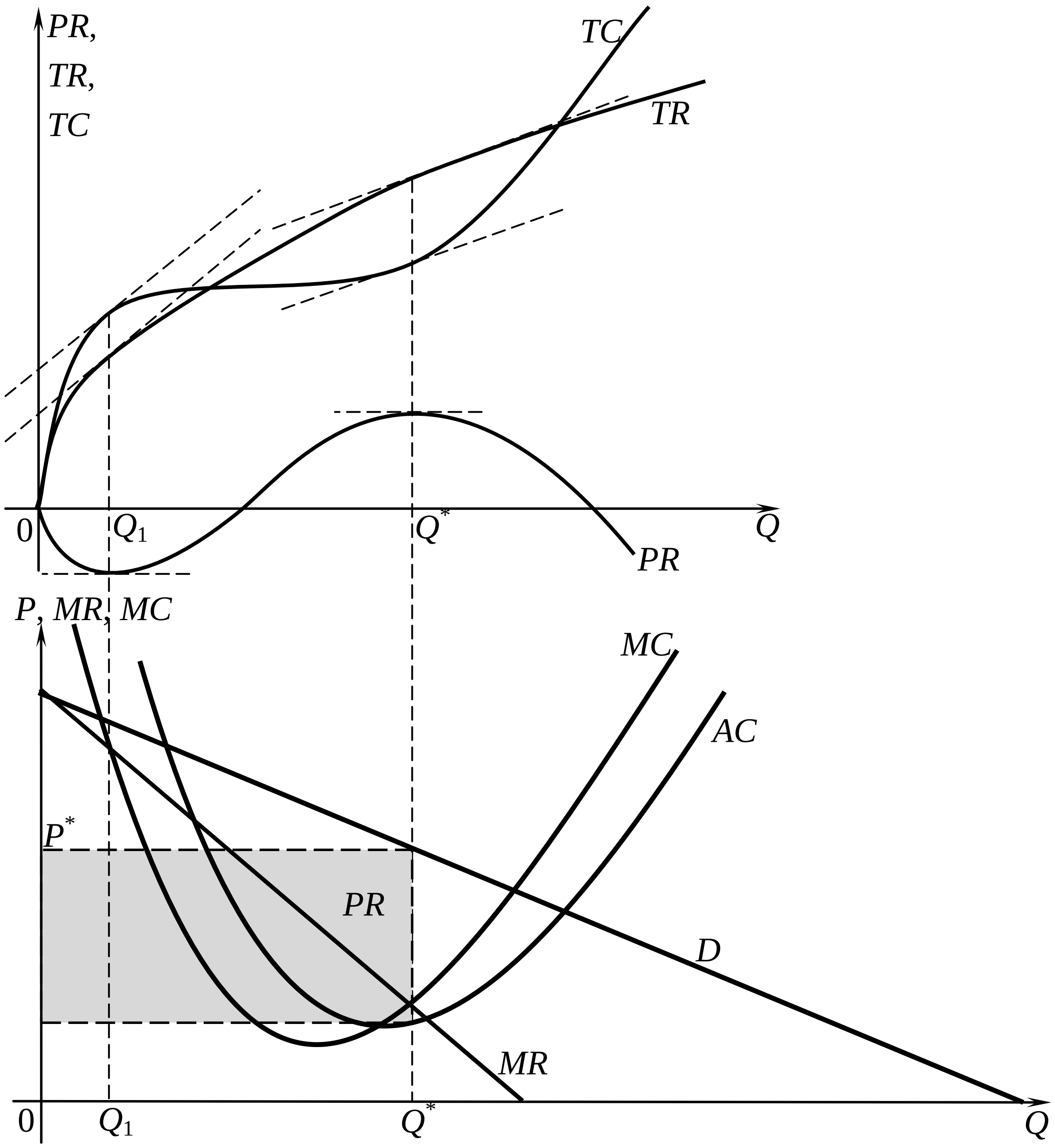
From now on we are going to take into consideration economic profit and costs.

One can calculate profit using average revenue and costs:

$$PR = Q(AR - AC) = Q(P - AC).$$

It is represented by a rectangle with the output as the horizontal side and the difference between market price and average cost as the vertical side (shaded area at the figure below).

The profit maximizing rule $\frac{dPR}{dQ} = 0$ implies that marginal revenue is equal to marginal cost at the profit maximizing output: $MR=MC$. If the equation does not hold, for instance, if $MR>MC$, an additional unit of output yields a gain in profit for the firm. So it pays for the firm to increase output until $MR=MC$. And vice versa, if $MR<MC$, an additional unit of output incures losses for the firm, and consequently it pays to decrease output until $MR=MC$.



Profit maximization by a firm (under imperfect competition)

Maximizing profits a firm is searching for the output which corresponds to the greatest positive distance between TR and TC curves. First order condition says that at this output the tangent line to PR curve should be horizontal, and TC and TC curves should have equal slopes (see the upper graph of the figure above). This corresponds to intersection of MR and MC curves on the lower graph at the figure above.

It should be noted that there are two levels of output (Q_1 and Q^*) which satisfy the first order condition: MR=MC. But the smaller output Q_1 yields maximum losses, and Q^* is the sole profit-maximizing output. To prove it consider the second order condition of profit maximization:

$\frac{d^2PR}{dQ^2} = \frac{d(P-MC)}{dQ} = -MC' \leqslant 0$, i.e. $MC' \geqslant 0$. It means that the profit maximizing output corresponds to increasing (at least non-decreasing) segment of MC curve (see the figure above).

5.4. Producer's surplus in short run

Producer surplus in short run is equal to a sum of profit and fixed cost. It can be measured as a difference between total revenue (S_{0PAQ}) and variable cost (S_{0CBQ}), or the difference between the market price and average variable cost (AB), multiplied by the corresponding optimal output ($0Q=CB$): $PS=PR+FC=TR-VC=Q(P-AVC)$, i.e. S_{CPAB} (see the figure at the left below).

Variable cost can be measured as an area below marginal cost curve (S_{0PDAQ}):

$$\int_0^Q MC(x)dx = \int_0^Q \frac{dVC(x)}{dx}dx = VC(Q) - VC(0) = VC(Q).$$

Thus the producer surplus can be calculated as the difference between total revenue (S_{0PAQ}) and variable cost (S_{0PDAQ}), i.e. as the area above the marginal cost curve bounded from above by the price level (S_{PAD}) (see the figure at the middle below).

Finally, taking into consideration that marginal cost is equal to average variable cost for the first, extremely small unit of output, the two methods of calculating producer surplus can be combined. It can be viewed as sum of rectangle S_{FPGE} and area S_{AEG}. That is, producer surplus is the area to the left of the segment of MC curve above minimum of AVC bounded from above by the market price level: S_{FPAE} (see the figure at the right below).

Producer's surplus in short run