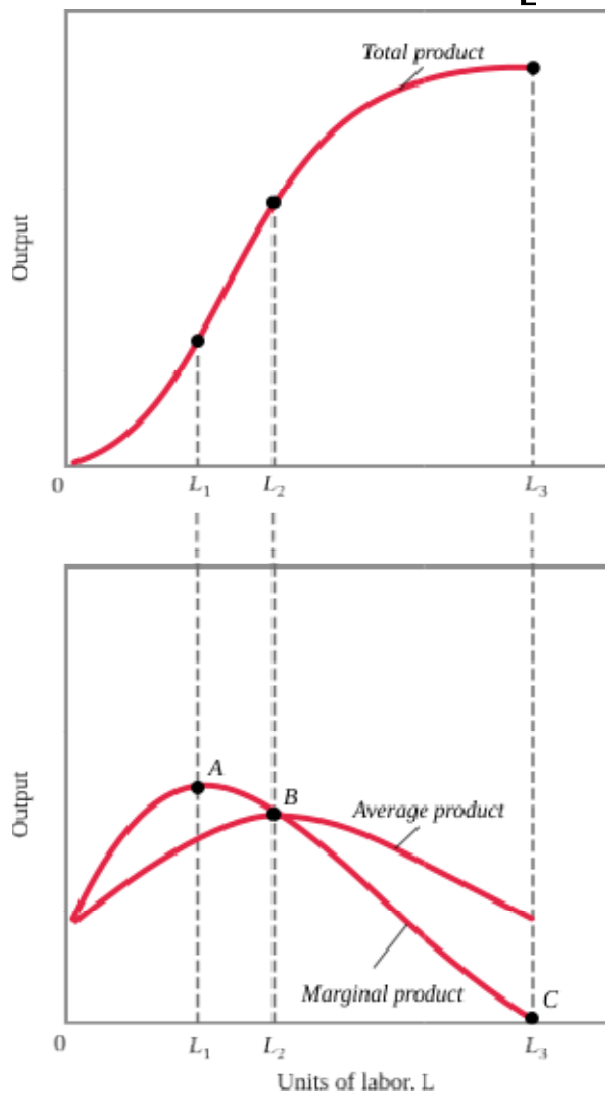


Production function[UNIT-II]



Graph of total, average, and marginal

product

In [economics](#), a **production function** gives the technological relation between quantities of physical inputs and quantities of output of goods. The production function is one of the key concepts of [mainstream neoclassical](#) theories, used to define [marginal product](#) and to distinguish [allocative efficiency](#), a key focus of economics. One important purpose of the production function is to address allocative efficiency in the use of factor inputs in production and the resulting distribution of income to those factors, while abstracting away from the technological problems of achieving technical efficiency, as an engineer or professional manager might understand it.

For modelling the case of many outputs and many inputs, researchers often use the so-called Shephard's distance functions or, alternatively, directional distance functions, which are generalizations of the simple production function in economics. ^[1]

In [macroeconomics](#), aggregate production functions are [estimated](#) to create a framework in which to distinguish how much of [economic growth](#) to attribute to changes in factor allocation (e.g. the accumulation of [physical capital](#)) and how much to attribute to advancing [technology](#). Some [non-mainstream economists](#), however, reject the very concept of an aggregate production function.^{[2][3]}

The theory of production functions

In general, economic output is *not* a (mathematical) function of input, because any given set of inputs can be used to produce a range of outputs. To satisfy the mathematical definition of a [function](#), a production function is customarily assumed to specify the *maximum* output obtainable from a given set of inputs. The production function, therefore, describes a boundary or frontier representing the limit of output obtainable from each feasible combination of input. Alternatively, a production function can be defined as the specification of the minimum input requirements needed to produce designated quantities of output. Assuming that maximum output is obtained from given inputs allows economists to abstract away from technological and managerial problems associated with realizing such a technical maximum, and to focus exclusively on the problem of [allocative efficiency](#), associated with the *economic* choice of how much of a factor input to use, or the degree to which one factor may be substituted for another. In the production function itself, the relationship of output to inputs is non-monetary; that is, a production function relates physical inputs to physical outputs, and prices and costs are not reflected in the function.

In the decision frame of a firm making economic choices regarding production—how much of each factor input to use to produce how much output—and facing market prices for output and inputs, the production function represents the possibilities afforded by an exogenous technology. Under certain assumptions, the production function can be used to derive a [marginal product](#) for each factor. The profit-maximizing firm in perfect competition (taking output and input prices as given) will choose to add input right up to the point where the marginal cost of additional input matches the marginal product in additional output. This implies an ideal division of the income generated from output into an income due to each input factor of production, equal to the marginal product of each input.

The inputs to the production function are commonly termed [factors of production](#) and may represent primary factors, which are stocks. Classically, the primary factors of production were land, labour and capital. Primary factors do not become part of the output product, nor are the primary factors, themselves, transformed in the production process. The production function, as a theoretical construct, may be abstracting away from the secondary factors and intermediate products consumed in a production process. The production function is not a full model of the production process: it deliberately abstracts from inherent aspects of physical production processes that some would argue are essential, including error, entropy or waste, and the consumption of energy or the co-production of pollution. Moreover, production functions do not ordinarily model the [business processes](#), either, ignoring the role of strategic and

operational business management. (For a primer on the fundamental elements of microeconomic production theory, see [production theory basics](#)).

The production function is central to the marginalist focus of neoclassical economics, its definition of efficiency as allocative efficiency, its analysis of how market prices can govern the achievement of allocative efficiency in a decentralized economy, and an analysis of the distribution of income, which attributes factor income to the marginal product of factor input.

Specifying the production function

A production function can be expressed in a functional form as the right side of

where Y is the quantity of output and X_1, X_2, \dots, X_n are the quantities of factor inputs (such as capital, labour, land or raw materials). For $Y > 0$ it must be $X_i > 0$ since we cannot produce anything without inputs.

If Y is a scalar, then this form does not encompass joint production, which is a production process that has multiple co-products. On the other hand, if Y maps from X to Y then it is a joint production function expressing the determination of m different types of output based on the joint usage of the specified quantities of the n inputs.

One formulation is as a linear function:

where a_0, a_1, \dots, a_n are parameters that are determined empirically. Linear functions imply that inputs are perfect substitutes in production. Another is as a [Cobb–Douglas](#) production function:

where A is the so-called [total factor productivity](#). The [Leontief production function](#) applies to situations in which inputs must be used in fixed proportions; starting from those proportions, if usage of one input is

increased without another being increased, the output will not change. This production function is given by

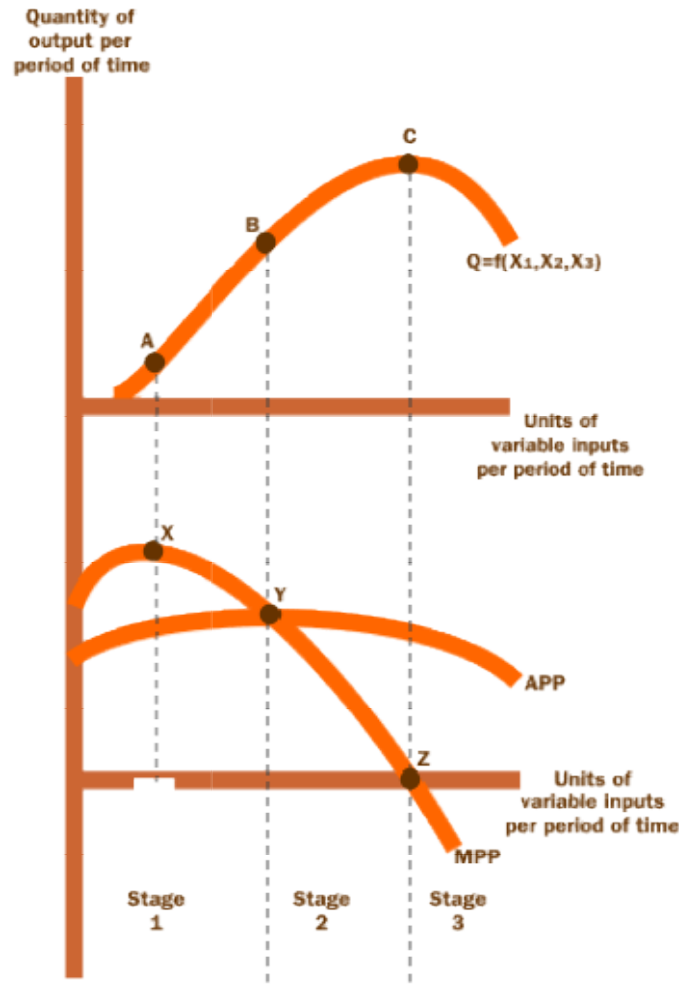
Other forms include the [constant elasticity of substitution](#) production function (CES), which is a generalized form of the Cobb–Douglas function, and the quadratic production function. The best form of the

equation to use and the values of the parameters () vary from company to company and industry to industry. In the short run,

production function at least one of the 's (inputs) is fixed. In the long run, all factor inputs are variable at the discretion of management.

Moysan and Senouci (2016) provide an analytical formula for all 2-input, neoclassical production functions.^[4]

Production function as a graph



Quadratic production

function

Any of these equations can be plotted on a graph. A typical (quadratic) production function is shown in the following diagram under the assumption of a single variable input (or fixed ratios of inputs so they can be treated as a single variable). All points above the production function are unobtainable with current technology, all points below are technically feasible, and all points on the function show the maximum quantity of output obtainable at the specified level of usage of the input. From point A to point C, the firm is experiencing positive but decreasing marginal returns to the variable input. As additional units of the input are employed, output increases but at a decreasing rate. Point B is the point beyond which there are diminishing average returns, as shown by the declining slope of the average physical product curve (APP) beyond point Y. Point B is just tangent to the steepest ray from the origin hence the average physical product is at a maximum. Beyond point B, mathematical necessity requires that the marginal curve must be below the average curve (See [production theory basics](#) for further explanation and Sickles

and Zelenyuk (2019) for more extensive discussions of various production functions, their generalizations and estimations).

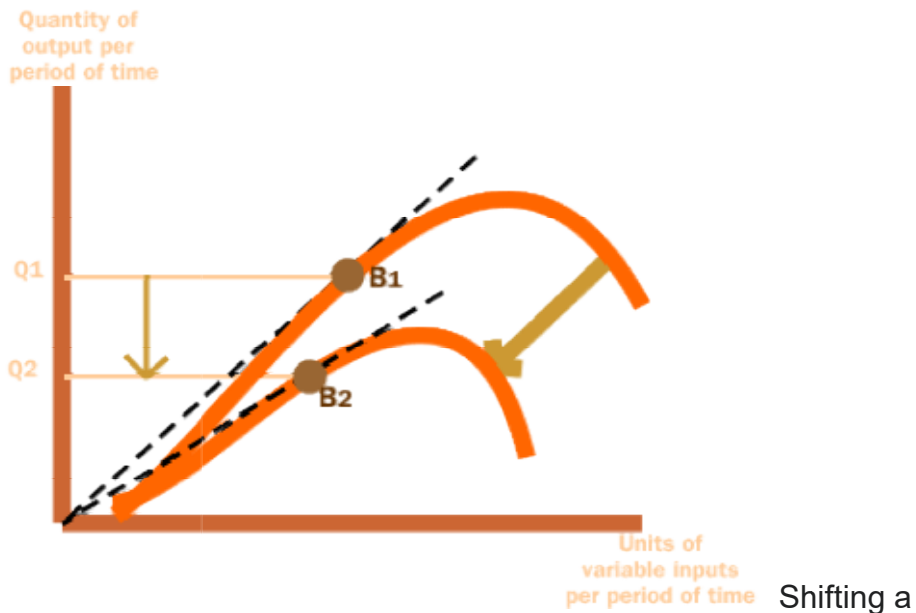
Stages of production

To simplify the interpretation of a production function, it is common to divide its range into 3 stages. In Stage 1 (from the origin to point B) the variable input is being used with increasing output per unit, the latter reaching a maximum at point B (since the average physical product is at its maximum at that point). Because the output per unit of the variable input is improving throughout stage 1, a price-taking firm will always operate beyond this stage.

In Stage 2, output increases at a decreasing rate, and the average and [marginal physical product](#) both decline. However, the average product of fixed inputs (not shown) is still rising, because output is rising while fixed input usage is constant. In this stage, the employment of additional variable inputs increases the output per unit of fixed input but decreases the output per unit of the variable input. The optimum input/output combination for the price-taking firm will be in stage 2, although a firm facing a downward-sloped demand curve might find it most profitable to operate in Stage 2. In Stage 3, too much variable input is being used relative to the available fixed inputs: variable inputs are over-utilized in the sense that their presence on the margin obstructs the production process rather than enhancing it. The output per unit of both the fixed and the variable input declines throughout this stage. At the boundary between stage 2 and stage 3, the highest possible output is being obtained from the fixed input.

Shifting a production function

By definition, in the long run the firm can change its scale of operations by adjusting the level of inputs that are fixed in the short run, thereby shifting the production function upward as plotted against the variable input. If fixed inputs are lumpy, adjustments to the scale of operations may be more significant than what is required to merely balance production capacity with demand. For example, you may only need to increase production by million units per year to keep up with demand, but the production equipment upgrades that are available may involve increasing productive capacity by 2 million units per year.



production function

If a firm is operating at a profit-maximizing level in stage one, it might, in the long run, choose to reduce its scale of operations (by selling capital equipment). By reducing the amount of fixed capital inputs, the production function will shift down. The beginning of stage 2 shifts from B1 to B2. The (unchanged) profit-maximizing output level will now be in stage 2.

Homogeneous and homothetic production functions

There are two special classes of production functions that are often

analyzed. The production function is said to be homogeneous of

degree α , if given any positive constant k , $f(kx_1, kx_2, \dots, kx_n) = k^\alpha f(x_1, x_2, \dots, x_n)$. If $\alpha > 1$, the function exhibits increasing returns to scale, and it

exhibits decreasing returns to scale if $\alpha < 1$. If it is homogeneous of

degree $\alpha = 1$, it exhibits **constant** returns to scale. The presence of increasing returns means that a one percent increase in the usage levels of all inputs would result in a greater than one percent increase in output; the presence of decreasing returns means that it would result in a less than one percent increase in output. Constant returns to scale is the in-between case. In the Cobb–Douglas production function referred to

above, returns to scale are increasing if $\alpha > 1$, decreasing if $\alpha < 1$, and

constant if $\alpha = 1$.

If a production function is homogeneous of degree one, it is sometimes called "linearly homogeneous". A linearly homogeneous production function with inputs capital and labour has the properties that the marginal and average physical products of both capital and labour can be expressed as functions of the capital-labour ratio alone. Moreover, in this case, if each input is paid at a rate equal to its marginal product, the firm's revenues will be exactly exhausted and there will be no excess economic profit.^[5] pp.412-414

Homothetic functions are functions whose marginal technical rate of substitution (the slope of the [isoquant](#), a curve drawn through the set of points in say labour-capital space at which the same quantity of output is produced for varying combinations of the inputs) is homogeneous of degree zero. Due to this, along rays coming from the origin, the slopes of

the isoquants will be the same. Homothetic functions are of the form

where f is a monotonically increasing function (the derivative of

is positive ($f' > 0$)), and the function g is a homogeneous function of any degree.

Aggregate production functions

[

In [macroeconomics](#), aggregate production functions for whole nations are sometimes constructed. In theory, they are the summation of all the production functions of individual producers; however there are methodological problems associated with aggregate production functions, and economists have debated extensively whether the concept is valid.^[3]

Criticisms of the production function theory

There are two major criticisms^[which?] of the standard form of the production function.^[6]

On the concept of capital

During the 1950s, '60s, and '70s there was a lively debate about the theoretical soundness of production functions (see the [Capital controversy](#)). Although the criticism was directed primarily at aggregate production functions, microeconomic production functions were also put under scrutiny. The debate began in 1953 when [Joan Robinson](#) criticized the way the factor input [capital](#) was measured and how the notion of factor proportions had distracted economists. She wrote:

"The production function has been a powerful instrument of miseducation.

The student of economic theory is taught to write $Q = f(K, L)$ where Q is a

quantity of labor, a quantity of capital and a rate of output of commodities. [They] are instructed to assume all workers alike, and to

measure in man-hours of labor; [they] are told something about the index-number problem in choosing a unit of output; and then [they] are hurried on to the next question, in the hope that [they] will forget to ask in what units K is measured. Before [they] ever do ask, [they] have become a professor, and so sloppy habits of thought are handed on from one generation to the next".^[7]

According to the argument, it is impossible to conceive of capital in such a way that its quantity is independent of the rates of [interest](#) and [wages](#). The problem is that this independence is a precondition of constructing an isoquant. Further, the slope of the isoquant helps determine relative factor prices, but the curve cannot be constructed (and its slope measured) unless the prices are known beforehand.

On the empirical relevance

As a result of the criticism on their weak theoretical grounds, it has been claimed that empirical results firmly support the use of neoclassical *well behaved* aggregate production functions. Nevertheless, [Anwar Shaikh](#) has demonstrated that they also have no empirical relevance, as long as the alleged good fit comes from an accounting identity, not from any underlying laws of production/distribution.^[8]

Natural resources

: [Nicholas Georgescu-Roegen § Criticising neoclassical economics \(weak versus strong sustainability\)](#)

[Natural resources](#) are usually absent in production functions.

When [Robert Solow](#) and [Joseph Stiglitz](#) attempted to develop a more realistic production function by including natural resources, they did it in a manner economist [Nicholas Georgescu-Roegen](#) criticized as a "conjuring trick": Solow and Stiglitz had failed to take into account the [laws of thermodynamics](#), since their variant allowed man-made capital to be a complete substitute for natural resources. Neither Solow nor Stiglitz reacted to Georgescu-Roegen's criticism, despite an invitation to do so in the September 1997 issue of the journal [Ecological Economics](#).^{[2][9]: 127–}

¹³⁶[3][10]

Georgescu-Roegen can be understood as criticizing Solow and Stiglitz's approach to mathematically modelling factors of production. We will use the example of energy to illustrate the strengths and weaknesses of the two approaches in question.

Independent factors of production

[Robert Solow](#) and [Joseph Stiglitz](#) describe an approach to modelling energy as a factor of production which assumes the following:^[11]

- Labor, capital, energy input, and technical change (omitted below for brevity) are the only relevant factors of production,
- The factors of production are independent of one another such that the production function takes the general form $Y = A K^\alpha L^\beta E^\gamma$,

- Labor, capital, and energy input only depend on time such that $L = L(t)$, $K = K(t)$, and $E = E(t)$.

This approach yields an energy-dependent production function given

as $Y = A K^\alpha L^\beta E^\gamma$.^{[11][12]} However, as discussed in more-recent work, this approach does not accurately model the mechanism by which energy affects production processes.^[13] Consider the following cases which support the revision of the assumptions made by this model:

- If workers at any stage of the production process rely on electricity to perform their jobs, a power outage would significantly reduce their maximum output, and a long-enough power outage would reduce

their maximum output to zero. Therefore Y should be modeled as

depending directly on time-dependent energy input $E(t)$.

- If there were a power outage, machines would not be able to run, and therefore their maximum output would be reduced to zero.

Therefore Y should be modeled as depending directly on time-

dependent energy input $E(t)$.

This model has also been shown to predict a 28% decrease in output for a 99% decrease in energy, which further supports the revision of this model's assumptions.^[13] Note that, while inappropriate for energy, an "independent" modelling approach may be appropriate for modelling other natural resources such as land.

Inter-dependent factors of production

The "independent" energy-dependent production function can be revised

by considering energy-dependent labor and capital input functions

$L = L(t, E)$ and $K = K(t, E)$. This approach yields an energy-dependent production function

given generally as $Y = A K^\alpha L^\beta E^\gamma$. Details related to the derivation of a specific functional form of this production function as well as empirical support for

this form of the production function are discussed in more-recently published work.^[13] Note that similar arguments could be used to develop more-realistic production functions which consider other depletable natural resources beyond energy:

- If a geographical region runs out of the natural resources required to produce a given machine or maintain existing machines and is unable to import more or recycle, the machines in that region will eventually fall into disrepair and the machines' maximum output would be reduced to near-zero. This should be modeled as significantly

affecting the total output. Therefore, therefore should be modeled as depending directly on time-dependent natural resource

input .

The practice of production functions

The theory of the production function depicts the relation between physical outputs of a production process and physical inputs, i.e. factors of production. The practical application of production functions is obtained by valuing the physical outputs and inputs by their prices. The economic value of physical outputs minus the economic value of physical inputs is the income generated by the production process. By keeping the prices fixed between two periods under review we get the income change generated by a change of the production function. This is the principle how the production function is made a practical concept, i.e. measurable and understandable in practical situations.