MVŠO

ADVANCED MICROECONOMICS

STUDY MATERIAL FOR BLENDED LEARNING

ADVANCED MICROECONOMICS

Ing. Jaroslav ŠKRABAL, Ph.D.

Ing. Eva JÍLKOVÁ, Ph.D.

Moravská vysoká škola Olomouc, o.p.s., 2018, 2024

Autor: Ing. Jaroslav Škrabal, Ph.D.

Ph.D. Ing. Eva JÍLKOVÁ, Ph.D.

Olomouc 2018, 2024

Table of Contents

Preface

Economics, as a discipline, deals with the complex interactions between individual market actors and issues of allocation of scarce resources. It aims to provide analytical tools to understand the decision-making processes of citizens, firms, and government, and how these decisions affect everyday life and economic outcomes. In this way, economics becomes indispensable for informed decisions at the individual and societal levels. In the context of these ideas, the Advanced Microeconomics course focuses on deepening students' theoretical knowledge and skills and serves as preparation for other specialized disciplines in economics.

One of the key issues addressed in microeconomics is the contradiction between unlimited human needs and limited resources, which is fundamental to understanding the behaviour of individual market actors. This basic concept is reflected in the decision-making process of consumers and firms. In the second chapter, "Consumer Behaviour and Demand Shaping", students are introduced to utility theory and how consumers respond to price changes and create demand for goods and services. This analysis is key to understanding market dynamics and the impact of price changes on aggregate demand.

The third chapter, "Consumer decision-making under conditions of risk", focuses on situations where consumers face uncertainty. Students will learn how people make decisions in risk situations and how they adapt to different levels of risk and uncertainty. This is important for analysing consumer behaviour in the marketplace, where situations can be unpredictable.

The next part of the course focuses on firms. The chapter "Corporate Behaviour, Selection of a Technology, Costs and Revenue of a Firm" examines the factors that affect firms' choice of production technology and cost structure. Students will analyse how firms optimize their production processes to maximize efficiency. This includes examining different types of costs, such as fixed and variable costs, and their impact on production decisions.

In the chapter "Output Choice for a Perfectly Competitive Firm," students will learn how firms maximize their profits under perfect competition. Here they discuss how prices and quantities of output affect firms' profitability and how firms respond to changes in market conditions. This knowledge block is essential for understanding market dynamics and the behaviour of individual firms.

Another important topic is the "Firm Decision-Making as a Monopoly" analysis. In this unit, students learn about the characteristics of a monopoly and the challenges that monopolies face in setting prices and output. Concepts such as price discrimination, optimal output, and the effect of monopoly on overall market efficiency will be explored.

The chapter "Firm Equilibrium in Oligopoly Conditions" focuses on how oligopolistic markets affect the behaviour of firms and what strategies these firms choose. Students will learn to recognize different forms of oligopolistic behaviour and their effects on market dynamics. Oligopoly is an important aspect of the market because it has a significant impact on price stability and overall competition.

An important part of the study is also "Price Formation in Factor Markets", where students learn how prices are equalized in labour and capital markets. This chapter includes an analysis of wages, interest rates, and rents and their impact on the decisions of firms and consumers.

The following chapter, "Capital Market", deals with the specifics of the capital market and its role in the economy. Students will explore how investment financing works and how the capital market affects the allocation of resources in the economy. This is key to understanding the broader economic context and the importance of investment to economic growth.

In "Investment Decisions Making in Conditions of Risk and Uncertainty," students will learn how firms and individuals face risk when making investment decisions. This analysis is important for understanding how risk affects long-term investment strategies and how these strategies are shaped by expected returns and costs.

The next chapter, "General Equilibrium and Efficiency", focuses on how different markets are interrelated and how overall economic equilibrium is achieved. This analysis provides students with insight into how different markets interact and how changes in one segment of the economy can have far-reaching effects on other segments.

The chapter "Market Failure and the Macroeconomic Policy of the State" examines situations where markets are not functioning optimally and explores possible government interventions that can improve economic efficiency. Students will learn to recognize different forms of market failures, such as externalities, public goods, and information asymmetries, and discuss ways in which the state can intervene in market mechanisms to contribute to a more efficient economy.

The study of Advanced Microeconomics thus provides a thorough foundation for the analysis of various economic phenomena and problems. It requires the ability to apply theoretical concepts to practical situations and develops students' critical thinking skills. This course is a key step for all those who wish to specialise in economic analysis and contribute to solving current economic challenges.

Chapter 1

Introduction to the Study of Advanced Microeconomics

After studying the chapter you will be able to:

- define economics as a science;
- characterise the concept of rarity;
- use mathematical tools in economics.

Keywords:

economics, rarity, tools used in economics - graphs, functions, directives, slope of the curve, derivatives.

After studying the chapter you will be able to

- Define and analyse the optimal strategy of the company in different market structures.
- Explain how the factors of demand, supply, elasticity and marginal values influence the firm's decision making.
- Apply mathematical tools to optimize production and cost functions.
- Analyse the decision-making processes of firms in pricing strategy and quantity decisions.
- Identify and describe the basic principles of monopolistic, oligopolistic, and competitive behaviour.

Chapter preview

- This chapter deals with the importance of strategic analysis in the strategic management process. It introduces three basic questions of the planning process that are key to the formulation of strategy: "Where am I now?", "Where do I want to go?" and "How do I get there?". It focuses on the analysis of the internal and external environment of the organization, the understanding of which is essential to identify strengths and weaknesses as well as opportunities and threats that can affect the organization.

Chapter objectives

- To deepen knowledge of optimal decision making of firms in different market conditions.
- To enable students to apply mathematical and economic tools to analyse firm strategies.
- Understand the differences between different types of market structures and their impact on the firm.
- To provide tools for analysing pricing and strategic decisions of firms in monopolistic and oligopolistic markets.

Estimated study time

- The estimated time required to study the chapter is 300 minutes.

1.1 Graphs and Mathematical Formulas Used in Economics

„A picture is worth a thousand words."

Chinese proverb

This chapter provides a brief overview of some of the mathematical concepts that are used in a basic economics course. The material presented should serve as a sort of reminder of the definitions of various concepts used in the texts of a distance learning economics course. In no way should it be a mathematics tutorial. The definitions given will generally be the simplest forms, not the most rigorous.

1.1.1 Mathematical Tools in Economics

1.1.1.1 Function

A function is a certain rule that captures the relationship between two quantities. For each value **x**, the function assigns a ***single*** value **y** depending on the rule. Then a function can be described as something that is determined by a rule like "take a number and multiply it by two" or "take a number and multiply it by two" etc. Specifically, we would write these functions as
$\mathbf{y = x^2}$ and $\mathbf{y = 2x}$. Functions are also referred to in some cases as ***transformations.***

We often wish to imply that some variable **y** depends on some other variable **x**, but we are unable to specify the algebraic relationship between the two variables. In such a case, we write $\mathbf{y = f(x)}$, which means that the variable **y** depends on the variable **x** according to the rule **f**.

Given a function $\mathbf{y = f(x)}$, the variable is often called the ***independent variable***, while the variable **y** is called the ***dependent variable***. The idea behind this naming is that **x** varies independently while the value of **y** depends on the value of **x**.

The variable y often depends on several different variables $\mathbf{x_1, x_2}$, etc., and this can be expressed as $\mathbf{y = f(x_1, x_2)}$. This indicates that the two variables jointly determine the value of **y**.

1.1.1.2 Graphs

Thousands of words, lots of numbers and tables, mathematical formulas and illustrative examples, but also graphical representations can be used to explain the economic context. The graph of a function is a graphical representation of the behaviour of a function. Graphs are an extremely important tool in modern economics. They allow a quick visual presentation of data or a relationship between two variables.

Figure 1 shows two graphs of functions. In mathematics, the independent variable is usually plotted on the horizontal axis, while the dependent variable is usually plotted on the vertical axis. The graph then indicates the relationship between the independent and dependent variables.

In economics, however, it is common to graph a function with the independent variable on the vertical axis and the dependent variable on the horizontal axis. For example, the demand function is usually expressed with price on the vertical axis and quantity demanded on the horizontal axis.

Definitions:

A graphical representation of the behaviour of a function that allows a quick visual presentation of data or a relationship between two variables. In economics, a graph is an important tool for explaining economic relationships, often showing the independent variable (e.g. price) on the vertical axis and the dependent variable (e.g. quantity demanded) on the horizontal axis.

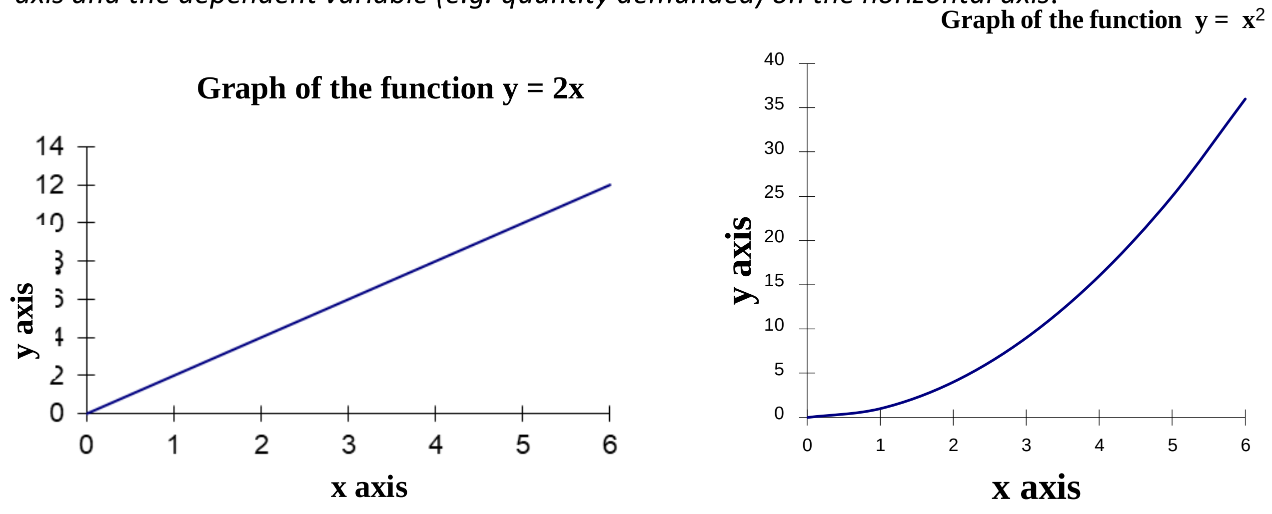

Fig. 1 Graphs of functions

1.1.1.3 Equations and Identities

We speak of an equation when a function is equal to a number. Examples of equations can be e.g. $\mathbf{2x = 8}$ or $\mathbf{x^2 = 9}$ or $\mathbf{f(x) = 0}$.

The solution to the equation is the value of x that satisfies this equality. The solution to the first equation is **x = 4**. The second equation has two solutions **x = 3** and **x = -3**. The third equation is a general equation. We won't know its solution until we know what rule

f represents. However, we can refer to the solution of this equation as x*. This simply means that x* is a number such that ***f*** **(x*) = 0**. Then we say that x* satisfies the equation ***f*** **(x) = 0.**

An identity is a relationship between variables that holds for all values of those variables. Here are some examples of identities:

$$(x + y)^2 = x^2 + 2xy + y^2$$

$$2(x + 1) = 2x + 2.$$

The special symbol = means that the left side and right side are equal for ***all*** variable values. Equality is valid only for some values of the variables, while identity is valid for all values of the variables. Identity often depends on the definition of the terms under consideration.

Linear functions

A linear function is a function that has the form **y = ax + b,** where a and b are constants. Examples of linear functions might be **y = 2x + 3** or **y = x - 99**, for example.

Strictly speaking, a function of the form **y = ax + b** should be called a ***related function*** and only the function y = ax should be called a linear function. However, we will not insist on this distinction.

Linear functions can be expressed implicitly in the form ax + by = c. In this case, we often solve for y as a function of x to obtain the "standard" form: **y = c/b - a/b.x.**

1.1.2 Slopes, Directives and Intersections

Most relationships in economics are not linear; they cannot be illustrated by a straight line but by a curve. Therefore, we must distinguish between the directive of a line and the directive of a curve.

The directive of a line is defined as the change of a variable on the vertical axis (y-axis) to the change of a variable on the horizontal axis (x-axis). This means that the line directive = $\Delta y / \Delta x$. The directive of a line can be positive or negative. A positive directive of the line expresses a directly proportional dependence of the variables and a negative one their inverse proportionality, i.e. an increase in x is accompanied by a decrease in y - the variables evolve in opposite directions. The directive of the line parallel to the x-axis is equal to zero, the directive of the line parallel to the y-axis is equal to infinity.

The rate of change of a function can be interpreted graphically as the **slope of** that function. **The slope of a line** is the absolute value of the directive of the line, that is, the absolute value of the change in the y-axis variable relative to the change in the x-axis variable. The slope (of a line or curve) is always expressed as a number. It indicates how much y changes if x changes by 1 unit. The slope of a line has a constant value. The term directive is sometimes confused with steepness. However, steepness is only a matter of the scale of the graph. The slope of a curve expresses whether x and y are directly or inversely proportional to each other. Direct proportionality means that the quantities move in the same direction (or that they both rise or fall together). Inverse proportionality occurs when they both change in opposite directions (or one rises while the other falls). Thus, a negative slope shows that x and y are inversely proportional. A curve is a line whose slope is different at different points. The absolute value of the curve directive gives the slope of the curve. It is necessary to distinguish between the directive slope of a curve at a point and between points. The direction or slope of a curve at a point is determined by drawing a tangent to the curved line at that point. By definition, a tangent to a curve is a line that does not intersect the curve but only touches it, namely at a single point. The slope of this tangent line measures the slope of the curve at that point. To calculate the slope, we use the usual rectangular measurement technique. It is often necessary to express not only the directive or slope of the curve at a point but also its directive or slope between two points. In this case, we draw a line joining these points. The directive of this line is the average directive of the curve between the two points. The slope of the curve between the points is calculated using the orthogonal measurement technique. This slope is called the average slope of the curve between the two points.

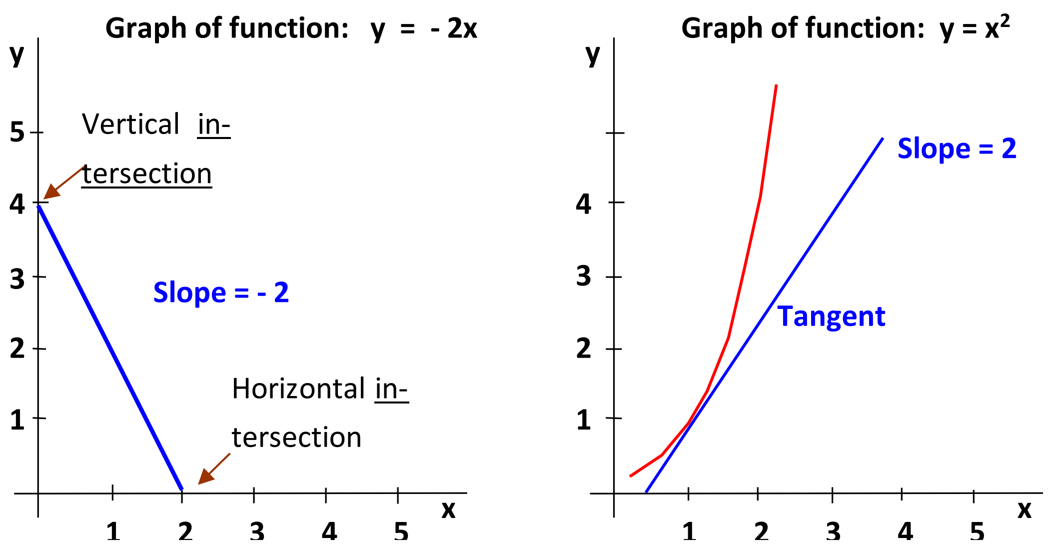

Fig. 2 Slopes and intersections

Figure 2 shows the function $\mathbf{y = -2x + 4}$ and the function $\mathbf{y = x^2}$. In general terms, if the linear function has the form $\mathbf{y = ax + b}$, the vertical intercept will be $\mathbf{y^* = b}$ and the horizontal intercept will be $\mathbf{x^* = -b/a}$. If the linear function is expressed in the form $\mathbf{a_1x_1 + a_2x_2 = c}$, then the horizontal intercept will correspond to the value $\mathbf{x_1}$, for which $\mathbf{x_2 = 0}$, corresponding to $\mathbf{x_1^* = c/a_1}$ and for the vertical intercept it must be the case that $\mathbf{x_1 = 0}$, corresponding to $\mathbf{x_2^* = c/a_2}$. The slope of this function **is** $\mathbf{-a_1/a_2}$.

A nonlinear function has the property that its slope changes as **x** changes. The tangent of a function at a point **x** is a linear function that has the same slope as that function at that point. The figure showing the function $\mathbf{y = x^2}$ shows the function $\mathbf{x^2}$ and its tangent at $\mathbf{x = 1}$.

If for every increase in **x**, **y** also increases, **Δy** will always have the same sign as **Δx**, so the slope of such a function will be increasing. If, on the other hand, **y** decreases when **x** increases or **y** increases when **x** decreases, **Δx** and **Δy** will have opposite signs so that the slope of such a function will be decreasing.

1.1.3 Absolute Values and Logarithms, Derivatives

The absolute value of a number is the function f(x), which is defined as follows:

$$f(x) = \begin{cases} x & if \quad x \geq 0 \\ -x & if \quad x \leq 0 \end{cases}$$

Then the absolute value of that number can be found by removing the sign of that number. The absolute value of a function is usually written as $|x|$.

The (natural) logarithm or **log x** describes a function x, which we write as **y = ln x** or **y = ln(x).** The logarithmic function is the only function for which it is true that

$$\mathbf{\ln(xy) = \ln(x) + \ln(y)}$$

for all positive numbers **x** and **y**

$$\mathbf{\ln(e) = 1}.$$

(In this last equation, **e** represents the base of the natural logarithm, which is equal to 2.7183...). In words, the logarithm of the product of two numbers is equal to the sum of the individual logarithms of those numbers. This property implies another important characteristic of logarithms:

$$\ln (x^{y}) = y \ln (x),$$

which implies that the logarithm of the power x^y is equal to **y-multiple of the logarithm x.**

1.1.3.1 Derivatives

The derivative of the function **y = F (x)** is defined as

$$\frac{df(x)}{dx} = \lim \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

The derivative is the limit of the rate of change of y with respect to x for x approaching zero. Derivation gives the exact meaning of the phrase "the rate of change of y with respect to x for small changes in x". The derivative of f(x) with respect to x is also called f´(x).

Therefore, for this constant linear function y = ax + b:

$$\frac{df(x)}{(dx)} = a.$$

For a nonlinear function, the rate of change of **y** with respect to x will usually depend on **x**. We could see that in the case of $f(x) = x^2$ it was true that $\Delta y / \Delta x = 2x + \Delta x$.

Applying the definition of derivative, we get

$$\frac{df(x)}{dx} = \lim 2x + \Delta x = 2x.$$

So, the derivative of x^2 by **x is 2x**. Using more complicated methods, we can prove that if **y = ln x**, then

$$\frac{df(x)}{dx} = \frac{1}{x}.$$

1.1.4 Relationships between Variables

A variable is an object of interest that can be defined and measured. Important variables studied in economics are prices, quantities, hours worked, acres of land, income in monetary units, number of employees, etc. In mathematics, as a rule, we plot the independent variable on the x-axis and the dependent variable on the y-axis. In economics, the axes are often swapped in this sense, i.e.: we put the dependent variable on the x-axis and the independent variable on the y-axis. A typical example and one of the most common is the supply and demand function.

Most graphical models (graphs) of economic relationships express the relationship between two variables. That is, it expresses only the dependence of one phenomenon on the other and abstracts from all other economic phenomena. When we talk about functional dependence, we say that a functional dependence between two variables exists if one variable is dependent on the other variable, i.e., the value of the dependent variable is determined by the value of the independent variable. For example, the price (independent variable) determines the quantity of the good you buy (dependent variable).

In general, there are three possible types of relationships between variables:

- **direct** - positive relationship, expressing that the growth of one variable is accompanied by the growth of the other variable (e.g. between price and quantity offered = supply function),
- **indirect** - negative relationship, expressing that the growth of one variable is accompanied by a simultaneous decrease in the other variable (e.g. between price and quantity demanded = demand),
- **neutral** - mutual independence of variables, i.e. the growth of one variable does not lead to a change in the other variable.

1.2 The Content of Economic Theory Research

Some economists argue that general economic theory is a social science whose inquiry must be based on the reality of the economic life of society and whose conclusions must be practically applicable to society. Others advocate an understanding of economic theory as more of a formally abstract discipline in which conformity with reality is not a condition of correctness and science. Economic theory always moves in a space defined by certain simplifying assumptions. These assumptions allow theoretical economists to formulate general principles for the functioning of the economic system, which are then modified by many circumstances in the real economic system. The dispute over the social or formal-logical character of economics boils down to the existence of two branches of economic theory:

- ***The mathematical branch*** promotes the view that the criterion of truth and science in economic theory is the possibility of mathematical proof. What cannot be proved mathematically cannot be proved at all and is therefore unusable.
- ***The social branch*** rejects this view on principle. It even refuses to use mathematics in economic theory at all. It bases its argument on the claim that economics is the science of human behaviour in production. And human behaviour cannot be reduced to formulas. If we open any textbook, it is easy to see that in some areas of economic theory, the use of mathematics is very useful, but it cannot cover everything.

Human society has productive resources at its disposal, which it uses in production to transform them into products and services. Since productive resources are scarce, there is a need for efficient management of these resources and a need for economic research on this basic economic problem - the scarcity of scarce resources.

Factors of production are all resources that are either consumed directly or used as inputs into production, where they are processed and used to obtain the final output. Obviously, the quantity of these factors is limited. In contrast, the volume of human needs is unlimited, they are never fully satisfied, and the set of human needs is constantly changing and growing.

The problem of the scarcity of resources and the limitlessness of human needs can be considered as a subject of economic science. Economics examines how economic agents deal with scarce resources, how they use them in production, how final production is divided between present and future consumption or between individuals or groups.

The name economics comes from ancient Greece, from the Greek words oikos (house) and nomos (law). Within general economic theory, there are two related and inseparable fields - microeconomics and macroeconomics.

Microeconomics focuses on the analysis of the behaviour of individual elements of the economy (individuals, firms, households), it deals with e.g. the determination of the price of a product, consumer behaviour, firm behaviour, etc. and **macroeconomics** examines the factors that determine the level and interrelationships in the development of aggregate variables (GDP, NI, unemployment, inflation. Microeconomics explains how and why markets work; macroeconomics tries to answer the question of why markets do not work satisfactorily.

We consider the two fields to be distinct levels of inquiry that can be studied separately, but there is an important interrelationship between them that cannot be ignored. Macroeconomics forms the framework for the operation of microeconomic phenomena and processes, and the study of microeconomics forms the basis for the study of macroeconomics.

The above-mentioned division refers to economic theory, which, moreover, must be distinguished from economic practice, i.e. economics from economics. The confusion between the two terms is tolerated in general parlance but should not be used by students and practitioners of economics. While economics is a social science that studies economic phenomena, economics represents the field of economic practice or economic reality. On a microeconomic scale, it is the corporate economy, while on a macroeconomic scale it corresponds to the national economy.

Economic reality is usually viewed through two basic approaches. These are:

- **positive approach** - enables the examination of reality, description of a given state, helps to understand economic relations, reveals the laws of the functioning of the economy,
- **normative approach** - observes the economic phenomenon and its development, but in addition tries to influence it, includes ethics and evaluative judgments, on the basis of which it draws up evaluative principles of what economic reality should be and establishes measures and norms.

The concept of factor (y) of production has already been mentioned. **Factors of production** are natural, human, financial and other resources that enter into production and help to generate final economic output. They are said to be scarce because their quantity is limited. These include land, labour, capital and technology.

Scarcity is the fact that resources (factors of production) are scarce and therefore the quantity of goods produced is not sufficient to satisfy all human needs. The category of needs is very broad, but economics does not address their scope, structure and breakdown. Within the framework of economic theory, the relevant needs are economic needs that are satisfied by the consumption of produced goods and services. The essence is the satisfaction of a feeling of lack of something that is desirable for a given consumer and the fact that there is never total satisfaction, since the fulfilment of one need gives rise to another, changes the intensity (urgency) of the need, the structure or hierarchy of needs, etc. The concept of consumption is related to the satisfaction of needs. Consumption of economic goods leads to the satisfaction of human economic needs, provided that the goods are useful, i.e. the goods are capable of satisfying the need.

To remember

Scarcity is the fact that resources (factors of production) are scarce and therefore the quantity of goods produced is not sufficient to satisfy all human needs.

The facts thus defined clearly show that it is necessary to consider the allocation of resources in the economy and the efficiency of the use of scarce resources. It is a matter of finding the best possible use of available resources among the various alternatives. There are always several, sometimes even many, options open to economic agents. Then the answers to the basic questions of economics must be resolved: "What?", "How?" and "For whom?" to produce. In a market economy, the mechanism that provides the answers is the market, "...through which buyers and sellers clash to determine the price of goods and the quantities to be bought and sold".

A function is a basic mathematical concept used to describe the relationship between two variables, whereby for each value of the input variable (often referred to as x), there is exactly one value of the output variable (often referred to as y). A function can therefore be interpreted as a rule that assigns one output to each input based on a given rule. An example of such a rule is the function "take a number and multiply it by two" or "take a number and multiply it by two". Algebraically, these relations can be written as $y = x^2$ and $y = 2x$. The function is sometimes referred to as a transformation because it describes how an input variable is transformed into an output.

It is often necessary to point out that one variable depends on another without being able to describe the exact algebraic relationship. In this case, the function is denoted as $y = f\ (x)$, which simply expresses that the value of the variable y depends on the value of the variable xxx via some rule f. In this notation x is referred to as the independent variable, while y is the dependent variabe, reflecting the fact that x can be freely changed while y is dependent on the value x. Sometimes the variable y may depend on more than one variable at the same time, which can be expressed by notation $y = f(x_1, x_2)$, which means that values $x_1 a\ x_2$together determine the output value y.

One of the most common ways to represent and analyze a function is to represent it graphically. A function graph allows you to visually see how the dependent variable changes as the independent variable changes. Typically in mathematics, the independent variable is plotted on the horizontal axis (x axis) while the dependent variable is plotted on the vertical axis (y axis). This graph then shows the relationship between the two variables and allows you to quickly understand their interdependence. However, in economics, we often see the axes swapped, especially in demand and supply charts. For example, a graph of the demand function may show the price on the vertical axis and the quantity demanded on the horizontal axis, making it easier to understand the relationship between price and demand.

Mathematically, the graph of a function is an important tool for identifying various characteristics of a function, such as its slopes, directives, or intersections with axes.

These concepts are especially crucial for linear functions that have the form $y = ax + b$, where α is the directive and b is the intersection with the y axis. The directive expresses how the value changes y when changing the value x. A positive directive means that the values y and x grow together (direct proportionality), while a negative directive means that the growth of x is accompanied by a decline y (inverse proportionality). In the case of non-linear functions, the directive changes at each point and the graph is curved. The rate of change of the function at a point is given by the derivative, which is a mathematical tool that accurately expresses the slope of the tangent to the curve at that point.

Derivation is a key concept not only in mathematics but also in economics, where it helps to analyse the rate of change of one variable relative to another. Derivation of the function $y = f\ (x)$ according to the variable x is referred to as $f'(x)$ and indicates how fast the value of y changes with a small change in the value of x. Derivation is defined as the limit of change y to change x as the change in x approaches zero. This concept is particularly useful when exploring the extremes of functions, for example, when looking for maximum or minimum profit points in economic models.

Another important concept is the logarithmic function, which has a wide range of applications in economics, for example in the study of exponential growth or in modelling elastic relationships between variables. The logarithm is defined as the inverse operation to an exponential function and has many useful properties that make it easier to work with exponential models.

In conclusion, functions, their graphs, derivatives and logarithms are the key building blocks of mathematical analysis, which economists use to understand the complex relationships between economic variables. These tools make it possible to analyse how different factors interact and provide the basis for many decisions in business, public policy or finance.

?

Control questions

1. What is the slope of the curve?
2. Draw the horizontal and vertical axes of the graph,mark the origin and describe the axes with x and y variables.
3. What is the difference in understanding the variables in terms of their dependence?
4. What types of relationships exist between variables?

5. Characterize the concept of scarcity to define the definition of economics as a scientific discipline.

Test questions

1. What is a function in mathematics?

a) The relationship between three quantities
b) A relationship between two variables where each value of the input variable corresponds to exactly one value of the output variable
c) A method of solving quadratic equations

2. What is the name of the function when we want to express that y depends on x, but we do not know the exact algebraic relationship?

a) $y = x^2$
b) $y = 2x$
c) $y = f(x)$

3. What is the directive u of the linear function $y = ax + b$?

a) Intersection with the y-axis
b) How the y-value changes as the x-value changes
c) The intersection with the x-axis

4. What is the derivative of a function used for in economics?

a) To determine the intersection of the function with the y-axis
b) To analyse the rate of change of one variable relative to another
c) To calculate the logarithm of a function

5. What is the meaning of the logarithmic function in economics?

a) Used only to solve differential equations
b) It has no use in economics
c) It has a wide range of applications, for example in the study of exponential growth or in modelling elastic relationships between variables

Answers:

1b, 2c, 3b, 4b, 5c

Practical task

1. Find a real-world economic problem that can be modelled using a function (for example, the relationship between price and demand). Design a suitable function and graphically illustrate how changes in price would affect the quantity demanded. Explain your findings.

2. Explore examples of functional relationships in economics. Choose a specific product or service and analyse how various factors (e.g., price, marketing, competition) affect the demand for that product. Write a short report that summarizes your findings and recommendations for how the firm should adjust its strategy to increase demand.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie*. 3. dopl. vydání. Praha: Management Press, 2003. s. 45 – 71. ISBN 80-7261-061-9. 548 s.

[2] HOLMAN, Robert. Ekonomie. Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[3] HOLMAN, Robert. *Mikroekonomie*. Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[4] SAMUELSON, P. A. a W. D. NORDHAUS. *Ekonomie*. 18. vydání. Praha: NS Svoboda, 2007. ISBN 978-80-205-0590-3

[5] JUREČKA, V. a kol., 2018. Mikroekonomie. 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[6] SOUKUPOVÁ, J., 2018. Mikroekonomie. 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 2

Consumer Behaviour and Demand Shaping

After studying the chapter you will be able to:

- define the basic concepts of consumer behaviour theory;
- characterise the shapes of indifference curves depending on preferences;
- explain the consumer optimum.

Keywords:

utility, preferences, indifference curve, budget line, consumer surplus.

After studying the chapter you will be able to

- Explain the concept of marginal rate of substitution in consumption (MRSc).
- Describe the budget constraint and how it affects consumer decision making.
- Give examples of special forms of indifference curves (perfect substitutes, complements).
- Analyse the optimal combination of goods according to ordinal and cardinal approaches.

Chapter preview

- In this section, we look in more detail at the analysis of consumer behaviour through indifference curves and budget constraints. The chapter focuses on modelling consumer preferences, optimal consumption basket choice and the relationship between utility and prices.

Chapter objectives

- Convey basic information about indifference curves and their shapes.
- To show, with examples, how consumers make choices depending on their preferences and budget.
- Understand the different shapes of indifference curves, for example for desirable and undesirable goods.
- Describe how a consumer determines his optimal combination of goods.

Estimated study time

- The estimated time required to study the chapter is 180 minutes.

2.1 Assumptions of Rational Consumer Behaviour

The behaviour of individuals, as well as of all economic agents, can be explained by comparing the effects of economic activity and the "harm" (expenditure, costs) associated with that activity. In the case of an individual, the effect is the utility derived from the consumption of particular goods, the "harm" is the expenditure of income on the purchase of these goods. A rational consumer maximises utility. However, he is limited in his decision-making by his income. The utility arises from the consumer's preferences.

The starting point of consumer theory is the notion that an individual chooses from different sets of goods or consumption baskets. The consumer's decision is then to choose the consumption basket that yields the maximum utility. The consumer compares the different consumption situations in terms of preferences. Preferences will be analysed using several simplifying assumptions which can be considered ***as axioms*** in terms of their meaning.

1. **The axiom of completeness of comparison.** We assume that any two baskets of goods can be compared in terms of consumer preference. For each two consumption baskets A and B, one of the following three situations must occur:
   - A is preferred to B, we denote this case by A B;
   - B is preferred to A, we denote A < B;
   - A and B are indifferent, equally "attractive", we denote A = B.

Simply put: the axiom of completeness of comparison implies the consumer's ability to decide which basket of goods to prefer, to rank the baskets according to preference.

2. **Transitivity axiom**. We assume that for every three baskets of goods A, B, C: if A is preferred to B and B is preferred to C, then A is also preferred to C, or: if A > B and at the same time B > C, then A > C.

These axioms are a simplification, in reality there may be situations where the axioms do not apply. The axioms of comparison and transitivity imply that consumption situations can be ranked according to consumer preferences. This arrangement is called a preference scale.

2.2 Measuring Utility

As expressed earlier, the consumer's goal is utility maximisation. **Utility** is a quantity indicating the direction of preferences, if the consumer finds the most preferred situation, he maximizes utility. Thus, we can derive a utility function from preferences. It is sufficient to assign a higher utility to the more preferred consumption basket. The specific amount of utility is irrelevant.

Since the emergence of utility theory, economic theory has been confronted with the problem of how to measure utility and whether utility is measurable at all. Depending on the approach to the measurability of utility, we distinguish between ***cardinal*** and ***ordinal*** versions of utility theory.

2.2.1 Cardinal Version of Utility Theory

The cardinal version regards utility as directly measurable, as a cardinal quantity. In this case, specific values of utility would be known. **Total Utility** (TU) expresses the total satisfaction of needs when consuming a given quantity of a good. **Marginal Utility** (MU) expresses the change in total utility induced by a unit change in the quantity of the good consumed.

If utility were measurable, it would be possible to construct a total and marginal utility curve as in Figure 3

Definition

The cardinal version of utility is an approach in economics that considers utility as a directly measurable (cardinal) quantity. According to this theory, utility can be quantified and expressed in concrete values. A distinction is made between total utility (TU), which represents the total satisfaction of needs when a given quantity of a good is consumed, and marginal utility (MU), which represents the change in total utility caused by a unit change in the quantity of a good consumed. This approach allows the construction of total and marginal utility curves.

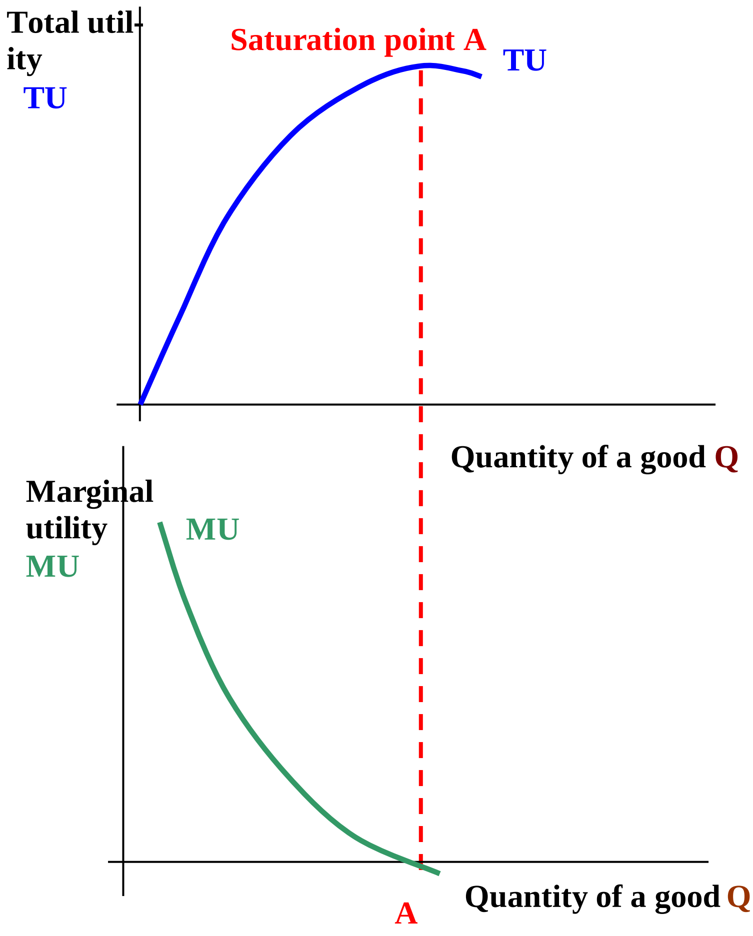

Fig. 3 Total and marginal utility

Total utility increases as the quantity of goods consumed increases, but the increments of utility slow down; marginal utility is thus decreasing. The assumption of decelerating increments of total utility is so important that we speak of the ***law of diminishing marginal utility***.

In the figure we see how total and marginal utility change as the quantity of the consumed good changes. From a certain quantity of a good consumed, its total utility can be decreasing and marginal utility negative. This point is called the **saturation point (point A**). At what quantity of the good the consumer reaches the saturation point, and whether it exists at all, depends on the nature of the good and the consumer's preferences.

Since total utility depends on the quantity of all goods, utility is, other things equal, a function of the quantity of goods consumed **U f (X1, X2, ..., Xn), where** X1, X2, ..., Xn are the quantities of each good.

For simplicity and graphical representation, in the following we will assume that the individual consumes only two goods, X and Y, and the utility is a function of these two goods:

$$\mathbf{U = f\ (X, Y)}$$

Considering two goods, MUx and MUy are partial derivatives of the utility function:

$$MUx = U/\delta X \text{ a } MUy = \delta U/\delta Y.$$

2.2.2 Ordinal Version of Utility Theory

Current economic theory tends to lean towards an ordinal version of utility theory, according to which utility is not directly measurable. The consumer is able to say which consumption situation he prefers, but not how great the utility is. Furthermore, it is possible to determine whether total utility increases as the quantity of the good consumed increases and the marginal utility is therefore positive, or whether total utility decreases and the marginal utility is negative.

It follows that the consumer is able to rank combinations of goods according to their utility but is unable to determine the magnitude of the utility of these combinations. Then the graphical representation is also different.

In this case, it is not possible to plot the total utility curve directly, but it is possible to ***connect points representing combinations with the same utility***. These are actually points equidistant from the base. To construct the figure, the direction of the preferences must be determined. We imagine the figure as a map of the combinations of X and Y goods. By determining the direction of preferences, we determine the "top of the utility hill". The curves representing the same utility are then the 'contours of the utility hill'.

Definition

The ordinal version of utility theory is an approach in economics that assumes that utility is not directly measurable. According to this theory, the consumer can rank combinations of goods according to preferences but cannot determine the specific magnitude of utility. Instead of a total utility curve, indifference curves are used to connect points representing combinations of goods with the same utility. This approach makes it possible to analyse consumer preferences without having to measure utility precisely and respects the non-saturation axiom.

For both goods, the combination with the higher quantity is preferred by the consumer to the combination with the lower quantity, i.e. the non-saturation axiom holds. We are therefore able to say that all combinations on the D*D* curve are equally useful. At the same time, combinations on D*D* are more useful than combinations on A*A* and less useful than combinations on G*G*. In doing so, we do not need to know the utility level of each combination. Curves showing combinations with the same utility are called indifference curves.

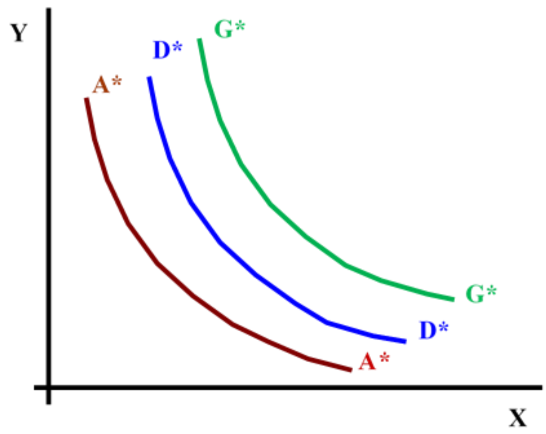

Fig. 4 Indifference map

An indifference curve is a set of combinations of goods X and Y with the same total utility. Since the specific amount of utility is not relevant, we can transfer the indifference curves into a two-dimensional figure (by projecting them onto the ground plane). Indifference curves can be used for the ordinal approach as a representation of combinations of two goods with the same utility. It follows from the above that indifference curves can be approached in two ways:

- based on utility (the indifference curve represents a certain level of utility),
- preference-based (indifference curves represent preferences).

2.2.2.1 Properties of Indifference Curves

In what follows, we will use indifference curves to analyse consumer behaviour under different preference conditions, so we first discuss their properties.

1. **Indifference curves are decreasing (have a negative directive).**

Assume that for both goods (X and Y) the larger quantity is preferred to the smaller (the axiom of non-saturation). Then combinations that imply more of both goods are preferred to combinations that imply less of both goods. This is illustrated in Figure 1.2a), where the regions marked + are combinations preferred over A (because both goods are more abundantly represented) and the regions marked - are combinations over which A is preferred (because both goods are less abundant). In the remaining two areas are all combinations where there is more X and less Y, or vice versa. So here we can look for combinations with the same utility as A, e.g. S and R. These points lie on the same indifference curve, which must be downward sloping (e.g. the indifference curve U_2).

2. **Indifference curves do not intersect.**

This requirement follows from the axiom of transitivity. Figure 5a) shows a situation where the indifference curves intersect. Points A and Q, representing combinations of goods X and Y, are on the same indifference curve and therefore have the same utility. This is also true for points R and A, also on the same indifference curve. Point R is further from the origin than point Q and therefore has a greater utility. Thus, A = Q and A = R and at the same time Q < R, which is of course a violation of the transitivity axiom.

3. **At each point in the figure showing the consumption situation there is an indifference curve.**

This condition follows from the axiom of completeness (or is its analogy). In order to compare the utility of different combinations of goods, each combination must lie on some indifference curve. This is a situation analogous to the real numbers, for which it is true that for every real numbers x and y, there exists a number z such that z > x and z < y.

4. **Indifference curves are convex with respect to the origin.**

This requirement means that the less the consumer of good X has relative to good Y, the more he is willing to sacrifice good Y to obtain an additional unit of good X.

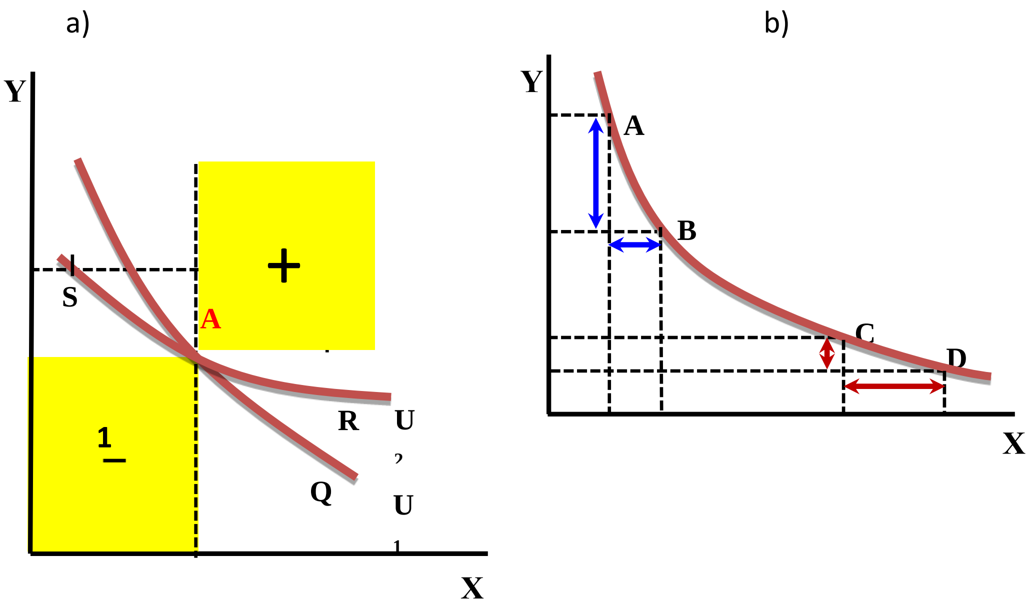

Fig. 5 Properties of indifference curves

This case is illustrated in Figure 5(b) by a shift from point A to point B. Conversely, the shift from roll C to point D represents a situation where farm X is represented relatively abundantly. In this case, the consumer is willing to substitute the same amount of good X with a smaller amount of good Y than in the case of the shift from point A to point B.

Unlike the previous properties of indifference curves, the convex shape is not a condition for rational consumer behaviour, but in the vast majority of cases we will take this shape of indifference curves into account.

2.2.2.2 Marginal Rate of Substitution in Consumption

Knowledge about some properties of indifference curves can be analysed using the indifference curve directive. It is called the marginal rate of substitution in consumption. The Marginal Rate of Substitution in Consumption (MRSc) is the ratio in which good Y is substituted for good X without changing the level of need satisfaction or total utility. (Sometimes the term Marginal Rate of Substitution of X for Y, MRScxy, is used.)

Thus applies

$$MRS_C = \frac{dX}{dY} \quad \Big| \quad U = \text{const.}$$

The marginal rate of substitution in consumption can be derived from utility. We consider a shift along the indifference curve. From the properties of indifference curves, we know that as X increases, Y must decrease in order for the level of total utility to remain the same. Suppose that the quantity of good X has increased by X. At the same time, the quantity of good Y has fallen by Y. We are therefore comparing two quantities:

- the "benefit" (gain in utility) resulting from an increase in the quantity of good X by ΔX, which can be expressed as $\Delta X * MUx$ (2.1)
- the "harm" (reduction in utility) caused by a decrease in the quantity of good Y by ΔY, which can be expressed as $\Delta Y * MUx$ (2.2)

These relations can be explained as follows: we know that MUx = ΔTU / ΔX, and therefore ΔTU = MUx * ΔX. Which is relation (2.1). We derive relation (2.2) for Y in a completely analogous way. Since on the indifference curve the utility is constant, the "benefit" and the detriment" must be equal, or ΔX *MUx = - ΔY *MUy (2.3) must hold. Minus means that Y moves in the opposite direction to X.

By modifying equation (2.3) (dividing ΔX * MUy and after truncation) we get:

$$-\frac{\Delta Y}{\Delta X} = \frac{MU_X}{MU_Y} \quad \textbf{and since it is} \quad -\frac{\Delta Y}{\Delta X} = \mathbf{MRS}_C$$

applies

$$\mathbf{MRS}_C = \frac{MU_X}{MU_Y}$$

In most cases, the marginal rate of substitution in consumption decreases as one moves along the indifference curve to the right (as the volume of goods on the x-axis increases). The decreasing marginal rate of substitution is reflected in the convexity of the indifference curves discussed earlier.

2.2.2.3 Special Shapes of Indifference Curves

Until now, we have assumed that both goods are desirable to the consumer; utility increases with the amount consumed. We call such goods desirable goods or goods with positive preferences (Goods).

However, there are also goods with a different direction of preference. It may be that a desirable good necessarily has a negative effect. In the context of society as a whole, this is the case for choosing a particular combination of industrial production volume and pollution. Even in consumer behaviour we can find cases where consumers prefer a smaller quantity of a good to a larger one. An example is the choice of portfolio structure (i.e. the choice between different types of securities). The utility function then has two variables: the return on the securities and the risk. The security holder always considers the return to be a desirable good. However, a higher return usually implies a higher risk. Risk, however, is an undesirable good for economic agents (not for all, as we will see later), or a good with a negative preference (Bad). This means that they prefer lower risk to higher risk. The indifference curves then have an atypical shape, being increasing (their directive is positive), as in Figure 6. On the x-axis the farm is undesirable, on the y-axis desirable. The direction of preferences is shown by the arrows.

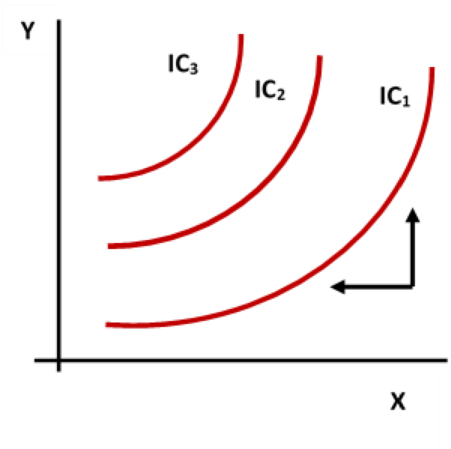

Fig. 6 Good x is undesirable

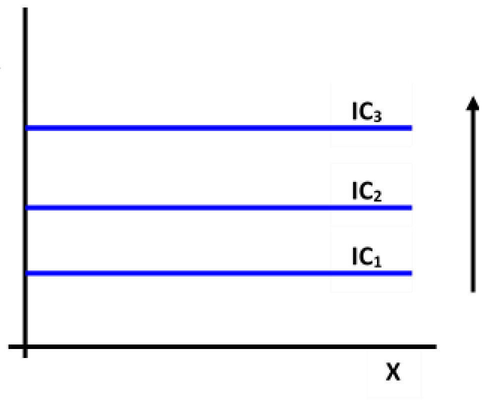

Fig. 7 Good x is indifferent

In addition to desirable and undesirable goods, there are also goods that do not affect the consumer's utility; the quantity consumed is indifferent to the consumer. Such goods are called indifferent goods or neutral goods (Neuters). The indifference curves then take the form of a straight line, as in Figure 5, where the good on the x-axis is the indifferent good and the good on the y-axis is the desirable good. In this case, the indifference curves are parallel to the x-axis. Here again, the arrows show the direction of preferences.

In reality, there may also be a situation where the direction of preferences changes as the quantity of the good consumed changes. Consider a good that is desirable up to a certain amount, but changes to undesirable from a certain amount. Figure 8 shows this good on the x-axis. The indifference curves therefore 'break' at a certain point. The indifference map can then be divided into two zones, positive and negative preference.

The indifference map can then be divided into two zones, positive and negative preference. There are cases where goods X and Y are perfectly substitutable for each other, they are perfect substitutes. The ratio in which the consumer is willing to substitute such goods, or MRSc, is constant and the indifference curves are straight lines.

It is obvious that if we allow the utility measurability, from the "break point" onwards the marginal utility of X would be negative and total utility would decrease as the quantity of X increases.

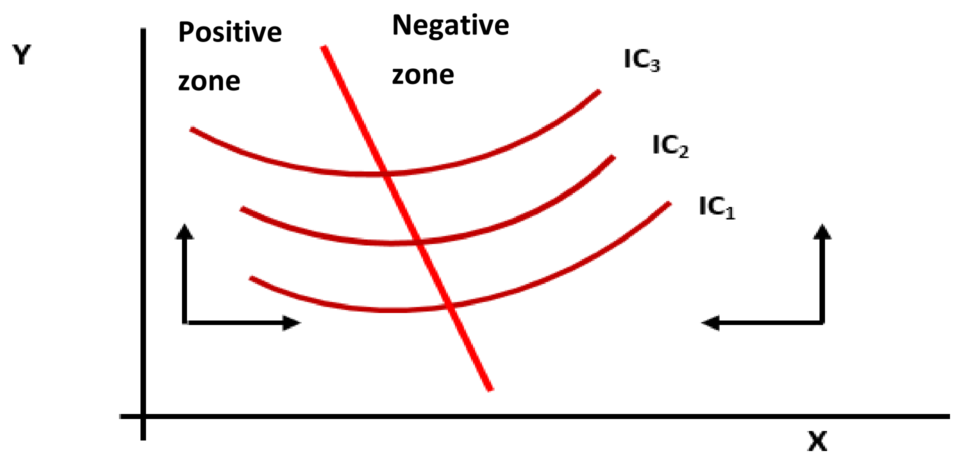

Fig. 8 The direction of consumer preferences is changing

Only if both goods X and Y are desirable can we consider the axiom of non-saturation. The shape of the indifference curves may also be affected by the relationship of goods X and Y in terms of preferences.

There are cases where goods X and Y are perfectly substitutable for each other, they are perfect substitutes. The ratio in which the consumer is willing to substitute such goods, or MRSc, is constant and the indifference curves are straight lines (Figure 9).

An example of perfect substitutes is the following situation: the consumer does not care whether he writes with a blue or a black pen.

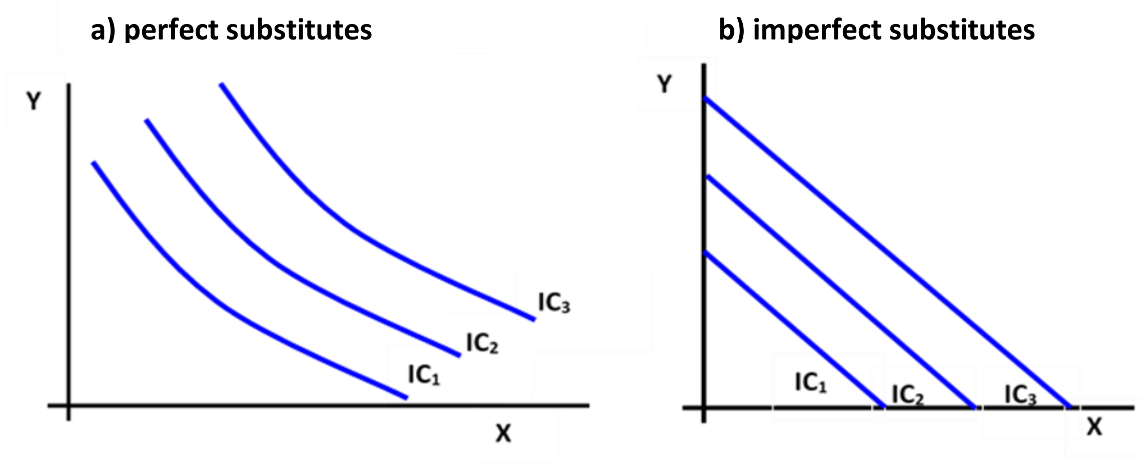

Fig. 9 Perfect and imperfect substitutes

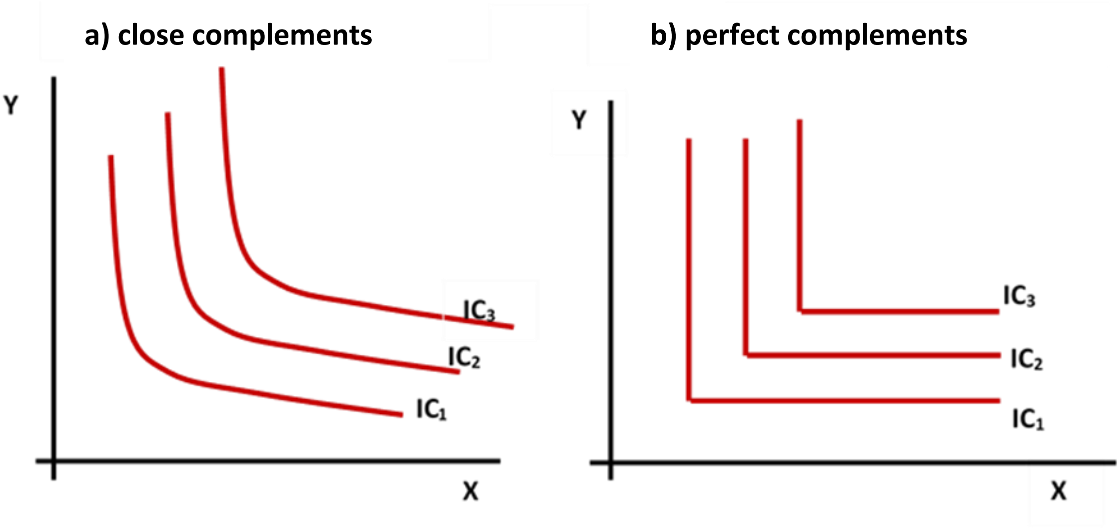

Fig. 10 Close and perfect complements

2.3 Budget Lines (Budgetary Constraints)

So far we have been concerned with consumer preferences and utility. However, in making a decision to purchase a good, the consumer is constrained by the level of his income and the prices of goods. Throughout what follows, we will assume that the prices of goods do not depend on the quantity that the consumer buys. Assume that the consumer spends all of his income on goods X and Y.

Thus applies

$$Px * X + Py * Y = I,$$

where I is the income of the consumer, Px is the price of good X and Py is the price of good Y.

Graphically, this equation is represented by a straight line, which we call the revenue line or budget constraint line. The area under this line (triangle OHJ) then represents all available combinations for which $Px * X + Py * Y \leq I$, or the Market Opportunity Set.

The intersection with the x-axis (point J, here $X = I / Px$) represents the situation where the consumer spends all of his income on the purchase of good X, the intersection with the axis (throw H, here $Y = I / Py$) represents the situation where he buys only good Y.

As with the indifference curve, we will be interested in the directive budget line. In the case of the budget line, we call it the Marginal Rate of Substitution in Exchange (MRS_E). It is the ratio at which the consumer can exchange goods X and Y in the market while spending a whole lot of income.

To remember

The optimal consumer choice occurs at the point where the indifference curve touches the budget line. This point represents the combination of goods that maximizes consumer utility under a given budget constraint. At this point, the marginal rate of substitution (MRS) on the indifference curve equals the directive of the budget line, i.e. the ratio of the prices of the goods. Graphically, this point is represented as the tangent point between the highest achievable indifference curve and the budget line. The consumer's optimal choice thus represents the best possible satisfaction of the consumer's preferences within his financial means.

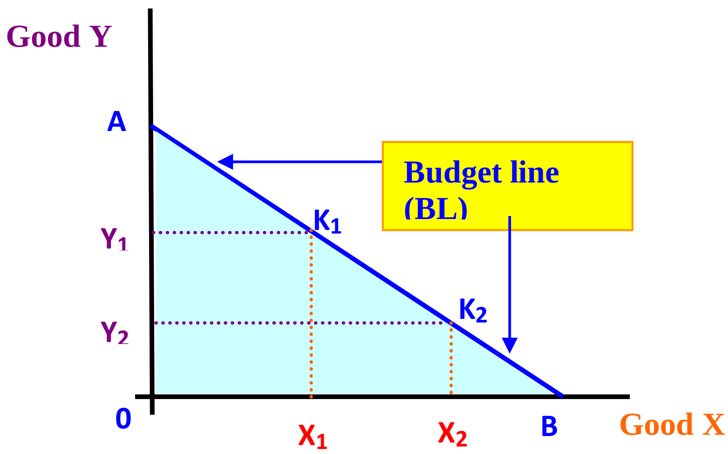

Fig. 11 Budget line (budget constraint)

2.4 Consumer Optimum

The consumer chooses the optimal combination of goods depending on his preferences and his market opportunities. These options are influenced by both his income and the market prices of the goods. How the consumer's optimum is determined depends on the possibility of measuring utility. According to the cardinal approach, the optimal quantity of a good is that for which the marginal utility equals the price of the good consumed: $\mathbf{MU_X = P_X}$

The optimal combination of two goods is one for which:

$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y}$$

The ordinal approach assumes that utility is not directly measurable. Therefore, the marginal utility ratio is used to determine the optimal combination: the marginal rate of substitution. While this depends on the utility function, it is not necessary to measure the utility directly to determine it. Recall that the marginal rate of substitution in consumption indicates the ratio in which the consumer is willing to substitute Y for X in his consumption basket.

The second question is in what proportion the consumer is able to exchange goods in the market. If we consider income to be constant, the proportion in which goods can be substituted depends on the ratio of their prices. This ratio is called the marginal rate of substitution in exchange.

The optimal combination of goods X and Y will be one for which the ratio in which the consumer is willing to substitute one good for the other is the same as the ratio in which he can exchange them in the market, or

$$\mathbf{MRS_C = MRS_E}$$

and therefore

$$-\frac{dY}{dX}\bigg|_{U,I\,=\,\text{const.}} = \frac{MU_X}{MU_Y} = \frac{P_x}{P_y}$$

The graphical representation of the optimal combination of goods X and Y is the point of contact between the indifference curve and the budget line. At the point of contact, the directives of the indifference curve and the budget line are identical. Thus, we again arrive at the equation (the situation is illustrated in Figure 12):

$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y}$$

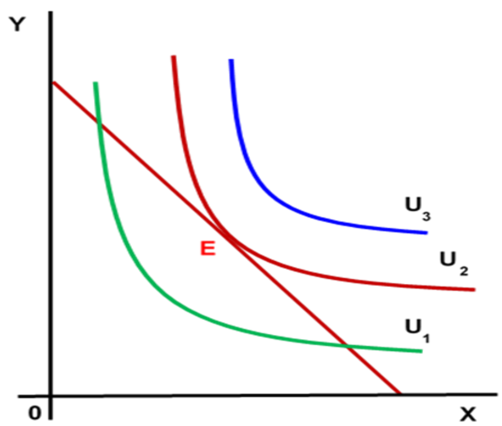

Fig. 12 Consumer optimum

Figure 12 shows that the consumer's optimum is located at the point of contact between the indifference curve and the budget line, i.e. at point E, where the slopes of the curve and the line are equal.

2.5 Consumer Surplus

A consumer's surplus is the difference between the total utility derived from the quantity of a good consumed and the cost of obtaining it (the total amount paid for it), or its market value.

Example: *we know the utility of the product and we also know that its price is CZK 5. Assume a cardinal approach and the utility is measured in monetary units.*

Tab. 1 Consumption of the good X

Purchased quantity	1	2	3	4	5	**6**
Total utility	10	19	27	34	40	**45**
Marginal utility	10	9	8	7	6	**5**

At a price of CZK 5 the consumer buys 6 units (this follows from the condition of equality of marginal utility and price).

*The total utility is 45 (and is actually the sum of marginal utilities or for X = 6 is TU = 10 + 9 + 8 + 7 + 6 + 5 = 45). The expenditure on the purchase of this good (market value) is **5 times 6 = 30**.*

From these data we can find the consumer surplus: 45 -30 = 15. Or the consumer is willing to pay CZK 45 for 6 units of good X, but pays CZK 30, so the consumer surplus is CZK 15.

In the following Figure 13 the consumer surplus is shown by the shaded area.

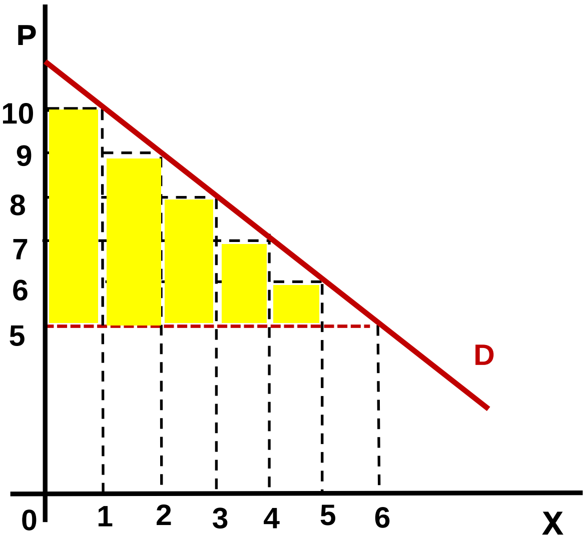

Fig. 13 Consumer surplus

Consumer theory is a key element of economic analysis, focusing on the behaviour of individuals in making decisions about the consumption of goods and services. Its foundations lie in several key axioms that define consumer preferences, their utility functions and their interaction with budget constraints. Consumers' preferences express their perceptions of what is satisfactory to them. These preferences are often expressed in terms of utility functions that measure the degree of satisfaction (utility) from consuming different combinations of goods. The basic axioms of consumer theory include completeness, which means that consumers can compare and evaluate all possible combinations of goods, and transitivity, according to which if a consumer prefers combination A to B and B to C, then he prefers A to C.

There are two main approaches to measuring utility that form the basis of consumer theory. The first is the cardinalist approach, according to which utilities are measurable and can be expressed in numerical form. Consumers have a clear idea of how much utility each combination of goods provides. The second approach is the ordinalist approach, which argues that utilities cannot be measured in absolute terms but can be ordered by preference. Consumers may say that combination A is better than B, but they will not say how much.

Indifference curves are a graphical tool that illustrates consumer preferences. Each indifference curve shows combinations of two goods that provide the same utility to the consumer. The curves are usually downward sloping, reflecting the substitution effect, i.e. the consumer's willingness to substitute one good for another while maintaining the same level of utility. The further away from the origin of the coordinates, the higher the utility represented by the curve. Indifference curves must never cross, as this would contradict the principle of transitivity.

The budget constraint is another important concept in consumer theory. It expresses the quantity of goods that a consumer can afford to buy based on his income and the prices of the goods. Graphically, this constraint is represented as a straight line on the same coordinate system as the indifference curves. The point where the budget line touches the indifference curve represents the optimal combination of goods that the consumer chooses. This point shows the maximum utility that the consumer can afford at a given budget.

The marginal rate of substitution (MRS) is a key concept that describes the rate at which a consumer is willing to substitute one good for another without changing the total utility. The MRS is calculated as the slope of the indifference curve and expresses the relationship between two goods. Optimal consumer choice occurs when the MRS is equal to the ratio of the prices of the goods. In this way, the consumer achieves maximum utility while fully exploiting his budget constraint.

The substitution effect and the income effect also play an important role in consumer theory. The substitution effect occurs when a change in the price of one good leads the consumer to change his preferences towards a cheaper good. The income effect refers to the change in a consumer's disposable income due to a change in prices, which affects his ability to purchase goods. These effects work together to influence consumers' overall decision-making, which is key to understanding market dynamics.

Overall, consumer theory and its concepts such as indifference curves, budget constraints and marginal rate of substitution provide a valuable framework for analysing consumer behaviour and decision-making processes in the economy. These tools allow economists and analysts to better understand how individuals react to price changes and how these reactions affect supply and demand in the market.

?

Control questions

1. What does the concept of preference transitivity mean? Can you give an example of a situation where preferences are not transitive?
2. What are the properties of indifference curves and how do they relate to the axioms of rational consumer behaviour?
3. Assume that indifference curves are not negatively sloped. What can you say about consumer preferences?
4. Explain the consumer's optimum condition.

Test questions

1. What do indifference curves express?

a) The combination of goods that the consumer can afford
b) The combination of two goods that provide the same utility to the consumer
c) The consumer's optimal choice for a given budget

2. What is the main difference between the cardinal and ordinal approaches to utility measurement?

a) The cardinal approach considers utility to be unmeasurable, the ordinal approach considers utility to be measurable
b) The cardinal approach measures utility in absolute terms, the ordinal approach merely orders preferences
c) The cardinal approach is used for luxury goods, the ordinal approach for ordinary goods

3. What represents the point where the budget line touches the indifference curve?

a) The consumer's minimum utility
b) Average consumer utility
c) The optimal combination of goods with maximum utility at a given budget

4. What is expressed by the marginal rate of substitution (MRS)?

a) The rate at which a consumer is willing to substitute one good for another while maintaining the same utility
b) The total utility of the consumer
c) The consumer's budget constraint

5. What does the axiom of transitivity in consumer theory express?

a) The consumer always prefers more expensive goods
b) If a consumer prefers A to B and B to C, then he prefers A to C
c) A consumer can compare all possible combinations of goods

Answers:

1b, 2b, 3c, 4a, 5b

Practical task

1. Draw a graph with indifference curves for two goods (e.g. food and clothing) and a budget line. Mark the point of optimal consumer choice. Then show how the situation would change if the price of one of the goods were to increase. Describe how this change will affect the consumer's optimal choice and explain the role of the substitution effect and the income effect.

2. Choose a specific example of a consumer decision (e.g. choosing between different types of holidays, choosing a mobile phone) and apply the concepts of consumer theory to it. Describe the consumer's preferences and budget constraints and try to determine the optimal choice. Explain how the consumer's decision might change if prices or income change. Write a short analysis (300-500 words) using concepts such as indifference curves, marginal rate of substitution, and substitution and income effects.

Literature:

[1] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005

[2] HOLMAN, Robert. Ekonomie. Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[3] HOLMAN, Robert. *Mikroekonomie.* Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[4] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. Mikroekonomie. 5. aktual. vydání. Praha: Management Press, 2010. s. 49 – 71. ISBN 978-80-7261-218-5. 574 s.

[5] JUREČKA, V. a kol., 2018. Mikroekonomie. 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[6] SOUKUPOVÁ, J., 2018. Mikroekonomie. 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 3

Consumer Decision-making under Conditions of Risk

After studying the chapter you will be able to:

- characterise risk and its impact on consumer decision-making;
- analyse the consumer's optimum under conditions of risk;
- address risk reduction.

Keywords:

risk, expected outcome, expected utility, optimal decision, relationship to risk, fair bet, diversification, portfolio.

After studying the chapter you will be able to

- Identify and describe the properties of indifference curves and their application in practice.
- Calculate the marginal rate of substitution and interpret its importance for consumer preferences.
- Apply the principles of consumer optimization to specific examples and scenarios in economics.

Chapter preview

- The chapter focuses on advanced aspects of microeconomics, including the properties of indifference curves, the marginal rate of substitution in consumption, and consumer optimization. We analyse how these concepts affect the decision making of individuals and firms, and how they can be applied to real economic situations.

Chapter objectives

- Understand the properties of indifference curves and their relevance to consumer behaviour.
- Explain the concept of marginal rate of substitution and its application in consumer optimization.
- Analyse the consumer optimisation process and its impact on decision making in different economic contexts.

Estimated study time

- The estimated time required to study the chapter is 120 minutes.

3.1 Expected Result

Risk is a situation where the decision maker knows all the possible consequences of his decision and is able to determine the probability of each of them. The consequences must be independent of each other and the sum of their probabilities must equal one under the given assumptions.

Probability expresses the possibility that an outcome will occur, and in decision theory under risk, not only ***objective but also*** and above all ***subjective probability*** is used. Objective probability is based on knowledge of the frequency with which certain events tend to occur.

In some situations, such as flipping a coin or drawing a card from a shuffled deck, the probability of each outcome can be derived logically (a priori probability). Such situations are not common in the economic system. In other situations, the probability can be derived from empirical data. For example, insurance companies dealing with a large number of people may collect a large amount of information to determine the probability of an event.

To remember

Risk in economics is a situation where the decision maker knows all the possible consequences of his decision and is able to determine the probability of each of them. These consequences must be independent of each other and the sum of their probabilities must be equal to one. Probability can be objective, based on the known frequency of occurrence of events, or subjective, based on personal judgement and experience. In making decisions for risk, people do not seek to maximize expected return, but expected utility, which takes into account the utility value of each possible outcome. This approach explains why different people may make different decisions in the same situation, because they have different ideas about the probabilities and utilities of each outcome.

Example

For example, suppose you are deciding whether to invest in shale gas exploration. If the research is successful, the share price will rise from CZK 300 to CZK 400, if it is unsuccessful, it will fall from CZK 300 to CZK 200. We are thus in a situation where there are two possible future outcomes - a share price of CZK 400 or CZK 200.

*Suppose we know that out of the last 100 exploratory wells, 25 have been successful and 75 have failed. Then the probability of success ¼ is **objective**, because it is based directly on the frequency of similar results.*

The subjective probability is some impression that the predicted outcome will occur. This impression may be based on a person's knowledge and experience (knowledge of the industry, the state of the economy, etc.). Because different people may have different information or different ability to use it, they may have different perceptions of the probability of different outcomes and therefore make different decisions.

Using probability, two quantities are calculated to describe and compare risky choices: **expected outcome** and **expected utility**.

The Expected Result (EX) is the mean of all possible outcomes, i.e. a weighted average where the probability of each outcome is taken as a weight. For a situation with n possible outcomes (outcomes X_1, X_2..., X_n), the expected outcome is:

$$EX = \sum_{i=1}^{n} X_i \,.\, \pi_i$$

where π_i is the probability that the consequence X_i occurs.

If, for simplicity, we assume a situation with two possible outcomes X_1 and X_2 with probabilities π_1 and π_2, then the expected outcome is as follows:

$$EX = X_1 \,.\, \pi_1 + X_2 \,.\, (1 - \pi_2)$$

However, people do not choose the option that would give them the highest expected outcome but evaluate the benefits of each possible outcome simultaneously in their decision-making. Thus, they behave as if they assign a utility value to each outcome. Under these assumptions, decision-making under risk is not guided by the desire to maximise expected return, but by the desire to maximise expected utility.

3.2 Expected Utility

The Expected Utility (EU) of random outcomes is the mean of the utility of individual outcomes weighted by their probabilities.

The expected utility of an action X, which has n consequences X_i that occur with probability π_i, is

$$EU(X) = \sum_{i=1}^{n} U(X_i) \,.\, \pi_i$$

Given two possible outcomes X_1 and X_2 ***with*** probabilities π_1 and π_2, the expected utility

$$EU(X) = U(X_1) \,.\, \pi_1 + U(X_2) \,.\, \pi_2$$

Since the sum of the probabilities must be equal to one, we can modify the previous formula as follows:

$$EU(X) = U(X_1) \,.\, \pi_1 + U(X_2) \,.\, (1 - \pi_2)$$

If expected utility becomes the criterion for decision-making under risk, and its amount depends on both the probabilities of the outcomes and the utility values of those outcomes, this means that we are assuming people's ability to assign numbers to those outcomes. Thus, we can say that people behave as if they have a cardinal utility function.

We know the axioms that are necessary for the existence of an ordinal utility function. Similarly, it is possible - assuming three possible outcomes of the decision X_1, X_2, X_3 - to state the axioms necessary for the use of a cardinal utility function:

1. **Completeness of comparison:** $(X_1 > X_2)$ or $(X_2 > X_1)$ or $(X_1 = X_2)$
2. **Transitivity:** $X_1 > X_2$ and $X_2 > X_3$, then $X_1 > X_3$
3. **Non-transitivity: if $X_1 > X_2$** then the consumer prefers X_1.
4. **Continuity** (probability is included here).

3.2.1 Derivation of the Utility Function

The continuity axiom allows us to derive utility values for each possible outcome of a decision, or to determine a set of numbers expressing the intensity of preferences regarding vigorous outcomes. The construction of the utility function, or the determination of the utility for each possible outcome, involves the following three steps:

1. ***Sort the results by preference*** - it is reasonable to assume that the sequence will be $X_1 > X_2 > X_3$ (see the axiom of completeness and transitivity).

2. Determine ***arbitrarily the utility value*** of the most preferred and least preferred outcome, which will allow for the establishment of a benchmark. Once the benchmark is set, the utility of any situation can be determined and compared with others. Since the determination of utility is *ordinal*, any value can be chosen for U (X1) and

   $$U (X3), \text{ if } U (X1) > U(X_3).$$

   By convention, it is usually determined $U (X1) = 1$ and $U (X3) = 0$

3. Calculate utility values for intermediate outcomes.

We know that there is some probability of the outcome X_1, at which one will be indifferent between ***X_2 with certainty*** and the ***risky alternative X_1*** (the most preferred outcome) or X_3 (the least preferred outcome) - see the continuity axiom. That is, with this probability, the

$$U (X2) = U (X1) \bullet \pi + U(X_3) \bullet (1 - \pi)$$

It remains to determine the probability of return X_1 such that the above equation holds. Then by simply plugging in the numbers after X_1 and X_3 we can calculate the value of the utility for a particular outcome.

3.3 The Attitude to Risk

A risk averse person requires a relatively high probability of the highest possible outcome of a risky alternative in order to be indifferent between the certain and risky alternatives. In risk aversion, a certain outcome is preferred to a risk with the same expected outcome.

In this case, the utility function is concave. As income increases, total utility increases, but at a decreasing rate, i.e. slower than the consumer's income. This reflects the declining marginal income.

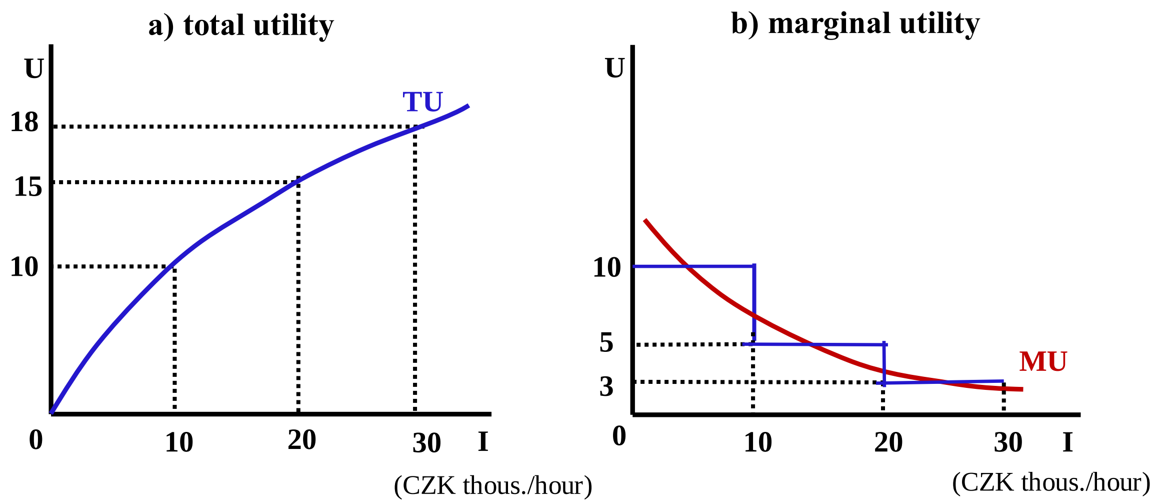

Fig. 14 Aversion to risk

The opposite approach to risk is ***risk seeking***. A risk-seeker is willing to take the risk of a relatively small probability of the highest possible outcome of a risky alternative. ***The utility function is convex, expressing the increasing marginal utility of income*** (Figure 15a) - utility increases faster than consumer income.

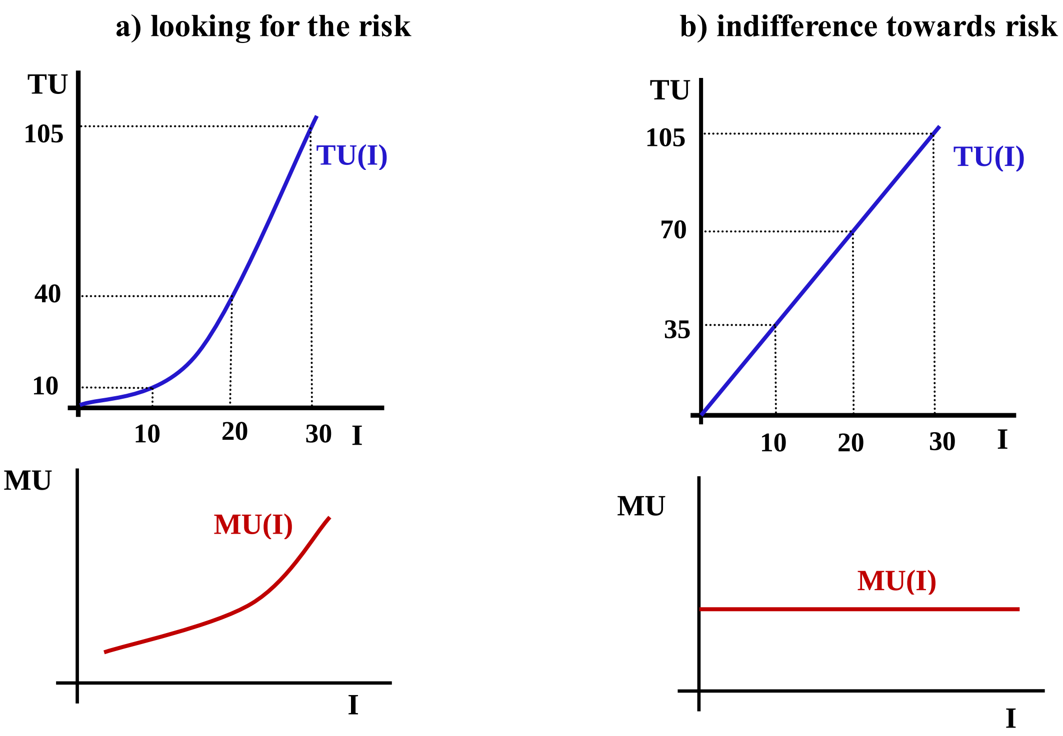

Fig. 15 Alternative attitudes to risk

If one is ***indifferent to risk*** (risk neutral), one is indecisive in choosing between the certain and risky decision alternatives if the certain outcome is the same as the expected outcome of the risky alternative. ***The utility function in this case is linear***, or a line passing through the origin and expressing a constant marginal utility of income.

3.3.1 Attitude to Risk and Willingness to Accept a Fair Bet

A person's attitude to risk can be inferred from the way he or she chooses between a certain amount of money and the fair bet alternative. A ***fair bet*** is considered to be a bet whose expected return is the same as the initial certain amount.

Example

We will assume that we have CZK 50 (= a certain sum of money). We are offered a fair bet in which we can win CZK 100 with 50% probability or lose all our money with the same probability (50%), i.e. have CZK 0.

Figure 16 below allows us to compare the two alternatives. The utility of the CZK 50 security is determined by point **D, the** utility of CZK 0 by point **A,** and the utility of CZK 100 by point **B.**

To compare the two alternatives graphically, we use a straight line connecting points **A** *and* **B***, because the decision maker considers both possible outcomes of his decision. The increasing probability of winning can be seen as moving along a straight line from point* **A (U = 0)** *towards point* **B** (U = 100). *If the probability of both outcomes were equal, the amount of utility is determined by point* **E** *in the middle of the line* AB.

The expected utility of the bet is determined by the point **E, which** is given by:

EU = 0,5. U(0) + 0,5 . U(100) = 0 + 10 = 10

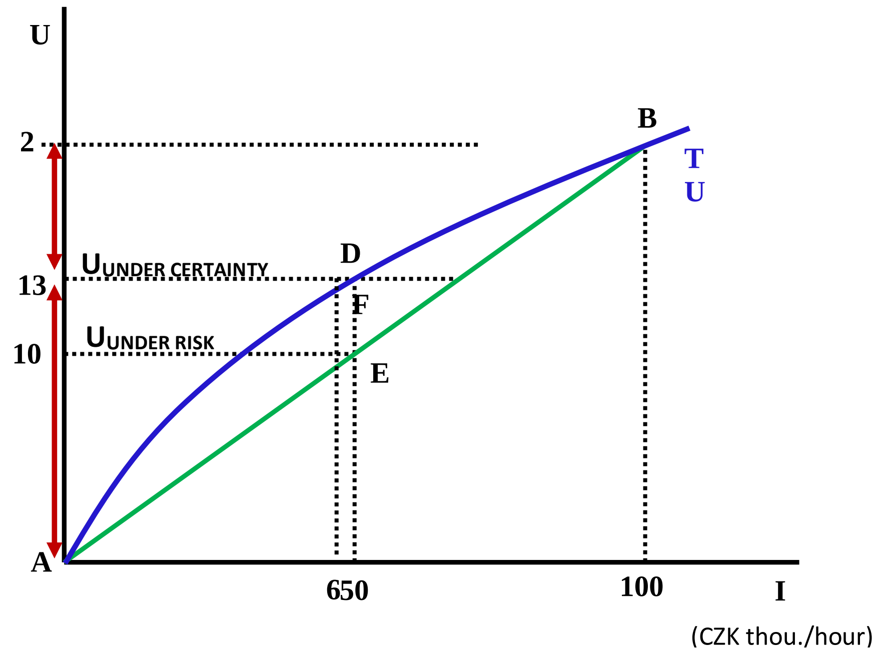

Fig. 16 Aversion to risk and fair bet

A risk averse person will prefer the alternative of certainty to a fair bet. The certain alternative will be preferred to the fair bet alternative in the case of a concave utility function of income - point **D** is associated with higher utility than point **E: U(50) > EU**.

Example

Jakub had the opportunity to win or lose CZK 50 with odds of 1:1. He rejects the offer because the expected benefit from it is less than the benefit from a certain CZK 50. The increase in utility from a

CZK 50 increase in income is less than the decrease in utility from a decrease in income of the same amount (see Figure 16). Since both options are equally likely, Jakub does not take the opportunity. As the probability of winning increases, he would be willing to consider participating in the game starting from the situation expressed by F, or at an expected outcome of 65, at which the utility of the game is the same as the utility of CZK 50.

The risk-seeker prefers the fair bet alternative to the sure bet alternative because it yields higher utility.: The sure alternative is preferred to the risky alternative for the convex utility function of income - point **E** in Figure 16(a) is associated with higher utility than point **D.**

Example

If Jakub were to participate in the game, he would be taking a risk, because the expected utility of the game is greater than the utility of a certain income of CZK 50. He values a CZK 50 increase in income more than a possible decrease in income of the same amount. Both options are equally likely, but Jakub expects a higher utility from participating in the game. He would accept even worse odds of winning than 1:1. Only when offered a game with an expected outcome of 27 (Figure 17a) would he hesitate and withdraw from the game when the probability of winning is further reduced.

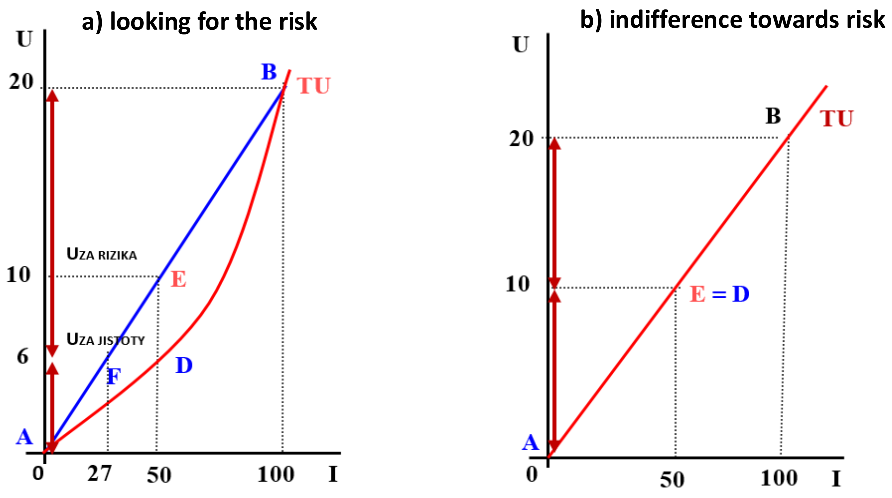

Fig. 17 Fair bet, looking for the risk and neutral attitude to risk

If the consumer is risk-seeking, then the possibility of increasing their income by a certain amount is very attractive to them, while the possibility of losing the same amount does not 'excite' them.

This means that he prefers to make a fair (and to some extent unfair) bet in order to secure a higher income.

The obvious proof of the existence of risk-preferring people is the fact that many people gamble. Equally, some offenders can be included in this category, especially when their criminal activity is associated with a relatively high chance of being caught and punished. With the exception of these special cases, few people seek out risk.

A person with a neutral attitude to risk will be indecisive in choosing between the alternative of certainty and a fair bet.

The most common attitude to risk is risk aversion. This fact is demonstrated by the enormous number of risks against which people insure themselves. Most people not only insure their car, house or health, but also seek employment with a relatively stable income.

3.4 Optimal Decisions Under Risk

The simplified model of decision making under risk assumes only two possible situations determining the outcome of a particular decision alternative: $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$. The graphical representation quantifies on the axes the returns that can be achieved in each of the two situations considered: on the x-axis the returns X_1 for the S_1 situation and on the y-axis the returns X_2 in the S_2 situation. Under these assumptions, essentially the same framework used by conventional consumer decision theory under certainty, i.e. ***indifference curves and budget lines***, can be used to model decision making under risk ***but their interpretation is different.***

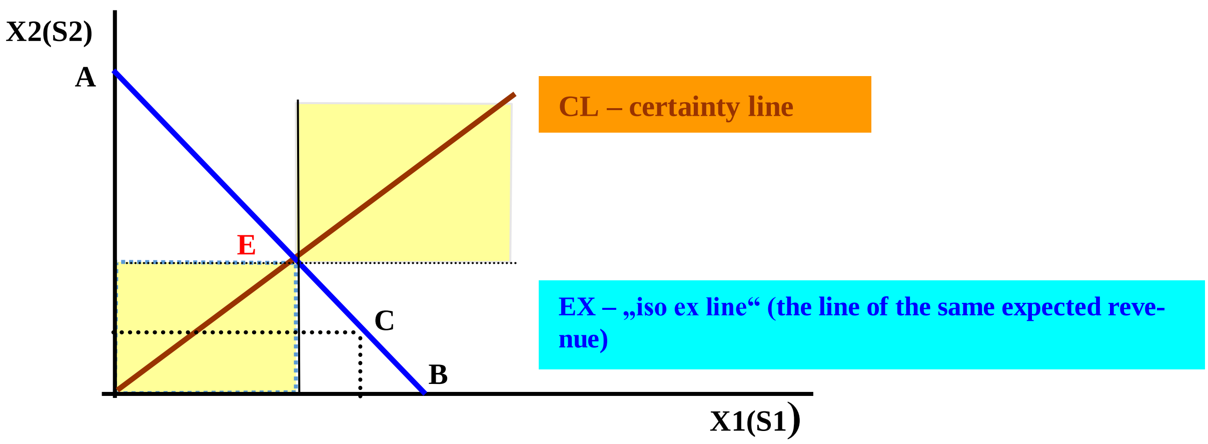

Fig. 18 The line of the same expected revenue and the certainty line

The line starting from the origin at an angle of 45°represents the same return in both situations considered and is therefore ***called the Certainty Line*** (CL) - whether ***the situation S_1 or S_2 occurs, the return will be the same***.

The budget line in this model is formed by the set of points that represent the ***same expected return in both situations***, i.e., it ***is a line of equal expected return*** (Iso - EX Line, EX). Recall that the equation for EX is

$$EX = X_1 * \pi_1 + X_2 * \pi_2 .$$

The slope of the equal expected return line is determined by the relative probability of the two situations π_1 / π_2, which can be derived by rewriting the equal expected return equation in directive form:

$$X_2 = \frac{EX}{\pi_2} - \frac{\pi_1}{\pi_2} . X_1$$

3.4.1 Shape of Indifference Curves and Relation to Risk

3.4.1.1 Aversion to Risk

Point C in Figure 24 represents a ***risky*** and ***point E*** a safe alternative decision. One can decide whether or not to take a certain risk. If not, he stays at point E. If he engages in risky activity, he may gain more or less, and ***the payoff is determined by point C***. This is a *fair offer*, because *both point E and C* lie on the *same line of equal expected return, and* so the expected return of the risky alternative is the same as the return of the certain decision alternative.

The risk-averse person will not accept the risky alternative, and it follows that point C must lie on a lower indifference curve than point E. The risk-averse person's indifference curve is convex to the origin. This is illustrated in Figure 19a.

At the same time, the indifference curve touches the ***EX*** line ***at point E*** (on the line of certainty) and therefore has a slope identical to the slope of the ***EX***: π_1 / π_2 line at this point

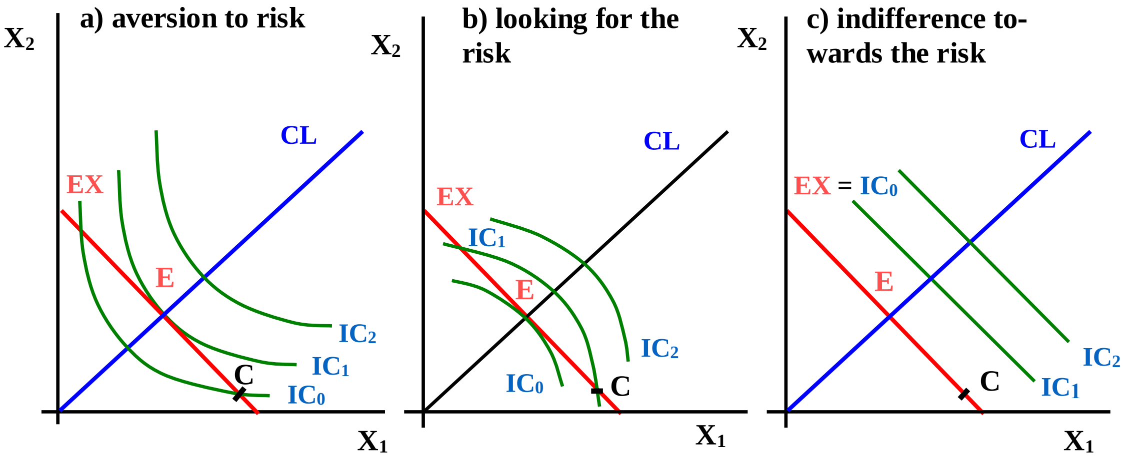

Fig. 19 Relationship to risk and shape of the indifference curves

3.4.1.2 Looking for the Risk

Those who prefer risk will prefer the risky alternative. Therefore, point C must lie on a higher indifference curve than point E - indifference curves are concave to the origin. This case is illustrated in Figure 19(b)

The indifference curves have a slope of π_1 / π_2 at the point of contact with the EX line on the certainty line

3.4.1.3 Neutral Attitude to the Risk

If one is indifferent to risk, both options (represented by points E and C) provide the same benefit. The indifference curves are parallel lines moving away with increasing utility from the origin. This case is illustrated in Figure 19c.

In this case, the indifference curve has a slope π_1 / π_2 throughout its course**.** Therefore, in the risk-neutral case, the line of equal expected return is identical to a certain indifference curve.

3.5 Risk Reduction

People generally tend to have aversion to risk even though in some cases a person chooses the more risky alternative. People therefore try to reduce the risk, which can be done basically in three ways: by obtaining more information about the alternative choices and their results, by insurance, and by diversification of activity.

3.5.1 Insurance

People with negative attitude towards risk are willing to give up some part of their income in order to avoid the risk. Insurance enables them to do so. A person with negative attitude towards risk will definitely insure themselves if the costs related to insurance are identical to the expected loss or lower (insurance with the expected loss of CZK 10,000 would cost CZK 10,000).

If a person has negative attitude towards risk, the marginal utility of income is downward sloping and the income utility curve is concave, he/she sees the possible decrease in income by a certain amount so "unappealing" that he/she prefers to pay the insurance premium so high that he/she actually pays more to the insurance company than it pays to him/her. The insurance premium is disadvantageous in the sense that he/she pays more for the insurance than he/she would lose, on average, if the insured event has occurred (he/she has been robbed), but the insurance gives him/her a certain income and this amount means to him/her a higher utility than the expected utility in the risk situation.

If the value of one's property or welfare (W) in the situation of certainty achieved by insurance is the same as the expected value of welfare in the risk situation (without insurance), the costs of such insurance is referred to as a fair insurance. In such case, the insurance coverage is identical with the expected loss, or the certain income equals the expected income. If the economic subject is insured, he/she secures the same income for himself/herself no matter if the loss occurs or not.

The guarantee of the identical income for any result represents for a person with negative attitude towards risk a higher utility than he/she would have if he/she chose the risky alternative (higher income with no loss or low income in case of loss).

Choosing insurance will not change the expected welfare but it can balance both possible results. This generates the higher level of the expected utility. The marginal utility is the same in both cases for the insured person - with or without loss (because the welfare remains the same). If the person is not insured, the marginal utility in case of loss is higher than with no loss (remember that in the case of negative attitude to risk, the marginal utility is decreasing). Transfer of welfare from the no-loss situation to the loss-making situation therefore must increase the total utility. This is enabled through insurance.

The aversion to risk explains ***the demand for insurance. The supply*** is explained by the law of large numbers.

The ability to avoid a large risk is based on the law of large numbers, which says that even if one event might be accidental and basically unpredictable, we can predict the average result of a large number of similar events. This law works only if the distribution of probability is statistically independent and identical for all people insured. In such cases, it is possible to change the uncertainty of one subject to certainty for a large group as a whole.

Consumers usually get an insurance from companies specialising in this activity. Insurance companies are firms that maximise profit. Why do they offer insurance? They know that if they split the risk, they take over only a small risk. Insurance companies work with a sufficiently large number of clients. This way they can ensure that due to a large number of events, the total amount of money paid out will be the same as the money they receive or lower.

Definition

Risk aversion is an attitude where people are willing to give up part of their income to avoid risk, which explains the demand for insurance. Insurance provides income security that yields a higher than expected utility in the face of risk, even though the premium may be higher than the average expected loss. People with risk aversion have diminishing marginal utility of income and a concave utility curve, which means that they perceive the possibility of income reduction as very unattractive. Insurance companies offer insurance because of the law of large numbers, which allows them to convert individual uncertainty into relative certainty for a large group of clients. The maximum premium that a risk averse person is willing to pay is determined by the difference between his current wealth and the level of wealth at which the utility of certainty is the same as the expected utility without insurance.

The utility connected to the insurance is thus higher than the expected utility of an uninsured person (in Fig. 20 $U^* > EU$). A person with negative attitude to risk therefore insures himself/herself if there is a possibility of fair insurance. He/she insures himself/herself even if the insurance exceeds the expected loss as long as the achieved utility does not exceed the utility of the uninsured person, i.e., as long as he/she achieves the welfare W - M or higher. If the insurance exceeds the level W-M, the utility connected to the certainty is lower than the expected utility if he/she refuses the insurance. **W_M - W therefore represents the maximum insurance a person is willing to pay.**

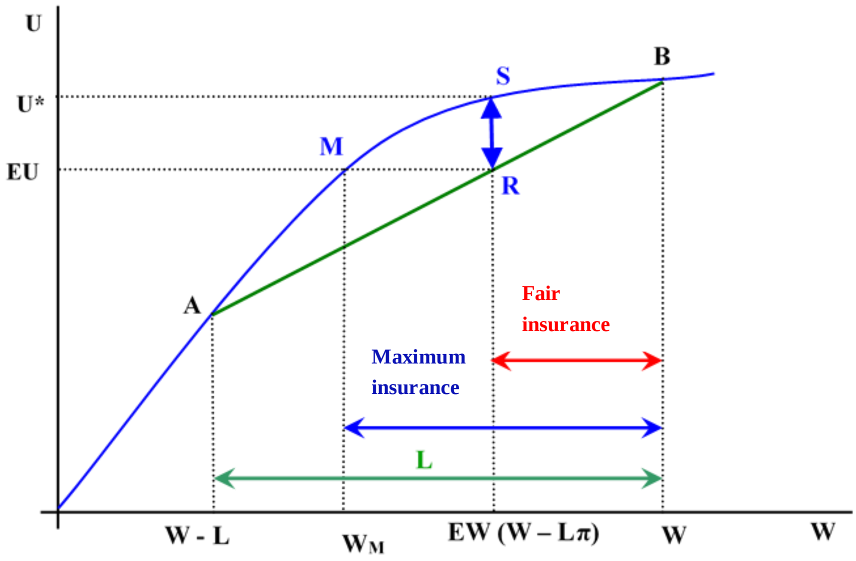

Fig. 20 Utility function and insurance

The maximum insurance is such a level of insurance for which the utility connected to the certainty (achieved through insurance) is the same as the expected utility connected to the risky alternative (without insurance). We can prove in a similar way that a person looking for a risk will not insure himself/herself at a certain level of insurance.

3.5.1.1 Optimal Choice on Insurance Using the Indifference Curves

If we label the situation with no loss S_1 (property is not damaged by fire) and the situation in which loss occurs S_2 (a fire occurs), the point C in Fig. 21 represents the *initial risky situation of an uninsured person* in which X_1 is the extent of the welfare of the economic subject in case when there is no fire and X_2 is welfare in the case that a fire occurs.

In the situation under consideration (represented by the point C), the economic subject is on the line of the identical expected revenue EX0. If the person was fully compensated in the case of damage caused by fire (without having spent any costs), his/her welfare would be represented by the point B. In such case, he/she would have the same assets in both situations - with or without fire (X_1' in S_1 and X_2' in S_2). However, since being insured against fire means that the person must pay the insurance, *the point B is unachievable*.

The case when a person insured himself/herself and the insurance is fair (the monetary value in a certain situation achieved through insurance is the same as the expected revenue of the risky prospects) is represented by the point E, which lies on EX0 as well as the point C. In that case, the person pays the insurance in the extent $X_1' — X_1^*$ and achieves a certain level of welfare - the point E lies on the certainty line providing him with X_1* in the situation S_1 and X_2* in S_2. (In case of loss, he/she will get a compensation in the extent X_1* — X_2" from the insurance company).

Would a person like to insure himself/herself under these conditions? If he/she has aversion to risk, his/her indifference curve is convex in the origin and it has the slope 1/ 2 in the intersection with the certainty line. It therefore touches the line of identical expected revenue at a point through which runs the certainty line. The indifference curve running through the point E (curve U1), which represents the certain amount of assets achieved through insurance, lies above the indifference curve running through the point C (curve Uo), which represents the risky prospects of an uninsured person. He/she thus insures himself/herself if the insurance is fair. We can also see in the chart that the person is willing to insure himself/herself even if the insurance is higher than the fair insurance (determined for instance by the point A, i.e., the extent X_1'—X_1) - see the idea of maximum insurance mentioned above.

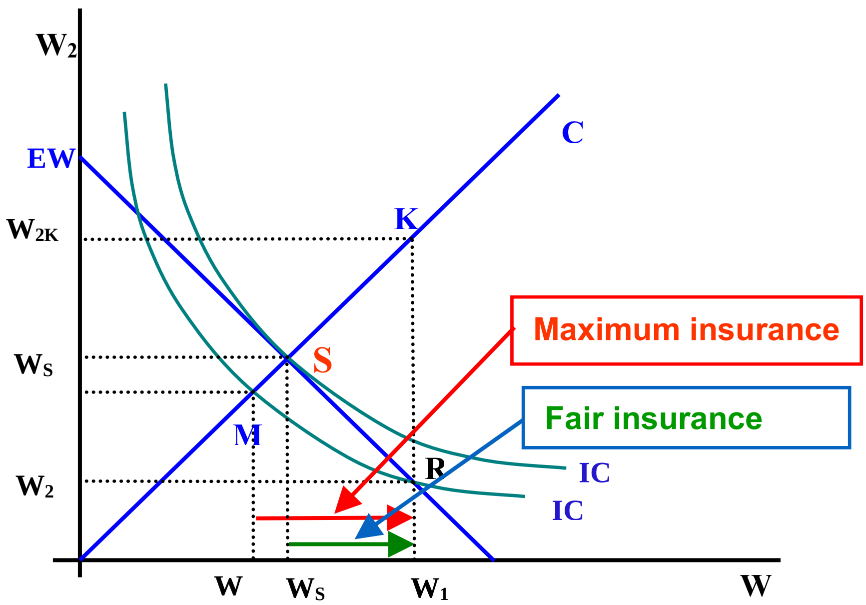

Fig. 21 Choice of insurance

The maximum insurance is such a level of insurance for which the utility connected to the certainty (achieved through insurance) is the same as the expected utility connected to the risky alternative (without insurance). We can prove in a similar way that a person looking for risk will not insure himself/herself at a certain level of insurance.

3.5.1.2 Optimal Choice on Insurance Using the Indifference Curves

If we label the situation with no loss S_1 (property is not damaged by fire) and the situation in which loss occurs S_2 (fire), the point C in Fig. 22 represents the *initial risky situation of an uninsured person* in which X_1 is the extent of the welfare of the economic subject in case there is no fire and X_2 is welfare in the case that a fire occurs.

In the situation under consideration (represented by the point C), the economic subject is on the line of the identical expected revenue EX0. If the person was fully compensated in the case of damage caused by fire (without having spent any costs), his/her welfare would be represented by the point B. In such case, he/she would have the same assets in both situations – with or without fire (X_1' in S_1 and X_2' in S_2). However, since being insured against fire means that the person must pay the insurance, *the point B is unachievable.*

The case when a person insured himself/herself and the insurance is fair (the monetary value in a certain situation achieved through insurance is the same as the expected revenue of the risky prospects) is represented by the point E, which lies on EX0 just like the point C. In that case, the person pays the insurance in the extent X_1' — X_1^* and achieves a certain level of welfare – the point E lies on the certainty line, providing him/her with X_1^* in the situation S_1, and X_2^* in S_2. (In case of loss, he/she will get a compensation in the extent X_1^* — X_2'' from the insurance company).

Would a person like to insure himself/herself under these conditions? If he/she has aversion to risk, his/her indifference curve is convex in the origin, and it has the slope 1/2 in the intersection with the certainty line. It therefore touches the line of identical expected revenue at a point through which runs the certainty line. The indifference curve running through the point E (curve U1), which represents the certain amount of assets achieved through insurance, lies above the indifference curve running through the point C (curve Uo), which represents the risky prospects of an uninsured person. He/she thus insures himself/herself if the insurance is fair. We can also see in the chart that the person is willing to insure himself/herself even if the insurance is higher than the fair insurance (determined for instance by the point A, i.e., the extent X_1'—X_1) – see the idea of maximum insurance mentioned above.

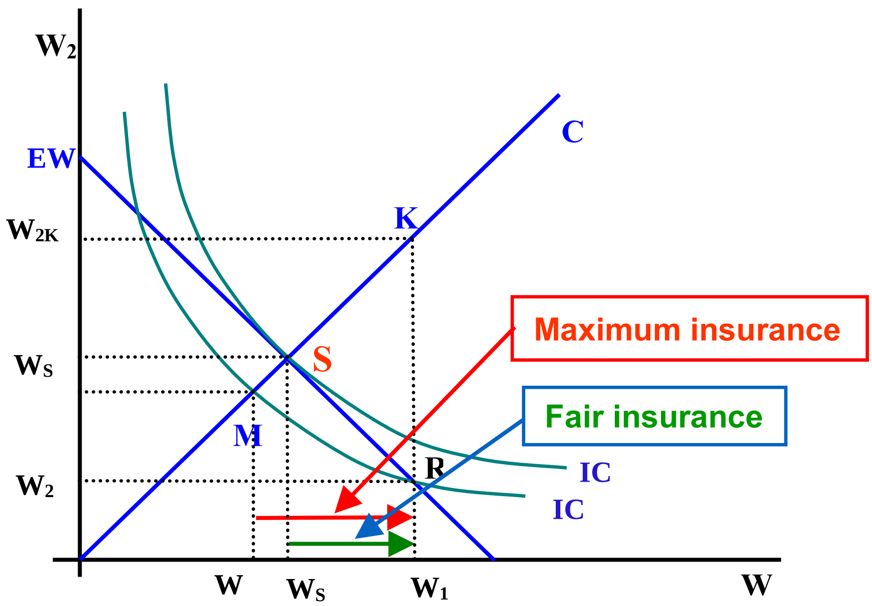

Fig. 22 Choice of insurance

3.5.2 Diversification

Diversification allows the reduction of risk if the effort can be split among various activities the results of which are not closely related. Insurance companies reduce the risk by insuring a lot of independent individuals, investors by investing their funds into many independent investment projects.

3.5.2.1 The Relation Between Risk and Revenue

Imagine a person with a certain initial level of welfare who has the opportunity to increase this welfare through two variants a) with risk (R) and b) without risk, or certain (C). He/she can choose a combination of these two options or only one of these. As we will see, this situation is analogous to the issue of splitting the consumer's income to purchase two goods.

The result of the variant with no risk will be labelled **X_J** and the expected result from the risky variant **EX_R**. **X_p** ***represents the revenue of the portfolio resulting from the combination of the certain and risky variant.*** When choosing, all possible results and their probability are known but which one of them will occur remains unknown. The risky variant must have higher expected revenue than the one without risk (**$EX_R > X_j$**); people with aversion to risk would otherwise choose only the certain variant.

The level of risk is represented by the range of probability or variability. It is assumed that large differences (whether positive or negative) among the various possible results and the expected result (called deviations) indicate a greater risk. ***The variability of results can be quantified by the standard deviation of revenues σ*** (see mathematical appendix).

Generally, the higher the revenue of an activity, the higher the risk as well. The person with negative attitude to risk therefore compares the expected revenue with risk.

All the possible combinations of revenues from the choice and the risk connected to these are graphically represented by the budget line, which, in this case, is the line with positive level of revenue – a (see mathematical appendix). The higher the revenue of a certain activity, the higher the risk usually is, too. The person with negative attitude to risk therefore compares the expected revenue with risk. All the possible combinations of revenues from the choice and the risk connected to these are graphically represented by the budget line, which, in this case, is the line with positive direction, because the riskiness or the standard deviation increases with the increase in revenue (see Fig. 22a).

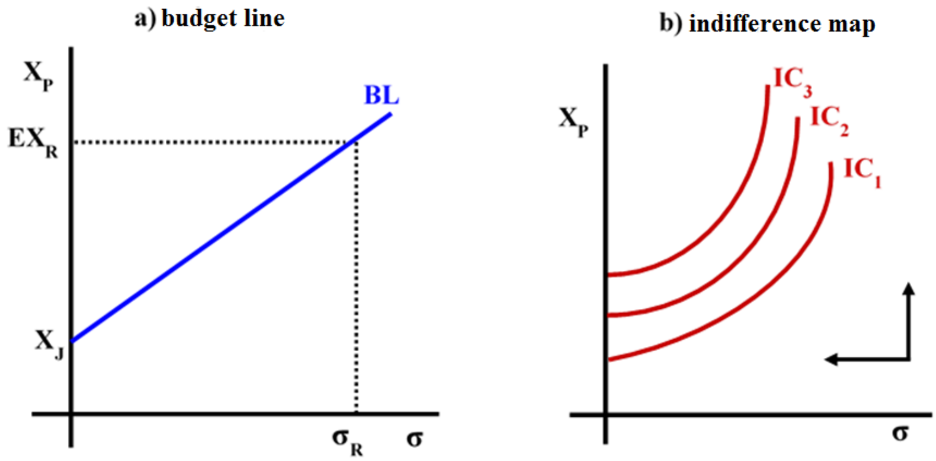

Fig. 23 Budget line and indifference map

If a person does not want to risk, he/she can choose the certain variant and get the revenue $\mathbf{X_J}$ ($\boldsymbol{\sigma}$ = 0). If he/she wants to get a higher expected revenue, he/she must undergo a certain risk. If he/she chooses the risky variant, he/she can get the expected revenue $\mathbf{EX_R}$, but he/she must undergo the risk represented by the standard deviation and $\boldsymbol{\sigma_R}$. If he/she chooses a combination of both of these variants, he/she gets a revenue in the extent **X_1 and EX_R** and undergoes a risk lesser than $\boldsymbol{\sigma_R}$.

If the choice depends only on the expected revenue and the standard deviation, the indifference curves can be used to illustrate the preferences regarding the revenue and risk.

In this case, the indifference curve represents the combination of risk and revenues that bring the same utility.

Since we are working with a negative attitude to risk, the higher expected revenue increases the utility, but the higher standard deviation decreases it. That means that the expected revenue represents “desirable goods”, and the standard deviation represents “undesirable goods”; the indifference curves thus have a positive direction as shown in Fig. 22b. The curve **IC_3** represents the highest level of satisfaction, IC_1 the lowest.

3.5.3 Optimal Choice of Portfolio Structure

As in the case when a consumer decides between two products, he/she will also optimise his/her choice by choosing such a combination of risk and revenue which corresponds to the point where the budget line touches the highest indifference curve (see point E in Fig. 23).

Optimal choice brings the expected revenue **EX*** and means undergoing a risk represented by the standard deviation **σ***.

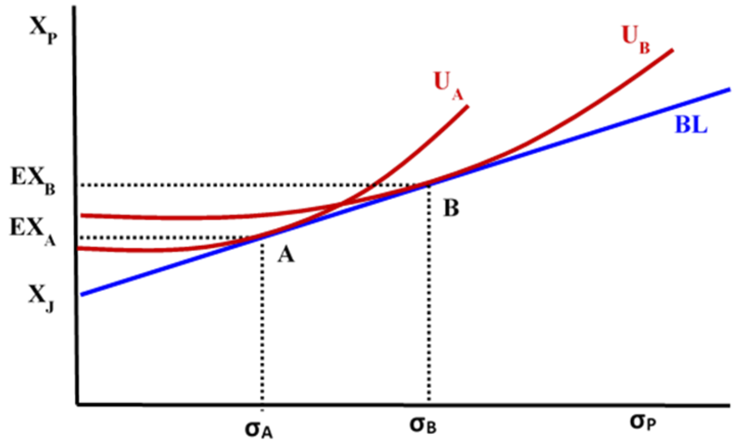

Fig. 24 Choosing between risk and revenue

People differ in their attitude to risk. Fig. 24 shows the differences in the choice of two people in the same situation (represented by a certain budget line). Consumer **A** is more worried about the risk than consumer **B**: the indifference curve **U_A** touches the budget line at the point of low risk (point **A**), because he/she will focus on an activity close to certainty. He/she will thus get a low expected revenue **EX_A***, which is only slightly higher than the revenue without risk. On the other hand, consumer **B** chooses to risk more and will get a higher expected result **EX_B**, but also a higher standard deviation **σ_B.**

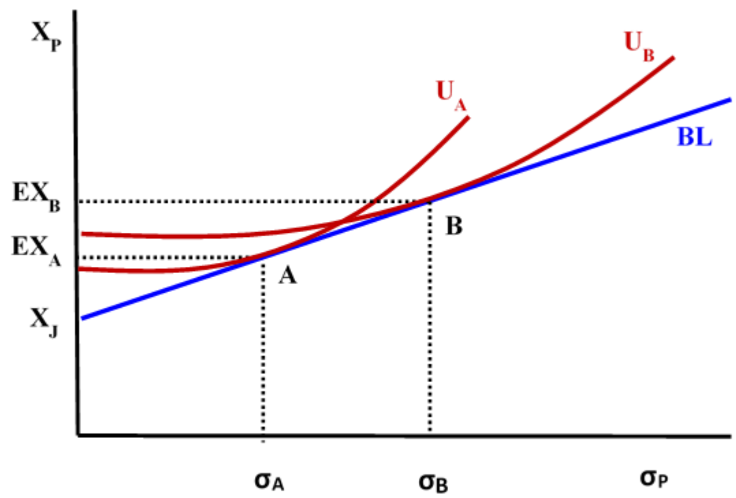

Fig. 25 The choice of two people with different attitude to risk

As in the simplified theory of consumer choice where we worked with only two goods, we will again simplify by choosing between only two options. The conclusions, however, also apply to a wider range of options.

Σ

In economic theory, risk and uncertainty represent key factors influencing the decision-making of individuals and firms. Risk occurs in situations where the possible outcomes of a certain decision are known, but which one will actually occur is unknown. In contrast, uncertainty arises when even the probabilities of possible outcomes are unknown, making decision-making significantly more complex.

Decision-making under risk requires quantifying the probabilities of various outcomes, enabling individuals and firms to analyse the potential consequences of their decisions. Probabilistic analysis models and the concept of expected utility are often used in decision-making under risk. Economists use expected value, a weighted average of outcomes where each outcome is assigned a probability. Expected utility extends this concept by considering not only the probability of occurrence but also the individual's preferences regarding risk. Some individuals may be risk-averse, meaning they prefer certainty even if a riskier option might offer higher expected returns. Risk-neutral individuals make decisions purely based on expected returns, while risk-seeking individuals prefer uncertain outcomes with potentially higher rewards.

Uncertainty complicates the decision-making process, as individuals and firms lack sufficient information about the probability of possible outcomes. In these cases, they often rely on subjective judgment or heuristics, which can lead to suboptimal decisions.

One tool to manage risk and uncertainty is diversification, which involves spreading risk across multiple investments or decisions. This approach reduces exposure to the negative effects of any single outcome. For example, in finance, investors often diversify their portfolios to minimise the risk of losses from any one investment.

Another approach to managing risk is insurance. Both firms and individuals can insure their activities to reduce the financial impact of adverse events. Insurance allows them to transfer part of the risk to another party (the insurer) in exchange for regular payments (premiums).

Risk and uncertainty are integral to everyday decision-making, whether in investment strategies, business decisions, or personal choices. Understanding risk management principles, including the use of tools such as diversification and insurance, can provide a sense of security when facing uncertain future outcomes and contribute to better decision-making.

?

Review questions

1. What does it mean if we say that the consumer has aversion to risk, is looking for risk or has neutral attitude to risk?
2. How would you explain these different attitudes to risk using the willingness to accept a fair bet?
3. Why do some people refuse to risk and some prefer it?
4. How would you explain the following statement: An individual maximises the expected utility?
5. Could you find some situations in which an individual could not maximise his/her expected utility?
6. Why do insurance companies behave more like subjects with neutral attitude to risk even though their managers have aversion to risk?
7. How does diversification reduce risk?
8. Imagine an investor chooses among three business activities whose probabilities and revenues are as follows (in thousands CZK):

PROBABILITY	0.2	0.4	0.4
REVENUE	100	50	-25

a) What is the expected value of the investment?

b) What is the variance?

9. The equation of income utility of the consumer Joe is expressed by the equation U(I) = 21-0.1 I2. Joe has the opportunity to participate in a game that can earn him 4 or 8 thousand CZK with the same probability.

a) What is the expected utility of this game for Joe?
b) What utility would Joe achieve if he would certainly win the same amount as the expected monetary value of the game?

Test questions

1. What is the main difference between risk and uncertainty?

a) Risk only applies to financial decisions, while uncertainty applies to all others.
b) With risk, the possible outcomes are known; with uncertainty, even the probabilities are unknown.
c) Risk only applies to companies, while uncertainty only applies to individuals.

2. What is expected value?

a) The most probable outcome of a decision.
b) The weighted average of outcomes, with each assigned a probability.
c) The maximum possible profit from a decision.

3. How does a risk-averse individual behave?

a) Prefers certainty, even if the risky alternative could yield a higher expected return.
b) Makes decisions purely based on expected returns.
c) Prefers uncertain outcomes with potentially higher rewards.

4. What is the main goal of diversification?

a) To increase potential profit.
b) To reduce exposure to the negative consequences of any one specific outcome.
c) To eliminate all risks.

5. What is the role of insurance in risk management?

a) To eliminate all risks.

b) To increase potential profit.
c) To allow transferring part of the risk to another party in exchange for regular payments.

Answers:

1b, 2b, 3a, 4b, 5c

Practical task

1. Propose an investment portfolio for a small investor who has €10,000 available. Explain how you would use the principle of diversification to minimise risk. Describe which types of investments you would choose (e.g., stocks, bonds, real estate) and in what proportions. Justify your decision and explain how this strategy helps to manage risk and uncertainty.

2. Imagine you are the owner of a small café. Identify three main risks that your business faces (e.g., coffee price fluctuations, changes in customer preferences, economic recession). For each risk, propose a strategy to mitigate or insure against it. Explain how you would use concepts from risk and uncertainty theory (e.g., diversification, insurance) to protect your business. Write a brief report (300-500 words) describing your analysis and proposed strategies.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. Mikroekonomie. 5. aktual. vydání. Praha: Management Press, 2010. s. 117 – 154. ISBN 978-80-7261-218-5. 574 s.

[2] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005.

[3] MACÁKOVÁ, Libuše; SOUKUPOVÁ, Jana. *Mikroekonomie – repetitoriu*. Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[4] HOLMAN, Robert. Ekonomie. Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[5] HOLMAN, Robert. *Mikroekonomie.* Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[6] JUREČKA, V. a kol., 2018. Mikroekonomie. 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[7] SOUKUPOVÁ, J., 2018. Mikroekonomie. 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 4

Corporate Behaviour, Selection of a Technology, Costs and Revenue of a Firm

After studying this chapter, you will be able to:

- explain the basic assumptions for the analysis of a firm;
- characterise the selection of a technology and production in both the short run well as the long run;
- explain the costs of a firm in both the short run and the long run and explain the relation between them;
- define the revenue of a firm.

Keywords:

Firm, production and technology, production function, costs (total, variable, fixed, average, marginal), revenue (total, average, marginal), profit.

After studying this chapter, you will be able to

- Differentiate between total, average, and marginal revenue.
- Describe the impact of market type (perfect vs. imperfect competition) on a firm's revenue.
- Graphically illustrate total, average, and marginal revenue under different market conditions.
- Explain the relationship between demand elasticity and a firm's revenue.

Chapter preview

- This chapter focuses on a firm's revenue as a key aspect of economic analysis. It begins with a definition of revenue and its significance for profit maximisation. It then discusses the various types of revenue – total, average, and marginal – and their characteristics in conditions of perfect and imperfect competition. The chapter also examines the relationship between demand elasticity and a firm's revenue, including graphical representations and mathematical formulas.

Chapter objectives

- To provide a comprehensive view of a firm's revenue and its role in economic decision-making.
- To teach students how to analyse a firm's revenue in different market conditions.
- To develop the ability to interpret graphs and mathematical formulas related to a firm's revenue.
- To deepen understanding of the relationship between demand elasticity and revenue.

Estimated study time

- The estimated time required to study this chapter is 320 minutes.

4.1 Theory of the Firm

4.1.1 Basic Assumptions for the Analysis of a Firm

The main reasons for the institutional organisation of production in the form of a firm are usually:

- the advantages of team work,
- the reduction of costs related to signing contracts.

A firm is usually characterised as a subject specialising in production, i.e., transformation of resources (inputs) into goods (output). This means that a firm focuses on 3 main activities:

1. purchase of services of factors of production,
2. organisation of the transformation of these into output,
3. sale of output.

An important question related to the existence of a firm is its aim. Economists usually assume that ***the aim of the firm is to maximise profit***, i.e., the maximisation of the difference between revenue and costs. Economists consider ***economic profit***, which differs from accounting profit in terms of how costs are understood. ***Accounting profit*** is the difference between revenue and explicit costs, i.e., the costs that were actually spent on the purchase of factors of production owned by other subjects. ***Economic profit is the difference between revenue and economic costs, which are the sum of explicit and implicit costs.*** In other words, economic profit is accounting profit minus alternative returns from all sources owned by a given subject.

4.1.2 Selection of Technology

Corporate behaviour is limited mainly by the technological possibilities of production and financial capacity of a firm. The theory of the firm therefore deals firstly with the technological circumstances mentioned above related to the transformation of inputs into outputs and subsequently with the cost limitations.

To analyse the decisions of a firm whose main activity is transforming inputs into outputs, i.e., production, it is useful to create an abstract model that depicts the relationship between inputs and outputs as simply as possible. ***The production function*** serves as this model. It represents the relationship between the quantity of inputs used in production over a given period and the maximum volume of output generated by those inputs during that period.

The "traditional" inputs used in production are labour, land, and capital. Some economists also consider the entrepreneurial spirit as an "untraditional" input. For the purpose of our analysis, we will simplify the real situation and we will assume that goods X are being produced (with the output labelled Q) from two inputs – capital (K) and labour (L), sufficient for the realisation of X. Same as with the production of goods which is considered to be the flow of output, inputs are also considered to be flow in the production process.

- K/t = machine hours per unit of time,
- L/t = hours worked per unit of time.

After these simplifications, we can write the production function in the following form

$$Q = f(K, L)$$

where Q = output, K = input of capital per unit of time, L = input of labour per unit of time. A production function defined this way has the following properties:

a) it expresses the fact that an output can be produced through different combinations of inputs;

b) it shows the technological constraints of production because it depends on the level of technology;

c) it does not assume unnecessary or inefficient production processes, which stems from the emphasis on maximum output as per the definition, i.e., firms use the most efficient combination of inputs to create output.

If a firm uses the most efficient technology available, its output will depend mainly on:

- the amount of inputs used, and
- the efficiency of their use.

For further analysis of a firm's behaviour, the time horizon in which the firm operates is important.

Definition

The production function is an abstract model of production that captures the relationship between the quantity of inputs used in production and the maximum volume of output that these inputs can generate in a given period. For simplification, two main inputs are often considered: capital (K) and labour (L), with output (Q) being a function of these inputs: Q = f(K, L). The production function expresses that output can be produced using various combinations of inputs, illustrates the technological constraints of production, and assumes the use of the most efficient combination of inputs. This *function primarily depends on the quantity of inputs used and the efficiency of their utilisation, with its specific form differing in the short run and long run.*

The short run (SR) is characterised as a period in which the services of at least one factor of production used by a firm are fixed due to prior choices. In the case of two factors of production, capital is typically considered the fixed input because it exists physically, for instance, in the form of machinery that is fixed at a specific location. A firm can either own or lease this capital but cannot change its volume to alter output. In contrast, the volume of labour involved in the production process can be easily adjusted—reduced or increased—if necessary, usually through short-term employment contracts. Therefore, we consider labour a variable input in the short run.

Since ***in the short run*** there is ***at least one fixed input – in this case, capital –*** the relationship between input and variable output at a given level of capital is characterised by a short-run production function. This function illustrates how output changes in response to changes in one input, namely labour. Thus, ***the returns from this single variable factor of production define the production function in the short run***.

The long run (LR) is a period sufficient for changes in the amount of all inputs used, meaning that all inputs are variable. In the long run, a firm can mutually substitute between the two inputs employed. A long-run production function illustrates the relationship between changes in the volumes of both inputs and the subsequent change in output. When we focus on a proportional increase in the volume of all inputs and the corresponding change in output within the long-run production function, we are dealing with returns to scale. The key characteristics of a production function in the long run are:

a) the substitution of inputs;
b) returns to scale of inputs.

Definition

The short run is characterised by the fact that at least one factor of production (usually capital) is fixed, while the others (typically labour) are variable. Therefore, in the short run, output changes as a result of changes in only one input, leading to returns from only one variable factor of production. In contrast, the long run is a period sufficient for all used inputs to be changed, meaning that all inputs are variable. In the long run, a firm can mutually substitute inputs, and the production function captures the relationship between changes in the volumes of both inputs used and the subsequent change in output. The key characteristics of the production function in the long run are input substitution and returns to scale.

4.2 Production in the Short Run (Short-run Production Function)

When analysing production in the short run, we assume that the amount of capital used is constant and the volume of labour and the size of output change. If we label the constant amount of capital as K_1, the short-run production function can be expressed as

$$Q = f(K_1, L)$$

Other concepts crucial for the analysis of production of a firm are total, average, and marginal product. The ***total product*** (TP) represents the ***output produced through the given inputs*** (i.e., TP = Q). Since its size is expressed in physical units, it is sometimes referred to as the ***total physical product*** (TPP). ***The curve of the total product shows various levels of output that can be produced by combining various amounts of variable input with constant amount of fixed input*** (assuming technology is invariable).

We will use a concrete example for the general production function, for instance:

$Q = 7L + 3l^2 – (1/3) L^3$

Average product (AP) represents the ***output per unit of input.*** Its size can be determined by dividing the total output by the number of inputs used for its production. The average product of a variable input of labour is $AP_L = \frac{Q}{L}$

For our production function, it is: $AP_L = \frac{7L + 3L^2 - (1/3)L^3}{L} = 7 + 3L \ \ (1/3)L^2$

The average product of fixed input of capital is defined as $AP_K = \frac{Q}{K_1}$

Marginal product (MP) represents the ***change in total product as a result of change in input by one unit, assuming that the amount of other inputs is constant***. If the changes in the variable input are very small, the marginal product can be expressed as the first partial derivation of the production function based on the variable input.

Marginal product of labour is defined as follows $MP_L = \frac{\delta Q}{\delta L}$

For our function $\boldsymbol{Q = 7L + 3l^2 – (1/3)\,L^3}$,

the marginal product of labour has the following form $\boldsymbol{MP_L = \dfrac{\delta Q}{\delta L} = 7 + 6L - L^2}$

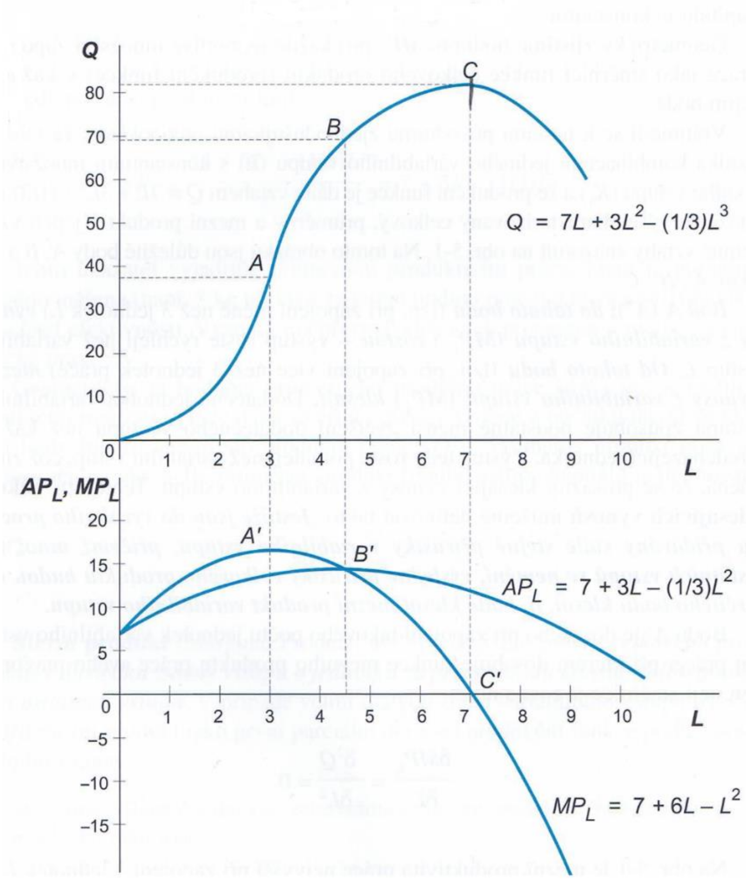

Fig. 26 Total, average, and marginal product

The marginal product of capital in the short run is not defined, because the amount of capital is constant.

Geometrically, we can find the value of MP_L for each individual amount of labour involved as the direction (slope) of the function of total product (production function) at all its points.

If we return to our original simplified assumption that output is produced by combinations of one variable input (L) and the constant amount of fixed input (K_1) and that the production function is $Q = 7L + 3L^2 — (1/3)L^3$, we can depict the total, average, and marginal product characterised above and their mutual relations in Figure 25. The points A, B, and C are important in this picture.

Point A (A'): the returns from the variable input (MPL) increase up to this point (i.e., employing less than 3 units of L) – output increases faster than the variable input L. From this point (i.e., employing more than 3 units of labour), the marginal returns from the variable input (MP_L) decrease. The additional unit of variable input causes significantly smaller increase in additional output. Output thus increases more slowly than the variable input, which means that diminishing returns from variable input are gaining ground. This so-called ***law of diminishing returns can be defined as follows: If we keep adding the same increments of the variable input in the productive process while holding all other inputs constant, the total increments of total product will start to decrease from a certain point, i.e., the marginal product of variable input will decrease***.

The point A' is reached if such a number of units of variable input of labour is involved so that the function of marginal product reaches its peak, i.e., its slope equals zero:

$$\frac{\delta MP_L}{\delta L} = \frac{\delta Q}{\delta L^2} = 0$$

In Fig. 25, the marginal productivity of labour reaches its peak when 3 units of L are involved.

Point B (B'): It is the point where the **average product of variable factor (AP_L) reaches its peak.** When less than 4.5 units of L are involved, the average product of labour increases, when more than 4.5 units of L are involved, the average product of labour decreases. The downward sloping curve AP_L intersects the MP_L curve at its peak, which is caused by the general relations between the marginal and average quantities.

Point C (C'): At this point, the **maximum output** is achieved. Any further increase in the variable input decreases the total product. At the given point, the slope of the production function equals zero and since the slope of the production function geometrically expresses the function of marginal product, the marginal product also equals zero. The MP_L curve intersects the x-axis when 7 units of L are involved, which ultimately lead to maximum output.

The shape of the short-run production function in Fig. 25 reflects the situation in production when with a small amount of variable input of labour (0-3 units of L), the efficiency of additional unit of labour increases. When more than three units of labour are used, efficiency – reflected in the marginal product of labour – begins to decrease. In other words, the described production function comprises two components. The first one mainly consists of the returns from variable input, the second one of the diminishing returns from variable input. In the next chapter, we will explore how the short-run production function and unit product functions (average and marginal) would look like if only increasing, diminishing, or constant returns from variable inputs were present.

4.2.1 Production Phases in the Short Run

Figure 26 shows the levels of variable input at which a firm is most likely to benefit from production.

When less than 4.5 units of labour are involved, we speak about **production phase I.** From the firm's perspective, this phase is positive: the average product increases the whole time and it can serve as a sufficient criterion for efficiency because it expresses the efficiency of all units of labour involved (as opposed to the marginal product, which expresses only the efficiency of the additional unit of labour).

The efficiency of the fixed input increases in this phase because $AP_K = Q/K$ and the output increases while the amount of capital is constant.

The efficiency of the variable input also increases. With less than 3 units of labour involved, the output increases faster than the volume of labour, so the marginal productivity of labour is increasing. From 3 to 4.5 units of labour involved, the law of diminishing returns from variable input manifests itself, however, the marginal product of labour stays positive and higher than the average product of labour. This development of marginal productivity of labour is reflected in the development of its average productivity ($AP_K = Q/K$), which increases up to 4.5 units of labour.

A firm in this production phase tends to increase the number of units of variable factor involved, because in this situation, when its fixed inputs are probably not fully employed, the increase in the number of units of labour increases the efficiency of both inputs. It aims to produce an output corresponding to 4.5 units of variable input involved, i.e., reaching the border between phase I and phase II, where AP_L reaches its peak.

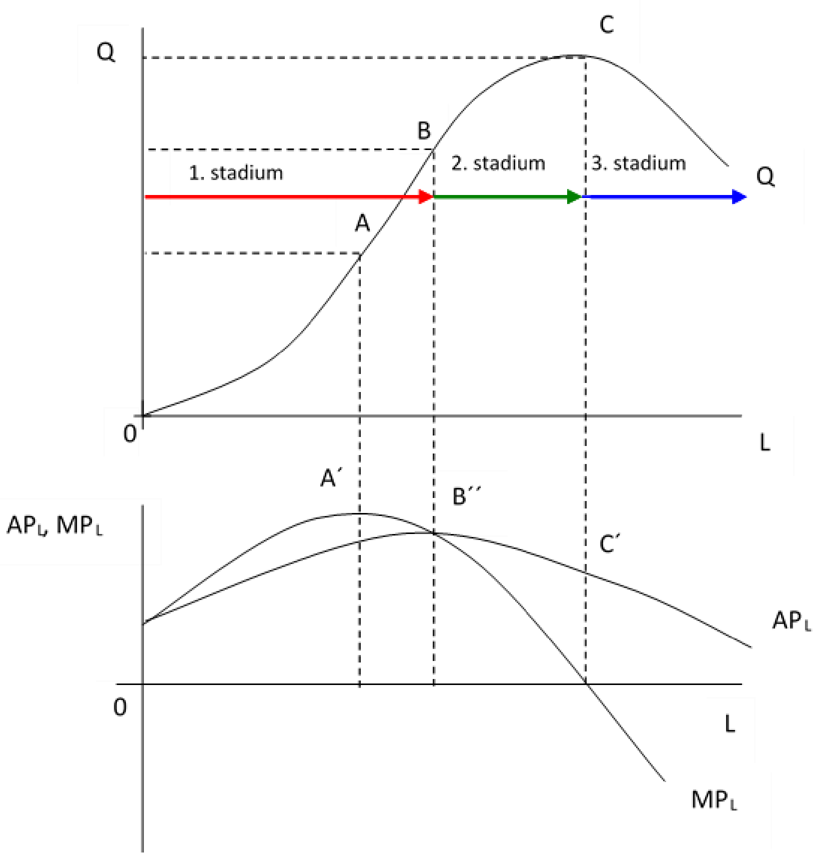

Fig. 27 Production phase in the short run

Production phase II represents the increase of output from the point B to the point C on the production function, caused by the increase in variable input from 4.5 to 7 units of labour. The efficiency of the fixed input in this phase increases because the volume of output increases and the amount of the fixed input K is constant. The efficiency of the variable input decreases because, even though the output increases, its rate is slower than for variable input (i.e., the marginal product of labour decreases but remains positive). The additional unit of labour in this phase thus increases the efficiency of capital but decreases the efficiency of labour.

On the border between phase II and III, the **short-run output reaches its peak at point C** as well as the efficiency of fixed inputs peaks at this point.

Production phase III occurs when more than 7 units of labour are involved; the increase in labour involved causes the decrease in output, which decreases both AP_L and AP_K efficiency. In economic reality, this is a situation in which there is a large amount of variable input relative to the given amount of fixed inputs. **The marginal product of labour in this phase takes on negative values.**

When finding out which phase is optimal for a firm, it is obvious that it is not phase III – a firm that would increase the number of units of variable input when output decreases would not behave rationally.

A negative feature of phase I is the relatively low use of fixed input; a firm thus usually wants to increase the efficiency by increasing the number of units of variable input and involving them in production in order to move from phase I to phase II.

The optimal stage can thus be considered to be phase II, during which the highest efficiency is achieved, and which culminates in maximum short-run output.

4.2.2 Short-run Production Functions – Types of Returns from Variable Input of Labour

The production function mentioned above with first increasing and from a certain point diminishing returns from variable input is considered to be typical. Other types of production function may exist both in theory and in reality.

4.2.2.1 Increasing Returns from Variable Input

As Fig. 27 shows, a situation may occur in which each additional unit of labour is more efficient (productive) than the previous unit. This is reflected in the output which increases faster than the variable input, which can be formally expressed as

$$Q = a+bL+cL^2,$$

where Q = volume of input, L = variable input of labour, a, b, c = constants.

In this case, these are the increasing returns from variable input of labour. If we assume that no output can be produced without the minimum volume of units of L involved and with a = 0, we get a modified production function $Q = bL+cL^2$

The average product of labour is then

$$AP_L = \frac{Q}{L} = \frac{bL + cL^2}{L} = b + cL$$

The marginal product of labour has the following mathematical form

$$MP_L = \frac{\delta Q}{\delta L} = b + 2cL$$

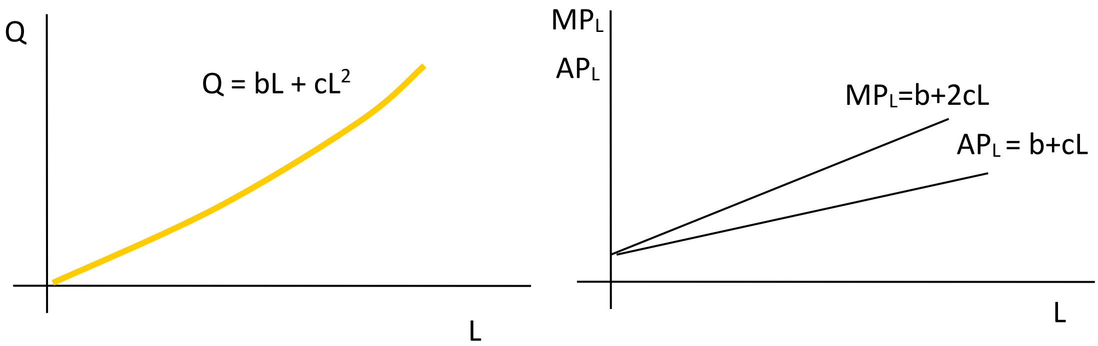

Fig. 28 Increasing returns from variable input

The slope of the MP_L curve (= 2c) is twice the slope of the AP_L curve (= c). Since $MP_L > AP_L$, the increasing MP_L "pulls up" the APL curve, too. This case does not occur very often in reality; when it does occur, it usually happens with a small number of units of variable input involved.

4.2.2.2 Diminishing Returns from Variable Input

Figure 28 represents a situation in which the output increases as a result of the additional involvement of labour, but it increases slower than the variable input. This shape of the production function shows the diminishing returns from variable input (labour). The only way of expressing this situation formally is the equation

$\boldsymbol{Q = a + bL - cL^2}$, or if $\boldsymbol{a=0}$, then $\boldsymbol{Q = bL - cL^2}$

The average product of labour is then

$$AP_L = \frac{Q}{L} = \frac{bL - cL^2}{L} = b - cL$$

The marginal product of labour can be expressed as

$$MP_L = \frac{\delta Q}{\delta L} = b - 2cL$$

The MP_L curve (with the slope of -2c) decreases twice as fast as the AP_L curve (with the slope of -c). Concurrently, as $MP_L < AP_L$, the decreasing MP_L "pulls down" the AP_L curve, too.

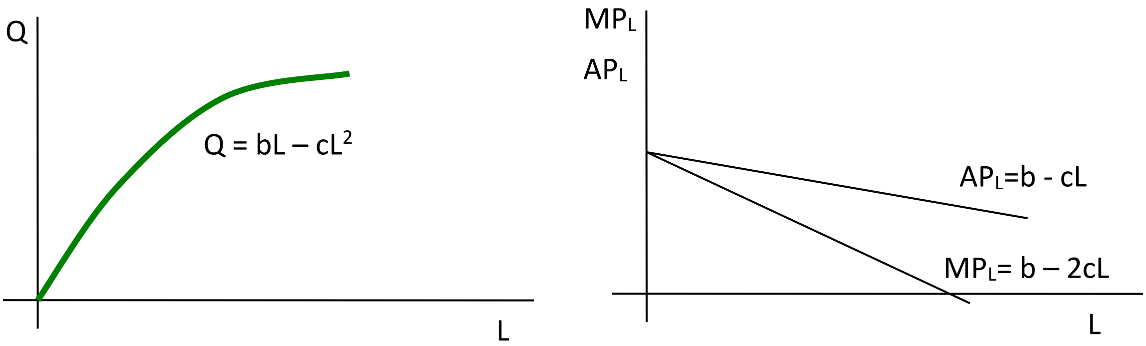

Fig. 29 Diminishing returns from variable input

4.2.2.3 Constant Returns from Variable Input

Another alternative are ***constant returns from variable input*** (Fig. 29), which show that with the increase in variable input, the output increases at a constant rate. Such production function can be expressed as follows

$$\boldsymbol{Q = a + bL}, \text{ or if } \boldsymbol{a=0}, \text{ then } \boldsymbol{Q = b^2}$$

If $a = 0$, then $Q = bL$

The average product of labour is then $AP_L = \frac{Q}{L} = \frac{bL}{L} = b$

Marginal product of labour $MP_L = \frac{\delta Q}{\delta L} = b$

Since each additional unit of L contributes to the increase in output just as much as any previous unit of L, all units of L are equally productive, i.e., the marginal product of all units of L is the same. This is reflected in the horizontal curve MP_L. If each unit of L is equally productive, the average product of labour is then constant and, at the same time, its amount is equal to the amount of the marginal product of labour.

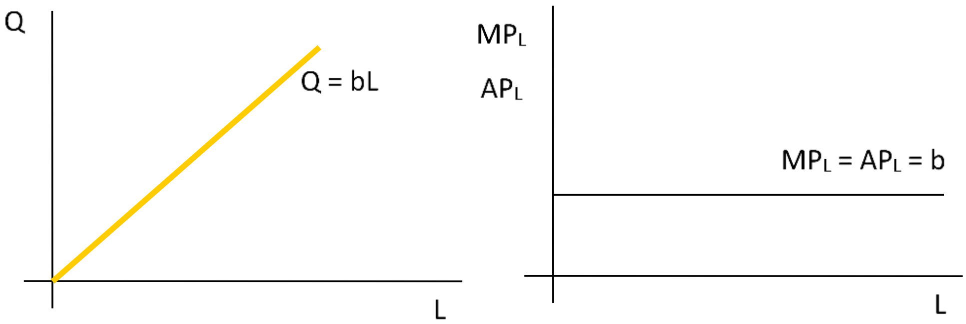

Fig. 30 Constant returns from variable input

Most of the microeconomic analyses assume that in the production process, with a relatively small involvement of variable input, the increasing marginal productivity of this input comes in first. If more units of variable input are involved, the fixed capital stock hinders further increase in the marginal productivity of labour. Returns in the form of marginal product are diminishing. The production function expressing such character of production can be expressed as

$$Q = a + bL + cL^2 - dL^3 \text{ and if } a = 0 \quad Q = bL + cL^2 - dL^3$$

$$AP_L = \frac{Q}{L} = \frac{bL + cL^2 - dL^3}{L} = b + cL - dL^2$$

$$MP_L = \frac{\delta Q}{\delta L} = b + 2cL - 3dL^2$$

4.2.3 Production in the Long Run (Long-run Production Function)

Production in the long run is characterised by the fact that the firm can change the amount of all inputs used in production. In our limited case of two inputs (labour and capital), the long-run production function can be expressed as follows: $Q = f(K, L)$

For graphical representation of the long-run production function, the isoquant map is used. In the following text, we will focus on two basic determinants of the long-run production function, i.e., the ***substitution of inputs*** and ***returns to scale.***

4.2.3.1 Isoquant

A long-run production function can be represented using a three-dimensional image. This image allows us to represent the changing amount of labour and capital on two axes of the horizontal plane and the amount of output corresponding to the various combinations of inputs used for its production on the vertical axis.

The curve made up by all combinations of inputs resulting in the production of equal output is called an isoquant (iso = equal, quant = quantity). An isoquant is always connected with a concrete level of output: for instance, the isoquant C'D' could represent the output Q_1 = 10 buns, which could be produced through various combinations of labour and capital. If we label the level of output for instance as Q_1, the isoquant is the set of combinations of inputs K and L satisfying the relation

$$f(K, L) = Q_1$$

The level of output increases as we shift upwards to the right: for instance the isoquant A'B' could represent the output Q_2 = 20 buns. Since there can be an infinite amount of outputs, there is also an infinite number of isoquants, which creates the so-called isoquant map. The isoquant map informs us about the maximum achievable output produced through any combination of inputs.

Definition

An isoquant is a curve that represents all combinations of inputs (typically labour and capital) leading to the same quantity of output. In a spatial representation of the long-term production function, the isoquant lies on a horizontal plane, where the axes represent the quantities of labour and capital, while the vertical axis indicates the level of output. Mathematically, an isoquant can be expressed as a set of input combinations K and L satisfying the relation $f(K, L) = Q$, where Q is a given level of output. The output level increases as the isoquant shifts upward to the right, with infinitely many isoquants forming what is known as an isoquant map. The isoquant map provides information about the maximum achievable output for any combination of inputs in the long run.

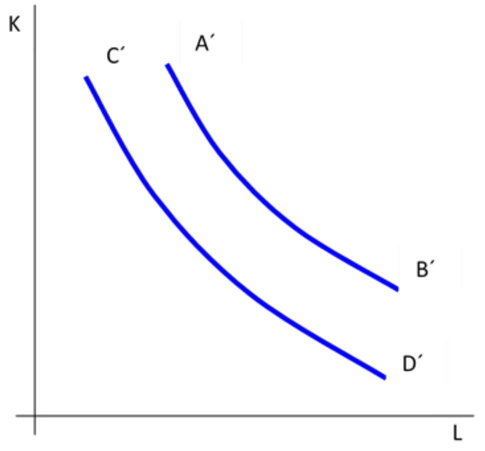

Fig. 31 The representation of production using the isoquant

The isoquant plays the same role in the production theory as the indifference curve in the consumer theory. The indifference curve shows various combinations of two goods that result in the equal total utility for a consumer; the isoquant shows various combinations of two inputs that result in the equal output of a firm. These two curves are also similar in their ordering: indifference curves order the level of satisfaction of needs in a northeast direction from the lowest to the highest, isoquants order the level of output likewise. The difference between an isoquant and an indifference curve is that each isoquant is connected with a specific level of output, but it is not always possible to link an indifference curve to a concrete level of utility (the indifference curves are meaningful from the ordinal perspective).

We can simplify it by representing the isoquants as smooth curves or more precisely continuous functions. The technology used in economic reality usually does not allow the existence of an infinite number of combinations of inputs for the production of an equal output. If an isoquant was supposed to accurately reflect the real economic processes, it would rather be a series of points leading to the specific level of output than a continuous curve.

Basic properties of isoquants:

- the individual isoquants are ordered in a northeast direction in the isoquant map, i.e., the isoquant closer to the origin represents the combinations of inputs leading to lower output than the isoquant located farther from the origin,
- isoquants have cardinal ordering,
- isoquants do not intersect, if they did, the assumption of efficiency that comes with the production function would be broken, because different outputs could have been produced through the same combination of inputs. That would mean that the firm uses the factors of production inefficiently.
- isoquants are downward sloping and convex to the origin. To explain this, let's go back to the production plane in the shape of a production hill and imagine we are looking at this hill from the top. We would see an isoquant map in the form of concentric curves.

4.2.3.2 Marginal Rate of Technical Substitution

The downward sloping direction of the isoquant is the consequence of the first property of the long-run production function mentioned above, i.e., substitution of inputs, which means that a firm can decrease the volume of one input and increase the quantity of another input without changing the amount of output. The slope of the isoquant is important mainly because it provides us with the information about the technical capabilities of mutual substitution of inputs. The rate of such substitution of inputs is referred to as the marginal rate of technical substitution.

The marginal rate of technical substitution (MRTS) ***expresses the extent to which the firm can substitute capital with labour without changing the amount of output.***

If, in the production, a relatively large amount of capital is substituted by a relatively large amount of labour with equal output (i.e., we move along one isoquant), the MRTS expresses the direction of the isoquant between the two points

$$MRTS = \frac{-\Delta K}{\Delta L} \bigg|_{Q = Q_1}$$

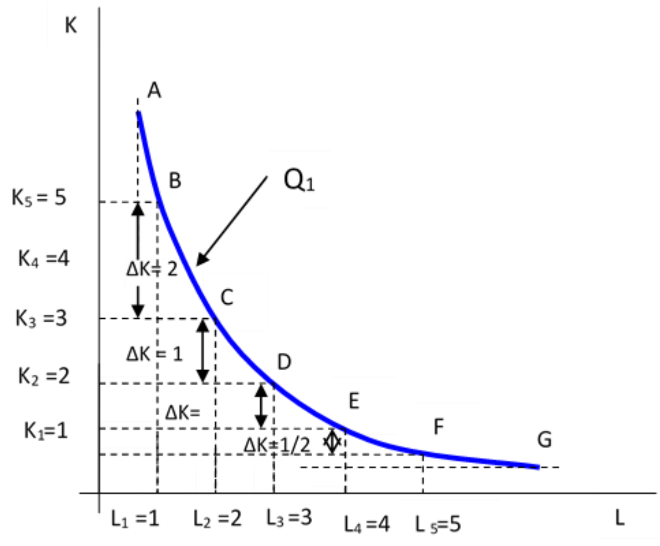

Fig. 32 Isoquant

For very small changes in the amount of capital and labour without any change in the output, MRTS is expressed as the slope of the isoquant at any of its points

$$MRTS = \frac{-dK}{dL}\Bigg|_{Q = Q_1}$$

We need to specify that these relations express the fact that the firm changes the inputs along the isoquant by ***decreasing capital*** (therefore ***- ΔK***, or ***-dK***) and increasing labour, i.e., this is the ***marginal rate of substitution of capital by labour***.

Imagine what will happen if we move along the isoquant Q_1 in Fig. 31 from the point B to the point C. The amount of output will remain the same because we stay on the same isoquant. However, the combination of inputs used for the production of this output changes; K_3L_2 is used instead of K_5L_1. The shift from K_5L_1 to K_3L_2 means that the decrease in output caused by the involvement of a smaller amount of capital is compensated by the increase in output produced with more labour. If we express the decrease in output caused by the decrease in capital as

$$-\Delta K \, . MP_K$$

and the increase in output caused by the increased involvement of labour as

$$\Delta L \, . MP_L$$

then the compensation mentioned above means that

$$-\Delta K \, . \, MP_K = \Delta L \, . \, MP_L$$

After transformation we get

$$-\frac{\Delta K}{\Delta L} = \frac{MP_L}{MP_K}$$

Since

$$-\frac{\Delta K}{\Delta L} = MRTS$$

It is true that $$MRTS = \frac{MP_L}{MP_K}$$

For very small changes in inputs, it is true that

$$-\frac{dK}{dL} = MRTS = \frac{MP_L}{MP_K}$$

The negative dK/dL is caused by the substitution of one input by another and by the negative slope of the isoquant. We usually add a negative sign to the K/L ratio in order to get a positive marginal rate of technical substitution (a similar approach was used in chapter 2 with indifference curves – MRSC analysis). However, from an economic perspective, the size of this ratio is what matters most.

Figure 31 shows that if the firm increases labour to the detriment of capital and keeps the same amount of output (Q_1), the rate of mutual substitution of capital by labour decreases. The absolute value of MRTS is relatively high at point B, which is reflected in the fact that technological circumstances of production enable the firm to significantly decrease the capital used and substitute it by labour whose amount is greater than the amount of capital used by one unit. On the contrary, at point F, where the firm uses a relatively large amount of labour to produce output Q_1, it can give up only a small amount of capital in exchange for an additional unit of labour; MRTS is low. The marginal rate of technical substitution of capital for labour decreases from 2 to 1, then to 2/3, and finally to 1/2 at successive points along the Q1 isoquant. It may seem that this is caused by the relation between the amount of a given input and its efficiency, i.e., marginal productivity: MP_L decreases with increasing labour, and MP_K increases with decreasing capital. Because MRTS = MP_L/MP_K, the marginal rate of substitution of capital by labour along the isoquant must decrease.

However, the efficiency (marginal productivity) of a given input also depends on the amount the other input(s) used. The marginal product of labour is thus influenced not only by the amount of labour, but also by the amount of capital, and the situation is analogous for the marginal product of capital.

4.2.3.3 Optimal combination of inputs

Similarly to the indifference curves representing the various combinations of goods a consumer would like to purchase, regardless of any restrictions he/she might encounter, the isoquants represent various combinations of inputs a firm would like to buy, without being limited in any way. The firm faces the same problem as the consumer – the limitation of the price of the inputs used. The firm must solve the issue of choosing such a combination of factors of production that will enable the production of a given output (Q1) at minimal costs, since the production function is Q = f (K, L). Imagine a firm that can purchase various combinations of labour and capital at the given prices and within the given total costs. **The line containing all the combinations of labour and capital that can be purchased at the given total costs** is called the line of equal costs or ***isocost.*** The equation expressing the isocost is

$$TC = w\,L + r\,K,$$

where TC = total costs, L = the volume of the labour used, w = the price of the unit of labour, K = the amount of capital used, r = the price of the unit of capital. The expression for the isocost is an equation of a line (with variables K, L). Let's modify it into the slope-intercept form:

$$\frac{TC}{r} - \frac{w}{r} \cdot L = K$$

The expression w/r expresses the direction of this line (isocost); the minus sign indicates that it is downward sloping (see Fig. 32).

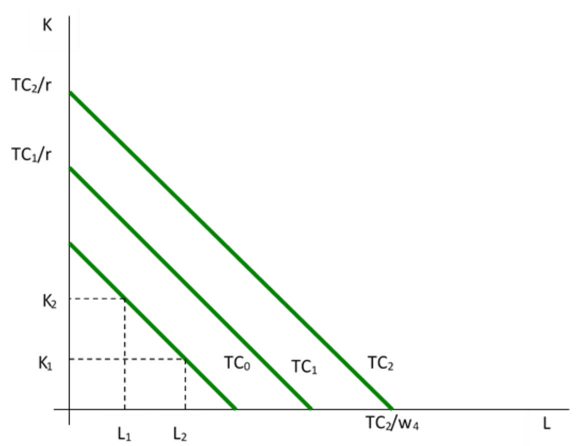

Fig. 33 Isocost

In the following text, we will work with minor changes in prices of labour and capital, and we will get the direction of the isocost using the first derivative with reference to L:

$$-\frac{dK}{dL} = \frac{w}{r}$$

The slope of the isocost is analogous to the slope of the consumer's budget line: while the slope of the budget line is set by the relative prices of two goods (Px/Py), the slope of isocost depends on the relative prices of inputs (w/r).

The slope-intercept form of the isocost equation shows that a change in total costs results in a parallel shift of the isocost, whereas a change in the price of just one input will alter the slope of the isocost.

For the **optimal combination of inputs**, the technical capabilities of substitution of capital by labour must be in accordance with the economic capabilities of this substitution. In other words, ***the rate at which the firm is technically capable of substituting capital by labour (MRTS) equals the rate at which it is capable of implementing this substitution on the market (w/r).***

This so-called **cost optimum** is defined as follows:

$$MRTS = \frac{w}{r}$$

Since we know that MRTS = MP_L/MP_K, we can also characterise the cost optimum of a firm as

$$\frac{MP_L}{MP_K} = \frac{w}{r}$$

This means that the firm should split its costs between labour and capital so that the ratio of their marginal products equals the ratio of their prices. Through the modification of the last relation, we get

$$\frac{MP_L}{w} = \frac{MP_K}{r}$$

This modified relation expressing the cost optimum shows that the firm should purchase inputs in such a way that it gets the same increase in output from each additional CZK 1 spent on the purchase of labour and capital. In other words, ***the firm minimizes its costs when the marginal product of each CZK 1 spent on inputs is the same across all inputs used.*** This rule is known as the ***least cost rule.***

Geometrically, the cost optimum is the point of contact between the isoquant and isocost, where their slopes are equal, i.e., when the MRTS equals w/r. When graphically representing the optimal combination of inputs, we encounter the so-called dual problem. In other words, the cost optimum of a firm can be represented in two ways (see Fig. 33).

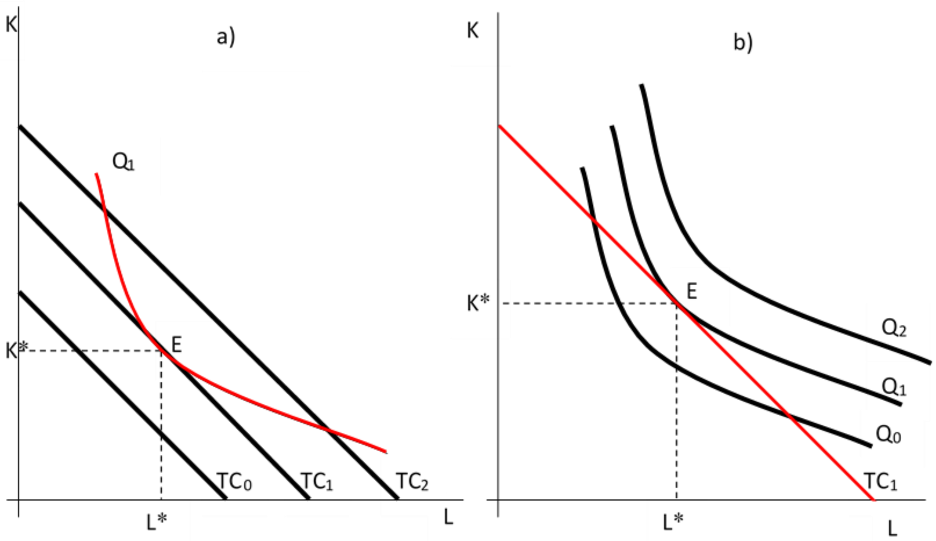

Fig. 34 Cost optimum of a firm

Figure 33a shows the situation where the firm aims to produce output Q1 and seeks to determine the minimum costs required for its production. In other words, it is the ***minimisation of costs*** in the production of a given output, represented by the combination K*L*.

Figure 33b shows the situation in which the firm tries to find out what is the maximum output that could be produced with the given level of its costs. The optimal combination K*L* represents a cost optimum in which a firm produces ***the maximum output with the given costs***.

4.2.4 Increasing Output Curve

Let us now assume there is excess demand in the market for one of the given goods, and the firm wants to increase its output. If we also assume that the price of inputs remains unchanged with changes in the quantity of inputs demanded (i.e., the w/r ratio is constant), the firm will identify the minimum costs required for producing each output. This is shown in Figure 34.

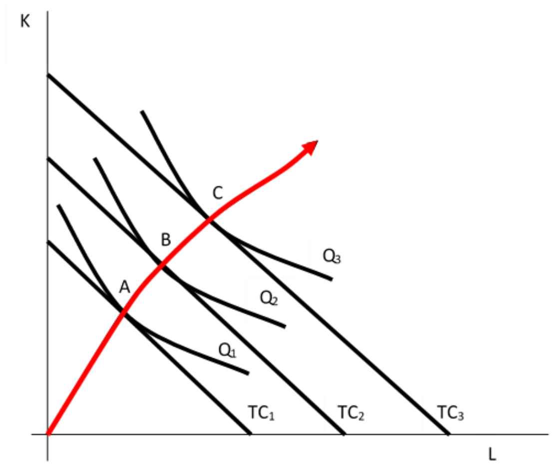

Fig. 35 Increasing output curve

If we connect the points A, B and C representing the minimum costs for the individual levels of output, we get the so-called curve of increasing output. The curve of increasing output thus represents a set of combinations of outputs with which the firm minimises its costs while producing various amounts of outputs (compare with income-consumption curve ICC).

To minimise costs, MRTS = w/r and the ratio of the prices of inputs is constant as a result of invariable prices of inputs, the marginal rate of technical substitution also remains constant along the curve of increasing output. The size of output is the only element that changes when moving in the north-easterly direction.

It is very likely that with the increasing output, the firm will purchase more of both inputs, which will be reflected in the positive direction of the curve of increasing output. We can draw the following conclusions from the shape of the increasing output curve (see Fig. 35):

a) the increase in output is connected to a relatively high involvement of capital (capital-intensive production);
b) the costs of increasing output are minimised by increasing both inputs in the same proportion;
c) the increase in output is connected with a relatively higher involvement of labour rather than capital.

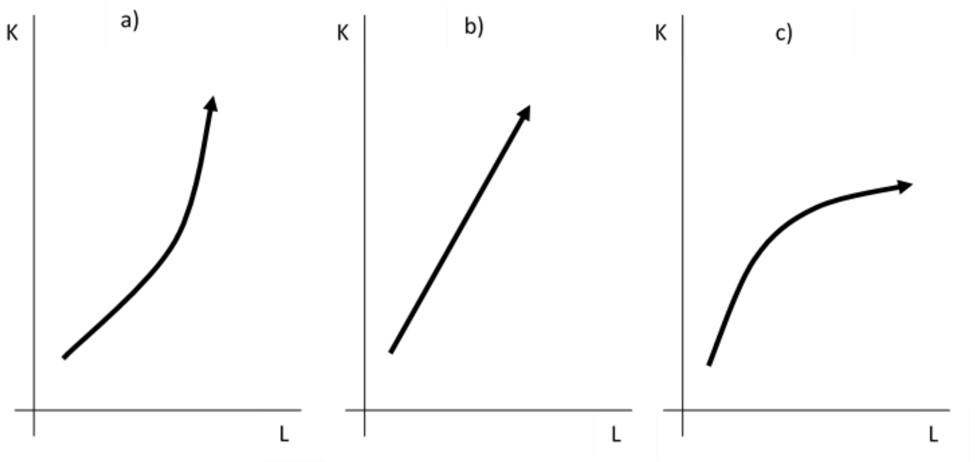

Fig. 36 Alternative shapes of the increasing output curve

4.2.5 Returns to Scale

Generally speaking, returns to scale express the relation between the changes in inputs and the change in output. **Returns to scale**, however, do not describe the whole long-run production function because they cover the links between the proportional change in inputs and the change in output caused by that. If we assume the production function Q = f(K, L) and multiply both inputs used by the same positive constant, for instance "t" (t > 1), three situations may occur:

A. An increase in the amount of both inputs by "t" percent causes the output to increase also by "t" percent. ***Constant returns to scale*** prevail in the long-run production function. In this case

$$f(tK, tL) = tf(K, L) = tQ$$

B. An increase in the amount of inputs used by "t" percent" causes the output to increase by more than "t" percent. In this case, we talk about ***increasing returns to scale.*** Formally

$$f(tK, tL) > tf(K, L) = tQ$$

C. An increase in the amount of both inputs by "t" percent causes the output to increase by less than "t" percent. In this case, the long-run production function exhibits ***diminishing returns to scale.*** More precisely

$$f(tK, tL) < tf(K, L) = tQ$$

If we take into account the long-run production function with n inputs, i.e., $Q = f(X_1, X_2, ..., X_n)$, and multiply all inputs by a positive constant "t", we get

$$f(tX_1, tX_2,...,tX_n) = t^c f(X_1,X_2 ..., X_n)= t^c Q$$

The value of the exponent k then signals the character of returns:

- **if k = 1,** constant returns to scale prevail in the long-run production function;
- **if k > 1**, we presume that the returns to scale are increasing;
- **if k < 1,** the long-run production function exhibits diminishing returns to scale**.**

4.2.6 Technical Advance

The production function reflects the technological constraints of production because it depends on the level of technology. The development of human knowledge leads to improvements in production processes. This raises the question of whether we can capture this technical advance in the production function. Firstly, we will explain it using the Figure 36.

The initial isoquant is labelled Q_1. If we assume that technical advance is applied in production, the output Q_1 can be produced with fewer inputs, causing the isoquant Q_1 to shift down and to the left, where it is labelled Q'_1. (That means that both Q_1 and Q'_1 represent the same amount of output.) The amount of capital K'_1, combined originally with the amount of labour L_1, enables the firm to produce the same output with less labour L'_1 thanks to the technical advance. The productivity of labour increased from Q_1/L_1 to Q_1/L'_1.

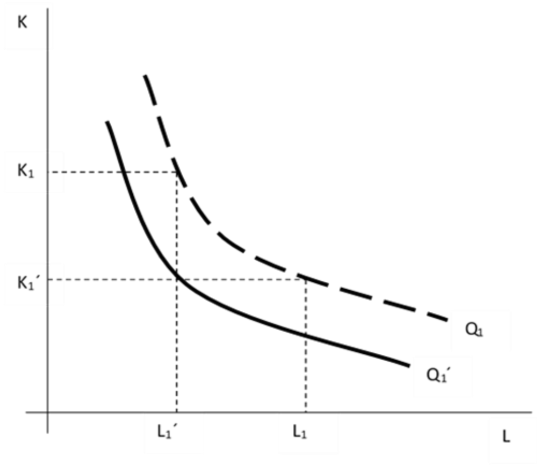

Fig. 37 Technical advance

If we look at the issue of technical advance from a very simplified formal perspective, we will use the production function

$$Q = A(t) \cdot f(K, L)$$

where A(t) represents the technical advance over time; we assume that **dA/dt >0.**

The influence of technical advance on the individual inputs may result in three cases:

1. ***Neutral technical advance*** which affects all the inputs equally

$$Q = A(t) \,.\, f(K, L).$$

2. ***Capital-intensive technical advance*** which affects only capital – using a new technology makes the capital more and more productive over time

$$Q = f[A(t) \,.\, K, L].$$

3. ***Labour-intensive technical advance*** which affects only the productivity of labour

$$Q = f[K.A(t) \,.L].$$

4.3 The Cost Structure of a Firm

Before we analyse the costs, let's recall the difference between the *economic* and *accounting* concepts of costs. In the narrower sense, ***accounting costs are the actual expenses incurred by the firm, the movement of which is recorded in the financial books.*** These are ***explicit costs.*** *The economic concept of costs is broader: economists take into account not only the explicit costs but also implicit costs.* ***Implicit costs*** are costs that the firm does not actually pay. The existence of these is based on the principle of alternative costs, i.e., opportunity costs. Implicit costs represent the costs the firm loses by using limited resources in a certain way and not any other. For better understanding, let's have a look at the specific differences between accounting and economic concepts of labour and capital costs (the inputs used in our analysis).

As for the ***labour costs***, there is little difference between the accounting and economic concepts, as both approaches consider these to be explicit costs. From an accounting perspective, labour costs are part of the actual expenses incurred by the firm. From an economic perspective, labour costs are determined by the wage rate, which forms part of the employment contract. It is assumed that this wage rate is equal to the best alternative return of this input for its owner.

Capital costs are perceived completely *differently* by accountants and economists. From an accounting perspective, capital costs are determined by the price of capital goods, which is used to calculate the specific share of capital costs in the total costs of a given output. These are thus the actual, ***explicit costs*** incurred. Economists, however, consider the capital costs to be implicit; their amount per hour is determined by the price anyone would be willing to pay for using the given capital goods if he/she rented it for an hour. The firm thus loses the alternative return from renting the capital goods to someone else, i.e., the rent, because it is the only subject using the capital goods. The capital costs are thus determined by the amount of rent based on the best alternative usage of the given capital goods.

When analysing the costs, we use the simplified situation in which the firm produces only goods X and uses only two inputs for the production, the prices of which do not change with the quantity purchased (in other words, we assume that there is perfect competition in the labour and capital markets), assuming completely homogeneous labour and completely homogeneous capital. Speaking of the prices of inputs, remember that these are considered to be flows in the production process. Based on what has already been said about labour and capital costs:

- ***the price of labour is the wage rate*** (w), the amount of money for one hour of work,
- ***the price of capital is the rent*** (r) corresponding to the amount of money for one machine hour. The firm can compare the price of capital to the interest it could have earned by depositing the money used to purchase the capital goods in a bank. This interest therefore represents the opportunity cost of owning the capital goods.

We should also mention the so-called ***sunk costs*** in the form of costs the firm cannot get back in any way. For instance, if the firm purchases a special device that could be used only for a given purpose with no other possible usage whatsoever. In this case, the ***alternative costs are zero.***

The starting point for the analysis of costs is the functional relation between the costs and the output per a unit of time. Since we know that the amount of output is a function of the inputs used and if we know the prices of the inputs used by the firm in the production process, we can calculate the costs of production for a specific output. The level and development of the costs due to changes in the output of the firm thus depend on two important factors:

1. The **character of the production function in question** (which determines the shape of the cost functions of the firm).
2. The **prices of inputs** (these determine the costs).

A cost function can be expressed as:

$$TC = f(Q, w, r)$$

If we assume that the firm behaves rationally, this cost function expresses the minimum costs of a firm to produce various amounts of output, using various combinations of labour and capital.

The character of costs in the short run is different in many ways from the character of costs in the long run. Since the firm cannot increase output in the short run by changing production facilities or technology, it can only increase output by adjusting the variable inputs, such as labour or raw materials. To differentiate between the short-run and long-run costs, we sometimes use the letters "S" and "L" and put these in front of the labels of individual costs (e.g. short-run total costs are STC and long-run average costs are LAC). We will first focus on the costs in the short run.

In most cases, the total costs increase together with the increase in output. We will therefore begin the analysis of costs by examining the links between the total costs and output, assuming that the prices of labour and capital are fixed. Later on, we will examine how changes in input prices affect the firm's costs.

Points to remember

When analysing the costs, we use the simplified situation in which the firm produces only goods X and uses only two inputs for the production, the prices of which do not change with the quantity purchased (in other words, we assume that there is perfect competition in the labour and capital markets), assuming completely homogeneous labour and completely homogeneous capital. Speaking of the prices of inputs, remember that these are considered to be flows in the production process. Based on what has already been said about labour and capital costs:

- the price of labour is the wage rate (w), the amount of money for one hour of work,
- the price of capital is the rent (r) corresponding to the amount of money for one machine hour.

4.3.1 The Costs of a Firm in the Short Run

The total costs TC can be defined as a sum of labour costs L and capital costs K.

$$TC = wL + rK$$

In the case of two inputs used, considering capital (at level K_1) to be the fixed factor of production in the short run, total costs can be defined as follows:

$$STC = wL + rK_1$$

Since the amount of capital is constant in the short run, the costs incurred for capital do not change with the increase in output. These are, therefore, the **fixed costs** (FC). Fixed costs exist even if the amount of output is zero and the firm does not use the capital at the given moment. This is caused by the fact that the firm must pay insurance, protection of facilities, etc., and it also spends the costs that change with the increase in output in the short run – **variable costs** (VC). Variable costs represent the costs of wages, materials, etc. If the output is zero, variable costs are zero as well. Total costs in the short run are the sum of fixed and variable costs

$$STC = FC + VC$$

In the graphical representation of the cost curves, the curve of fixed costs is a straight line parallel to the x-axis. ***The development of variable costs is therefore a significant element for the development of total short-run costs.*** The shape of the curve of variable costs reflects the returns from variable input.

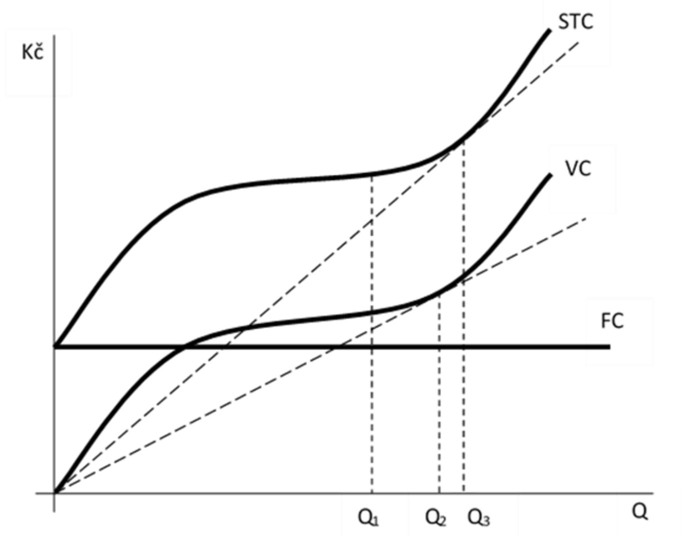

Fig. 38 Total costs in the short run

In the case of diminishing returns from variable input, the additional unit of labour creates lower increase in output than any previous unit, so with constant wage rate, the total costs will increase faster than the output. This applies to output greater than Q_1.

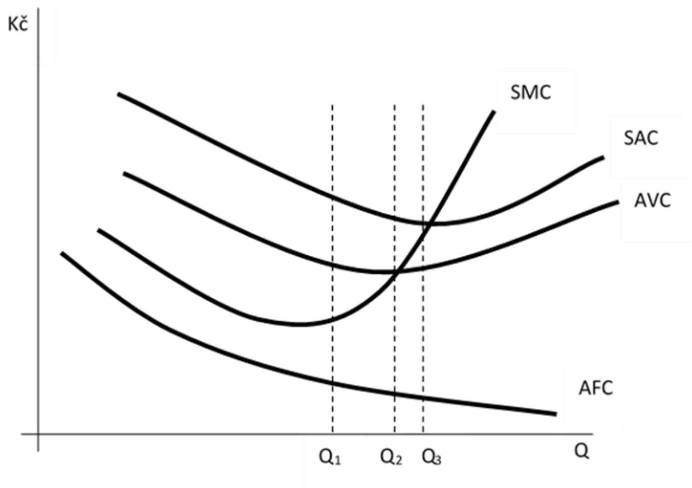

Fig. 39 Unit costs in the short run

Until reaching the output Q_3, the firm increasingly utilises its fixed capital capacity by hiring additional units of variable labour input, resulting in decreasing average costs. The fixed amount of capital, however, hinders further increase in the marginal productivity of labour, so the increasing output (higher than Q_3) is produced with increasing average costs. Assuming initially increasing and then decreasing returns from the variable input, and assuming constant input prices, the SAC curve takes a U-shape.

Another type of costs are **average costs** (AC), which are calculated by dividing the total production costs for a given output by the amount of that output.

$$SAC = \frac{STC}{Q}$$

The equation can be modified into the following form

$$SAC = \frac{STC}{Q} = \frac{FC + VC}{Q} = \frac{FC}{Q} + \frac{VC}{Q}$$

where FC/Q are average fixed costs,
VC/Q are average variable costs.

Average fixed costs (AFC) are fixed costs per unit of output.

$$AFC = \frac{FC}{Q} = \frac{rK}{Q} = r\frac{1}{AP_K} = \frac{r}{AP_K}$$

Average variable costs are variable costs per unit of output.

$$AVC = \frac{VC}{Q} = \frac{wL}{Q} = w\frac{1}{AP_L} = \frac{w}{AP_L}$$

The vertical summation of the AFC and AVC curves for each level of output results in the SAC curve because:

$$SAC = AFC + AVC.$$

The quantity of incremental costs is represented by the marginal costs (MC), defined as the increment of total costs caused by the increase in output by one unit.

$$SMC = \frac{dSTC}{dQ} = \frac{dVC}{dQ}$$

After modification, we get

$$SMC = \frac{dVC}{dQ} = \frac{wdL}{dQ} = w\frac{1}{MP_L} = \frac{w}{MP_L}$$

4.3.2 Costs and Returns from Variable Input

When analysing the individual costs in the short run, we worked with the fact that the production function reflects first the increasing and then the diminishing returns from the variable factor, as we already know, so we can distinguish the individual cases of production functions. Since the shape of the cost function is determined by the shape of the production function, we will try to find mutual links between them.

4.3.2.1 Costs under Increasing Returns from Variable Input

If we have the production function in the form

$$Q = a + bL + cL^2$$

and if $a = 0$, then $Q = bL + cL^2$. The relation between this production function and costs is depicted in Fig. 39.

1. **Fixed costs** are, since the fixed inputs are constant, also constant – we assume that their size is a constant "a". The equation of fixed costs is thus as follows:

$$FC = a$$

The constant a represents the money influenced by the amount of fixed inputs used and their prices. This also applies in all the following cases.

2. The shape of the curve of **variable costs** is derived directly from the production function. The relation between Fig. 39a and 39c shows that ***while output increases at an accelerating rate, variable costs decrease at a decelerating rate***. The fact that output increases faster than variable costs implies the simplest equation of variable costs

$$VC = bQ - cQ^2$$

3. **Total costs** are $STC = FC + VC$,

and since FC = a and VC = $bQ - cQ^2$, STC can be expressed as STC = a + bQ – cQ2.

Graphically, the curve of total costs is a vertical summation of the functions of fixed and variable costs.

4. **The average fixed costs** can be calculated as follows

$$AFC = \frac{FC}{Q} = \frac{a}{Q}$$

5. For **average variable costs**

$$AVC = \frac{VC}{Q} = \frac{bQ - cQ^2}{Q} = b - cQ$$

Because with diminishing returns from variable inputs (L), both MP_L and AP_L decrease. The relation $AVC = w/AP_L$ then shows that ***if the price of labour is constant and the average product of labour decreases, the average variable costs increase with the increase in output.***

6. For **average costs** is true

$$SAC = \frac{STC}{Q} = \frac{a + bQ - cQ^2}{Q} = \frac{a}{Q} + b + cQ$$

7. **Marginal costs**

$$SMC = \frac{dSTC}{dQ} = \frac{dVC}{dQ}$$

If we use the function of total costs

$$STC = a + bQ - cQ^2\text{, then}\quad SMC = b - 2cQ$$

The SMC curve is linear with negative slope. That means that marginal costs under increasing returns from variable input diminish with the increase in output. In the case of increasing returns from labour, MP_L increases and since

$$SMC = \frac{w}{MP_L} \Bigg|_{w = const.}$$

marginal costs must decrease (w is assumed to be constant).

4.3.2.2 Costs under diminishing returns from variable input

If we have the production function in the form

$$Q = a + bL - cL^2$$

and if ***a = 0,*** then ***$Q = bL - cL^2$.*** The relation between this production function and costs is depicted in Fig. 40.

1. For **fixed costs**, it is true that ***FC = a***

2. The shape of the curve of **variable costs** is influenced by the character of the production function. If we assume that the price of variable factor w = CZK 100 per unit, then the first unit L produces output Q_1 (Fig. 4.18a), which is connected with variable costs CZK 100 (Fig. 1.18c). Every additional unit L increases variable costs by CZK 100, but the increase in output is slower ($Q_1Q_2 < 0Q_1$; $Q_2Q_3 < Q_1Q_2$ etc.). So ***if output increases with a decreasing pace, variable costs increase with increasing pace.*** The equation of variable costs has the following form

$$VC = bQ + cQ^2$$

3. **Total costs** are

$$STC = FC + VC = a + bQ + cQ^2$$

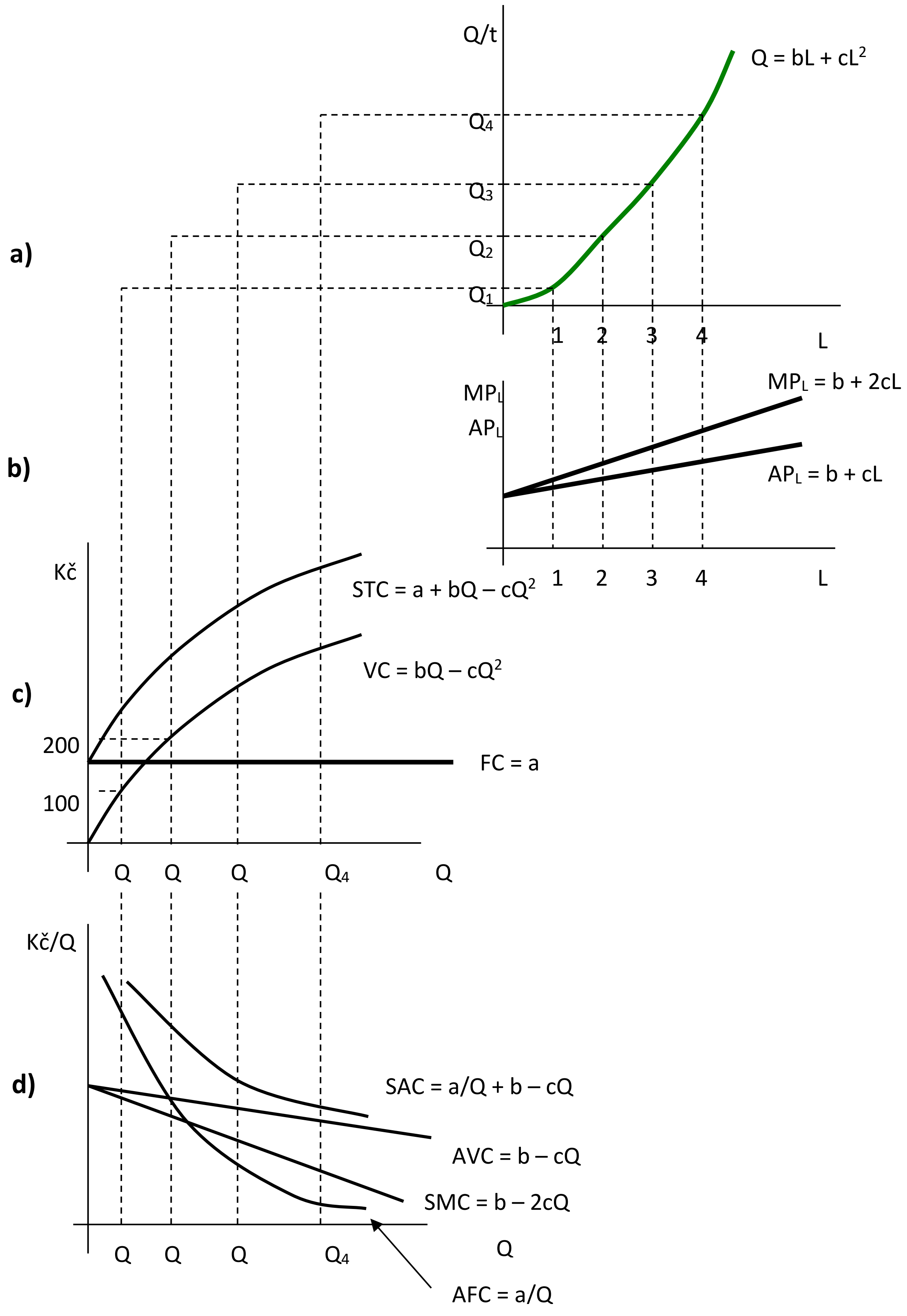

Fig. 40 Costs under increasing returns

4. **Average fixed costs** are ***AFC = a/Q***

5. **Average variable costs** are calculated as

$$AVC = \frac{VC}{Q} = \frac{bQ + cQ^2}{Q} = b + cQ$$

As in the case of increasing returns from variable factor, we will link Fig. 40b and 40d. The relation AVC = w/AP_L then shows that ***if the price of labour is constant and the average product of labour decreases, the average variable costs increase with the increase in output.***

6. For **average costs** is true

$$SAC = \frac{STC}{Q} = \frac{a + bQ + cQ^2}{Q} = \frac{a}{Q} + b + cQ$$

The SAC curve is the vertical summation of the AFC curve and the AVC curve. Since AFC diminish with the increasing output and AVC increase with the increasing output, the final shape of the SAC curve depends on the mutual relation between the diminution of AFC and increase in AVC. Figure 40d shows that:

- with low output (up to Q_2), the diminution of AFC is greater than the increase in AVC, so the AC curve is downward sloping;
- when producing the output Q_2, the diminution of AFC is levelled by the increase in AVC, and the AC curve reaches its minimum;
- with output higher than Q_2, the increase in AVC is greater than the diminution of AFC, and the AC curve therefore turns upwards.

7. **Marginal costs are**

$$SMC = \frac{dSTC}{dQ} = \frac{dVC}{dQ} = b + 2cQ$$

The equation implies that the graphical representation of the curve of marginal costs is the linear function with a positive slope = 2c. The slope of the SMC curve (=2c) is twice the size of the slope of the AVC curve (=c), the SMC curve intersects the SAC curve from below (Fig. 40d).

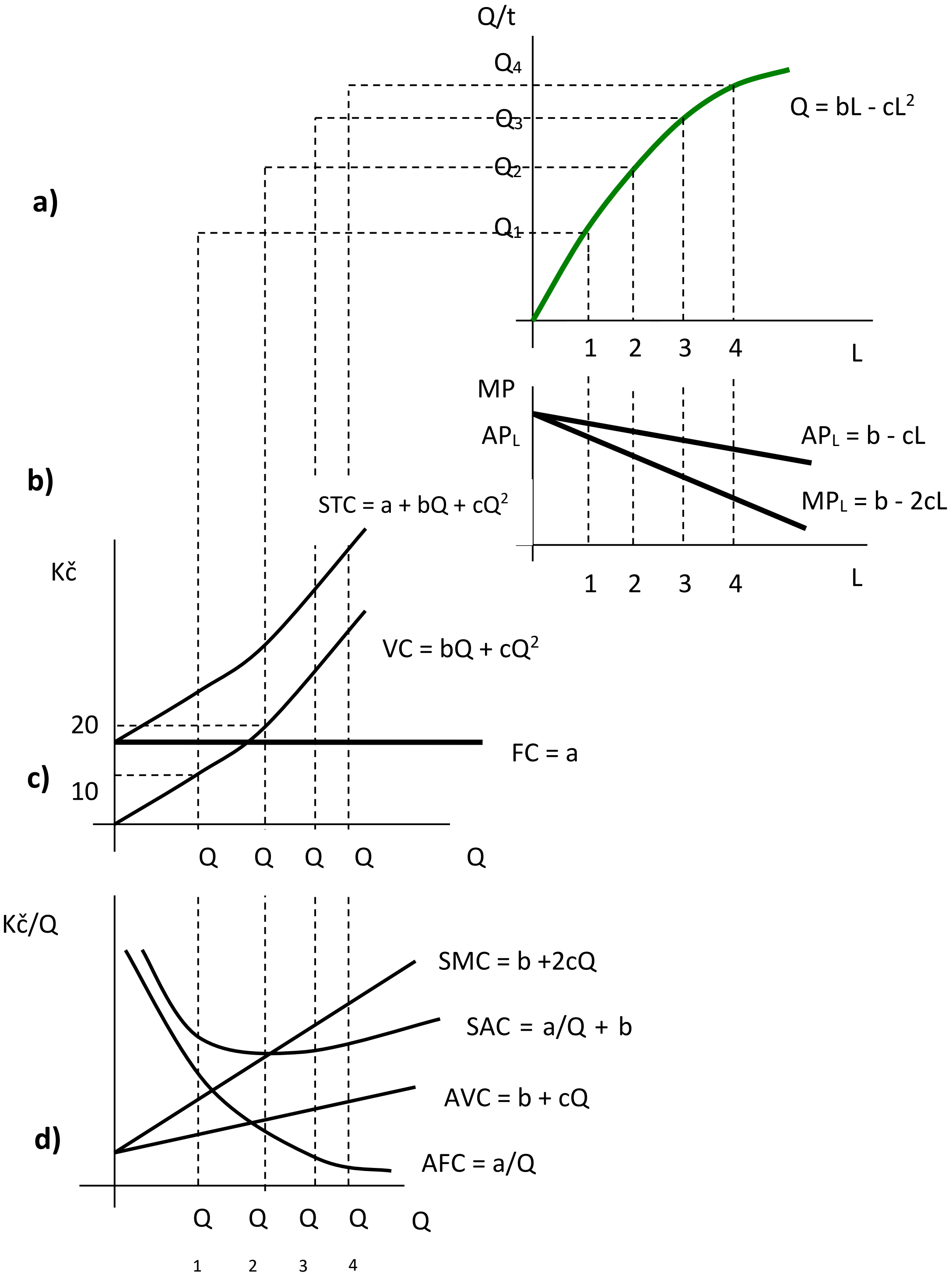

Fig. 41 Costs under diminishing returns

4.3.2.3 The Development of Costs under Constant Returns from Variable Input

The production function is defined as

$$Q = a + b;\ \text{if}\ a = 0,\ \text{then}\ Q = bL - cL^2.$$

4. For **fixed costs**, it is true that $FC = a$
5. The curve of variable costs is determined by the fact that each additional unit of variable input causes equal increase in output (MP_L is constant). If we assume that the price for one unit of labour w is constant, the first unit of L produces output Q_1 (Fig. 41a) and represents the costs of w (Fig. 41c). The second unit of L produces output Q_2 and is connected with variable costs 2w. The marginal product of the first unit of labour is equal to the marginal product of the second unit. Since variable costs constantly increase by the same amount (w), the VC curve is linear, and its general equation is $VC = bQ$

3. **Total costs** $STC = FC + VC = a + bQ$

4. **Average fixed costs** are again $AFC = a/Q$

5. **Average variable costs** $AVC = \dfrac{VC}{Q} = \dfrac{bQ}{Q} = b$

6. **Average costs** $SAC = \dfrac{STC}{Q} = \dfrac{a + bQ}{Q} = \dfrac{a}{Q} + b$ $(= AFC + AVC)$

7. **Marginal costs** $SMC = \dfrac{dSTC}{dQ} = \dfrac{dVC}{dQ} = b$; where b is constant.

Since $SMC = w/MP_L$ and both w and MP_L are constant, under conditions of constant returns to the variable input, marginal costs are also constant.

The development of costs first under increasing and then diminishing returns from variable input

The predominant type of production function in the short run has the general equation

$$Q = a + bL + cl^2 - dL^3$$

Up to L_1 in Fig. 41a, the output increases at an increasing rate, but from L_1 to L_4, it increases at a decreasing rate.

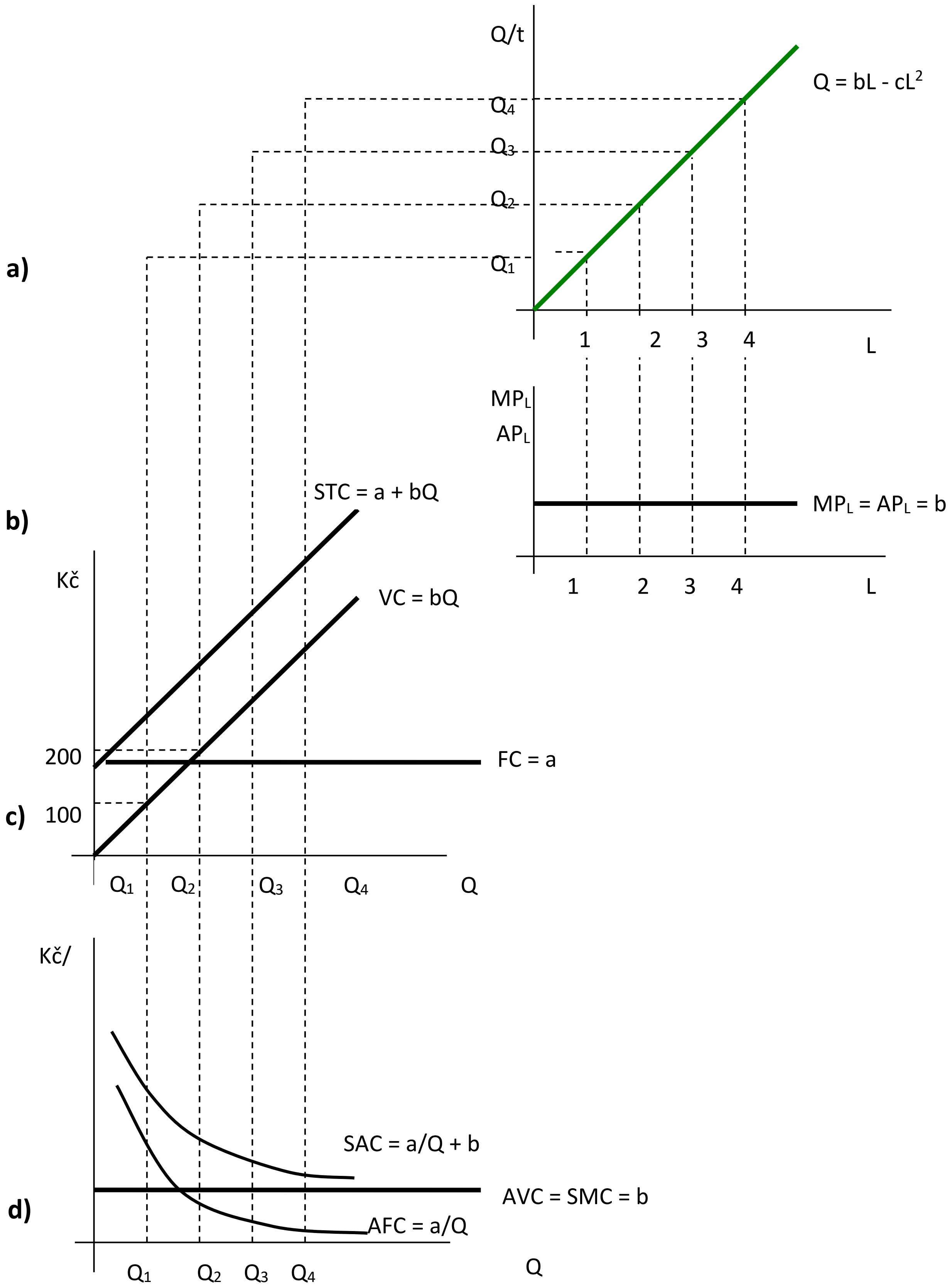

Fig. 42 Costs under constant returns

1. **Fixed costs** can be expressed again as $FC = a$
2. The curve of **variable costs**

The curve of variable costs in Fig. 4.17c increases between output 0 and Q_1 at a decreasing rate (which is reflected in increasing returns from variable input in Fig. 4.17a and 4.17b), and between the outputs Q_1 and Q_3 at an increasing rate (as a result of diminishing returns from variable input). The general equation of variable costs is

$$VC = bQ - cQ^2 + dQ^3$$

3. **Total costs** are

$$STC = FC + VC = a + bQ - cQ^2 + dQ^3$$

With the increase in output, STC changes in the same way as VC (Fig. 4.17c), which results in the same shape of the STC and VC curves.

4. **Average fixed costs** are depicted in Fig. 4.17d, AFC = FC/Q = a/Q as in the previous cases; AFC diminish with the increase in output.
5. **The average variable costs can be calculated in the same way as in the previous cases:**

$$AVC = \frac{VC}{Q} = \frac{bQ - cQ^2 + dQ^3}{Q} = b - cQ + dQ^2$$

If we follow the relation AVC = w/AP_L and assume that w is constant:

- if AP_L increases (Fig. 4.17b), AVC diminish (Fig.4.17d);
- if AP_L diminishes (Fig. 4.17b), AVC increase (Fig. 4.17d);
- AVC reach the minimum level when Q_2 output is produced (Fig. 4.17d), i.e., the output at which AP_L reaches its peak (when involving L_2 units of labour in Fig. 4.17b).

6. **Average costs**

$$SAC = \frac{STC}{Q} = \frac{a + bQ - cQ^2 + dQ^3}{Q} = \frac{a}{Q} + b - cQ + dQ^2$$

7. **Marginal costs** are

$$SMC = \frac{dSTC}{dQ} = \frac{dVC}{dQ} = b - 2cQ + 3dQ^2$$

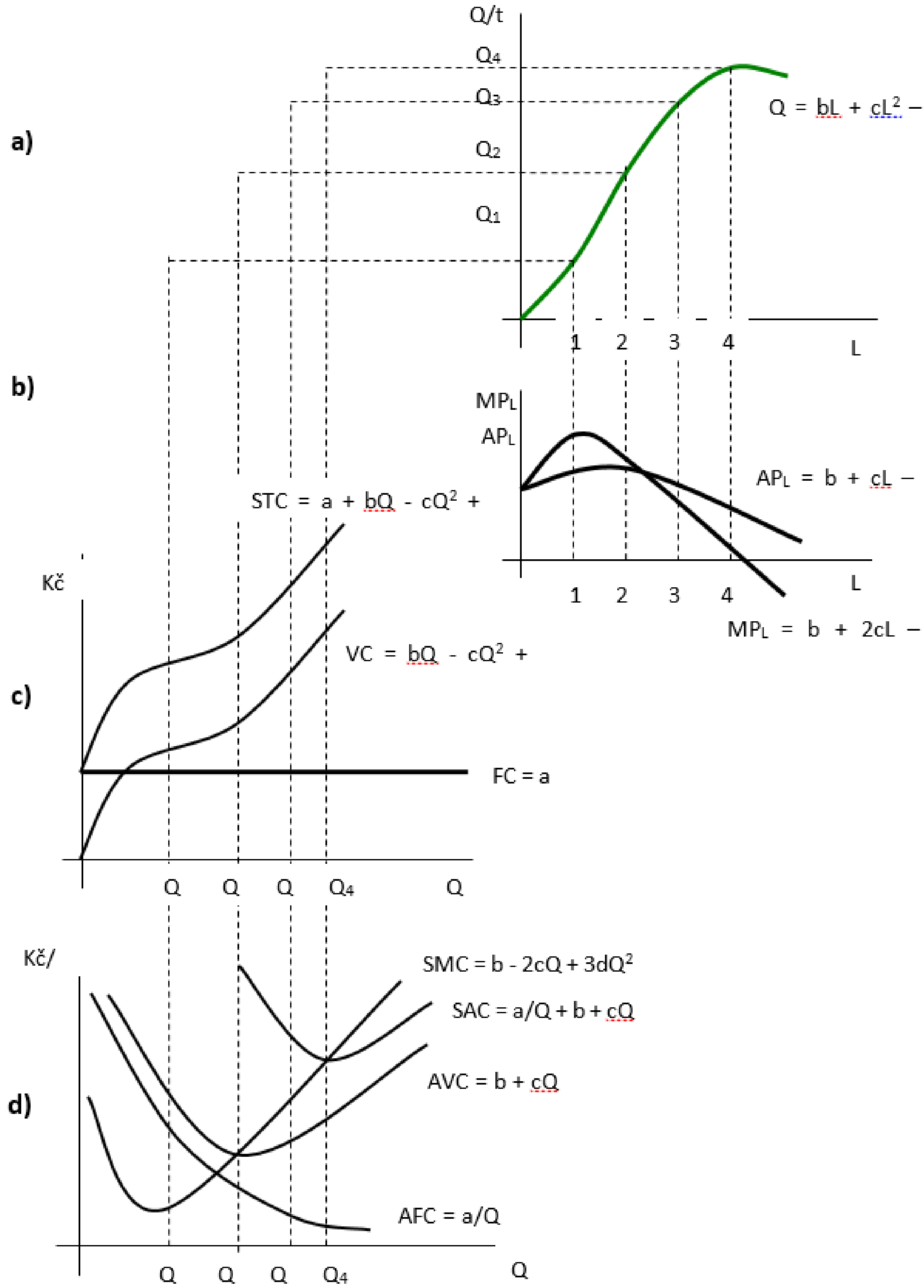

Fig. 43 Costs first under increasing and then diminishing returns

4.3.3 The Costs of a Firm in the Long Run

In the long run, all inputs a firm uses are variable, i.e., with two factors of production, the firm can adjust both labour and capital. While the same types of costs exist as in the short run – total costs and unit costs (average and marginal) – total costs in the long run cannot be divided into fixed and variable, as all costs are variable.

The curve of **total costs in the long run** LTC is derived in the same way as the curve of total costs in the short run STC. The fundamental difference, however, is that the shape of the STC curve was influenced by returns from variable factor, but the **shape of the LTC curve is determined by returns to scale**.

In the case of **constant returns to scale**, the total long-run costs increase at the same rate as the output (the percentage increase of which is equal to the percentage increase of both inputs). **The LTC curve therefore has the shape of an upward sloping line**.

If **increasing returns to scale** prevail in the long-run production function, the **LTC curve increases with the increase in output at a decreasing rate** (because if the prices of labour and capital are constant, the total long-run costs increase more slowly than the output).

Conversely, under **decreasing returns to scale**, **long-term costs (LTC) will increase faster than output**.

Definition

In the long run, all inputs a firm uses are variable. This means that, with two factors of production, the firm can adjust both the amount of labour and the amount of capital. The same types of costs exist in the long run as in the short run: total costs (however, these cannot be divided into fixed and variable as in the short run, since all costs in the long run are variable) and unit costs (average and marginal). The long-term total cost curve (LTC) is derived in the same way as the short-term total cost curve (STC). However, the key difference is that, while the shape of the STC curve is influenced by returns to the variable factor, the shape of the LTC curve is determined by returns to scale.

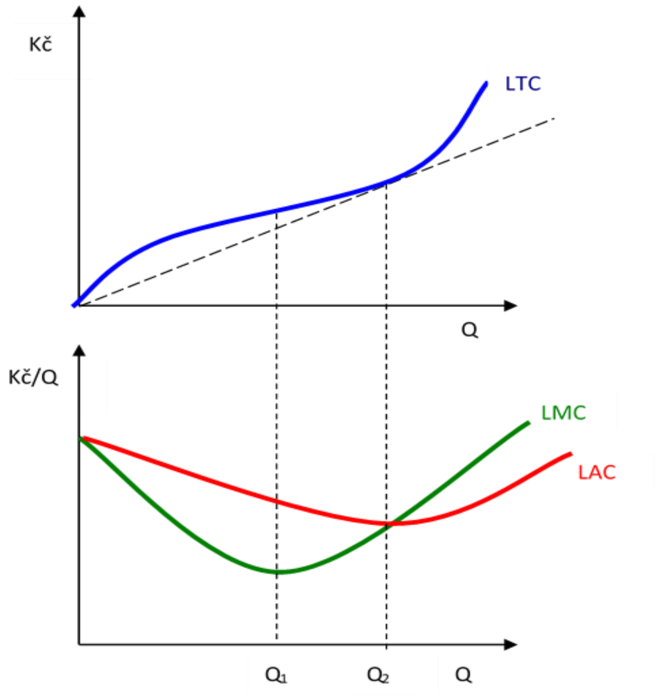

Fig. 44 Total and unit costs in the long run

Figure 43 shows that the LTC curve starts from the baseline, which is due to the absence of fixed costs in the long run.

In a similar manner to the short-run period, we can derive unit cost curves, specifically the long-run average cost (LAC) and long-run marginal cost (LMC) curves, from the total cost curve in the long run.

The long-run average cost (LAC) is represented geometrically by the slope of a line drawn from the origin to a point on the long-term total cost (LTA) curve. This demonstrates that as output increases, LAC initially decreases before eventually rising. It reaches its minimum at output level Q1, where the line from the origin becomes tangent to the long-term total cost (LTC) curve. Figure 43 illustrates that long-run marginal costs reach their minimum at a lower output level than long-run average costs.

4.3.4 Relationship between short-term and long-term costs

In concluding the firm's cost analysis, let's attempt to compare the relationships between costs in the short-run and long-run periods. Generally, short-term costs tend to be higher than long-term costs. This is primarily due to the existence of fixed costs in the short run, which prevent the firm from optimizing its input combinations as output changes. It is therefore important to keep in mind that short-run costs do not represent the minimum possible costs for varying levels of output.

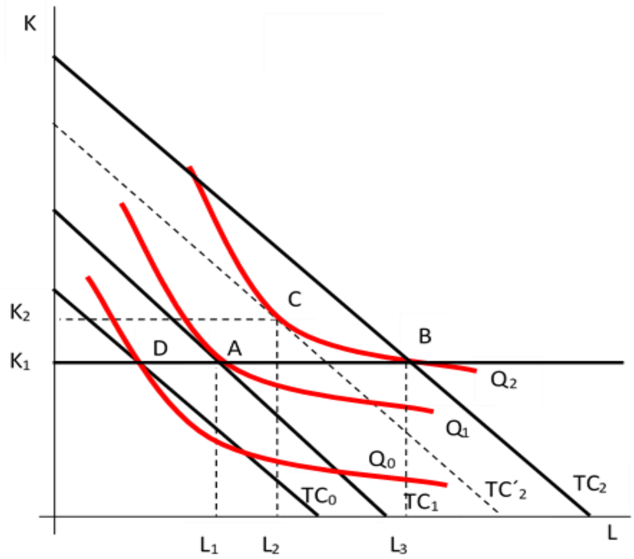

Fig. 45 Differences in production costs for the same output in the short run and long run

In the long run, on the other hand, a firm can produce increasing output with a changing mix of inputs (because they are all variable). A rational firm will try to produce each increased level of output at minimum cost. As mentioned at the beginning of this section, **long-run cost analysis assumes cost minimization by the firm**, i.e., movement along its increasing output curve.

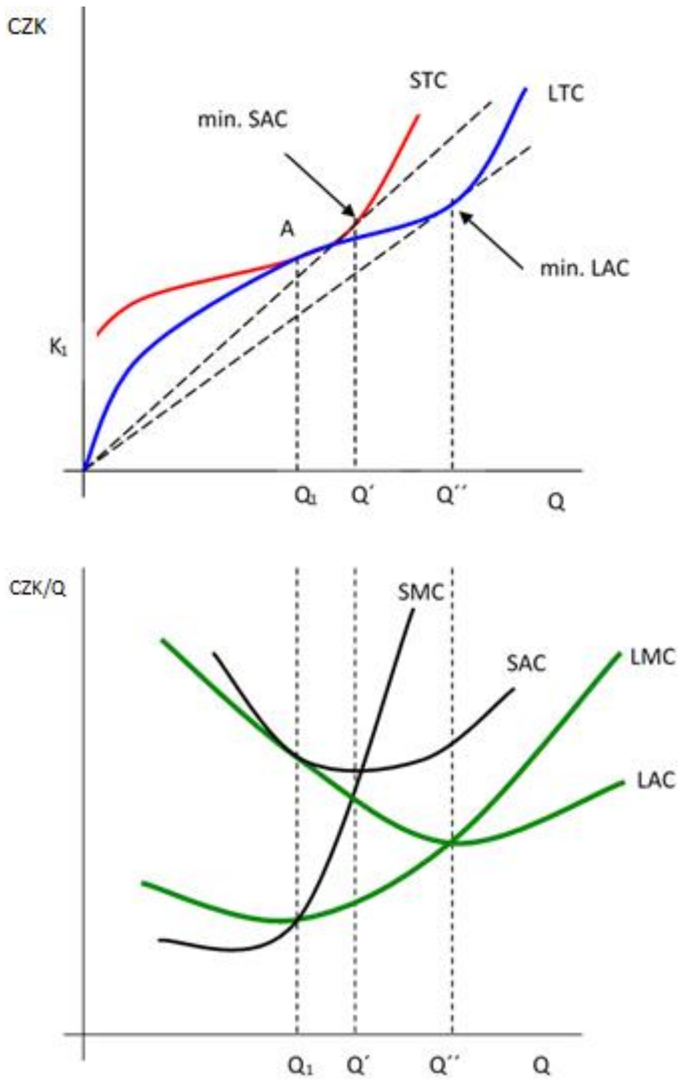

Fig. 46 Total and unit costs in the short and long run

Let us assume a firm that produces output Q_1 in the short run. If the capital costs are fixed at the level K1, the firm will produce this output using K_1 units of capital and L1 units of labor (point A on the diagram). The combination of inputs K1 and L_1 is also the combination that allows the firm to minimize the cost of producing output Q1 in the long run (at point A, the MRTS equals the ratio of input prices, so it represents the cost-optimal point).

.

$$STC\ (Q_1, K_1) = min.\ LTC$$

If the firm wants to produce Q_2, then in the short run, when capital is fixed, it can achieve this goal by using K1 units of capital and L_3 units of labour (point B). K_1 and L_3 is the optimal solution because the isoquant TC_2 is the intercept of the isoquant Q_2.

Figure 45 allows us to derive the relationships between short-run and long-run marginal costs (SMC and LMC) and short-run and long-run average costs (SAC and LAC). The **short-run and long-run marginal costs** (SMC and LMC) are equal when the firm's fixed costs are best used to produce that output. In the figure, this output is Q_1 when the directives of the two total cost curves (STC, LTC) are the same.

In the production of Q_1, the short-run marginal cost (SMC) equals the long-run marginal cost (LMC). This corresponds to the point where the SMC and LMC curves intersect. For output levels below Q_1, the short-run marginal cost is lower than the long-run marginal cost (the slope of the short-term total cost (STC) curve is less steep than that of the long-term total cost (LTC) curve). Conversely, for output levels greater than Q_1, the firm’s short-run marginal cost exceeds its long-run marginal cost (the slope of the STC curve is steeper than the slope of the LTC curve).

4.3.5 Envelope curves

Assume that additional units of capital (K_1, K_2, K_3) are introduced into production, each associated with levels of fixed costs FC_1, FC_2, FC_3. The total costs are then represented as STC_1, STC_2, STC_3, as shown in Figure 46.

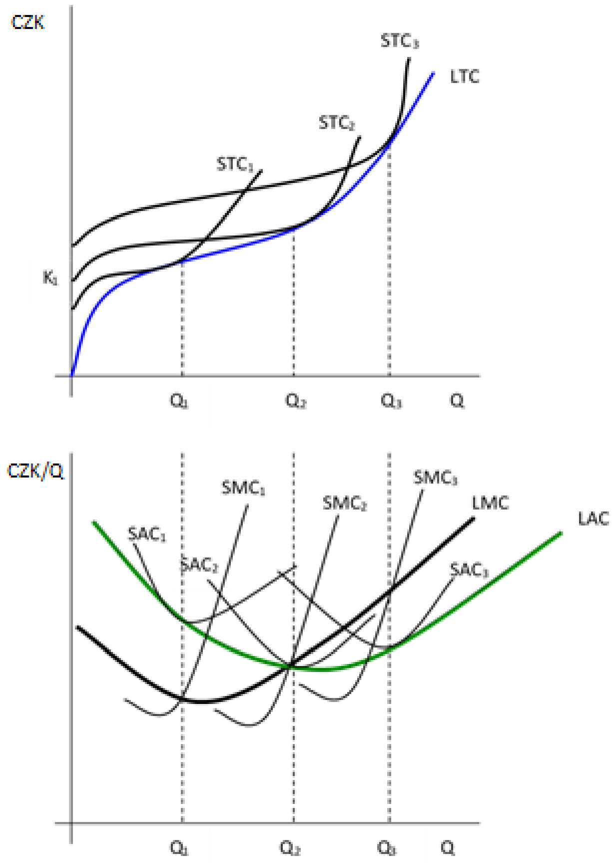

Fig. 47 Envelope curves

From the long-term total cost (LTC) envelope curve depicted in Figure 46, we can derive the long-run unit costs. For each short-run total cost curve (STC_1, STC_2, STC_3), there is a corresponding short-run average cost curve (SAC_1, SAC_2, SAC_3) shown in Figure 46, which is intersected from below by the short-run marginal cost curve (SMC_1, SMC_2, SMC_3) at its minimum.

Average costs are equal in both the short-run and long-run for levels of output where the quantity of fixed capital used allows for total cost minimization. Graphically, this is represented by the tangency points of the SAC curves for various levels of output with the LAC curve. The set of points where SAC equals LAC for varying output levels forms the LAC envelope curve.

Marginal costs in the short-run and long-run are equal at the level of output where the amount of fixed capital allows total cost minimization. Graphically, this is illustrated by the intersections of the SMC curves and the LMC curve at varying output levels.

4.3.6 Impact of input price changes on firm costs

Let us assume that **the prices of both considered inputs change proportionally**, for example, they increase by a factor of "t." Consequently, the total costs for producing a given output will also increase by a factor of "t."

Labour and capital costs for a given output

$$TC_1 = wL_1 + rK_1$$

After the prices of labour and capital increase by a factor of "t," the total costs (TC) for the given output will also increase by "t."

$$TC_2 = twL_1 + trK_1 = t\left(wL_1 + rK_1\right) = tTC_1$$

Although the firm's total costs increase due to changes in input prices, the optimal combination of labour and capital units for producing the given output remains unchanged. However, the average and marginal costs will also increase by a factor of "t."

$$TC_2 = tTC_1$$

Therefore $$AC_2 = TC_2\ /Q = t\left(TC_1\ /Q\right) = tAC_1$$

and also $$MC_2 = \delta TC_2\ /\delta Q = t\left(\delta TC_1\ /\delta Q\right) = tMC_1$$

When the costs of all inputs rise proportionally, the firm has no incentive to change its optimal input combination.

4.4 Firm revenues

Revenues represent the sum of money a firm earns from selling its output, which is why some authors use the analogous term of "sales."

If we consider the firm's goal of profit maximization, one approach to achieving this is maximizing revenues alongside minimizing costs.

The development of revenues is influenced by the nature of the market on which the firm sells its products. In a perfectly competitive market, many small independent producers and consumers buy and sell an identical product, meaning no single participant can influence the market price. Conversely, in an imperfectly competitive market, the number of participants is limited, and either sellers, buyers, or both may have a privileged position allowing them to influence the market price to their advantage.

Revenues can be categorized as **total**, **average**, and **marginal** revenues.

Points to remember

Revenues are the money earned from selling products, often referred to as sales. To maximize profit, a firm aims to minimize costs and maximize revenues. Revenue trends depend on market type – firms cannot influence prices in a perfectly competitive market, while in an imperfectly competitive market, they may have the ability to do so. Revenues are divided into total, average, and marginal categories.

4.4.1 Total firm revenues

Total revenue (TR) is the total amount of money a firm receives from the sale of its products. Its size is found by multiplying the unit price by the quantity sold:

$$TR = PQ$$

This relationship implies that the graphical representation of total revenue will depend on the type of competition.

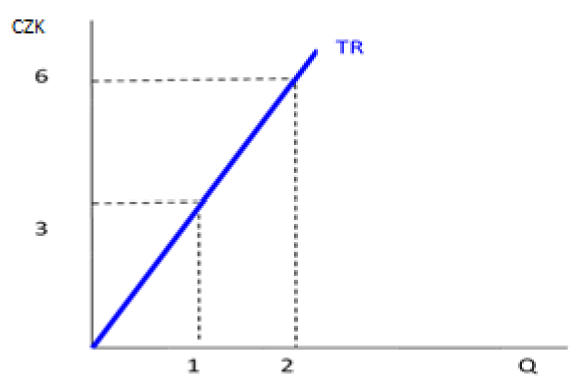

Fig. 48 Total revenue under perfect competition

In perfect competition—due to certain assumptions—a firm cannot influence the price, as it is treated as **an exogenous constant.** Therefore, under these conditions, **the total revenue (TR) curve is a function of the quantity sold** and can be graphically represented as an upward-sloping line originating at the origin, with a slope equal to the price (Figure 47).

Under imperfect competition - the situation is more complex. The price is not constant but falls as output rises and the demand curve is therefore downward sloping.

The elasticity of demand affects the specific shape of the total income curve. When an imperfectly competing firm lowers price to sell more output, the percentage reduction in price may be less than, equal to, or greater than the percentage increase in quantity realized, so that total income may rise, stay the same, or even fall.

- If demand is elastic (coefficient $\mathbf{e_{PD} < -1}$), the percentage increase in the quantity sold is greater than the percentage decrease in price, resulting in an increase in total revenue.
- If demand is unit elastic (coefficient $\mathbf{e_{PD} = -1}$), the percentage increase in the quantity sold will equal the percentage decrease in price, meaning total revenue remains unchanged.
- If demand is inelastic (coefficient $\mathbf{e_{PD} > -1}$), the percentage increase in the quantity sold is smaller than the percentage decrease in price, leading to a decrease in total revenue.

A graphical representation of the total revenue function can take various forms under conditions of imperfect competition (see Fig. 48).

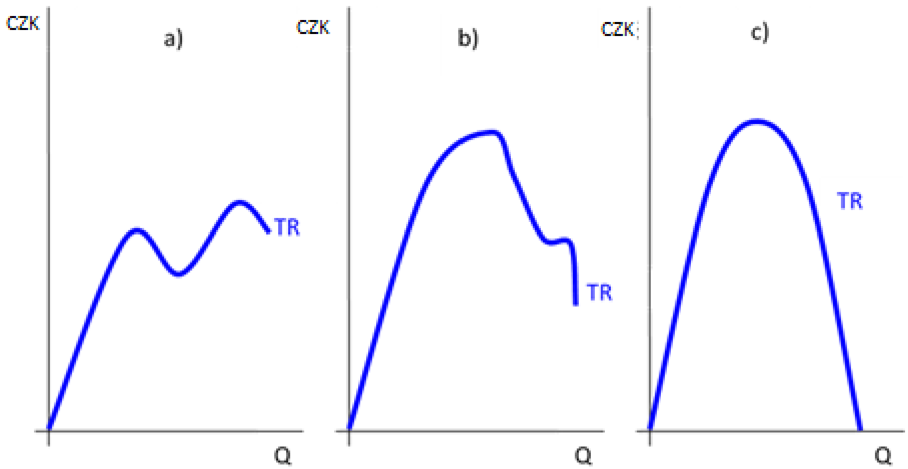

Fig. 49 Variants of the total revenue function under imperfect competition

To simplify, if we assume a linear demand function, its graphical representation forms a downward-sloping straight line, expressed by the equation:

$$P = a - b\,Q$$

By substituting this price expression into the total revenue equation

TR = PQ, we obtain ***TR = (a – bQ)Q*** which simplifies to ***TR = aQ – bQ²***

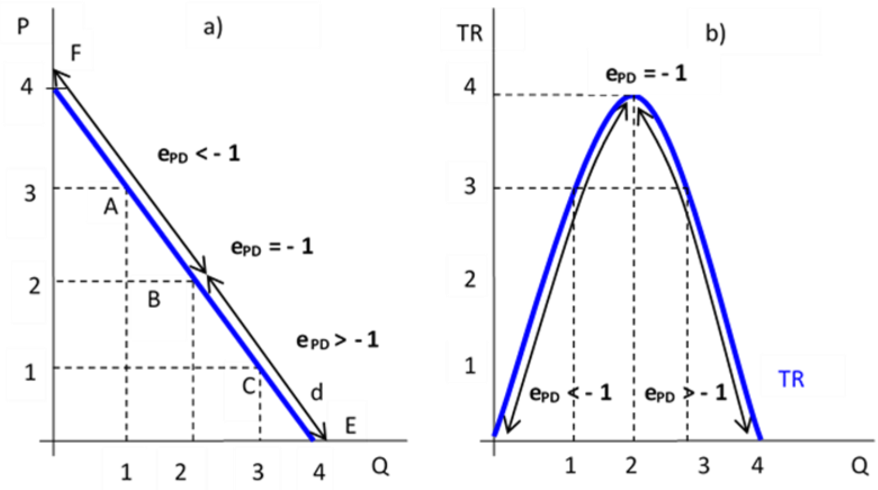

Fig. 50 Total Revenue Evolution

Figures 49(a) and 49(b) illustrate the relationship between price elasticity of demand and total revenue.

In Figure 49(a), the slope of the demand curve is represented as **dP/dQ = -1**. To calculate the price elasticity of demand, we consider the inverse slope of the demand curve, resulting in **dQ/dP = 1/-1 = -1**.

Consequently, segment FB of the demand curve is price elastic, point B represents unitary elasticity, and segment BE is price inelastic. The corresponding sections of the total revenue curve TR shown in Figure 49(b) reflect the characteristics of this demand curve.

4.4.2 Average and Marginal Revenue of a Firm

Another important type of revenue is **Average Revenue (AR)**, which refers to the revenue a firm earns from selling one unit of output. It is calculated by dividing total revenue by the quantity of output:

$$AR = \frac{TR}{Q} = \frac{PQ}{Q} = P$$

The **average revenue curve** under **perfect competition** is parallel to the x-axis at a level that corresponds to the price (AR = P = constant). However, under conditions of imperfect competition, it is downward-sloping (AR = P = a - bQ).

Marginal Revenue (MR) is defined as the change in total revenue resulting from a one-unit change in output (sales).

$$MR = \frac{dTR}{dQ} = \frac{d(PQ)}{dQ} = P + Q \cdot \frac{dP}{dQ}$$

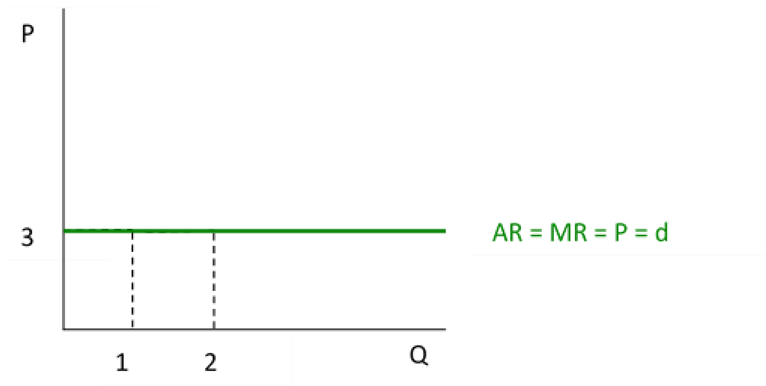

Fig. 51 Average and marginal revenue

Under **perfect competition**, with all firms identical and none of them influences the market price, the individual demand curve is parallel to the x-axis, implying that its slope ***(dP/dQ) = 0.*** Consequently, marginal revenue equals price: ***MR = P.***

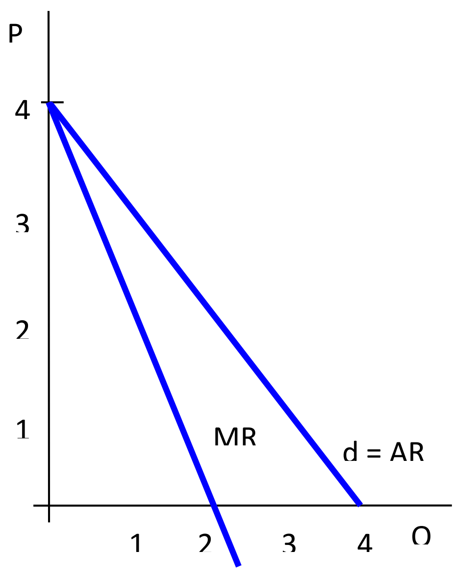

Fig. 52 Average and marginal revenue

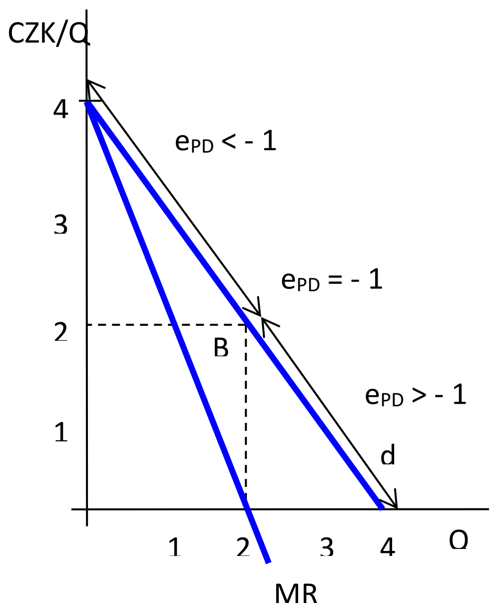

Fig. 53 Elasticity and marginal revenue

A key characteristic of **imperfect competition** is a downward-sloping individual demand curve. To sell an additional unit of output, the firm must reduce its price. This means that in the relationship where ***dP/dQ***, represents the slope of the individual demand curve, the value is negative (excluding the case of Giffen goods). As a result, when output is greater than zero, marginal revenue will be less than price ***MR < P.***

Under conditions of imperfect competition and assuming a linear individual demand curve, the total revenue function can be represented as $TR = aQ - bQ^2$.

The marginal revenue (MR) is calculated as follows: $$MR = \frac{dTR}{dQ} = a - 2bQ$$

Since the slope (or directive) of the marginal revenue function (-2b) is twice as steep as that of the demand curve (-b), the marginal revenue curve declines at twice the rate of the demand curve (see Figure 4.26).

Similar to the connection between total revenue and price elasticities of demand, there is also a link between marginal revenue and price elasticities of demand. This relationship can be expressed by the following equation:

$$MR = P . \left(1 + \frac{1}{e_{PD}}\right)$$

From this equation and as illustrated in Figure 52, it can be observed that marginal revenue behaves as follows:

- it is positive when $e_{PD} < -1$, meaning demand is elastic;
- negative when $e_{PD} > -1$, demand is inelastic;
- equal to zero when the price elasticity of demand is unitary $e_{PD} = -1$

Σ

A firm's revenue is a critical component of economic analysis, offering insights into the behaviour of firms across various market structures. This chapter delves into three key revenue types: total revenue (TR), average revenue (AR), and marginal revenue (MR), while exploring their characteristics under conditions of both perfect and imperfect competition.

Total revenue (TR) represents the total income a firm earns from selling its products and is calculated as the product of the unit price (P) and the quantity sold (Q). In a perfectly competitive market, the graphical representation of TR appears as an upward-sloping linear line originating at the origin, with its slope determined by the price level. This occurs because a firm in perfect competition is unable to influence market prices and thus functions as a price-taker. Conversely, in markets characterized by imperfect competition, the shape of the TR curve is dictated by the demand's elasticity. Specifically, TR rises when demand is elastic, remains unchanged when demand is unit elastic, and decreases when demand is inelastic.

Average revenue (AR) is the revenue earned by a firm per unit of output sold, calculated as the total revenue divided by the quantity sold. Under perfect competition, the AR curve is a horizontal line parallel to the x-axis, reflecting a constant price level. This constancy arises because firms in perfect competition are price-takers, unable to influence the market price. In contrast, under conditions of imperfect competition, the AR curve is downward-sloping, reflecting the firm's ability to influence price by varying output levels.

Marginal revenue (MR) measures the change in total revenue resulting from the sale of one additional unit of output. In perfect competition, MR is equal to the market price, and thus the MR curve aligns with the AR curve, both appearing as horizontal lines parallel to the x-axis. Conversely, in imperfect competition, MR is always less than the price due to the downward-sloping demand curve. As a result, the MR curve declines more steeply than the AR curve.

The price elasticity of demand (ePD) significantly influences revenue dynamics. When demand is elastic ($ePD < -1$), total revenue increases with a decrease in price. If demand is unit elastic ($ePD = -1$), total revenue remains unchanged, while with inelastic demand ($ePD > -1$), total revenue decreases as price falls. The relationship between MR and the elasticity of demand is captured by the equation $MR = P \times (1 + 1/ePD)$. This relationship indicates that MR is positive when demand is elastic, zero at unit elastic demand, and negative when demand is inelastic. This interplay highlights how price changes impact a firm's revenue-generating potential in different market conditions.

To gain deeper insights into these relationships, graphical representations are invaluable. Graphs can visually demonstrate total revenue under perfect competition, variations in the total revenue function in conditions of imperfect competition, and the evolution of total revenue as a function of demand elasticity.

Additionally, they can illustrate average and marginal revenue curves under different market structures, as well as the connection between demand elasticity and marginal revenue.

Understanding a firm's revenue and how it interacts with demand elasticity is essential for analysing firm behaviour across diverse market types. Mastery of these concepts enables a better comprehension of firm decisions regarding pricing and output levels, directly impacting their strategies for profit maximization. This analysis requires a clear distinction between perfect and imperfect competition conditions, as these market structures have a significant influence on the nature and shape of revenue functions.

In practical terms, firms apply their knowledge of revenue and demand elasticity to fine-tune their pricing strategies. For example, a firm operating in a market characterized by elastic demand can boost total revenue by reducing its prices, as the percentage increase in quantity demanded will outweigh the price decrease. Conversely, a firm in a market with inelastic demand can enhance total revenue by raising prices, given that the drop in quantity demanded is relatively small. Nonetheless, such strategies must always be evaluated considering the firm's costs and the broader market environment to ensure effective decision-making and optimal profitability.

In conclusion, analysing a firm's earnings is a complex topic that requires an understanding of various economic concepts and their interrelationships. The ability to interpret graphs and mathematical formulas related to a firm's earnings is essential for deeper economic analysis and decision making. This chapter provides a basic framework for understanding these relationships and applying them to real economic situations.

?

Review questions

1. Do decreasing marginal returns to a variable factor necessarily imply decreasing average returns to that variable factor? Explain.
2. Compare an isoquant to an indifference curve. What are the defining properties of an isoquant?
3. Which of the following factors would lead to a parallel shift of the isocost line?

a) a change in the firm's total expenditure,
b) a reduction in the price of one of the firm's inputs,
c) a proportional change in the prices of both inputs?

4. At TRIOLA, each seamstress can operate only one sewing machine at a time. What value is likely to be observed for:

a) the marginal rate of technical substitution,
b) the elasticity of substitution coefficient.

5. What equals the value of implicit costs? Select the correct answer:

a) the amount paid for the purchase of inputs,
b) zero, because the firm does not actually pay it,
c) the price of the products that the same inputs would have produced if they had been used differently or elsewhere.

6. Draw five short-run average cost (SAC) curves, ensuring that the minimum point of each SAC curve touches the long-run average cost (LAC) curve. What conclusions can you draw from this graph?
7. Discuss how and why the graphical depiction of a firm's total revenue varies under perfect versus imperfect competition.
8. Under conditions of imperfect competition, why does marginal revenue decrease at twice the rate of average revenue? Explain.

?

Test questions:

1. Which of the following situations is likely to lead to an increase in the firm's total revenue?

a) Lowering the price when demand is elastic
b) Increasing the price when demand is elastic
c) Lowering the price when demand is inelastic

2. Under which market condition is marginal revenue (MR) equal to price?

a) Monopoly
b) Perfect competition
c) Monopolistic competition

3. How does total revenue change if demand is unit elastic and the firm increases the price?

a) Total revenue will increase
b) Total revenue will decrease

c) Total revenue will remain unchanged

4. Which of the following equations correctly represents the relationship between price (P) and marginal revenue (MR) under imperfect competition conditions?

a) MR = P
b) MR > P
c) MR < P

5. At what level of price elasticity of demand is marginal revenue (MR) equal to zero?

a) ePD = -1
b) ePD = -2
c) ePD = 0

Answers:

1.a, 2.b, 3.c, 4.c, 5.a

Practical task

1. *Choose three popular mobile apps or games that offer microtransactions (e.g. in-game currency purchases, premium features or cosmetic upgrades). Record the prices of the various offers in these apps. Then conduct a survey of classmates or friends who use these apps. Ask them which items they buy most often and how their buying behaviour would change if prices were changed up or down by 10%. Based on the information gathered, estimate the elasticity of demand for different types of offers. Finally, suggest an optimal pricing strategy for one of these applications that could increase total revenue, and present your suggestions to the class, explaining how you used the concept of elasticity of demand.*

2. *Visit a local café and take note of the prices of its most popular drinks and snacks (e.g., coffee, tea, cakes). Observe which items are sold most frequently during different times of the day. Conduct a simple survey for café customers, asking how they would adjust their purchases if prices increased or decreased by 10%. Use the survey results to estimate which products have elastic demand and which have inelastic demand. Take into account factors such as time of day, seasonality, or local competition that may influence demand elasticity. Propose improvements to the café's pricing strategy aimed at increasing overall revenue, such as introducing happy hours, a loyalty program, or seasonal promotions. Present your recommendations to the class, explaining how you applied the concept of demand elasticity in your analysis.*

Literature

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. Mikroekonomie. 5. aktual. vydání. Praha: Management Press, 2010. s. 157-245. ISBN 978-80-7261-218-5. 574 s.

[2] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005.

[3] MACÁKOVÁ, Libuše; SOUKUPOVÁ, Jana. Mikroekonomie – repetitorium. Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[4] HOLMAN, Robert. Ekonomie. Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[5] HOLMAN, Robert. *Mikroekonomie.* Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[6] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[7] SOUKUPOVÁ, J., 2018. *Mikroekonomie.* 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 5

Output Choice for a Perfectly Competitive Firm

After studying this chapter, you will be able to:

- characterize the assumptions of the perfect competition model;
- explain the supply of a perfectly competitive firm in the short run and the long run;
- describe the equilibrium of a perfectly competitive industry and how market equilibrium is established in the long run;
- discuss the efficiency of perfect competition.

Keywords:

Perfect competition model, output decision-making, supply of a perfectly competitive firm, supply of a perfectly competitive industry, equilibrium of a perfectly competitive industry in the short run.

After studying this chapter, you will be able to

- Analyze the graphical representation of producer surplus and its influence on market supply.
- Understand the differences in producer surplus between the short and long run.
- Evaluate how economic profit and industry output change over the long term.
- Identify the role of factor owners in the context of producer surplus and changes in market supply.

Chapter preview

- This chapter focuses on perfect competition and the analysis of producer surplus in the short and long term. Producer surpluses are a critical component in economics and play a significant role in firms' decision-making processes. The chapter examines how producer surplus changes in response to economic conditions and its impact on market dynamics. We will explore graphical representations of these surpluses and explain their significance for producers and owners of production factors.

Chapter objectives

- Explain the concept of perfect competition.
- Understand how producer surplus is expressed in the short term and how it changes in the long run.
- Analyze the implications of producer surplus on firm decision-making and market behavior.
- Discuss the impact of input costs and prices on producer surplus in the long run.

Estimated study time

- Ca 240 min.

A profit-maximizing firm chooses its inputs and outputs to achieve ***maximum economic profit***. Zero economic profit implies that inputs generate as much value in their current use as they would in their best alternative use. In such a case, there is no incentive for the firm to reallocate its inputs to another industry.

Ekonomic profit (***π***) is the difference between total revenue and total costs:

$$\pi = TR - TC \text{ or } \pi = P \,.\, Q - (w \,.\, L + r \,.\, K).$$

For the maximum difference between total revenue and total costs, the following applies: $\frac{dTR}{dQ} = \frac{dTC}{dQ} \approx MR = MC$

The equality of marginal revenue and marginal costs is referred to as *the* ***golden rule of profit maximization***. It can be interpreted as follows: to maximize profit, a firm should choose an output level where marginal revenue equals marginal costs. It is called the "golden rule" because it guides any profit-maximizing firm, regardless of the type of market structure in which it operates.

5.1 Assumptions of the Perfect Competition Model

Perfect competition is one of the oldest and most thoroughly developed models of market structures within firm theory. This model is built upon the following **assumptions**:

- There is a ***large number of buyers and sellers*** in each market, none of whom is powerful enough to influence the price or the industry's output.
- All goods are ***homogeneous***.
- There is ***free entry and exit*** in all markets.
- All producers and consumers have ***perfect information*** about the prices and quantities traded in the market. Additionally, firms have free access to information about technology.
- Firms aim to ***maximize profit***, while consumers aim to ***maximize utility***.

In reality, few markets meet all of these conditions. Examples of markets that come close to perfect competition include agricultural products, certain raw materials, and money markets. Although these conditions rarely exist in the real world, the significance of the perfect competition model is undeniable: it serves as a foundation for analyzing market structures under conditions of imperfect competition.

A perfectly competitive firm is often described as a "price taker," meaning it accepts the market price as given. This term reflects the fact that for a firm, the price of its output and inputs is **exogenous—an external factor** beyond its control. Consequently, the firm cannot influence either of these prices.

The fact that there are a large number of small firms in a perfectly competitive market means that each firm contributes only a very small share to the total market output, so its sales volume does not affect the market price. This results in the demand for its output being perfectly elastic, which can be graphically represented as a horizontal line parallel to the x-axis. The demand for the output of a small agricultural producer (individual demand) is illustrated in Figure 53b, where 6,200 CZK is the price per ton of rapeseed. The market demand, shown as D in Figure 53a, reflects the quantities of rapeseed that all consumers purchase at varying prices. This demand curve has a negative slope, indicating the well-known fact that consumers are willing to buy larger quantities at lower prices.

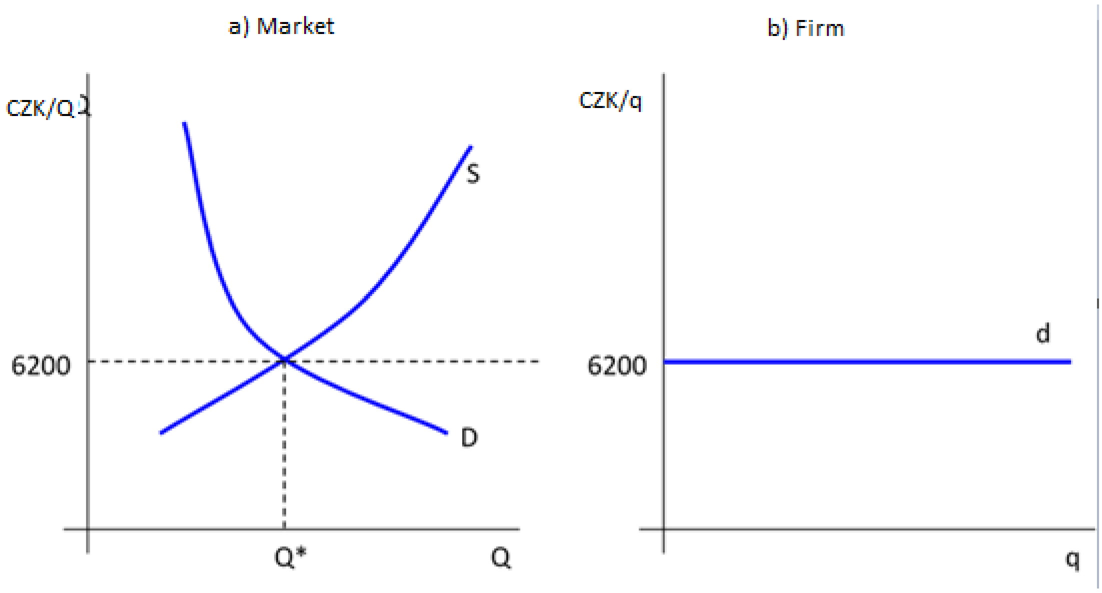

Fig. 54 Perfect competitive market and firm

5.2 Firm's Output Decision in the Short Run

A firm's optimal output can be determined in two ways:

a) by analyzing the difference between total revenue and total costs,
b) by identifying the point where marginal revenue equals marginal cost.

Profit is maximized when the difference between total revenue and total costs is greatest, corresponding to the largest vertical distance between the TR and TC curves. This occurs when the curves have the same slope, at output level q* (Figure 54a). The curves also have the same slope at output level q', but in this case, the TC curve lies above the TR curve, resulting in maximum loss for the firm.

Since the slope of the total revenue curve represents marginal revenue and the slope of the total cost curve represents marginal cost, the equality of these slopes indicates that marginal revenue equals marginal cost. The point where the MR and MC curves intersect, q*, represents the firm's optimal output. At q*, the firm produces the truly optimal output because the sufficient condition is met: the marginal cost curve is increasing (it crosses the marginal revenue curve from below). This condition is not satisfied at output level q'. Figure 54b illustrates these relationships.

The following applies to the identified output levels:

- q*: This represents the optimal output where the firm maximizes its profit; at this point, the slopes of the TR and TC curves are equal, meaning MR = MC, and the marginal cost curve is increasing;
- q': At this output, the slopes of the TR and TC curves are equal, meaning MR = MC. However, since total revenue is less than total cost, the firm maximizes its loss; in this case, the marginal cost curve is decreasing;
- q_1: The line segments drawn from the origin to the TR and TC curves have the same slope, meaning AR = SAC. At this point, the TC curve stops declining and starts rising, indicating that marginal revenues are at their minimum;
- q_2: The line segments drawn from the origin to the TR and TC curves again have the same slope, meaning AR = SAC.

If the firm produces an output greater than q1 and less than q*, it will generate profit (since TR > TC in Figure 54a and AR > SAC in Figure 54b) but not the maximum profit. It can increase profit by increasing output. Similarly, for output levels greater than q* but less than q_2, the firm can maximize profit by reducing output towards the optimal level q*. Since, for output between q_1 and q_2, the slope of the total cost curve (or marginal cost) is increasing, it suggests that the firm will produce at an output level where diminishing returns to the variable factor are evident (assuming its price remains constant).

Because a perfectly competitive firm maximizing profit bases its output decision on the equality of marginal revenue and marginal cost (MR = MC), and since for a price-taking firm, marginal revenue is equal to the market price (MR = P), it will produce at an output level where marginal cost equals the market price (P = MC) in the short run.

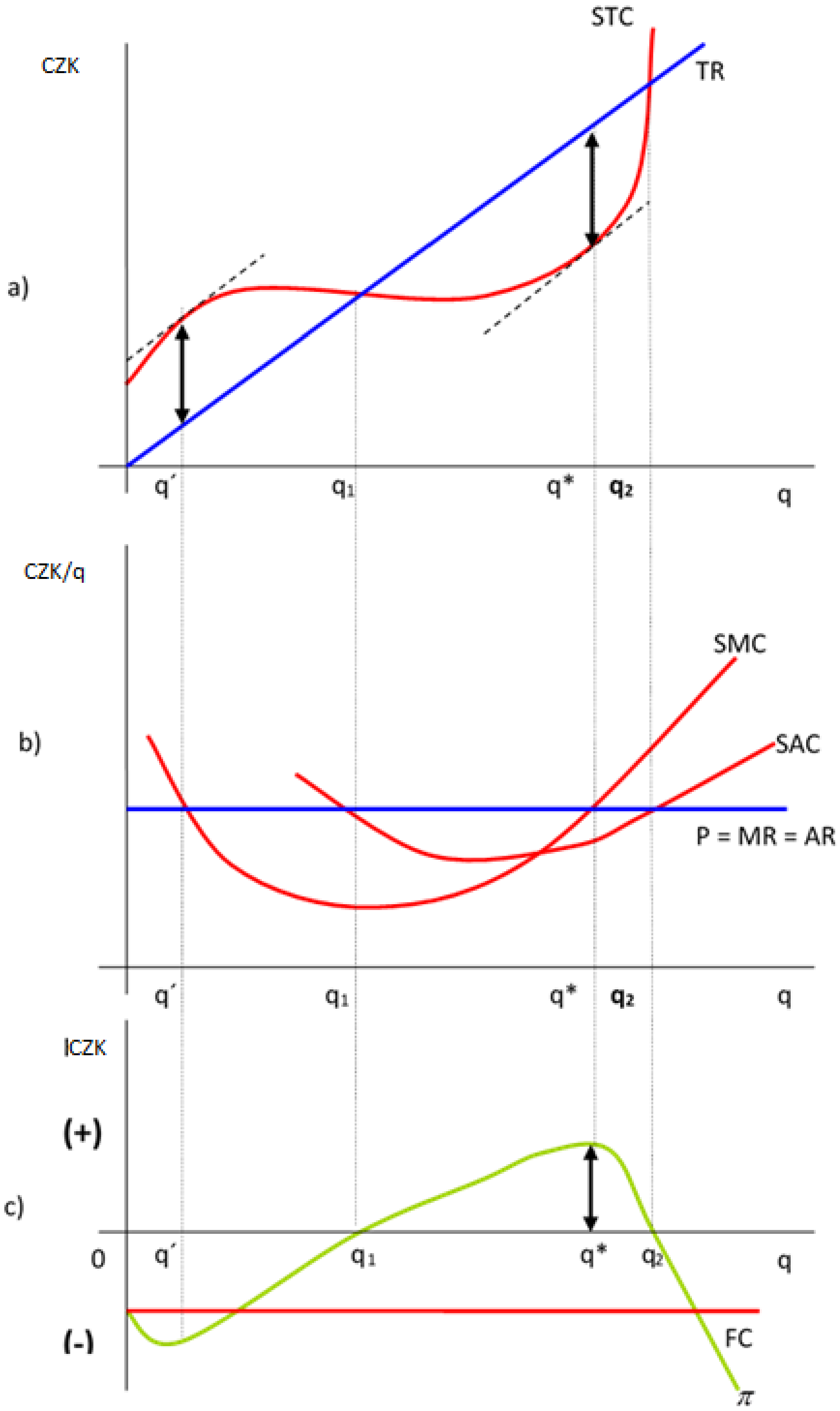

Fig. 55 Alternative output derivation in which the firm maximizes profit

5.3 Supply of a Perfectly Competitive Firm in the Short Run

An increase in the market price results in an upward movement of the intersection of marginal revenue (MR) and marginal cost (MC) along the marginal cost curve, leading to an increase in the quantity supplied. Conversely, a decrease in market price causes a downward movement along the marginal cost curve, resulting in a reduced quantity supplied. Therefore, it is clear that the short-run supply curve of a firm is represented by the upward-sloping portion of its marginal cost curve.

Up to this point, our analysis has assumed that the firm is capable of producing at least some output where total revenue exceeds total costs (the production range between q_1 and q_2 in Figure 55).

But what happens if, at any level of production, the firm's total revenue is less than its total costs (TR < STC)? While it might seem that halting production would be the obvious solution, in the short run, a firm is unlikely to take this step immediately. This is because, in the short run, the firm must still pay its fixed costs, even if output is zero. If production stops, the firm's losses will amount to its fixed costs. However, if it continues producing and generates total revenue that exceeds its variable costs, it will be able to cover all of its variable costs and a portion of its fixed costs, thereby reducing its losses. This loss-minimization condition (TR > VC) can also be expressed using per-unit metrics as TR/q > VC/q, meaning P > AVC (Figure 54b).

Thus, if a firm in the short run finds itself in a situation where total revenue is less than total costs (and average revenue, or price, is lower than average total cost), it uses the criterion of average variable costs to decide whether to stop production. If the price per unit of output exceeds the average variable cost of producing that unit (P > AVC), as shown by price P_1 in Figure 5.3b, the firm will minimize its losses by continuing to produce at output level q*. However, if the price is equal to or below average variable costs (P ≤ AVC), the firm will minimize its losses by shutting down production. The critical point here is when price equals average variable costs. In Figure 5.3b, the firm will cease production at output q**, where P_2 equals the minimum AVC, which also equals MC (known as the shutdown point).

Points to remember

When the market price rises, the firm increases the quantity of its offered production, whereas when the price falls, the firm reduces its supply. In the short run, the firm does not necessarily stop production even in the case of a loss, as long as total revenue covers variable costs and part of the fixed costs. The firm will continue production if the price exceeds average variable costs (P > AVC). If the price falls below this level, the firm will halt production, which is known as the shutdown point.

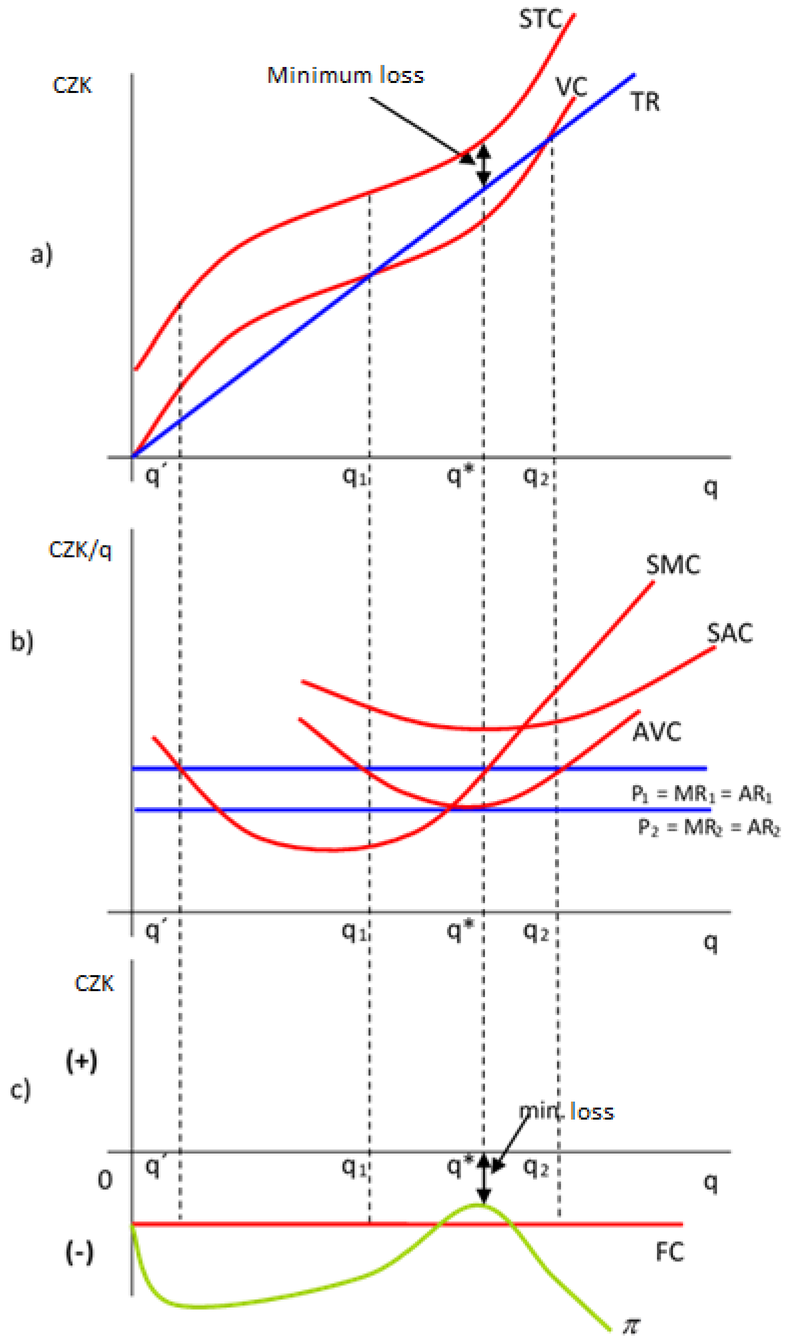

Fig. 56 The firm minimises losses by continuing production

Now we can refine the definition of the **firm's supply curve**: the short-run supply curve of the firm represents the relationship between the changing output price and the quantity of output the firm produces. In the case of a perfectly competitive firm that takes the market price as given, the short-run supply curve is formed by the **rising part of the marginal cost curve, with the lower bound being the minimum average variable cost (min. AVC).**

Given the assumptions of the perfect competition model, in which the firm treats the price of its output, as well as the prices of labor and capital it purchases in the factor markets, as fixed, the supply function of the profit-maximizing perfectly competitive firm can be expressed as:

$$q_s = q^* (P, w, r)$$

This defined supply function expresses the firm's decision-making about its optimal output, depending both on the price of the output it produces and on the prices of the inputs used to produce it.

5.4 Supply of a Perfectly Competitive Industry in the Short Run

Since the formation of the equilibrium price, which is an exogenous variable for the firm, occurs based on the intersection of supply and demand forces in a perfectly competitive market in the short run, it is necessary to derive the short-run supply curve for the entire perfectly competitive industry.

Note: *Due to firm specialization, each given product is produced by a limited number of firms. This group of firms is referred to as an industry or a market. The aggregate of the supply curves of all firms producing a given good forms the industry's supply curve, or the market supply curve.*

In the short run, we assume a constant number of firms in the industry. However, each firm is capable of changing the size of its output in response to changes in market conditions. This change is made by adjusting the quantity of the variable input. The construction of the supply curve for a perfectly competitive industry is based on the fact that the total amount of production offered by the entire industry is the sum of the amounts offered by individual firms. Therefore, the industry supply curve is the horizontal sum of the short-run supply curves of all firms in the industry at any given price. This conclusion holds true only in our simplified case, where one of the basic assumptions is constant input prices. The formation of the short-run supply curve of a perfectly competitive industry, under the simplified assumption of only two firms in the industry, is illustrated in Figure 56.

At price P_1, the first firm will offer 2 units, and the second firm will offer 3 units of output. Thus, the total output of the industry at price P_1 will be 5 units. Similarly, we can calculate the total market quantities for prices P_2 and P_3. Due to the increasing marginal cost function of each individual firm, the industry's supply curve will also have an increasing shape.

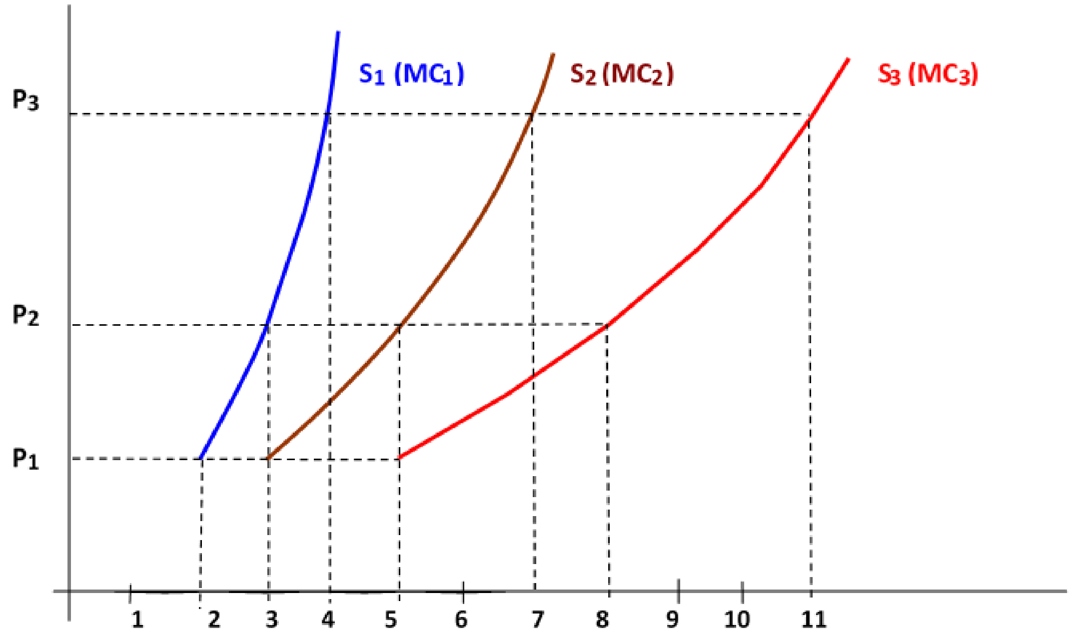

Fig. 57 Supply of a perfectly competitive industry in the short run

The factors influencing the short-run supply curve of a perfectly competitive industry include primarily the number of firms in the industry, the output level of each firm, and the determinants of marginal costs for each firm.

Let us take a closer look at the last of these factors. In the analysis of a perfectly competitive firm, we implicitly assume that the firm is a price taker not only in the product market but also in the input markets from which it rents production factors. Each perfectly competitive firm is assumed to be so small that changes in its output do not affect the prices of inputs due to increased demand for them. However, when considering the output change of the entire industry, this change in size affects the demand for production factors and subsequently their prices.

The effect of rising input prices on the market supply is illustrated in Figure 57. The initial price of the good, and industry output Q_1 determine one point on the industry supply curve, S. We might assume that if the price of the good rises from P_1 to P_2, each firm will increase its production, and the total industry output at price P_2 will be Q_3, and the market supply will be represented by this function. It can be expected that the growth in industry output, and hence the increase in demand for inputs, will push input prices upward. Higher input prices will lead to an increase in the marginal costs of each firm, which can be graphically represented by a shift of the MC curve (or individual supply curve, S) upward and to the left. As a result, at the price P_2, the market quantity will be derived from the sum of the shifted supply curves of all producers in the industry. The actual short-run market supply curve, S, assuming rising input prices, will be relatively steep.

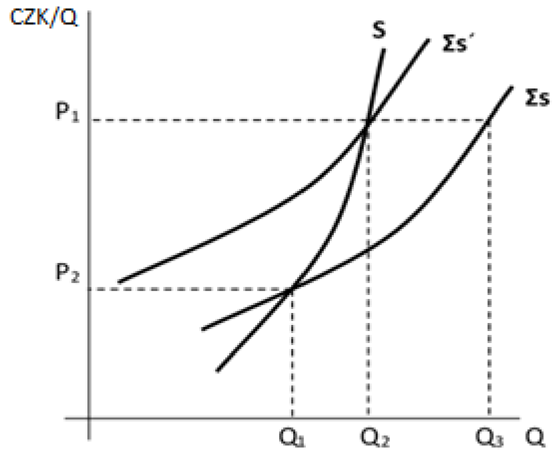

Fig. 58 Impact of input price increases on industry supply

If we examine the extent to which firms in an industry adjust their output in response to changes in market prices, we can construct an indicator of short-term supply elasticity. It is very similar to the demand elasticity indicator discussed in Chapter 3. The coefficient of supply elasticity, denoted as ePS, is defined as the ratio of the percentage change in the quantity supplied to the percentage change in price:

$$e_{PS} = \frac{\frac{\delta Q}{Q}}{\frac{\delta P}{P}} = \frac{\delta Q}{\delta P} . \frac{P}{Q}$$

Given that the quantity supplied is an increasing function of price (**the ratio *δQ/δP* is positive**), the coefficient of elasticity for short-term market supply will have positive values.

5.5 Equilibrium of a Perfectly Competitive Industry in the Short Run

The equilibrium of a perfectly competitive industry in the short run is achieved when the market "clears," meaning that at the short-run equilibrium price P* , the quantity demanded and the quantity supplied of a given good are equal (Q*) – see Figure 58b. Neither buyers nor sellers have any incentive to change this quantity. The market supply S and market demand D are determined by the horizontal summation of individual supply and demand curves. The combination P*Q* represents the equilibrium between the demands of all individuals and the costs of all firms. The equilibrium price serves two important functions:

1. It acts as a signal for producers in their decision-making regarding the output level. Profit-maximizing firms will produce an output level where the equality P* = MC holds. The total industry output will be Q*.
2. It serves as a signal for individual buyers. At the market equilibrium price P* , utility-maximizing individuals decide how much of their income to allocate to the purchase of the good. The market quantity demanded will be Q*

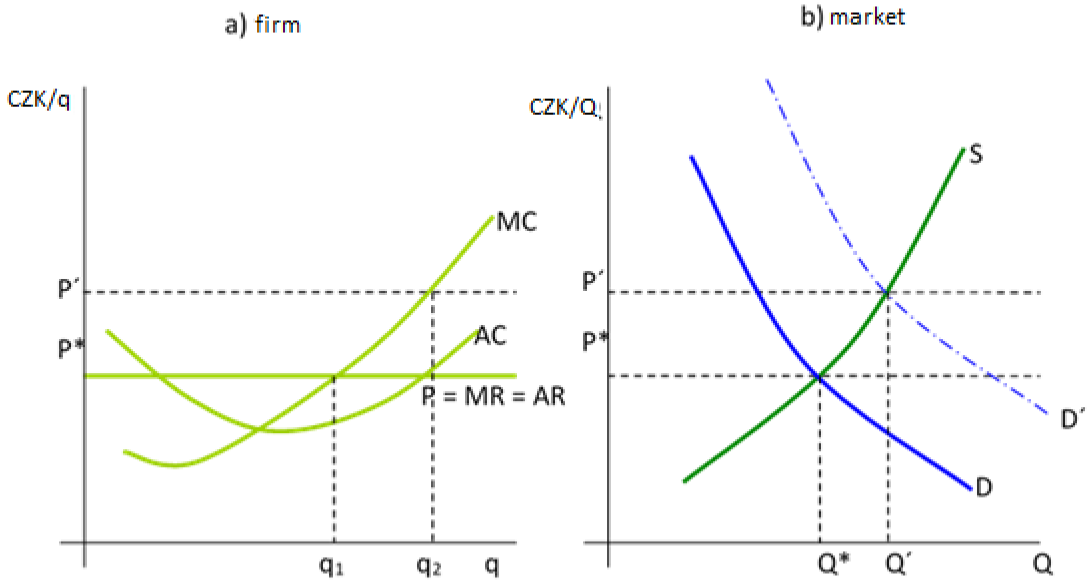

Fig. 59 The equilibrium of a perfectly competitive market and the firm's optimal output

The impact of the equilibrium price on the decision-making of a representative perfectly competitive firm is illustrated in Figure 57a: the firm accepts the price P∗ and produces the optimal output q_1. This price is higher than its short-run average costs, meaning the firm earns an economic profit.

If market demand increases, the demand curve shifts to D´ , resulting in an increase in both the equilibrium price and quantity in the market (P´, Q as shown in Figure 58b. At the same time, the output offered by an individual firm will increase to q_2. It is important to note that, all else being equal, the new equilibrium price P´ allows each firm to earn a higher profit.

The primary factors causing changes in market supply (shifting the supply curve) include changes in the prices of inputs used by firms, technological advancements, producer expectations, and changes in the number of firms in the market. Assuming that the market demand function remains unchanged, a downward shift of the market supply curve due to one of the aforementioned factors will influence changes in the equilibrium price and quantity, depending on the elasticity of market demand:

- In the case of highly inelastic market demand, a rightward shift in the supply curve will lead to a relatively large decrease in the equilibrium market price and a small increase in the equilibrium quantity. For example, if the supply of salt significantly increases, consumers would not substantially increase their quantity demanded even though the market price would be much lower.
- In contrast, with highly elastic market demand, a rightward shift of the supply curve will result in a relatively small decrease in the equilibrium market price and a considerable increase in the equilibrium quantity.

The main factors causing changes in market demand (manifesting as a shift in the demand curve) include changes in consumer incomes, preferences, expectations, and the prices of substitutes and complements. Assuming no changes in the market supply function, an upward shift of the demand curve due to these factors will influence changes in equilibrium price and quantity depending on supply elasticity:

- If market supply is inelastic, an increase in market demand will lead to a substantial increase in the equilibrium market price and only a small increase in the equilibrium quantity.
- Conversely, with highly elastic market supply, an increase in market demand will lead to a relatively small rise in the equilibrium price and a significant increase in the equilibrium quantity

When simultaneous changes occur in supply and demand in a perfectly competitive market, the resulting change in market equilibrium will depend on the elasticity of both supply and demand and the relative magnitude of their shifts.

Definition

Equilibrium in a perfectly competitive industry in the short run occurs when the quantity demanded equals the quantity supplied Q at the equilibrium price P*. Firms produce the quantity at which price equals marginal cost (P* = MC), and individuals maximize their utility based on the price P*. Changes in supply or demand affect the market equilibrium price and quantity, depending on the elasticity of demand and supply.*

5.6 Firm Decision-Making on Output in the Long Run

When considering the long-term optimal output of a perfectly competitive firm, we apply the "golden rule" of profit maximization under long-term conditions. It is assumed that the firm can change the volume of all inputs it uses. Therefore, ***the optimal output of the firm in the long run is derived from the equality of marginal revenue and long-run marginal cost*** (P = MR = LMC).

In the long run, the firm's optimal output is also influenced by another crucial factor: the free entry or exit of firms into or out of the industry. Free entry into the industry means that firms wishing to invest capital and compete with existing firms can do so. There are no barriers such as secret agreements among producers, trademarks, patents, or licenses. The model of perfect competition assumes that entry or exit of firms entails no costs. The long-run period represents a genuine timeframe for the emergence of new firms and the dissolution of existing ones. Consequently, the number of firms in the industry is determined not only by shifts between different industries but also by processes of formation and closure of individual enterprises.

If firms within an industry are generating economic profit, it will serve as an incentive for other firms to enter. A larger number of firms in the industry will lead to a greater volume of output, which will result in an increase in market supply, shifting the industry supply curve downward to the right. Under otherwise unchanged conditions, the market price and firm profits will decline. The influx of new firms will continue until the market price drops to the level of average costs (P = AR = LAC) and economic profit reaches zero. At this point, no new firms enter and none leave the industry, as the alternative use of their resources in another sector would yield the same return. The number of firms in the industry can thus be considered at equilibrium.

A similar process occurs if firms in the industry experience a short-term loss. Many firms will choose to leave the industry for this reason. As a result, the overall industry output decreases, and the market supply curve shifts upward to the left. Under otherwise unchanged conditions, the market price will rise. When the price reaches the level of average costs, firms will achieve zero economic profit and will have no reason to leave the industry, as they would earn the same rate of return elsewhere.

Thus, in a perfectly competitive industry, it is not possible for total revenues to exceed total costs indefinitely, nor for total revenues to remain below total costs over the long term. In the long run, a firm moves toward equilibrium between total revenues and total costs, or zero economic profit. This point of revenue-cost equilibrium is referred to as the break-even point.

Points to remember

In the long run, a firm operating under perfect competition achieves optimal output when its marginal revenue equals its long-run marginal costs (P = MR = LMC), with the ability to adjust all inputs. Free entry and exit of firms into and out of the industry ensure that market supply and prices adjust such that economic profit stabilizes at zero in the long term. If firms generate a profit, new firms enter the industry, increasing supply and reducing the price until it reaches the level of average costs (P = AR = LAC). Conversely, if firms incur a loss, some firms exit the industry, decreasing supply and increasing the price until it stabilizes at a level covering average costs. The result of these processes is that, in the long run, firms reach a point where total revenues cover their total costs, known as the break-even point.

The above shows that the optimal output of a perfectly competitive firm in the long run is determined by two conditions:

1. The necessary condition for profit maximization: ***P ≡ MR = LMC***;
2. The condition of zero economic profit: ***P ≡ AR = LAC***.

Note: *At first glance, these two conditions seem contradictory. This is because they are of different natures. The profit maximization condition expresses the assumed goal of the firm, i.e., the firm actively strives for the highest possible economic profit. Zero profit is not the firm's goal but is something firms are compelled to accept due to their free movement between industries.*

Satisfying both conditions implies the equality of long-run marginal and average costs (LMC = LAC), which occurs at the minimum point of the LAC curve. ***Thus, for the long-run optimal output of a perfectly competitive firm, the firm produces at the minimum of its long-run average costs.***

Given that all firms in a perfectly competitive industry are price-takers (i.e., the price is the same for all firms) and that we assume identical cost curves for all firms, it follows that the minimum of the LAC curve is the same for all firms.

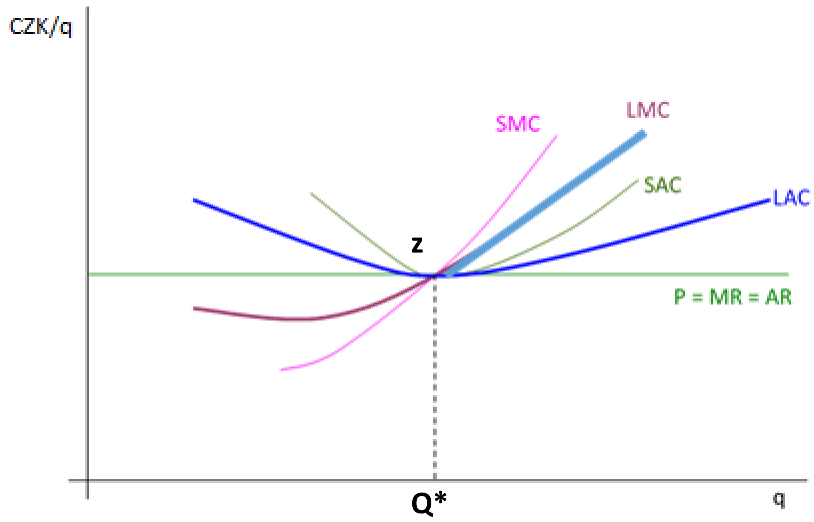

Fig. 60 Optimal Output of a Perfectly Competitive Firm in the Long Run

Figure 59 shows that the optimal output of the firm in the long run is characterized by:

a) The firm produces at its short-term optimal output, where its marginal revenue equals marginal costs (MR = SMC). Since the price equals marginal revenue (P = MR), this also means that the price equals marginal costs (P = SMC). Therefore, the firm has no incentive to either increase or decrease its output.
b) The average costs of producing the chosen optimal output in the short term are at the level of minimum average costs in the long run (min. SAC = min. LAC).
c) The firm has no incentive to cease production and exit the industry as long as, when producing at the optimal output level, both its long-run and short-run average costs equal the price (P = SAC = LAC), resulting in zero economic profit.

The firm's long-run supply curve is therefore identical to the upward-sloping part of the long-run marginal cost curve, with its lower boundary at the minimum of long-run average costs.

5.6.1 Supply of a Perfectly Competitive Industry in the Long Run

When constructing the **Long Run Industry Supply Curve, LIS**, we cannot proceed in the same way as with the short-run industry supply curve, which is obtained by horizontally summing the individual supply curves of the firms. This is because, in the long run, individual firms react to changes in market price and the existence (or non-existence) of positive economic profits by entering or exiting the industry, resulting in a highly variable industry output over time. When constructing the supply curve of an entire perfectly competitive industry in the long run (Long Run

Industry Supply Curve, LIS), we cannot proceed in the same way as with the short-run industry supply curve, which is obtained by horizontally summing the individual supply curves of the firms. This is because, in the long run, individual firms respond to changes in market price and the existence (or absence) of positive economic profits by entering or exiting the industry, resulting in significant variability in industry output over time. The long-run supply curve of a perfectly competitive industry (LIS) is constructed as a ***set of long-run equilibrium points of the industry***, which emerge at the intersections of the shifting demand curve and short-run supply curves. To graphically represent the LIS curve, at least two long-run equilibrium points are needed, which can be connected to form the LIS curve. The initial market equilibrium represents one of these points, while the second is a newly established market equilibrium resulting from changes in market demand and the subsequent response of firms.

A change in market demand will be considered as the stimulus leading to changes in the supply of individual firms and the industry as a whole, ultimately creating a new equilibrium point. The reaction of the firm and the industry to changes in market demand can be viewed from both a short-term and a long-term perspective. In the short run, firms respond to changes in market demand by adjusting the quantity supplied to match the price with ***marginal costs***. In the long run, firms enter or exit the industry, causing shifts in the short-run market supply and the establishment of a new equilibrium price at which firms realize zero economic profit. In such a situation, the movement of firms between industries ceases, and the number of firms in the industry can be considered in equilibrium.

5.6.1.1 The Long-Run Industry Supply Curve (LIS) with Constant Input Prices

The formation of the LIS curve is illustrated in Figure 60, where part (a) shows the cost curves of a typical firm, and Figure 60b shows the short-run market supply (S) and market demand (D) curves.

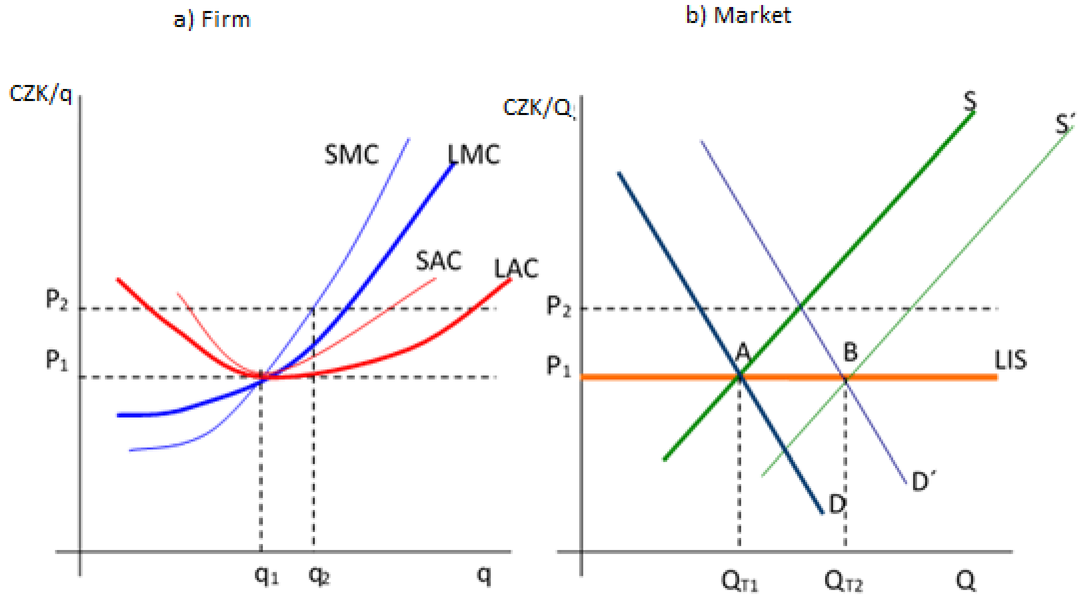

Fig. 61 Long-run supply of a perfectly competitive industry with constant input prices

In the short run, the industry is initially in long-run equilibrium (point A in Figure 60b). At equilibrium price P_1 , a typical firm produces a short-run optimal output q_1 (viz obr. 60a), (see Figure 60a), as the condition P = SMC is satisfied. The output q_1 also represents the firm's **long-run optimum** since P = LMC and the **long-run optimality condition *P = min. LAC*** is met. The firm's economic profit is zero, and the number of firms in the industry remains stable.

Now, let us assume an increase in market demand, which shifts the demand curve to the right (D' in Figure 60b). This triggers the following reactions:

a. In the short term, the market price rises to P_2. Each firm increases its output from q_1 to q_2 and earns an economic profit of $(P_2 – SAC) * q_2$.
b. In the long run, the existence of economic profit attracts new firms into the industry, increasing industry supply and shifting the market supply curve to the right. The short-run market supply curve will continue to shift, and the market price will decrease until firms once again earn zero economic profit. This occurs at the intersection of the new demand curve D' and the shifted supply curve S´při at the equilibrium price P_1 and market quantity Q_{T2}. Thus, a new industry equilibrium is established at point B.

The result of these processes is that ***the market price, after temporarily increasing, returns to its original level (P1) in the long run, while industry output increases (from Q_{T1} to Q_{T2})***. For individual firms, positive economic profits eventually turn into zero economic profit, and their optimal output returns to q_1. (An analogous process would occur if market demand decreased, leading to a decrease in industry output).

A perfectly competitive industry is in long-run equilibrium when profit-maximizing firms have no incentive to enter or exit the industry. This state occurs when the number of firms is such that each firm realizes its long-run optimal output, where P = LMC = LAC and their long-run average costs are minimized.

By connecting equilibrium points A and B, we find that ***the long-run industry supply curve (LIS) takes the form of a line parallel to the x-axis***. Its distance from the x-axis is determined by the price level, which equals the minimum long-run average costs.

The LIS curve, represented as a horizontal line at the price level P_1, is one of three potential scenarios. It illustrates the long-run supply of an ***industry with constant costs***. Such an industry is characterized by a key assumption: an increase in industry output does not lead to a rise in input prices. This means that the entry of new firms in the long run does not raise the costs of existing firms. This situation occurs, for example, when firms use only a small portion of existing and available resources, or when those inputs are widely used across many sectors. Thus, firms entering the industry already possess these production factors, and their entry does not significantly increase demand, prices, or the costs for firms already in the industry. The ability to expand industry output without rising costs implies that the industry's production volume can continue to grow. However, this growth is limited by market conditions, especially the level of market demand. Therefore, ***a key characteristic of a constant-cost industry is that its output can grow or shrink in the long run without a change in market price***, which is also reflected in the perfectly elastic LIS curve.

5.6.1.2 LIS Curve in the Case of Increasing Input Prices

In most industries, as output increases over the long run, production costs rise. Existing and newly entering firms purchase inputs that are limited in supply, which drives up their prices. An increased number of firms in the industry may also lead to additional external costs, such as environmental pollution. These industries represent the second of the three scenarios, known as increasing-cost industries. o derive the LIS curve for an increasing-cost industry, we follow a similar approach as in the previous case. Starting from a long-run industry equilibrium formed by the intersection of the market demand curve D and the short-run supply curve S (point A in Figure 61b), this intersection determines the equilibrium price P_1 and industry output Q_{T1}. At this price, a firm will produce output q_1 (Fig. 61a).

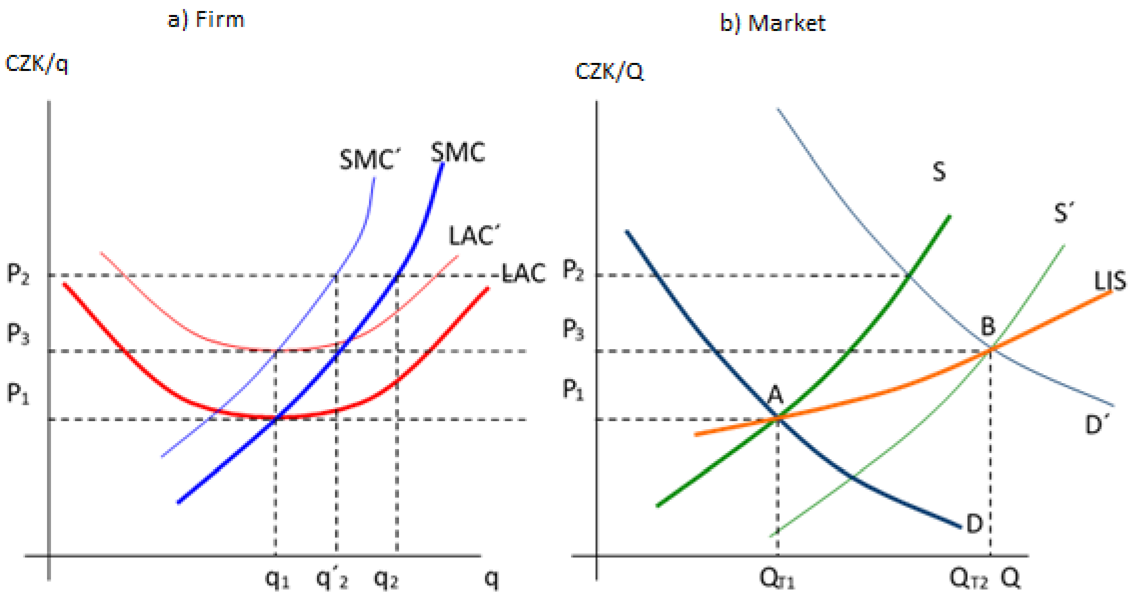

Fig. 62 Long-Run Supply of a Perfectly Competitive Industry in the Case of Rising Input Prices

If there is an increase in demand for a given good, it will be reflected in a rightward shift of the market demand curve D’D'D’ and the emergence of a short-run equilibrium price P_2. Each firm will increase its output along the marginal cost curve until it reaches the level q_2, where P_2 = MC. At the same time, firms will realize economic profit, which serves as an incentive for new firms to enter the industry.

The growing number of firms in the industry will lead to an increased demand for inputs, which in turn drives up input prices. The rise in costs for each firm will cause all cost curves to shift upward, as shown in Figure 61a. Under these circumstances, the representative firm will not produce output q_2, but instead will produce output $q_{2'}$. Even though this output allows the firm to earn economic profit, it is smaller than in the case of constant costs. The influx of firms into the industry will continue until the market supply curve, shifting to the right, reduces the equilibrium market price to a level where economic profits are zero. At this point, no firms will enter or exit the industry, meaning that at price P_3 and market quantity Q_{T2} , a new long-run equilibrium point is reached (point B in Figure 61b). By connecting the long-run equilibrium points of the industry, we obtain the upward-sloping LIS curve.

5.6.1.3 LIS Curve Under Falling Input Prices

In some industries, an increase in output can lead to a reduction in costs over the long run. This can occur due to **external economies of scale**, which are not directly controllable by individual firms, in contrast to economies of scale. For example, as an entire industry develops, improvements in transportation networks and connections might contribute to a reduction in firms' costs. Falling costs can also be observed in newly developing industries or those introducing entirely new products. Initially, only a few firms may be involved in the production of the new product, competing for limited production resources. However, once it becomes clear that there is significant market demand for the product, more firms will emerge, offering specialized inputs to companies in the industry, which can lead to a reduction in production costs.

In this case, we are dealing with ***an industry experiencing decreasing costs***. The graphical representation follows a similar process as in the previous case. The starting point is the long-run market equilibrium point (point A in Figure 62b), which is the intersection of the market demand curve (D) and the short-run market supply curve (S). At the equilibrium price P_1, each firm will produce the optimal output q_1 (as shown in Figure 62a).

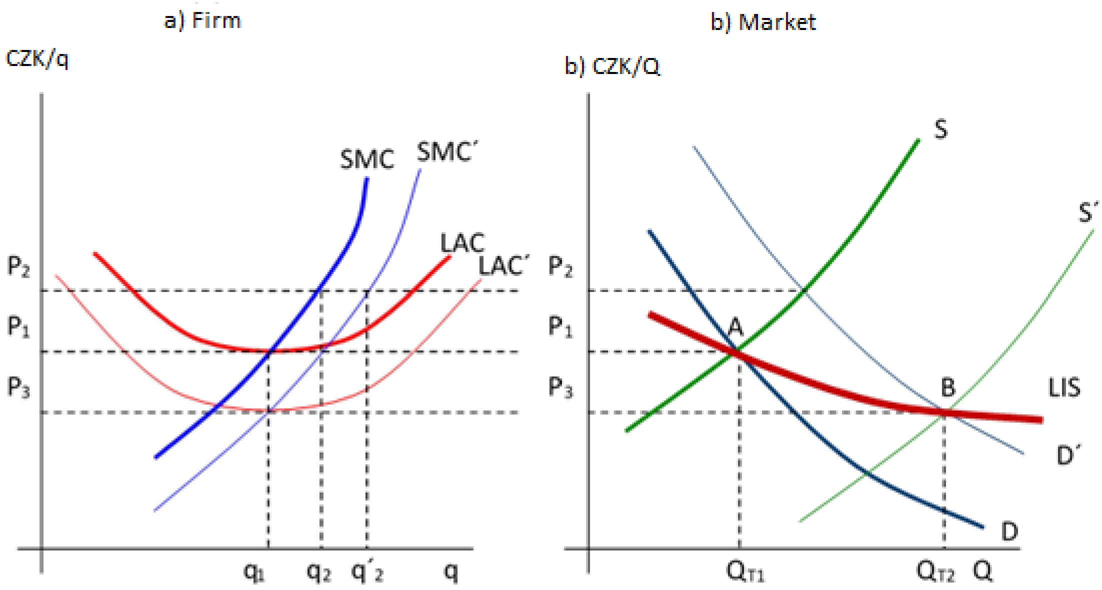

Fig. 63 Long-run supply of a perfectly competitive industry in the case of falling input prices

f there is an increase in market demand (D'), the short-run rise in market price (P2) will allow firms in the industry to achieve economic profit by producing output q_2, and new firms will enter the industry. However, the expansion of the industry will cause a decrease in costs. The cost curves of a typical firm will shift downward. The optimal output of the firm at price P_2 will not be q_2, but q'_2. The substantial economic profit will serve as a strong incentive for new firms to enter the industry. This process will continue until the downward shifting short-run market supply curve reduces the market price to the level where the firms' long-run average costs are equal to the market price, thus ensuring zero economic profit for each firm. A new long-run equilibrium of the industry will then occur at price P_3 and market quantity QT_2 (point B).

By connecting the equilibrium points A and B, we get a decreasing LIS curve. Thus, in industries with decreasing costs, an increase in industry output is accompanied by a decrease in prices.

5.6.1.4 Elasticity of Market Supply in the Long Run

The LIS curve reflects both the internal adjustment of firms to price changes and the changing number of firms in the industry, along with the nature of cost conditions. All these factors are captured by the long-run elasticity of supply, which expresses the ratio of the percentage change in industry output to the percentage change in price. Formally, it is the same expression as in the case of short-run supply elasticity.

However, the conclusions about the value of the long-run supply elasticity coefficient are not the same as those for short-run supply elasticity, which is always positive.

a) ***For industries with constant costs***, where the LIS curve is perfectly elastic, the value of the long-run supply elasticity coefficient is infinite, because the output of the industry can change without affecting the market price.
b) ***For industries with increasing costs***, where the LIS curve is upward sloping, an increase in market price will lead to an increase in market quantity (see Figure 61b), and the long-run supply elasticity coefficient will be positive.
c) ***For industries with decreasing costs***, the LIS curve is downward sloping. An increase in output is accompanied by a decrease in market price (see Figure 62b), so the long-run supply elasticity coefficient will be negative.

5.7 Efficiency of Perfect Competition Mechanism

In answering the question of the efficiency of perfect competition, it is necessary to distinguish between productive efficiency and allocative efficiency.

Productive efficiency means that output is produced at the minimum cost. In perfect competition, productive efficiency exists: the free movement of firms between industries ensures that each firm produces output at its minimum long-run average cost. Since all firms in perfect competition are considered identical, the entire industry, in long-run equilibrium, produces output at the minimum long-run average cost.

Allocative efficiency means that firms produce the output that consumers desire the most. It is not just about firms producing at the lowest possible cost, but also about producing goods that consumers are interested in. The answer to the question of whether allocative efficiency exists under perfect competition can be found in the supply and demand curves. The supply curve is derived from the increasing portion of the firm's marginal cost (MC) curve, as the equilibrium price equals the additional cost of producing the last unit sold (P = MC). The demand curve is based on the utility that the consumer derives from the last unit of the purchased good, meaning the price is determined by how much the consumer is willing to pay for that last unit (P = MU, where MU is marginal utility). At the point where the supply curve intersects with the demand curve, the condition MC = MU holds. This means that at the equilibrium price and quantity, the firm's cost of producing the last unit is equal to the utility that the consumer derives from consuming the last unit. The firm cannot increase output by reallocating inputs, just as the consumer cannot increase their utility by reallocating resources. ***The economy is in a state of allocative efficiency.***

Apart from agricultural markets, there are very few other industries that are perfectly competitive. That is, under the main assumptions mentioned, the demand for their production is perfectly elastic, the products they manufacture are homogeneous, and the movement of firms between industries is completely free. However, the significance of analyzing a perfectly competitive market lies in the fact that many markets are nearly perfect: the demand for production in these industries is highly elastic, and entry and exit from the industry are relatively easy. Firms in such markets typically decide to produce an amount of output where their marginal costs approximately equal the market price.

Definition

The efficiency of perfect competition is divided into two types – productive and allocative. Productive efficiency occurs when a firm produces at the lowest possible cost, which is the case in perfect competition. The free movement of firms between industries ensures that every firm produces at minimum long-run average costs, which applies to the entire industry in long-run equilibrium. Allocative efficiency occurs when firms produce exactly what consumers want the most. Under perfect competition, allocative efficiency is achieved when the price equals marginal costs (P = MC) and marginal utility (P = MU). The economy reaches a state where neither output nor consumer utility can be increased.

5.8 Producer Surplus

When a perfectly competitive firm decides to maximize its profit and produce a certain amount of output, it must be more profitable for the firm to produce than to produce zero output. ***The difference between what the firm earns from producing the optimal output and what it would earn if it produced zero output*** is called ***producer surplus***.

In the ***short run***, we can define this difference as:

$$\left[P^* q^* - TC(q^*)\right] - \left[P^* .0 - TC(0)\right]$$

Adjusting it $$TR(q^*) - TC(q^*) - 0 + FC$$

i.e. $$\pi(q^*) + FC$$

or by denoting the maximum profit as π*, we can formulate the ***producer surplus*** in the short run as $\pi^* + FC$

Since fixed costs are constant by nature, changes in producer surplus occur as a result of changes in the market price P** , which lead to changes in short-run profit. Graphically, we can represent these changes as changes in the area between the market price and the short-run individual supply curve (Figure 5.11).

The surplus of all producers in the short run is shown in Figure 63 as the area between the market price P* and the market supply curve S (the horizontal sum of individual supply curves).

In the long run, the producer surplus is defined as zero: economic profit is zero, fixed costs do not exist, therefore:

$$(\pi^* + FC) = (0 + neex.)$$

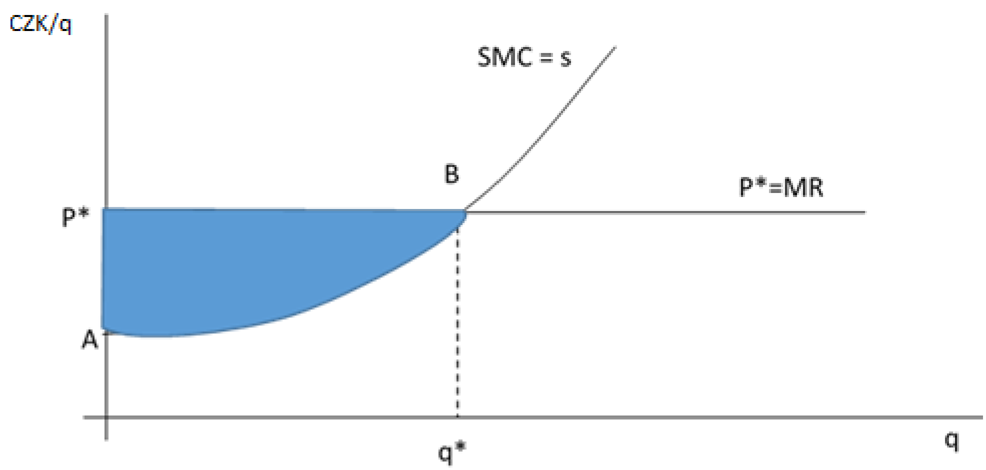

Fig. 64 Producer surplus in the short run.

What about the surplus of all producers in the long run? Since the economic profit of each producer is zero in the long run, it does not matter to them on which specific market they operate, as they could realize the same return on any other market. The parties who do care about the level of output produced in a specific industry are the owners of the factors of production. If we assume an industry with increasing costs, the entry of additional firms into the market will lead to an increase in the prices of certain inputs, and the situation for the owners of those inputs will improve.

The producer surplus in the long run represents the additional returns to the owners of the inputs used in the industry, above the returns that would be realized by these inputs if the volume of output in the industry were zero.

Graphically, it is represented similarly to the producer surplus in the short run, with the difference being that the market supply curve is now represented by the long-run market supply curve (LIS). It is the area above the long-run market supply curve and below the equilibrium market price P∗. In the case of an industry with constant costs, where the market supply is perfectly elastic, this area does not exist, and the owners of the inputs do not realize these additional returns. In the case of an industry with increasing costs, the market supply curve is upward sloping, so as the market output increases, the additional returns to the owners of inputs increase as well. The issue of producer surplus in the long run can be summarized by stating that, although we use the market supply function (LIS) to measure this surplus, it is not realized by the producers of market output but by the owners of the inputs used in production in the industry.

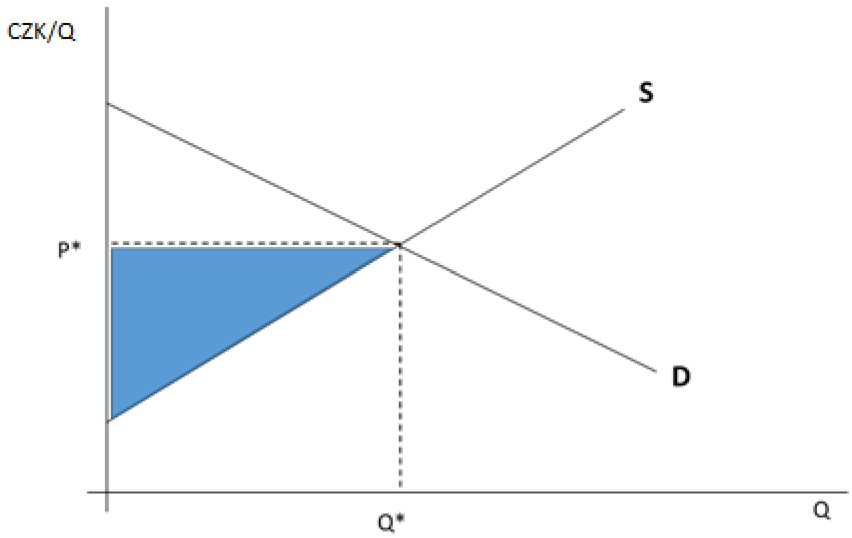

Fig. 65 Surplus of all producers in the short run

Σ

Perfect competition is a theoretical market model characterized by several key features. In such a market, a large number of small firms operate, selling a homogeneous (identical) product. No firm has enough market power to influence the price of the product, so all firms are "price takers." Entry and exit from the market are free, and there are no significant barriers. All market participants (both sellers and buyers) have perfect information about the prices and quality of the products.

Under conditions of perfect competition, firms aim to maximize their profit by producing the amount of output where marginal costs equal the price of the product (which is also equal to marginal revenue). In the long run, when new firms can enter the market, economic profits decrease to zero, and firms earn only normal profit. Perfect competition is considered an efficient market structure because it leads to optimal allocation of resources and maximization of social welfare. However, in the real world, pure perfect competition rarely occurs and serves more as a theoretical benchmark for comparison with other market structures.

In the short run, producers can enjoy a surplus if the market price is higher than their average costs. This situation is represented in a graph, where the area between the market price and market supply (formed by the horizontal sum of individual supply curves) represents the surplus of all producers. Thus, in the short run, producers realize economic profits, meaning they receive more for their products than what it costs to produce them.

Economic profit in this period may encourage existing producers to increase production and attract new firms into the industry. This process can lead to increased competition and potentially a reduction in prices as new producers try to capture market share. Therefore, the short run is crucial for analyzing market dynamics and its responses to changes in supply and demand.

However, in the long run, the situation changes. Once the market reaches equilibrium, the economic profit of all producers approaches zero. This is because new firms enter the market to take advantage of profit opportunities, leading to profit equalization across the industry. In this context, producers become indifferent as to which specific market they operate in, as they could achieve the same returns on any market.

From the producers' perspective in the long run, the question arises regarding the impact on the owners of the factors of production. Assuming an industry with increasing costs, the entry of additional firms into the market will cause the prices of some inputs to rise. This improves the situation of the owners of those inputs, leading to additional returns for them.

The producer surplus in the long run is defined as the additional returns to the owners of production factors, which arise above the returns these owners would receive if the output volume of the industry were zero. Graphically, the producer surplus in the long run is represented as the area above the long-run market supply curve (LIS) and below the equilibrium market price.

An important factor in analyzing producer surplus in the long run is the type of cost structure in the industry. In industries with constant costs, where the market supply curve is perfectly elastic, producer surplus does not exist. In such cases, input owners do not realize additional returns because prices remain stable, and economic profits approach zero.

In contrast, in industries with increasing costs, the level of additional returns to input owners grows as output increases. In this way, input owners can profit from growing production, which benefits the entire economy. Therefore, producer surplus in the long run represents a key mechanism through which the interests of producers and the owners of production factors are balanced.

The chapter on producer surpluses concludes an important aspect of economic analysis of market behavior. In the short run, producers have the opportunity to achieve economic profits, which can lead to higher competition and market dynamics. On the other hand, in the long run, profits are equalized, and the producer surplus is actually realized as additional returns to the owners of production factors. This dynamic is crucial for understanding how markets operate and how production processes evolve in response to changes in costs and competition.

?

Review questions

1. Explain why and how the graphical representation of a firm's total revenue differs under perfect and imperfect competition.
2. Explain why and how the graphical representation of marginal and average revenue differs under perfect and imperfect competition.
3. Explain why, under conditions of imperfect competition, marginal revenue decreases twice as fast as average revenue.
4. Explain and graphically represent the determination of price and output in a perfectly competitive market.

5. Explain under what circumstances a firm continues to produce even if it is making a loss.
6. The agricultural market is often cited as an almost perfectly competitive market. Suppose there is an increase in oil prices on the global market. Will this affect the shape of the long-run supply curve for agriculture?
7. Reflect on whether the following statement is true: total profit is maximized when the same output is produced where the maximum additional profit per unit of production occurs.
8. Graphically represent the supply curve of a perfectly competitive firm in the short run, assuming its short-run production function is characterized only by diminishing returns to a variable input (with a constant price of the variable input).

Example

Imagine that you are the manager of a watch manufacturing company. This company is one of many in a perfectly competitive market. Its production costs are given by the equation: $TC = 20{,}000 + Q_2$

a) If the market price of the watches is 500 CZK, how many watches would you suggest producing to maximize profit?

b) How large would the profit be?

? Test questions

1. In perfect competition, firms are:

a) Price makers
b) Price takers
c) Price manipulators

2. In the long run under perfect competition:

a) Firms make economic profit
b) Firms make only normal profit
c) Firms always make a loss

3. The producer surplus in the long run is defined as:

a) Economic profit of firms
b) Additional revenues of the owners of production factors
c) The difference between market price and marginal costs

4. In an industry with constant costs:

a) Producer surplus in the long run does not exist
b) Producer surplus is maximized
c) Producer surplus increases linearly with production

5. Short-run economic profit in perfect competition:

a) Is permanent
b) May attract new firms into the industry
c) Is banned by market regulators

Answers:

1.b, 2.b, 3.b, 4.a, 5.b

Practical task

Create a simple model of a perfectly competitive market using a spreadsheet (e.g., Excel). Define the demand and supply functions for a hypothetical product. Set initial conditions (e.g., number of firms, fixed and variable costs) and simulate how the market equilibrium changes in the short and long run with the entry of new firms into the market. Create graphs illustrating changes in price, quantity, and firm profits over time. Present your simulation to the class and explain how the model illustrates the key principles of perfect competition.

Explore the online e-book market (e-books) on at least three different platforms (e.g., Amazon Kindle, Google Books, Apple Books). Select 5 popular titles and compare their prices across platforms. Also, note the availability, format, and other relevant information. Based on your research, write a short report (500-700 words) evaluating the extent to which this market reflects the characteristics of perfect competition. Consider factors such as the number of sellers, product homogeneity, customer information, and barriers to entry. Discuss any deviations from the perfect competition model and their reasons.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie*. 5. aktual. vydání. Praha: Management Press, 2010. s. 117–154. ISBN 978-80-7261-218-5. 574 s.

[2] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005.

[3] MACÁKOVÁ, Libuše; SOUKUPOVÁ, Jana. *Mikroekonomie – repetitorium*. Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[4] HOLMAN, R. *Ekonomie*. Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[5] HOLMAN, R. *Mikroekonomie.* Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[6] JUREČKA, V. a kol., 2018. *Mikroekonomie. 3. akt. vyd.* Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[7] SOUKUPOVÁ, J., 2018. *Mikroekonomie. 6. akt. vyd.* Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 6

Firm Decision-Making as a Monopoly

After studying this chapter, you will be able to:

- Explain the principles of a firm's decision-making regarding output and pricing in a monopoly;
- Describe the different levels of price discrimination employed by monopolies;
- Clarify a firm's decision-making on output and pricing in monopolistic competition.

Keywords:

Monopoly, price maker, price discrimination, monopolistic competition, productive efficiency, allocative efficiency.

After studying the chapter, you will be able to

- Understand how a monopoly determines output and pricing in the context of imperfect competition.
- Identify barriers to market entry and their impact on a monopoly market structure.
- Analyze how a monopoly chooses optimal output and sets prices using markup pricing rules.
- Recognize different forms of price discrimination and their effects on market efficiency.

Chapter preview

- This chapter focuses on the decision-making processes of firms operating as monopolies, emphasizing the choice of optimal output and price setting. Special attention is given to the characteristics of imperfect competition, market entry barriers, and how monopolies exploit their market position to maximize profits. The chapter also explores various forms of price discrimination and their impact on market allocative efficiency. Additionally, it delves into the monopoly supply curve and the mechanisms leading to allocative inefficiency within a monopolistic environment.

Chapter objectives

- Familiarize yourself with the characteristics of imperfect competition and understand the role of market entry barriers in a monopolistic environment.
- Understand the process by which a monopoly determines optimal output and pricing.
- Comprehend markup pricing rules and their influence on a monopoly's decision-making.
- Learn to analyze monopoly profit and its impact on economic dynamics.

Estimated study time

- Cca 240 min.

6.1 The monopoly firm's output-price decision making

A firm's behavior differs when it can influence the price of the product it produces (i.e., it acts as a price maker) under conditions of imperfect competition. Imperfect competition generally refers to a market situation where there is at least one seller or buyer capable of influencing the market price. In the following analysis, we will assume imperfect competition on the producer side. We assume perfect competition on the buyer side, meaning a large number of buyers, none of whom can affect the price of the given good. Within imperfect competition on the supply side, we distinguish between three types of market structures: monopoly, oligopoly, and monopolistic competition. We will first analyze monopoly behavior, then monopolistic competition, and finally, the oligopolistic market structure.

6.1.1 Main Causes Leading to the Emergence of a Monopoly

A monopoly represents the opposite of perfect competition. While the supply on a perfectly competitive market is comprised of the production of many small firms, the supply of an industry with a monopoly is provided by only one firm. In the model of perfect competition, a large number of firms produce identical products; in contrast, the monopoly model assumes a single firm producing a product with no close substitutes. The existence of a monopoly as the sole supplier also implies the absence of competition. A monopoly is represented by a single firm in the market and is often emphasized with terms such as "pure monopoly" or "absolute monopoly."

A monopolistic market is characterized by the presence of a single supplier. At any given time, the monopolist can decide on the level of price or the quantity of output (but not both simultaneously). Within microeconomic analysis, it is commonly assumed that the firm primarily chooses the level of output that allows it to maximize profit, and from this optimal output level and market demand, it derives the market price it sets. There is a relatively simple reason why only one firm operates in a given market: other firms either do not want to enter or are unable to do so. The inability to enter the market is related to barriers to entry that become a source of monopoly power. ***The main barriers to entry include:***

a) ***The average costs*** of a particular firm reach their minimum at a larger output than the market demand (price is higher than average costs, thus the firm realizes profit). Production technology used by relatively large firms allows them to produce at relatively low costs. If the market demand were served by multiple firms, their individual demand curves would shift to the left, leading to an increase in average costs.
b) In an attempt to reduce average costs, firms would try to increase the volume of production, lowering prices. The winner of such a price war would eventually become a single firm—a monopoly. This situation, where one firm can satisfy market demand with lower average costs

than if there were multiple smaller firms in the industry, results in what is known as a **natural monopoly**.

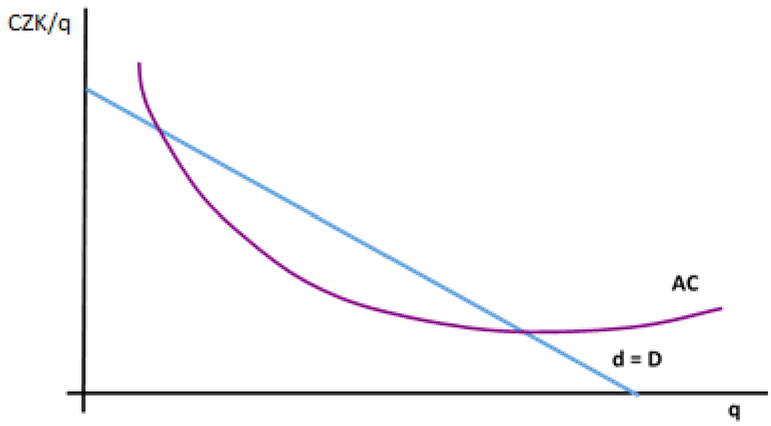

Fig. 66 Natural monopoly

c) ***Control of Resources Necessary for Production:*** A monopoly can emerge when a single firm controls the essential resources required for production. For example, Alcoa controlled all sources of bauxite, a key input for aluminum production, prior to World War II and was consequently the sole producer of aluminum in the U.S. for a long period.
d) ***Government Intervention:*** A monopoly may also form "artificially" as a result of non-market factors, particularly through state intervention in the economy. The government may grant a particular firm exclusive rights to produce and sell a good in a specific area. In return for such exclusive rights, the firm often agrees to regulatory measures imposed by the state concerning its conduct.
e) ***Legal Restrictions:*** Monopolies can arise from legal constraints, such as patents and copyright protections, which grant exclusive rights to produce or sell certain products or innovations.

6.1.2 Characteristics of a Monopoly

Since the output of a monopoly as the sole producer in an industry is the same as the total industry output, the ***individual and market demand curves are identical***. A monopolist can choose any combination of output and price along the demand curve that it finds most profitable. Its activities are constrained solely by the nature of demand for its product. Because ***there are no competitors in the industry it controls, a monopoly enjoys complete autonomy in its decision-making***.

Unlike a perfectly competitive firm, whose market position limits its decisions solely to choosing the output that maximizes profit, the decision-making process for a monopoly is more comprehensive.

By selecting a level of output, the monopoly simultaneously determines the price it will charge. While a perfectly competitive firm is a price taker, a monopoly is a price maker.

6.1.3 Choice of Optimal Output for a Monopoly

When deciding on the level of output that will maximize profit, a monopoly aims to maximize the difference between total revenue and total cost (by equating marginal revenue and marginal cost). A key aspect in analysing a monopoly's decision-making process is recognizing that revenues and their graphical representation under imperfect competition exhibit ***specific characteristics due to the downward-sloping demand curve.***

To sell an additional unit of output, a monopoly must lower not only the price of the last unit but of all units of output. Consequently, marginal revenue declines more quickly than the price. Even though a monopoly is the sole supplier in a particular goods market, it may be just one of many firms in the market for production factors, where it acts as a buyer and cannot influence input prices. Therefore, our conclusions about costs can be applied to monopoly analysis without modification.

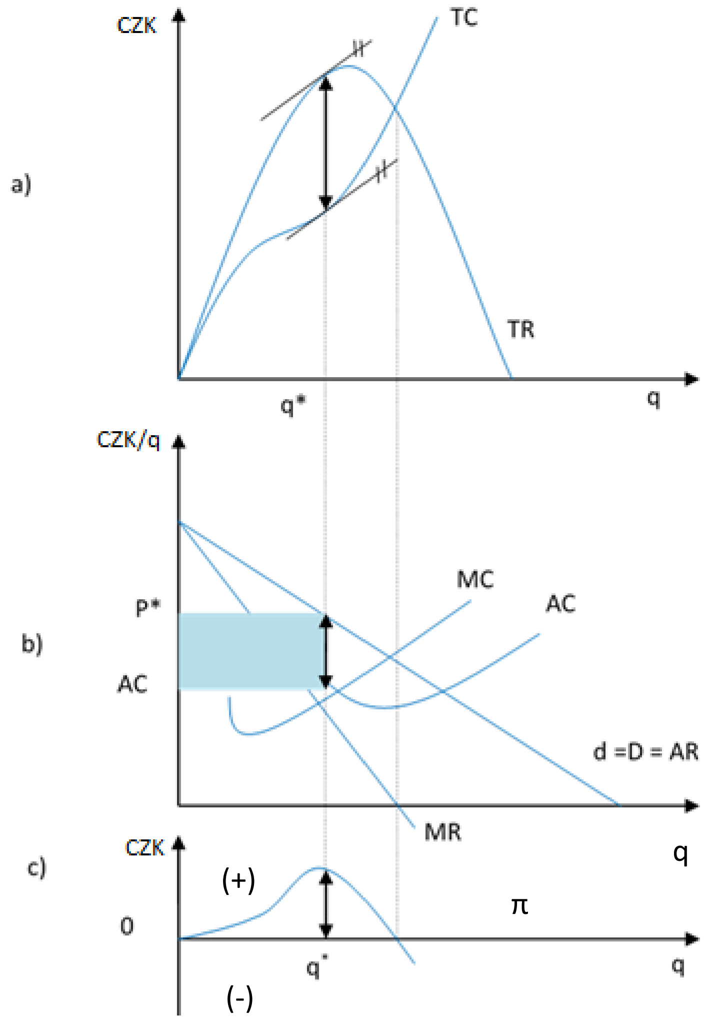

Fig. 67 Alternative Derivation of the Output at Which a Monopoly Maximizes Profit

The ***optimal output*** at which a monopoly maximizes profit can be determined using the same approach as with a perfectly competitive firm, either through an analysis of total quantities (total revenue and total cost) or through marginal quantities (marginal revenue and marginal cost), as shown in Fig. 66.

The necessary condition for maximizing a monopoly's profit is the equality of marginal revenue and marginal cost (see Fig. 66b). The optimal output is q*. If the monopoly were to produce slightly less than q*, its profit would decrease because the revenue loss due to reduced production (MR) would be greater than the cost savings (MC). Similarly, a decision to produce more than q* would also reduce profit, as the additional costs associated with increased output would exceed the additional revenue.

6.1.4 Price Setting by a Monopoly

After determining the optimal output, a monopoly then decides on the price level. Contrary to popular belief, a monopoly cannot set its price at any arbitrarily high level. The price at which it can sell the optimal output is constrained by consumers' willingness to pay. The monopoly must consider the demand curve (see Figure 66). From the figure, we can see that the monopoly will sell output q*at price P*. The quantity q*thus becomes the equilibrium quantity, and the price P* becomes the equilibrium market price. ***The price P*will exceed both the marginal revenue and marginal cost associated with the monopoly's optimal output.***

The inverse elasticity rule implies that the ***difference between price and marginal cost*** for the optimal output is influenced by the price elasticity of demand for the monopoly's product (identical to the elasticity of market demand), and this relationship is ***inversely proportional***. This leads to two implications for the monopoly's pricing strategy:

a) The monopoly should only produce an output level corresponding to the elastic portion of the demand curve, meaning ePD < −1. (If the monopoly produced an output tied to inelastic demand, it would result in negative marginal revenue, which could not equal marginal cost in potential production, as marginal costs are always positive).

b) The more elastic the market demand, the smaller the markup of price over marginal cost.

6.1.4.1 Price Setting Through Markup

Managers base their pricing considerations on covering the average costs associated with the usual output level. The price is then determined by adding a profit markup, "m," to the calculated average costs, defined by the formula:

$$m = P - AC/\ AC$$

We assume that the firm produces the long-term optimal output, achieving minimum long-run average cost (LAC). At this point, the average cost curve intersects with the upward-sloping marginal cost curve. At the minimum point of LAC, the equality LMC=LAC holds, allowing us to reformulate the equation as follows:

$$m = P - MC/MC$$

Applying the golden rule for profit maximization, MR=MC, at the optimal output, we can further modify the equation:

$$M = P - MR/MR$$

6.1.5 Monopoly Profit

A characteristic feature of a monopoly is that it can achieve profit even in the long run, unlike perfect competition, where there is a tendency toward zero economic profit. This is due to barriers to entry in the industry and the inability of potential competitors to enter the market.

Monopoly profit itself is not necessarily a consequence of specific monopolistic behaviour. A monopoly can also realize zero economic profit. Therefore, the actual amount of monopoly profit may not always be a reliable indicator of monopolistic power.

If a monopoly incurs a loss in the short term, it follows the same principles as a perfectly competitive firm. It will minimize the loss by continuing production, provided that the price is higher than the average variable costs.

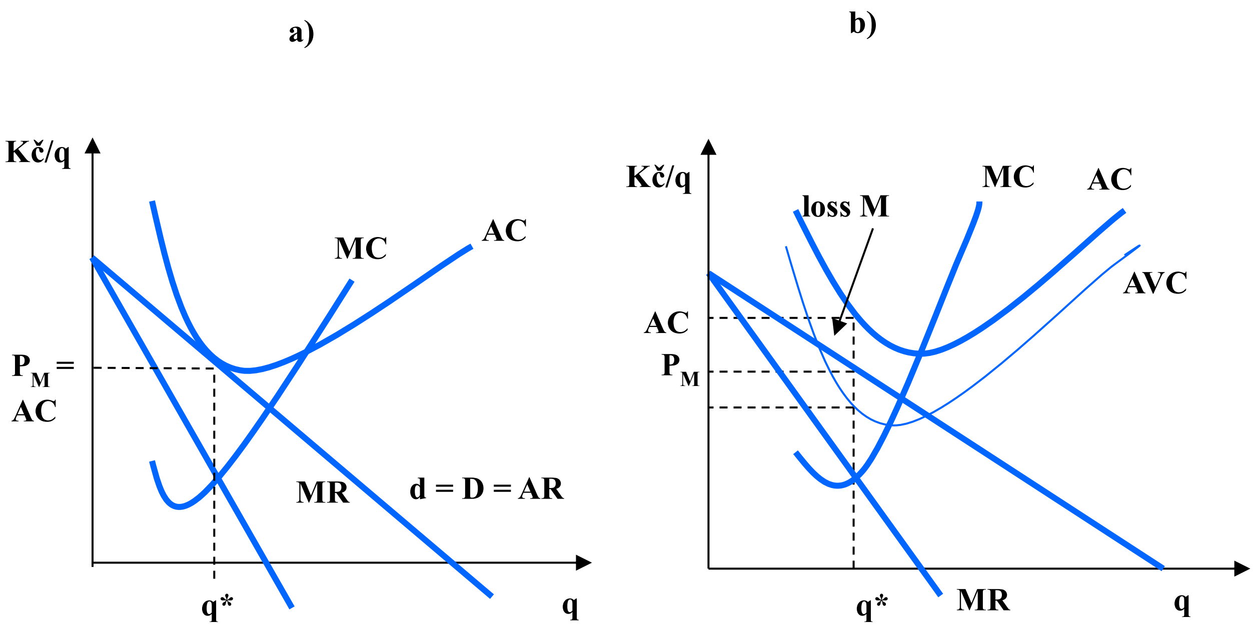

Fig. 68 Monopoly Achieving Zero Economic Profit (a) or Loss (b)

6.1.6 Monopoly Supply Curve

When constructing the supply curve, we encounter a problem. If we aim to create a long-term supply curve for a monopolistic industry, which would be identical to the monopoly's long-term supply curve, complications arise. In the case of forming a long-run supply curve (LRS) for a perfectly competitive industry, we based it on shifts in the market demand curve. These shifts were triggered by the responses of existing firms in the industry and, if firms were realizing profits, by the entry of new firms into the industry, ultimately leading to the formation of a new market equilibrium.

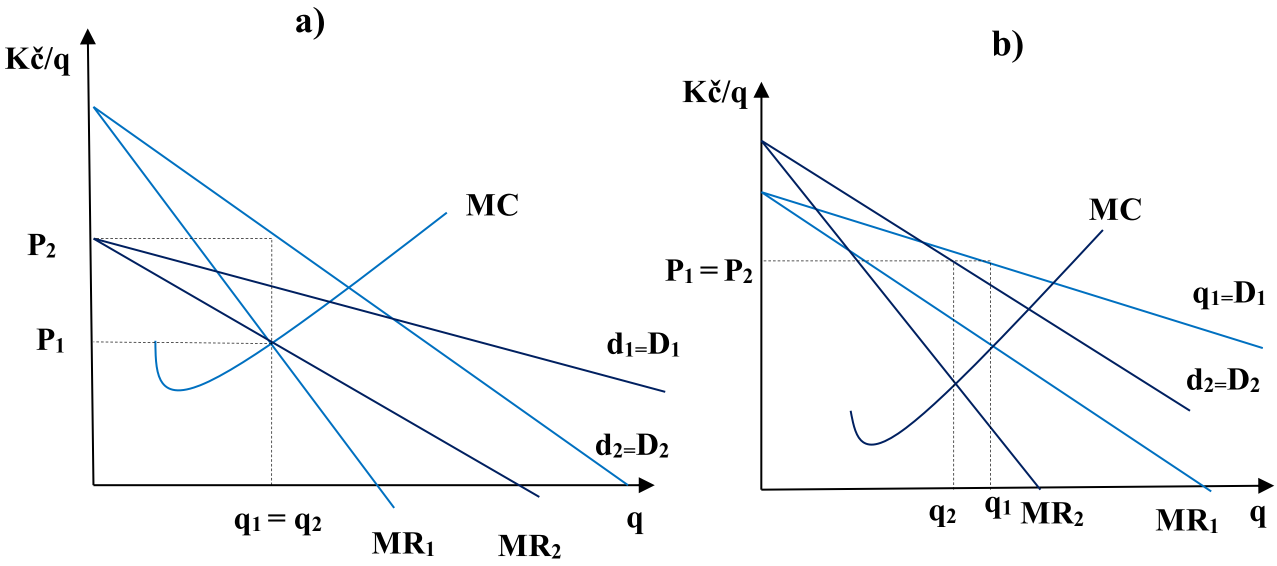

Fig. 69 Ambiguous Relationship Between Price and Output Under Monopoly Conditions

It is clear from the figures that under monopoly conditions, there is no unique relationship between price and quantity supplied. In figure 68a, we can see that the monopoly offers the same output at price P1 as at price P2. In figure 68b, the monopoly offers either output q1 or q2 at price P1. For the reasons described above, the monopoly's supply curve cannot be represented graphically, nor can it be equated with the rising portion of the marginal cost curve.

6.1.7 Price Discrimination

A monopoly possesses a certain degree of monopoly power, which allows it to use ***price discrimination*** as part of its pricing strategy. The goal is to capture consumer surplus and convert it into additional profit for the firm. ***The essence of price discrimination is setting different prices for the same products without cost-related reasons***. Price discrimination is characterized by the setting of different prices (for different consumers or for different quantities) based on factors other than cost considerations. There are various forms of price discrimination in economic reality. We will focus on first- to third-degree price discrimination.

6.1.7.1 First-Degree Price Discrimination

This represents a more or less theoretical situation of ***discrimination based on individual consumers***, where the ***monopoly sets the maximum price*** each consumer is willing to pay for each unit purchased. In doing so, the monopoly captures the entire consumer surplus for itself.

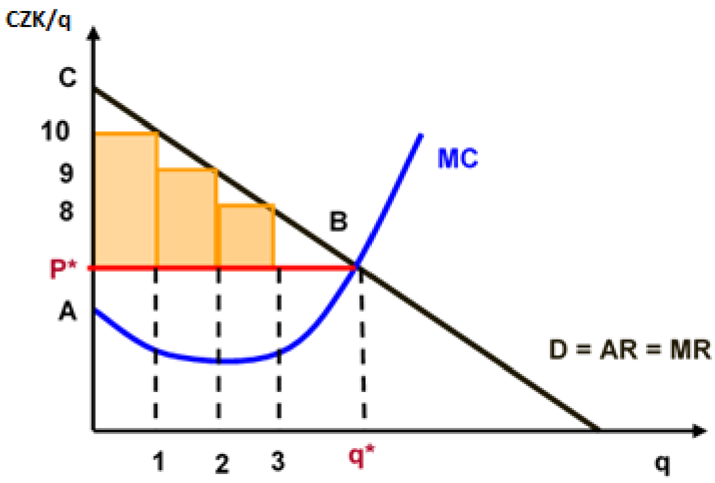

Fig. 70 First-Degree Price Discrimination

In practice, first-degree price discrimination remains only an abstraction, mainly for three reasons:

a) The firm usually does not know the maximum price each consumer is willing to pay for a unit of the goods.

b) Even if the firm were to ask each consumer, it would likely not receive truthful answers, as consumers have an interest in keeping the price as low as possible.

c) The transaction costs involved in determining these maximum prices would be so high that it would likely not be profitable for the firm to engage in such an approach.

Sometimes, however, a monopoly may resort to ***imperfect first-degree price discrimination***, based on estimates of the maximum prices that its consumers are willing to pay (for example, a tax advisor who knows the financial situation of his clients may estimate how much his client will be willing to pay for his services).

6.1.7.2 Second-Degree Price Discrimination

Second-degree price discrimination depends on the quantity sold. Since the monopoly sets different prices for a single consumer depending on the various "blocks" of the quantity purchased, some authors refer to second-degree price discrimination as "Multi-Part Pricing."

If the monopoly did not engage in price discrimination, it would set a price P* and produce an output Q*. Instead, the monopoly divides the produced quantity into blocks and sets different prices for each block. The price for the first block would be determined by the quantity Q1 at the price P1, etc.

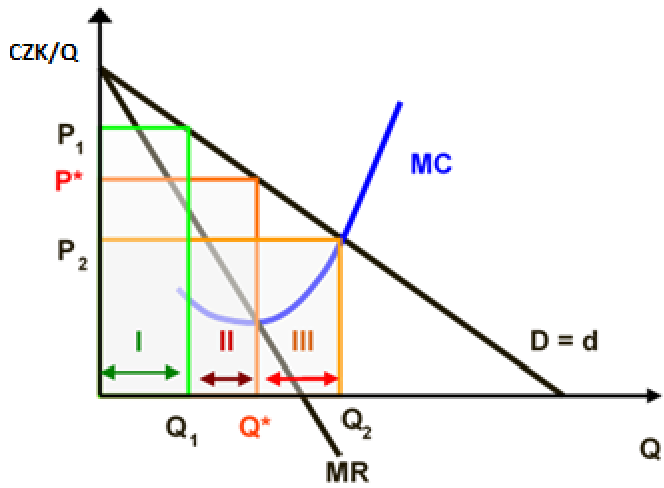

Fig. 71 Second-Degree Price Discrimination

Second-degree price discrimination in the case of a natural monopoly, where market demand is secured with decreasing marginal and average costs, allows for a larger output volume at reduced costs. As a result of the increased output, the firm achieves greater economies of scale, and its profit may rise despite the increase in consumer surplus. This happens because, alongside the decrease in prices, the electricity company realizes cost savings due to lower unit costs.

6.1.7.3 Third-Degree Price Discrimination

This form of price discrimination is similar to first-degree price discrimination in that it involves discrimination based on consumers. ***The essence of third-degree price discrimination is the division of consumers into two or more groups, each with its own demand curve***. In practice, this form is most commonly used, and we must understand its conditions:

a) There must be a criterion for dividing consumers into different groups (market segments). This criterion is typically significant differences in the price elasticity of demand for the product (which can be influenced by varying income levels of different groups, differing preferences, or the availability of substitute products).

b) There must be no possibility of interaction between consumers, because if such interaction existed, someone from the low-price group could sell the product to someone in the higher-price group, which would erase the price differences.

These conditions affect the extent to which a monopoly can use its monopoly power to set different prices for different consumers.

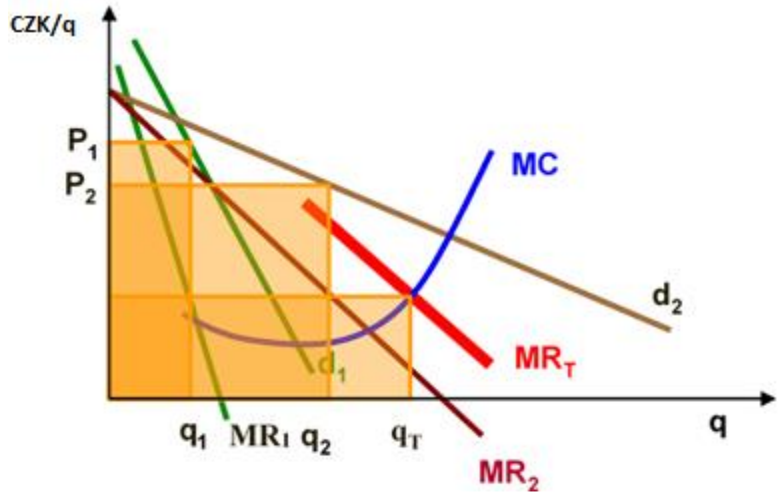

Fig. 72 Third-Degree Price Discrimination

We assume that the monopoly sells to two groups of consumers and decides how much of the total output to sell to each group and at what price. The firm's decision-making is determined by two factors:

a) The monopoly divides the total output between the groups of consumers in such a way that its marginal revenue from selling each part of the output to each group is the same:

$$MR_1 = MR_2$$

If this were not the case, and the marginal revenue from selling to the first group were higher than the marginal revenue from selling to the second group, the firm would shift part of the output from the second group to the first group, which would lower the price for the first group and increase the price for the second group.

b) The marginal revenue from selling to each group is equal to the marginal cost:

$$MR_1 = MC = MR_2$$

If condition (a) is met, and the marginal revenue from selling to both groups is the same, but still greater than the marginal cost (MR > MC), the firm could increase its profit by increasing output. This would lead to a reduction in prices for both groups of consumers, and the marginal revenue for both groups would decrease and align with the increased marginal costs.

The division of consumers into two or more groups, each with its own demand curve, occurs as follows: consumers are divided into two groups. D1 represents the demand curve for the first group, and MR1 is the marginal revenue from selling to the first group of consumers. D2 represents the demand curve for the second group, and MR2 is the marginal revenue from selling to the second group. D2 is shown as more elastic than D1.

An example of third-degree price discrimination could be a method used to boost sales via coupons in newspapers. A person cuts out the coupon and presents it to receive a discount on a purchase.

6.1.7.4 Other Forms of Price Discrimination

A form closely related to third-degree price discrimination is ***time-based price discrimination***. At different times, consumers, divided into groups based on the elasticity of their demand, are charged different prices.

A specific form of time-based price discrimination is ***peak pricing***. Unlike the previously described time-based discrimination, where marginal costs do not change over time, in this specific case, due to capacity constraints during periods of high demand, marginal costs increase. As a result, prices are also increased during peak periods.

6.1.8 Allocative Inefficiency of Monopoly

From the analysis so far, it is clear that, compared to perfect competition, the existence of a monopoly leads to a higher price and lower output. This negative impact of monopoly is expressed in terms of producer surplus and consumer surplus. For comparison, let us first look at the allocative efficiency of perfect competition. If the market were perfectly competitive, output would be at Q_{DK} as shown in Figure 72a), and its level would be derived from the equality of price and marginal cost.

- In a perfectly competitive market, the output level QDK would be determined by the equality of price and marginal cost. The difference between the total utility (area $AEQ_{DK}0$) and the total production cost (area $EQ_{DK}0C$) represents the optimal level.
- In the case of a monopoly, output is lower and price is higher than under perfect competition. Consumer surplus is reduced to area AP_MB, while producer surplus is increased to area P_MBGC.

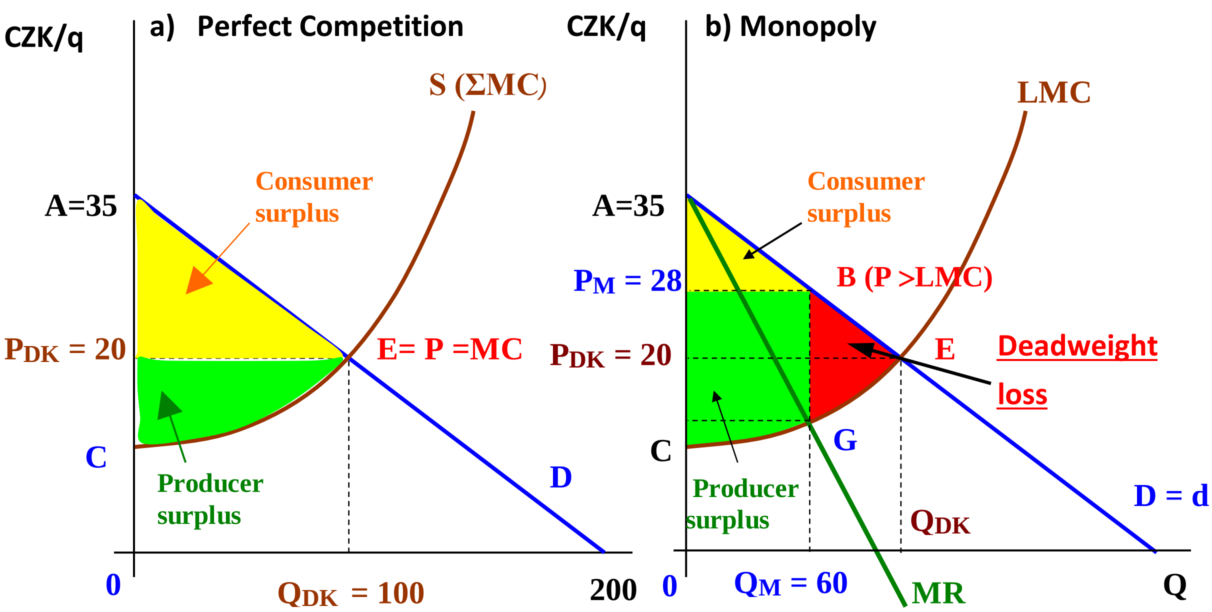

Fig. 73 Comparison of Allocative Efficiency in Perfect Competition (a) and Monopoly (b)

The output of a perfectly competitive market is derived from the equality P=MC. This means that no further improvements can be achieved through continued trade (neither for producers nor consumers, without worsening the position of one of the sides), resulting in what is known as a Pareto-optimal allocation of resources. In contrast, on a monopolized market, output is derived from the price exceeding marginal cost (P>MC), and the monopoly uses its power to prevent further exchanges that would lead to improved Pareto efficiency. In a monopoly that does not apply any form of price discrimination, a Pareto-efficient allocation of resources is not achieved.

6.1.9 Price Regulation of a Monopoly

The goal of monopoly regulation is to eliminate its inefficiency, embodied in deadweight loss. This regulation typically addresses industries such as transportation companies and similar enterprises. Regulation can pertain to price levels and structures, service quality, financial structure of companies, and so forth. A key issue in price regulation is determining the level of the regulated price. A common approach is to set the regulated price at the level of marginal cost (P=MC), applying the principle of perfect competition. However, when the regulated price is set at marginal cost for natural monopolies, it can lead to losses for these monopolies. This effect of price regulation is illustrated in the figure.

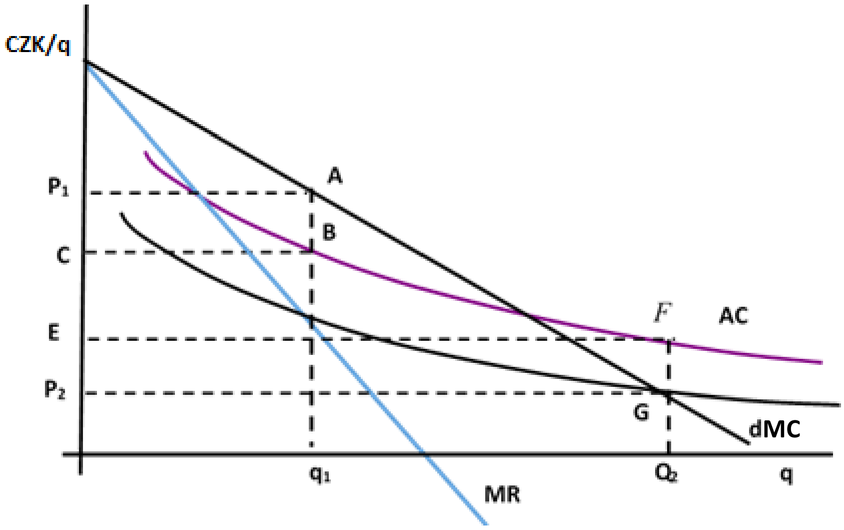

Fig. 74 Price Regulation of a Natural Monopoly

The solution is to set the regulated price at the level of marginal cost. If the monopoly were not regulated, it would produce output q_1 and sell it at price P_1. The profit would be P_1ABC. With the regulated price P_2, the quantity demanded is q_2, and the marginal cost of production is equal to the regulated price. Due to declining average costs, the marginal cost curve lies below the average cost curve. The monopoly incurs a loss in the size of EFG P.

6.1.10 Bilateral Monopoly and Monopsony

Let us abandon the assumption in monopoly analysis that the market demand, where only one producer manufactures and sells, is composed of such a large number of buyers that none of them can influence the market price. In other words, we will not assume perfect competition on the demand side. If the market supply is represented by the output of a single firm and demand by a single buyer, we call it a bilateral monopoly. The existence of a single buyer in the market is referred to as a monopsony. A bilateral monopoly thus represents a special case where there is only one seller (monopoly) and one buyer (monopsony) in the market. In economic reality, such cases are rare, but we encounter this situation more frequently in the labour market.

6.2 Monopolistic Firm Output and Price Decision Making

The perfect competition and monopoly are the two extreme poles of market structures. Between these two extremes, there is ***monopolistic competition and oligopoly*** – these types of market structures are predominant in real economy.

6.2.1 Monopolistic Competition Characteristics

The concept of monopolistic competition itself suggests that it is a model that has both the features of monopoly (monopolistic), as well as the elements of perfect competition (competition). The monopolistic market is characterized by the following characteristics:

a) ***A large number of manufacturers*** whose products are very close substitutes. A large number of companies in the industry means that each firm assumes that its own decisions on output and price are not taken into account by the other companies, so its actions are independent of the behaviour of other companies.

b) A significant characteristic of monopolistic competition is ***product differentiation***. This can consist e.g. of the location of the firm, the price, product quality, related services, credit conditions, packaging, etc. This product differentiation then affects the monopoly power of the manufacturer within the field of the given product. A monopolistic firm can, therefore, in a very limited sense may be the price maker. From this point of view, the situation is reminiscent of a monopoly, but with the difference that the individual demand curve after the production of the firm is not the same as the market demand curve after the product of the whole sector. The above-mentioned significant number of companies in the market causes the differentiation between the products of individual companies to be very small. For this reason, the price differentiation between them is also very small. In other words, the cross-elasticity of demand after the production of a single firm in relation to the prices of substitutes produced by other companies in the framework of the monopolized sector is very high. Also, the price elasticity of demand after each of the companies' production is very high.

c) The last significant feature of monopolistic competition is associated with the possibility of entry into the industry. This is reminiscent of perfect competition, where there are no barriers that would prevent companies from entering the sector. Due to the differentiation of the product, the monopolistic sector has certain barriers. The new firm must establish itself on the market and gain customers. To do this, it can use a wide variety of sales supports, which require spending, it should create a marketing strategy, etc. These financial and market circumstances constitute barriers connected with the entering of new companies to the monopolized competitive market. For most companies, however, these barriers are not insurmountable.

6.2.2 Short Term Monopoly Firm Profit Maximization and Loss Minimization

A firm in monopolistic competition produces a product different from the product from other companies, and the individual demand curve after its production is therefore declining. In the monopolistic market, there are minimal differences between the products sold and their prices. Therefore, if one firm reduces the price, it is likely that this will increase the volume of this firm's sales more than proportionally. Conversely, if the firm increases the price, its customers are going to buy from companies selling for an unincreased price, so that our firm is likely to sell a significantly smaller volume of its production. (On the contrary, in perfect competition, the price increase of one firm would lead to the loss of all of its customers, since the product is totally identical). This fact, enhanced by the existence of a large number of substitutes in a monopolized competitive market makes the ***individual demand curve very elastic.*** The price elasticity of demand for production of a firm in the conditions of monopolistic competition is therefore affected by the differentiation of products and the number of companies in the industry. The more substitutes the firm product has and the more companies there are in the market, the greater individual demand price elasticity can be expected. This also corresponds to the graphic representation of the firm production demand curve.

When determining the optimal output, the firm bases its assumptions on the famous "golden rule of profit maximization", which is equilibrium of marginal revenue and marginal costs. For a graphic representation of the choice of the optimal output, we will make use of the "traditional approach": from profit maximization, seen as the difference between the firm's total revenue and total cost, we will derive the relationship between the marginal revenue and marginal costs. We will then show the corresponding development of the revenue function.

To create short-term profit, the firm will maximize its production q^*, where it reaches the largest difference between the total revenue and costs (Fig. 74a)). The profit-per-unit also reaches its maximum (AR-AC Fig. 74b)). Total economic profit of the firm in the short-term is then determined by the shaded area (AR-AC). q^* and its development is shown in Fig. 74c). The firm can attain economic profit not only by the production of optimal output q^*, but also by the production of any output greater than q_1 and lesser than q_2. At the level of production of output q_1 or q_2 would the selling at prices acceptable by the demand mean an equilibrium of total revenue and total cost (the average revenue and average costs) and the firm would thus attain a zero economic profit.

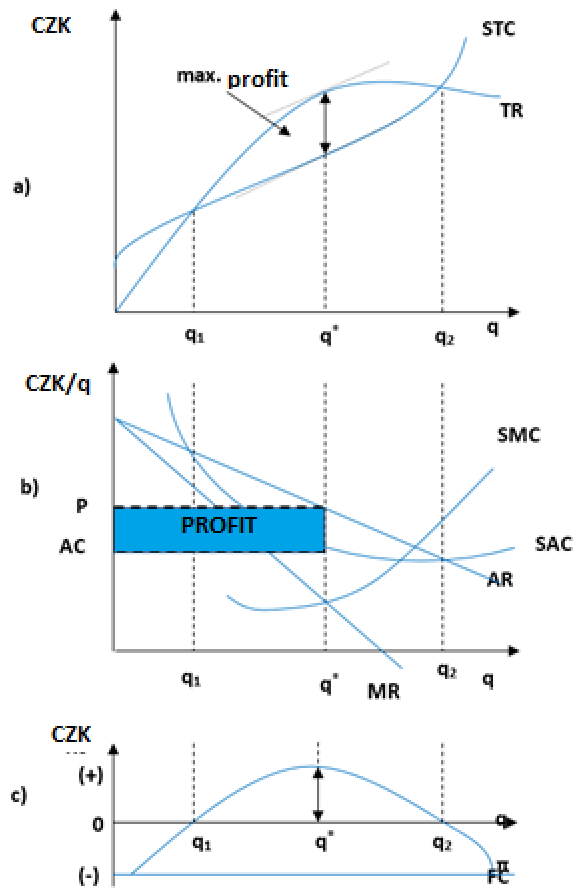

Fig. 75 Short-term firm profit maximization in the conditions of monopolistic competition

The firm, however, in the short-term can find itself in a situation where its products are not sold enough, which results in the total revenue falling below the level of the total cost. As we already know from the analysis of a perfectly competitive firm, the firm in such a moment faces the choice, whether it shall stop its production, or continue with it, even though it is creating a loss. Let us recall that the deciding criterion in this case is the amount of the variable, or the average variable costs. Or to be more exact: the decision is whether the total profit is greater than variable costs, or the average profit (price) is higher than the average variable costs. In Figure 75 a), the total profit excesses the variable costs in the sale of the output greater than q_1 and smaller than q_2. This means that the amount of total profit in excess of variable costs can be used to cover part of the fixed costs. The loss is smaller in the production of output q*, even if the firm creates a loss in the short term, and its production can therefore continue.

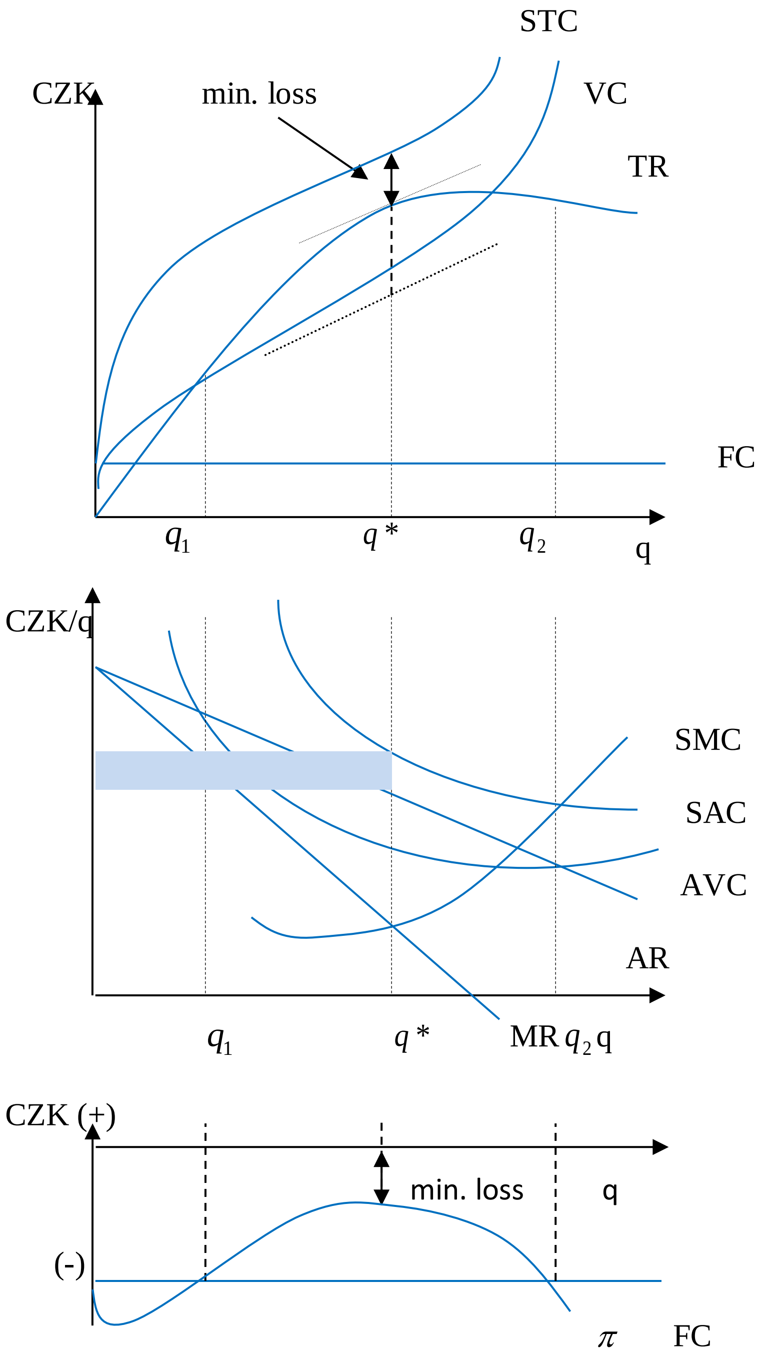

Fig. 76 The firm minimizes loss by continued production

Firm closure

If a firm is making a loss, where its total revenue at any level of output is less than its variable costs, the best solution for the firm is to cease production. The smallest loss will be realized by the firm when it produces zero output. Continuing production would lead to an increase in losses. The firm stops production and either exits the industry or remains in it, waiting for an improvement in market conditions.

6.2.3 Long-Term Firm Profit Maximization in Monopolistic Competition

The process of the formation of the firm's long-term optimal output in the conditions of monopolistic competition is analogous to the process described in the conditions of perfect competition. Let us assume that firms on a given monopolistic market will attain a short-term economic gain. Its existence will become an impetus for a number of other companies to enter the market (either new companies, or companies engaged in other sectors). Their influx occurs basically without any problems because the barriers of entry into the industry can be relatively easily overcome by the newcomers. Market demand is then secured by a larger number of companies, so that each of them has smaller part of the market. This results in a shift of the individual demand curve to the left. (We assume that due to the growing number of companies producing very close substitutes, the demand curve will become even more elastic). The individual demand curve will most likely move to the left and down, until the price and long-term average costs do not create an equilibrium.

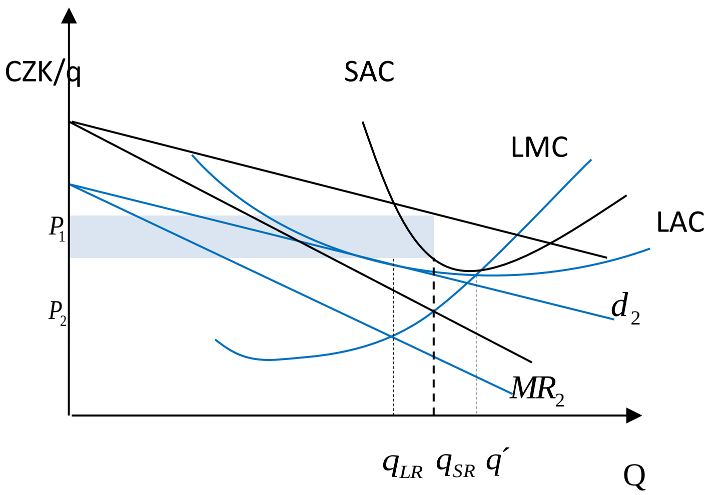

Fig. 77 Long-Term Firm Profit Maximization in Monopolistic Competition

At the short-term optimum output q^*_{SR} and the price P_1, the firm attains an economic profit. This stimulates the entry of new companies to enter the industry. This entry causes the offset of the individual demand curve of each firm to the left and downwards, with the unchanged average costs, reduces the profit of the companies. At the output q^*_{LR} and price P_2 there will be zero economic profit. Companies outside the industry are not interested in entering it.

Easy entry of companies into the monopolistic competitive sector, or their exit from this sector will in the long-term lead to the enforcement of a tendency to zero economic profits. It is a similar mechanism as in the conditions of perfect competition. And yet, especially due to the differentiation of the product and to the diversity of companies, the long-term optimum of all the companies in the long-term in the monopolistic competitive industry may not be characterized by zero economic profit. Therefore, we tend to speak more about the tendency of balancing the profit and normal rates of return.

6.3 Monopolistic Competition Efficiency

Lower production efficiency of a firm in monopolistic competition, when compared with the perfect competition, is often associated with an excess production capacity. In the production of long-term optimal output, the LAC curve is a tangent of the demand curve, i.e. the firm produces this output with decreasing long-term average costs. However, the LAC curve reaches its lowest level at an output that is higher (q') than the optimal output (q*). This can be interpreted in that companies operating in a monopolistic competitive industry are too small to produce with the lowest unit cost.

On a monopolistic market, there are more firms than there would be if all firms were producing at the minimum long-run average cost.

The allocation inefficiency in monopolistic competition is analogous to the case of monopoly. Both a monopoly and a monopolistic firm possess the monopoly power, which allows them to set a price above the marginal costs (P > MC). This results in allocative inefficiency in the form of so-called deadweight loss.

Despite the aforementioned inefficiencies of monopolistic competition, this type of market structure is considered desirable. Since none of the competing firms typically have significant monopoly power, the deadweight loss resulting from monopoly power should be small. High elasticity of demand curves will work to reduce excess production capacity. Furthermore, any inefficiency of monopolistic competition must be compared with its significant advantage in the form of product differentiation.

Points to remember

In monopolistic competition, firms produce at a lower level of efficiency than in perfect competition because their output is smaller than the level at which they would achieve minimum costs. This type of market structure also leads to allocative inefficiency, where the price exceeds marginal cost (P > MC), resulting in deadweight loss. Nevertheless, monopolistic competition is considered beneficial due to product differentiation, which provides consumers with a greater variety of choices.

A monopoly represents a specific market entity that is the sole provider of a particular product or service with no close substitutes. Due to its dominant position, a monopoly firm can freely set both the price and the quantity of production, which is a fundamental difference compared to a perfectly competitive environment where prices and quantities are determined by the market. A monopoly thus faces market demand that has a downward-sloping function, meaning that in order to increase the quantity sold, it must lower the price.

One of the key characteristics of a monopoly is imperfect competition, which is especially evident in entry barriers to the market. These barriers can take various forms, such as legal protection in the form of patents, high entry costs, or exclusive ownership of essential resources. Barriers to entry prevent the emergence of competition, allowing the monopoly firm to maintain its dominant position and achieve above-average profits.

A monopoly firm seeks to maximize profit by choosing the optimal level of output. This process is based on comparing marginal costs (MC) and marginal revenue (MR). The optimal output level is reached at the point where marginal costs equal marginal revenue (MC = MR). This point represents the optimal level of production that maximizes the firm's profit. In contrast to perfect competition, where price equals marginal costs, a monopoly sets the price higher than marginal costs. By doing so, the monopoly maintains a price differential, leading to the creation of a so-called monopoly markup.

Another key element in the behavior of a monopoly is price setting. Since a monopoly determines both the quantity of output and the price, it can set a price that maximizes its profit. The monopoly follows the rule of price markup, where the price is set in relation to marginal costs with a certain markup added. The markup depends on the price elasticity of demand; the less elastic the demand (i.e., the less consumers respond to a price change), the larger the markup the monopoly can afford.

Monopoly profit arises when the firm sells its products or services at a price that exceeds its costs. This profit is maintained primarily due to barriers to entry that prevent the emergence of new competitors. In the long run, a monopoly firm typically achieves above-average profits because no new players can enter the market to reduce its revenues.

Unlike firms in a perfectly competitive environment, a monopoly does not have a classic supply curve. While in perfect competition the supply curve is derived from marginal costs, in the case of a monopoly, there is no clear relationship between price and the quantity supplied. A monopoly responds to consumer demand, and its production depends on the profit it can maximize at a certain price.

As a result of the monopolistic market structure, various forms of price discrimination arise, where the monopoly sets different prices for different groups of consumers. The first degree of price discrimination, also known as perfect price discrimination, involves the monopoly charging each consumer the maximum amount they are willing to pay. The second degree of price discrimination involves setting different prices based on the quantity of units purchased. The third degree of price discrimination involves setting different prices for different market segments, such as students, seniors, or residents of different regions. Price discrimination can allow monopolies to increase profits, but it also has implications for the allocative efficiency of the market.

Monopolistic market structures often lead to allocative inefficiency. This inefficiency arises because the monopoly produces a smaller quantity of output at a higher price than would be the case in a perfectly competitive environment. As a result, consumers pay higher prices, and some are completely excluded from the market because they cannot or do not want to pay the higher price. This situation leads to losses in terms of total social value, which is referred to as deadweight loss. Deadweight loss is a measure of inefficiency that occurs when a monopoly sets prices above the level of marginal costs and restricts the amount of production below the optimal level.

Despite these drawbacks, a monopolistic market structure can have positive aspects, particularly in industries where there are high fixed costs to entering the market or where intensive research and development are necessary. In some cases, a monopoly may be more efficient than competing firms, especially if its size allows it to benefit from economies of scale. A monopoly can also invest in innovation and technological progress if it is motivated by long-term profits derived from these investments.

Review Questions

1. Clarify and graphically display the principles of decision making of a monopolistic firm about its output and price.
2. Assess the positive and negative features of price discrimination. Try to find examples of its application in the Czech Republic.
3. Compare the monopolistic competition with perfect competition and monopoly.
4. Try to explain why some sectors, such as telecommunications, airlines and financial markets, were considered natural monopolies and why in the 80s, in the framework of the so-called deregulation, these natural monopolies were broken in a number of developed countries.

5. Explain why the individual demand curve after the production of a firm in a monopolistic sector is highly elastic.
6. Display in a single graph the optimum of a firm in the conditions of perfect competition, monopoly and monopolistic competition.

?

Test Questions

1. A monopoly is characterized as:

   a) Many sellers in the market

   b) A single seller in the market with no close substitutes

   c) A few large firms in the market

2. A monopolistic firm maximizes profit when:

   a) Price = Marginal cost

   b) Marginal revenue = Marginal cost

   c) Price = Average cost

3. Third-degree price discrimination means:

   a) Charging different prices to different groups of customers

   b) Charging different prices based on purchase quantity

   c) Charging the maximum price each customer is willing to pay

4. Deadweight loss in a monopoly represents:

   a) Monopoly profit

   b) Loss of social welfare

   c) Market entry costs

5. A monopoly typically:

   a) Has a supply curve derived from marginal costs

   b) Does not have a traditional supply curve

   c) Has a perfectly elastic supply curve

Answers:

1. b, 2. b, 3. a, 4. b, 5. b

Practical task

1. Analyse a specific example of a monopoly in your area, such as a local water or electricity provider. Research the history of this company, the reasons for its monopolistic position, and the current market situation. Focus on the monopoly's pricing policy, its impact on consumers, and any regulatory measures by the government. Investigate whether there are any plans to liberalize this market or introduce competition. Based on your research, write an analytical report of 600–800 words in which you assess the advantages and disadvantages of this monopoly for society. Consider whether a monopoly is an effective solution in this particular case, or if greater competition would yield better results for consumers and the economy as a whole.

2. Create a simple economic model of a monopoly in a spreadsheet. Define the demand function and total cost function for a hypothetical firm. Based on these functions, derive the marginal revenue and marginal cost curves. Find the point where marginal revenue equals marginal cost and determine the optimal level of production and profit-maximizing price for the monopoly. Calculate the total profit of the monopoly and the deadweight loss. Create a graph displaying all relevant curves. Then, write a short report (400–500 words) explaining your calculations and their economic significance. Discuss how the situation would change if this market were perfectly competitive instead of monopolistic.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie.* 5. aktual. vydání. Praha: Management Press, 2010. s. 281-327. ISBN 978-80-7261-218-5. 574 s.

[2] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005.

[3] MACÁKOVÁ, L., SOUKUPOVÁ, J. *Mikroekonomie – repetitorium*. Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[4] HOLMAN, R. *Ekonomie*. Praha : C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[5] HOLMAN, Robert. *Mikroekonomie.* Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0

[6] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[7] SOUKUPOVÁ, J., 2018. *Mikroekonomie.* 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 7

Firm Equilibrium in Oligopoly Conditions

After studying this chapter, you will be able to:

- clarify the basic characteristics of oligopoly;
- explain the basic points of each type of oligopoly;
- analyse the various models;
- clarify the causes of emergence of alternative firm theories and characterize the individual models.

Keywords:

Oligopoly, duopoly, cartel, Cournot's model, oligopoly with a dominant firm, game theory, sales maximization, Baumol's firm model.

After studying this chapter, you will be able to:

- Define an oligopoly and identify its key characteristics.
- Distinguish between different types of oligopolistic structures.
- Apply basic oligopoly models to analyse market situations.
- Explain the principles of game theory and its relevance for oligopolistic markets.

Chapter preview

- The following chapter focuses on the oligopolistic market structure, characterized by a small number of large firms dominating the industry. It presents various types of oligopolies and analyses their behaviour using economic models. The chapter also explores alternative theories of the firm that diverge from the traditional profit maximization assumption. Emphasis is placed on understanding the strategic interactions between firms in an oligopolistic environment and their impact on market outcomes.

Chapter objectives

- Introduce the concept of oligopoly and its main characteristics.
- Analyse different types of oligopolistic structures and their specific features.
- Explain key oligopoly models, including the Cournot model and the dominant firm oligopoly.
- Introduce students to the basics of game theory and its application in an oligopolistic environment.

Estimated study time

- The estimated time required to study the chapter is 180 minutes.

7.1 Firm Optimal Output in Oligopoly Conditions

An oligopoly is a market structure that is different from the structures characterized to this point especially by the small number of companies and a relatively high degree of interdependence in the companies' decision making.

An oligopolistic market structure assumes the activity of only a few companies in the industry, where the production of each of these companies represents a significant market share and the companies' decision-making is dependent on each other. Each of them must consider the impact of their decisions on the behaviour of other companies in the sector, and to predict the other companies' reaction to their own decision. And it is this interdependence that complicates the analysis of the oligopoly: the companies mutually react not only on the price changes, but also on the changes in output, product quality, advertising etc. of each of the other companies. For these reasons, there are various models of oligopoly differing from each other in particular in the assumptions about the behaviour of the competing companies. These different models of oligopoly, however, agree on the following three assumptions:

1. A relatively small number of producers in the industry. Some models analyse the case of only two companies in a given market (duopoly), while others analyse an unspecified number of equally strong companies, other models assume one of the companies in a dominant position, etc.
2. The character of the manufactured product can be either homogeneous or differentiated. In the case of a homogeneous product, we can talk of a pure, or homogeneous, oligopoly. This type of oligopoly has a particularly strong interdependence of companies, e.g. even the slightest change in the price of one firm will significantly affect the other companies. If the companies in an oligopoly produce a differentiated product, we are talking about a differentiated oligopoly. The differences between the products of individual companies in the oligopolistic industry are generally insignificant, i.e., they are close substitutes. At this point, we may point out such industries as the manufacture of cars, laundry detergents, cleaning products, cosmetics.
3. There may exist barriers to entry the industry (e.g. in the form of economies of scale, product differentiation costs, legal restrictions, etc.), which prevent the entry of new companies in the industry. Every firm has such an economic force, that it can set the price above the marginal cost (P > MC).

Definition

An oligopoly is a market structure characterized by a small number of firms that are mutually dependent in their decision-making, with each firm needing to consider the reactions of others. Firms in an oligopoly may produce either homogeneous or differentiated products, and entry barriers may exist in the market, allowing firms to set prices higher than marginal costs (P > MC).

If such a barrier of entry is represented by the economies of scale, then every firm seeking entry into the industry should achieve in their production as low average costs as any firm already functioning in the industry. Due to the fact that many of the barriers are not insurmountable, we could assume a situation where after overcoming the initial obstacles, other companies may enter into the sector. Depending on their number, this could lead to the demise of the oligopolistic market structure. From this point of view, the existence of oligopoly is affected by the relationship between the market size and the optimal size of a firm (the size, which allows the firm to attain economies of scale). If the market is, in relation to the optimal size of companies in the sector, too small, then the market demand will probably be provided by a small number of companies and the oligopolistic market structure is maintained. If the market was, in relation to the optimal size of the companies in the sector, too large, there would be an influx of new companies and the oligopoly would probably cease to exist.

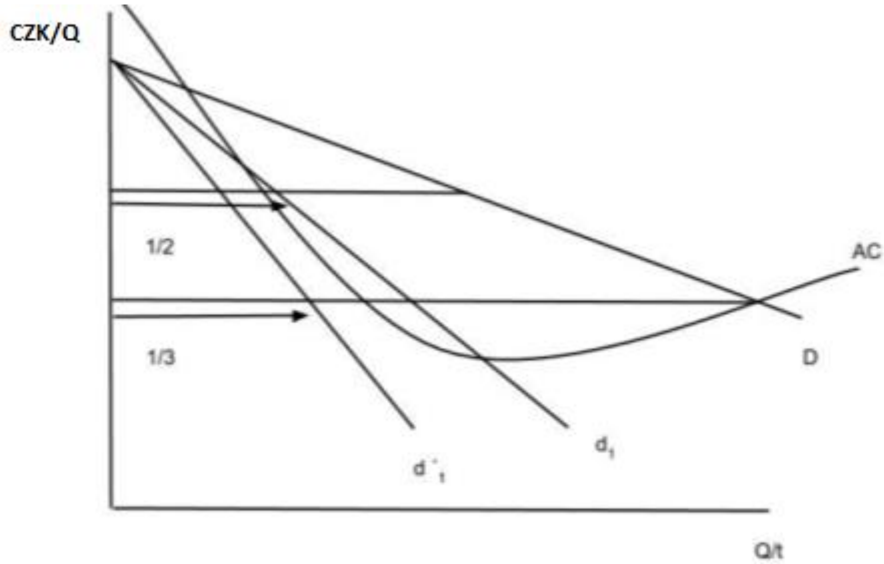

Fig. 78 Market size as the barrier of entry

Sometimes, the oligopoly companies can use the **limiting price** as a barrier of entry. The limiting price is set lower than the price, at which the oligopoly companies would maximize their profits, if they were not threatened by the entry of companies from other industries. A prerequisite for the application of the limiting price is the united price adjustment of oligopolistic companies.

7.1.1 Elementary Oligopoly Models

When constructing various models of oligopoly, we'll assume a fixed number of companies in the sector, which is referred to as n. In order to distinguish the output of individual companies from the entire output of the oligopolistic industry, we will refer to the output of the companies as qi (i = 1, 2, ..., n) and the output of the industry will be referred to as Q. A simplifying prerequisite is the existence of identical companies in the sector with identical costs, and therefore with identical amounts of optimal outcome. Similarly to our previous analyses of the previous market structures, we shall assume perfect competition on the side of demand, i.e. we assume the existence of such a large number of consumers that none of them is able to influence the market price. The inverse function of the demand for the production of the entire oligopolistic industry, that will be referred to as f(Q), expresses, at what price these consumers are willing to buy this sector's changing output volume: **P = f(Q)**

Most models of oligopoly are based on the assumption that the output of the given sector is completely homogeneous, so the market is determined by the **law of one price**. It then does not matter to consumers from which of the manufacturers they would buy the product. This, however, does not correspond to reality. The firm tries to distinguish its product from the products of its competitors, whether by improving the quality of the product, its design, its health aspects, the environmental consequences of its production and consumption, warranty periods, services provided, etc. The market then contains differentiated products by different companies and the law of one price cannot be enforced. In fact, the consumers do differentiate between the products of individual companies. In connection with the analysis of oligopoly, the product of which is differentiated, there are two major problems:

1. The question of market definition. Should we be analysing e.g. the whole "sound carrier" market, or rather the isolated tuner, car-radio, cassette player and portable player etc. markets? This issue can be tackled via the concept of product groups.
2. With the previously used tools, the analysis of oligopoly with a differentiated product is very difficult. The use of game theory seems far more preferable. In a further analysis of oligopoly, we will focus on some of the most famous models, most of which are based on the assumption of homogeneous firm production. At the end of the chapter, we will describe the method of setting prices in the product group, and the basic aspects of game theory when researching oligopoly.

7.2 Cartel

The cartel constitutes the so-called collusive (or also contractual) oligopoly, where the industry sector is represented by a group of several companies, which sells the same or similar products, the prices of which are approximately at the same level and a price war would actually weaken their position. The companies together enter into a secret (cartel) agreement, where they establish a monopoly price and each firm acts, within a defined framework, as a monopoly. The objective of the cartel is to maximize the total profit of the given sector.

Points to remember

A cartel is a form of collusive oligopoly in which several firms agree to set a monopoly price, thereby avoiding mutual price wars and maximizing the industry's total profit. In this model, companies behave like a monopoly with multiple plants, with each firm selling products based on a common marginal revenue and producing at costs below the socially efficient level.

The common marginal revenue MR(Q) is derived from the market demand curve (D). The sum of MC is calculated as the horizontal sum of the long-term marginal costs of member companies.

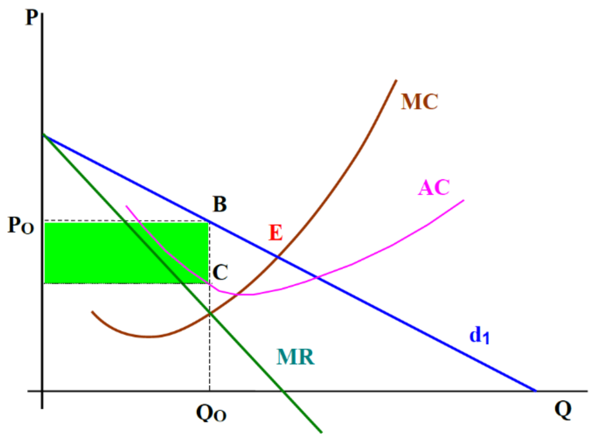

Fig. 79 Contractual oligopoly

Figure 78 illustrates the situation of one of the three companies that jointly agreed to each occupy one-third of the market.

The model assumes that all three companies have the same cost curve and the same pricing strategy, i.e. that they simultaneously increase or decrease the price, so that their market share remains unchanged. The demand of the first firm, covering one third of the market production is marked as d_1. Curve d_1 is relatively little elastic, reflecting the existence of agreements on the level of prices. The first firm of the oligopoly, as well as the second and the third, is in equilibrium at the point where marginal revenue is equal to marginal cost. The oligopoly will sell at price P_O, which is higher than the cost of production. The scope of the production will be compressed below the socially efficient level, which is similar to the case of monopoly.

7.3 Cournot's Model

This model is based on the assumption that there are only two companies (duopoly). These two companies produce a homogeneous product, have completely identical cost curve and know the market demand curve (which is both declining and linear). The starting point of the model is the assumption that the first firm considers, when deciding on the size of its output, the output of the firm's competitor (i.e. the second firm) to be constant. In other words, the first firm thinks that the other firm will not change its output in response to the change in the output of the first firm.

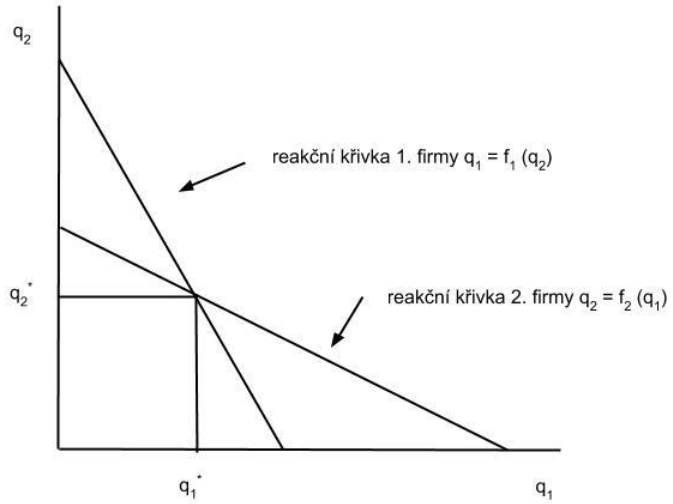

Fig. 80 Cournot's oligopoly model[1]

Cournot's equilibrium is represented by the intersection of the reaction curves of both companies. Cournot's equilibrium occurs at point (q_1*; q_2*) when both companies maximize their profits. None of them is motivated to a change in their output.

In the real economy, the case where a firm would repeatedly erroneously assume that its only (probably strong) competitor will not respond by changing their production to any output change of the first firm, would probably not occur. From this perspective, Cournot's model is static.

[1] Figure translation: „Reakční křivka 1. firmy": 1st firm reaction curve, „Reakční křivka 2. firmy": 2nd firm reaction curve

7.4 Dominant Firm Oligopoly

This type of oligopoly is also known as the oligopoly with a price leader. **Price leadership** can be characterised as a situation where one firm in the industry has the crucial initiative in setting prices, while other companies merely copy that price. Generally, two forms are described. One with a dominant price leading firm, and the second is the so-called barometric price leadership. A typical dominant firm is usually a firm whose only competitors are numerous smaller businesses on a competitive edge. These smaller companies are not able to fundamentally influence the market price through their decisions. Rarely is the dominant firm in the industry also complemented by several medium-sized companies and a larger amount of small businesses.

The situation is illustrated in Fig. 80. When determining the optimal output and prices, the dominant firm bases its calculations on the assumption that the market demand curve is that of D_T and its own demand curve is d_D. The dominant firm derives its optimum output and price from the golden rule of profit maximization; the intersection of its curve of marginal costs (MC_D) and the marginal revenue curve MR_D corresponds to the output Q_D and price P_D. At this price, the companies on the competitive edge supply the output Q_T - Q_D, while the total output of the industry is Q_T.

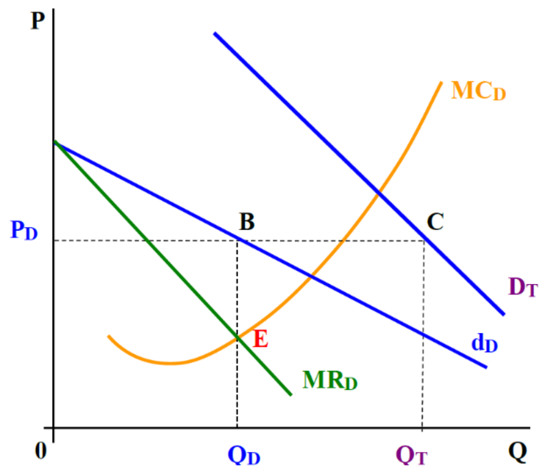

Fig. 81 Dominant firm oligopoly

It is essential that the companies on the competitive edge, behave like perfectly competitive companies: for the price specified by the dominant firm, they can sell any amount of their output and their individual demand curve is therefore, at the given price, horizontal. The necessary condition of these companies' profit maximization is the equality of the copied prices and of the marginal costs of each of these companies. The smaller companies in the market are forced to respect the dominant firm price, since due to their size they cannot attain economies of scale and their cost conditions are worse than in the case of a large firm. The pricing process is the same as in the conditions of a monopoly. If the companies decided to establish a price lower than the dominant firm, they would attract more customers, but due to their cost conditions, they would not be capable of attaining such production. If they set a price higher than the dominant firm, they would lose customers. Seeing their cost conditions, they cannot afford to attract more customers by lowering their prices in these conditions. Moreover, the dominant firm, having its stores in many regions, could reduce the price in the area far more. Therefore, the optimum solution for the companies on the competitive edge is the monitoring of prices of the dominant firm.

In the position of the price leader, whose pricing policies is pursued by other companies, can be a firm that is the largest in its industry, has the lowest cost, tradition, renowned name, etc.

The barometric firm model assumes changes in the position of the price leading firm. Such firm implements the first price changes and performs for other companies the role of a barometer of the market conditions. Whether the other companies in the sector will or will not follow its strategy depends on how far this strategy reflects the market conditions that apply to these other companies. Some empirical studies have shown that this type of price leadership often arises as a reaction to significant problems in the sector, which are related to the constant fluctuations of prices and also the ruthless competition, under which the companies in the industry have suffered significant losses – there is therefore the need for stabilization.

7.4.1 Game Theory Based Oligopoly Models

Economic games that model the behaviour of companies, can be divided into two categories: cooperative and non-cooperative. The basic difference between them lies in the possibility or impossibility of companies to enter into a mutual agreement. Most of the games, to which we will pay attention, will be of the non-cooperative sort. The models based on the game theory show very simplified strategic situations. Each game consists of three basic elements: players, strategies and results.

A player is any participant of the game, who autonomously decides on the choice of one of the many different strategies. For our purposes, we will mostly assume two players, being the companies A and B.

The strategy is seen as any of the possible activities for which a player may decide in a given game. Even though the firm can usually choose between a larger number of activities, we will for simplicity assume that each of our two companies will decide between the two strategies.

The results represent the final game revenue for each of the players. For our purposes, when there are the above-mentioned companies in the position of the players, we will be expressing the final revenue in monetary amounts. The results of two alternative strategies for each of our two companies will be described in a table which we will refer to as the so-called payoff matrix. We will assume that the companies are able to sort the results from the least to the most preferred and try to achieve the most preferred outcome.

We are using the game theory as an alternative (and in the conditions of strategic decision making of companies also more appropriate) oligopolistic market analysis tool. Since we are interested in the market's equilibrium, it is essential to answer the question, how this equilibrium is understood by the game theory. So far, we have understood the market equilibrium as a state where either the supplier or the buyer had any interest, at the given equilibrium price and quantity, in changing their behaviour.

We have so far analysed the so-called non-cooperative game: each player maximizes their custom result, regardless of what the effect on his opponent might be. The tendency of players to coordinate and cooperate can be observed in connection with the so-called repeated games. While some of the prisoners have only once in their lifetime the option of choosing between the strategies of pleading guilty or innocent, most companies repeatedly set the size of their output and their level of prices. In real economic life, therefore, the companies play repeated games, which allow them to evaluate their own position and to obtain information about the behaviour of their competitor. If a firm repeatedly cannot achieve the results it desires, it will understand that entering into an agreement with a competing firm could improve its situation. Some markets, therefore, can be infiltrated by secret agreements between manufacturers, which then lead to the cartel behaviour.

7.5 Alternative Firm Goals

7.5.1 Reasons for Alternative Theories of Firm

The explanatory power of the classical theory of the firm is in this context considered limited by some authors. The main reason is a complicated ownership structure (a firm may be owned by an individual, group of persons, by a different firm) and complex relations between the owners and management of the firm. This is also negatively supported by organizational issues, lack of information and constantly changing conflicts between different individuals and groups within the firm. One of the often-claimed arguments against the objective of maximizing the firm's profits is the behaviour of a number of companies that does not correspond to this very objective. The theory of profit maximization is then seen as a special case of the ownership, management and organizational structure of the firm i.e. it is not generally applicable. Even the answer to the question of why firms actually make a profit is not uniform. The various theoretical approaches to profit can be divided into three groups:

1. **Compensation and functional theories** see profit as payment to entrepreneurs for their activity and substantial risk. This approach can be applied to companies in the case of individual ownership. In the cases of multiple business owners and publicly listed companies, this approach is more problematic. In this case, the ownership is separated from the performance of the business function as such. However, due to the fact that shareholders are subjected to a risk, their share of the profits can also be understood.
2. **Monopoly theories** understand profit as a result of a benefit i.e. the unique position of the firm, that is, as the monopolistic profit.
3. **Technology and innovation theories** understand profit as a result of technical and technological improvement and innovation activities.

The next problem is encountered if we take into account the point of view of time. The traditional theory of the firm is mostly static. Therefore, according to one of the approaches, the aim of a firm is not to maximize immediate profit levels, but to maximize its flow in time, which can be expressed as **maximizing the present value of the expected profits**. Unlike the static concept, when the firm maximizes profit in a typical given period, the dynamic approach is based on the assumption that the firm maximizes its profit over time.

The main variable in this case are the expected profits. The aim of the firm is therefore to maximize their total. In terms of the decision making of the firm, it is necessary to take into account not only the amount of the profits, but also their present value. Profit maximization is not generally accepted as the objective of the firm. One of the main objections to the simple model of profit maximization is the lack of information in the context of the risk and uncertain results of decision making.

The uncertainty is due to the constantly changing conditions of both the market and the competition on the demand side, as well as the changes induced by technical progress. In this context, we should recall another previously mentioned problem: the existence of complex ownership and interest relations and imperfections in the field of the organization and inside information of companies, especially in the case of large companies.

All of the above-mentioned facts can be used as an argument in favour of the theory, according to which the aim of the firm is not profit maximization as such, but achieving a satisfactory amount of profit. The approach of the theories based on the principle of satisfaction is quite different from profit maximization.

In the case of these theories, we are not looking for an optimal volume production, a combination of production volume and price, but we are rather asking whether the given volume of production and prices will ensure a satisfactory level of profit. We will discuss certain aspects of the satisfaction-theory based approaches in context of behaviouristic theories.

The problems associated with the model of profit maximization also form the basis for the theories postulating ***other firm objectives:***

- A possible objective is to achieve a certain **market share.**
- Some economists consider the decisive motivation of the firm's efforts its **long-term survival.**
- **The growth and expansion of the firm** is another possible alternative to profit maximizing.

The satisfaction theory enables another approach to the behaviour of the firm: the firm's objective is understood not as a maximization of one variable, but as the simultaneous implementation of the objectives in several areas. The significant alternative theories of the firm also include the management and behaviouristic approaches.

7.5.2 Management Theories of the Firm

Management and behaviouristic theories of the firm arose as a response to a complex ownership and interest structure in the firm and the problems connected with this phenomenon. Their main feature is the proposition that the firm controlled by managers may have a different objective than profit maximization. The concept of the objectives of the firm in management theories of the firm is different from the traditional theory, in that the management theories emphasize the ***separation of ownership and management*** – the company's owners are the shareholders, but the executive management is exercised by the paid managers. Both groups pursue their own interests, which may be in mutual conflict. Supporters of management theory conclude from these presuppositions that the objectives of the firm controlled by managers is different from the objective of profit maximization. The individual authors define this objective as such in various manners.

Behaviouristic theories of the firm focus on the formulation and search for the objective of firms. In a sense, they go further than the management theories. They are based on the fact that the objectives of the company are the result of decision making and negotiating processes between individuals and groups interested in the firm.

7.5.2.1 Simple Management Model

The aim of the firm controlled by managers is then to maximize the benefits of managers. In the case of managers, the aim is not the optimisation of the consumption due to budgetary constraints. The benefit of the manager is given by their position in the firm. The individual models differ in the utility function variables they take into account. An example of a management model that is based on the benefit of the managers, is a model according to which the variables of the management utility function are both the profit of the firm on the one hand, and the secondary benefits and income of the managers on the other. The secondary benefits and income of the managers include e.g. a company car with a driver. The existence of these secondary benefits of the managers, however, means an increase in the total costs, and therefore the decrease in profit. In a simple management model, the firm will determine its output and price according to the rules of profit maximization.

The optimum point is the point at which the line of profit is tangent of the indifference curve A. Profit maximization corresponds to the situation where the profit is equal to the reported profit and M = 0 (point Z). In this case, the one variable – here represented by the secondary benefits and income of managers – is an analogy of a neutral object and does not affect the profit of the manager. The optimum is therefore the corner solution when all of the profit is being clearly reported and the secondary benefits and secondary income of the managers are zero.

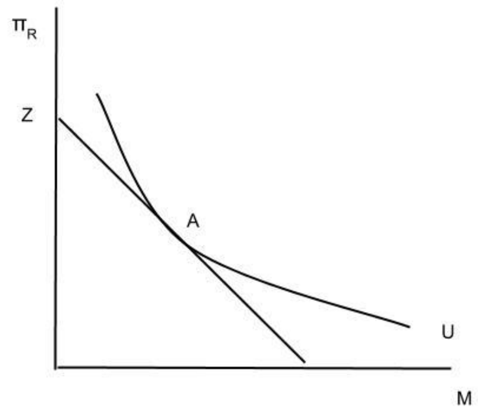

Fig. 82 Simple management model

The line of profit is growing further from the beginning with the growth of the difference between total revenue and total cost. The volume of production and price are in a simple management model identical to profit maximization but carry higher cost and lower reported profit. The main difference of a simple management model from the classic profit maximization is a greater freedom of choice of the combination of reported profit and the secondary revenue and benefits of the managers (in the context of various taxes).

7.5.2.2 Baumol's Model of the Firm Maximizing its Turnover

It is based on the assumption that the managers' goal is to maximize the firm's turnover. **TR = P x Q**

In doing so we assume imperfect competition and the existence of barriers of entry into the industry. The most important feature of imperfect competition is the fact that demand is not perfectly elastic and therefore price is not an independent parameter. Turnover, therefore, may not always increase with the volume of production. For the sake of simplicity, let us assume a linear demand function. In this case, the total income will first grow with the growth in production volume. From some point on, however, it starts declining. Our reflections, however, have been greatly simplified so far, because we did not incorporate the presumption of a minimum required profit, which is a factor limiting the decision making of managers.

The decision making of the managers is then more complicated. They strive to maximize turnover, but at the same time the profit must not fall below a certain level, i.e. below the ***minimum required profit***. The existence of the minimum required profit can be reasoned mainly by the fact that firm owners require income (dividends). If the income of the owners of the firm falls below a certain threshold, the firm and its managers will be threatened by a violent takeover. In addition, if we leave the framework of the simplified static model, the profit can be one of the resources for the further development of the firm.

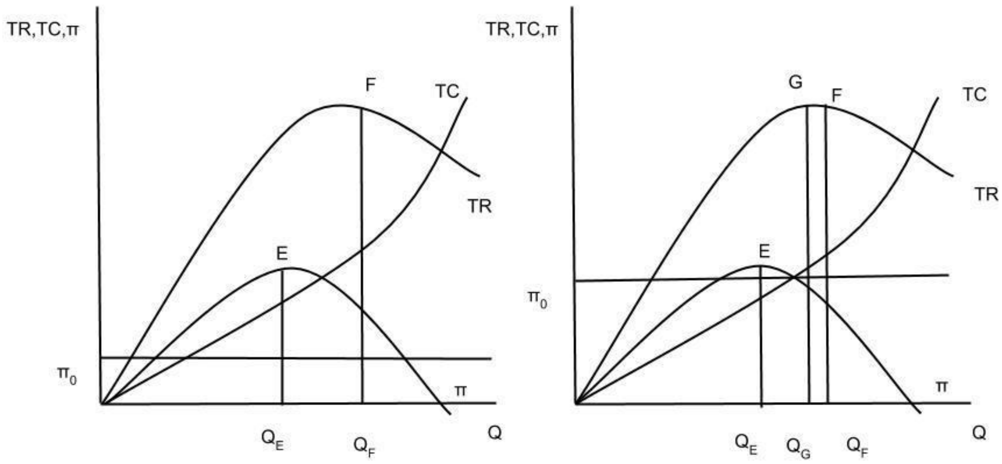

Fig. 83 Baumol's model (total magnitudes)

The firm that maximizes its turnover reaches a profit higher than the minimum required profit (the difference between TR and TC is higher than the π0). Therefore, the optimal output and price do not change.

Let's summarize the interpretation of Baumol's model: a firm maximizing its turnover always produces a greater volume of production at a lower price than the company maximizing its profit. If the marginal costs are non-zero, then the turnover-maximizing firm in its optimum point always fulfils that MR < MC. This means that the volume of production is higher than corresponds to the optimum of the firm that aims to maximize its profits. The mentioned characteristics of course relate to cases where the minimum required profit is lower than the maximum profit.

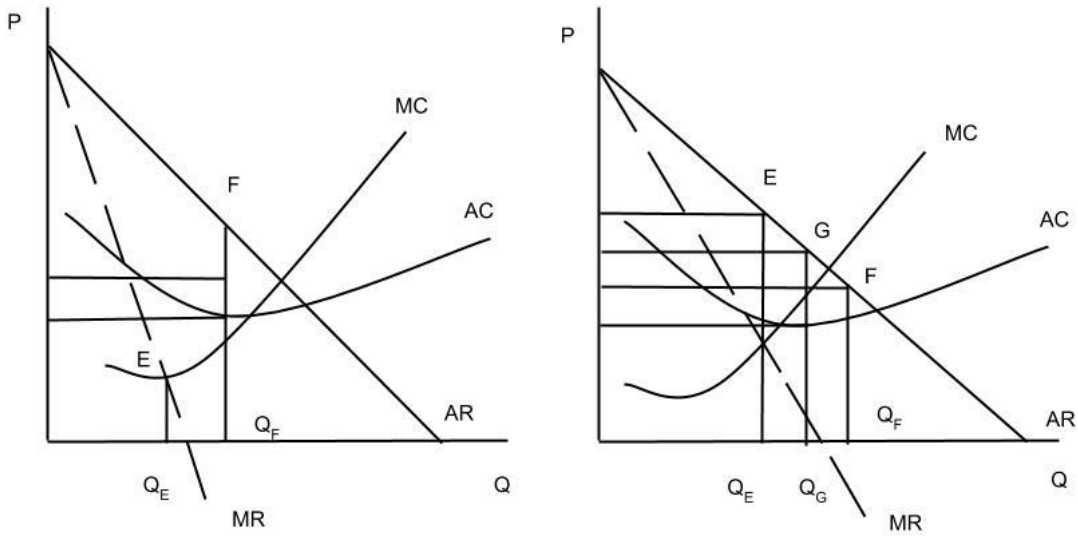

Fig. 84 Baumol's model (unit magnitudes)

Point E is the optimal point of the firm maximizing its profit. Point F is the optimal point of the firm maximizing its turnover. The picture shows the curves of total revenue (TR) and total cost (TC). The difference between total revenue and cost is profit (π). As explained in the preceding chapters, we can now find the optimum for the firm maximizing its profit (point E). In our case, the optimal output is QE. In the point of maximum turnover, π > π0. Point E is the point of maximum profit. Point F is the point of maximum turnover.

7.5.3 Other Alternative Theories of the Firm

7.5.3.1 Behaviouristic Theories of the Firm

The common feature of all the behavioural approaches is the recognition of the fact that it is very difficult to specify the objective of a large organization. Individuals have their objectives and, if they are in control of the firm, these are transferred also into the objectives of the firm. Behaviouristic theories leave behind the assumption that the objective of the firm is merely the concern of the owners (top managers). The firm is therefore a "coalition" of individual interest groups. Various conflicting groups can have their own representatives in the board of representatives of the firm. The conflict of interests is then reflected already in the outcome of the decision-making process. The objective of the company then may change with changes in the objectives of the individual groups and their relative power within the firm.

From these facts, it follows that in the final function of the firm, some of the maximizations of some of the variables are replaced by merely attaining a "satisfactory" amount. In this sense, the theory of satisfaction may be considered a starting point for the behaviouristic theories.

If there are different groups in the firm with different objectives, it is not possible that one group achieves its maximum goal. The assumption that all groups will seek to achieve an acceptable result is far more realistic. The behaviour of the firm controlled by a group, whose decision corresponds to the “satisfaction“ theory then follows this pattern: If the situation is unsatisfactory, the firm tries to find a change leading to the achievement of a satisfactory outcome.

If the result is satisfactory, the firm can maintain its status quo or try to further improve the result.

The decision making of the firm is subordinated to the objectives in five key areas:

- In the area of production, the firm objective can be formulated in two ways. Firstly, as the fluidity of the production, preventing excessive fluctuations by setting the maximum possible volume change of production between the various stages and, secondly, as achieving a certain volume of production.
- The objective in the inventory area, can be formulated in the form of their absolute level or as a desirable range of their movement. It is clear that the optimal amount of inventory must be sufficient for the continuity of production and sales, but on the other hand it must not bind too many resources.
- The objective in the area of sales can be formulated either in value- or physical- units.
- Market share is an alternative to the preceding objective, as it is the scale of the successful sales.
- One of the key objectives is also profit, which serves as an indicator of the competence of top managers. It can also be expressed as the ratio of profit to costs, investments, etc.

The fact that different companies behave differently under completely analogous conditions and respond to the same stimuli in distinct ways supports the case for behaviouristic theories.

7.5.3.2 Employees’ Firm Model

In employees’ firms, all of the employees take over the business role. Also the profit is be distributed among all the employees. For simplicity, we will assume that all employees have an identical share of the profits. Furthermore, we will assume that the employees do not own capital goods, but hire them at their market price. The main differences between the employees’ firm and the firm maximizing its profit may be derived from the target function. The goal of the employees’ firm is maximizing the income per employee. The income of a member of such a firm has two components:

- wage, the amount of which is determined by the labour market and is therefore independent of the firm as such,
- share of the profit, which is dependent on the amount of profit and the number of employees.

An important feature of the employees’ firm is the immediate dependency of the target function on the number of members of the firm. This amount then co-dictates the volume of the production produced. The discussion around the employee’s firms relates primarily to the comparison of its efficiency with the efficiency of firms with other forms of ownership.

The Importance of Profit in Alternative Theories of the Firm. Profit plays a role in a number of management models as a constraint in the form of a minimum required profit. An important condition required for the validity of the management theory is the condition that this minimum required profit is lower than the maximum profit. The main causes are the competition on the final production market as well as on the capital market. The company's competitive position and financial resources are directly or indirectly related to profit. The basic premise of managerial theories of the firm is the separation of ownership and control. In favour of the managerial theory is the fact that there is a large number of shareholders and the profit fluctuations lead to the fact that a large proportion of the shareholders prefer to split their investments into several companies, rather than invest in a single company. On the other hand, it is necessary to take into account the fact, that it is necessary to differentiate also among the shareholders – at least among the shareholders actually controlling the firm on one side and small shareholders on the other – and above all the fact that major shareholders are also various financial institutions.

The alternative theories of the firm are an important part of economic theory. Their significance lies in the fact that they provide the ability to analyze a much broader spectrum of issues associated with the theory of the firm than the traditional model.

Σ

Oligopoly is a market structure characterized by a small number of firms within a particular industry, which leads to a high degree of mutual dependence between these firms. This mutual dependence means that the decisions of one company directly affect other competitors in the industry, contributing to the dynamics of the market environment. Firms in an oligopoly usually aim to maximize their profit but may also set alternative goals, such as maximizing revenue, achieving a certain market share, or striving for long-term survival in a competitive environment.

The basis of Cournot’s model of oligopoly, often referred to as a duopoly, is the assumption that firms make production decisions simultaneously, and each firm considers its competitors' output as fixed. This model thus assumes that firms respond to competitors' decisions, which impacts total market supply and thus the market price. Within the Cournot model, each firm seeks to find the optimal level of output that maximizes its profit while taking into account competitors' reactions.

Price leadership is another important element of the oligopolistic structure. In this situation, one firm takes the initiative in setting prices, while other firms accept this price with little resistance. This dynamic can arise from the dominant position of a firm in the industry, which has significant market power and can set a price acceptable to other firms. Price leadership can lead to price stabilization within the industry but can also contribute to price discrimination, where a firm may set different prices for different market segments.

When analyzing oligopoly, it is also important to consider other factors, such as barriers to market entry, which can limit the possibility of new competitors entering the industry. These barriers can take various forms, including high investment costs, patents, access to distribution channels, or strong brand loyalty among customers. High barriers to entry often maintain the stability of an oligopolistic structure and prevent intense competition, which can lead to higher profits for existing firms.

Another important characteristic of oligopoly is the ability of firms to cooperate or form cartels. A cartel is an agreement between firms that seek to coordinate their actions to increase prices and maximize profits. Such agreements can have serious consequences for competition and consumers and are therefore considered illegal in many countries. However, there are cases where companies may collaborate on the development of new technologies or standards without violating antitrust laws.

Various forms of price discrimination are also common in an oligopolistic market, where firms charge different prices to different customers based on their willingness to pay. In this way, firms can maximize their profits while meeting the needs of different market segments.

An oligopolistic structure presents unique challenges and opportunities. Firms must carefully monitor and respond to the decisions of their competitors, which requires strategic thinking and adaptability. Since an oligopoly can influence market prices and product availability, it is important for consumers and regulators to understand the dynamics of this type of market and its implications for the economy as a whole.

Review questions

1. Describe the differences between the monopolistic competition, and oligopoly.
2. Explain the Cournot's solution to oligopoly equilibrium.
3. Explain the concept of price leadership using an example.
4. What alternative firm objectives do you know?
5. Under what conditions would a firm aiming to maximize its profit behave completely identically to a firm maximizing its turnover?
6. Explain the reasons why the alternative theories often include the requirement of a specified minimum level of attained profit.

Test questions

1. What is the main characteristic of an oligopoly?
   a) A large number of firms in the industry
   b) A small number of firms with a high degree of mutual dependence
   c) A single firm dominating the entire market
2. What is the basic assumption of Cournot's model?
   a) Firms make pricing decisions simultaneously
   b) Firms make production quantity decisions simultaneously
   c) Firms agree on a common price
3. What is price leadership in an oligopoly?
   a) All firms set prices together
   b) One firm takes the lead in setting prices, and others follow
   c) Prices are set by the government

4. What are entry barriers in an oligopolistic market?
   a) Physical barriers preventing entry to company buildings
   b) Factors that limit the possibility of new competitors entering the industry
   c) Laws prohibiting new firms from entering the industry
5. What is a cartel in the context of an oligopoly?
   a) A type of product sold by oligopolistic firms
   b) An agreement among firms to coordinate actions to increase prices and maximize profits
   c) A marketing strategy used in an oligopoly

Answers:

1. b, 2. b, 3. b, 4. b, 5. b

Practical task

1. Identify an oligopolistic market in your area, such as the mobile operator or banking services market. List three to five major firms operating in this market and find out their approximate market shares. Compare the prices of a basic product or service among these firms and observe whether the prices are similar or significantly different. Consider potential barriers to entry in this market, such as high initial investments or regulatory requirements. Observe whether there are signs of price leadership, where one firm typically initiates price changes and others follow. Finally, evaluate how this oligopolistic structure affects competition and consumers in the given market. Briefly describe your findings and state whether, in your opinion, this market structure brings advantages or disadvantages to customers.

2. Select two competitors in an oligopolistic market, such as two car manufacturers or two major retail chains. Monitor their marketing activities over two weeks, including advertisements, special offers, and promotional events. Note how quickly and in what way one competitor responds to the actions of the other. Observe whether the companies offer similar products or services and how they attempt to differentiate themselves from the competition. Look for signs of non-price competition, such as emphasis on service quality, product innovation, or brand building. Consider how the mutual dependence of these companies affects their strategic decisions. Finally, assess whether you observe more competitive behavior or indications of tacit cooperation between the firms. Summarize your observations and discuss how this dynamic between competitors in an oligopoly affects offerings and prices for consumers.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie*. 3. dopl. vydání. Praha: Management Press, 2003. s. 306 -348. ISBN 80-7261-061-9. 548 s.

[2] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[3] SOUKUPOVÁ, J., 2018. *Mikroekonomie*. 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 8

Price Formation in Factor Markets

After studying this chapter, you will be able to:

- characterize the specifics of the production factors market;
- explain the formation of demand and supply in the labour market;
- clarify the concepts of monopsony and bilateral monopoly.

Keywords:

Marginal product, individual and market labour demand, monopsony, bilateral monopoly, capital, consumer spending.

After studying this chapter, you will be able to

- Characterize the specifics of the factor market.
- Explain the formation of demand and supply in the labor market.
- Clarify the concepts of monopsony and bilateral monopoly.
- Analyze the relationship between the marginal product of labor and the demand for labor.

Chapter preview

- This chapter focuses on the factor market, particularly the labor market. Students will become familiar with the characteristics of this specific market and its differences from the goods and services market. The chapter will explain the processes of demand for labor formation by firms and labor supply by households. Special attention will be given to the concept of the marginal product of labor and its relationship to labor demand. Additionally, unique market structures, such as monopsony and bilateral monopoly, and their impact on the labor market will be introduced. The chapter provides students with a comprehensive view of the functioning of the labor market and its importance in the economy, including an analysis of factors affecting wages and employment.

Chapter objectives

- Introduce students to the specifics of the factor market and its differences from the goods and services market.
- Explain how labor demand is formed by firms and labor supply by households.
- Present the concepts of monopsony and bilateral monopoly and their impact on the labor market.
- Provide students with a deeper understanding of the functioning of the labor market and its importance in the economy.

Estimated study time

- The estimated time required to study the chapter is 240 minutes.

8.1 Specifics of Firms' Labour and Capital Demand Formation

Unlike the finished goods market, where the demand side consisted of individuals (households) and the supply side was formed by firms, the input market has the given input side created by individuals (households), whereas the firms enter this market as buyers.

By definition, we can understand the production factors as economic resources, which are transformed in the manufacturing process into the desired products and services. In other words, the factors of production are the inputs, which the firm uses to produce output desired by consumers and to deliver this output to the market. The objective of the firm is to achieve the maximum amount of economic profit.

The factors of production include:

- labour,
- land and natural resources (raw materials),
- capital,
- the level of technology (level of knowledge), or entrepreneurship.

The demand for a given input is the so-called **derived demand.** It is derived from the demand for the goods, which is produced using the given input. The firms maximizing their profit play a sort of integrating role: they join (put in contact) the individuals demanding a specific asset with the individuals capable of producing this asset with the involvement of the factor of production they own. If the asset is not demanded at all, the firm will not employ anyone and will not produce this asset.

The market of production factors is the point at which the demand for factors of production meets with the supply of factors of production. The demanders in this market are the firms, that - via the use of production factors - offer on the final production market a specific product in order to achieve profit. The suppliers on the market of production factors are the households, i.e. the owners of the factors of production. These households then rent out these factors in order to get income.

Points to remember

The factor market is the space where the demand of firms for production factors meets the supply from households that own these factors. Production factors, such as labor, land, capital, and technology, are economic resources that firms use to produce goods and services in order to maximize economic profit.

8.1.1 Terms of Profit Maximization on the Inputs Market

On the market of production factors, the golden rule of profit maximization exists in its modified form. When analyzing the behavior of a firm making its decisions about the optimal output, we assumed that the profit is the difference between total revenue and total costs, while both revenue and costs were directly associated with production volume. We therefore assumed:

$$\pi\ (Q) = TR\ (Q) - TC\ (Q)$$

We formed the necessary condition of profit maximization as **MR = MC**

The profit of the firm can be expressed as a function of the included inputs:

$$\pi\ (K,L) = TR\ (K,L) - TC\ (K,L)$$

The problem of a firm entering the market of production factors as a buyer and aiming to maximize profit is that the firm needs to hire their optimal amount. Under our assumptions, this meant involving in the production such quantity of labour and capital that enables the firm to realize its main objective, i.e. maximum profit.

In order to maximize its profit the firm should buy additional units of the given input until the additional cost per unit of input match the additional revenue that this unit will bring by its operation. The terms of profit maximization on the market of inputs therefore imply that for the determination of demand, the income from product manufacturing (MRP) and the cost of production factor (MFC) are crucial variables.

8.1.2 Income Variables in the Factors of Production Market Analysis

Marginal Product Revenue (MRP) of a given input can be generally defined as the change in total income, caused by the changing volume of the given factor of production by one unit. The revenue from the marginal product of capital (MRP_K) can be defined as total income change caused by changing the volume of the used capital by one unit. We can formally express the marginal capital product revenue in several ways:

$$MRP_K = \delta TR/\ \delta K$$

$$MRP_K = MR_A\ .\ MP_K$$

where MR_A = marginal revenue of the firm from the sale of an additional unit of product A, i.e. ($\delta TR/\ \delta QA$), MP_K = marginal capital product, i.e. ($\delta QA/\ \delta K$).

By analogy, the income from the marginal product of labour (MRP_L) represents a change of total income caused by changing the volume of the labour used by one unit. The formal expression is similar to the marginal capital product income:

$$MRP_L = \delta TR / \delta L$$
$$MRP_L = MR_A \cdot MP_L$$

where MR_A = marginal revenue attained by the company by the sale of additional quantities of the product A, i.e. $\delta TR / \delta QA$, MP_L = marginal labour product, i.e. ($\delta QA / \delta L$).

Average product revenue $ARP_{(L,K)}$ is the revenue per unit of factor of production (labour or capital).

8.1.3 Cost Variables in the Factors of Production Market Analysis

Marginal Factor Cost (MFC) represents the change in the total cost of the firm caused by the fact that the company has hired an additional unit of the given factor of production. Marginal capital factor cost (MFC_K) can be defined as the additional costs that the firm incurred by involving an additional unit of capital:

$$MFC_K = \delta TC / \delta K$$

In the same way we express the marginal factor cost of labour (MFC_L) that represents a change to the total cost of the firm caused by the change in the volume of labour used by one unit:

$$MFC_L = \delta TC / \delta L$$

If there is perfect competition in the labour market, then any firm entering the market as buyer hires each additional unit of labour at the constant wage rate w. This means that the additional cost of labour do not change when involving more labour and that the cost is still equal to the wage rate. In its graphic representation, this linear function would be shown as a horizontal line, i.e. it has a zero direction. Therefore we can say that $\mathbf{MFC_L = w}$. If there is imperfect competition on the labour market, the firm would pay a higher wage rate for each additional unit of labour. The function of labour supply would be growing, i.e. its direction would be positive.

Average Factor Cost (AFC) is the cost attributable to the unit cost of the given factor. Therefore: $\mathbf{AFC_K = r}$, where r is the cost of capital. Average factor cost of labour AFC_L is the cost of a labour unit used in production: $\mathbf{AFC_L = w}$, where w is the price of labour.

We know the condition of profit maximization of the firm coming to buy production factors on the market of these factors. This condition is known as equilibrium of marginal product revenue and marginal factor costs. Specifically, for the labour input, the following relationship applies:

$$MRP_L = MFC_L$$

We will now explain why the firms' demand for labour is derived from marginal product revenue of this entry to the market (MRP_L).

If the labour market does have perfect competition, a firm entering into this market would be one of a large number of companies. Therefore the firm will not be able to affect the price of labour (wage rate). This is reflected in the fact that the labour supply curve is perceived by the firm as horizontal (it pays the same wage rate for each additional unit of labour). The marginal costs of labour are identical to the constant wage rate ($MFC_L = w$). If the labour market only has an imperfect competition, the firm demanding labour is one of a small number of firms that enter into this market. Due to its favourable position, the firm will be able to influence the price of labour. In order to obtain an additional unit of labour, the firm will offer a higher wage, which will, however, be subsequently applied for all involved units of labour. It follows from this that the firm understands the labour supply as an increasing function and the marginal labour cost is higher for the firm than the wage rate ($MFC_L > w$).

8.2 Labour Supply

At this point, it is appropriate to draw attention to the two possible points of view on the individual labour supply:

- labour supplied to a single firm (i.e. in terms of a single demander, buyer);
- labour supplied by a single individual (i.e. in terms of a single supplier).

8.2.1 Individual Labour Supply

For the description of an individual's decision-making about the amount of hours they will supply for labour, we will use an analogy with the decision-making of a consumer. Let us recall the fact that we assumed that the consumer would be limited by their income and the prices of goods. The consumer would be deciding about the optimal distribution of their income between two goods. What choices does the owner of the production factor of labour therefore face? The answer is simple: they decide whether to work or not to work. More specifically, they decide what combination of labour and free time they choose in order to maximize their benefits. What are the limits of this individual's decisions? The limited number of hours in one day.

More specifically, the individual chooses between the two "assets": between consumption (C) and free time (herein referred to as H).

We assume that consumption can only be realized as a result of the individual's own labour (L). The total sum of hours of labour and free time during one day cannot be greater than 24 hours: **L + H = 24.**

The optimal distribution of time between labour and leisure time means that the individual in the combination of labour and leisure time maximizes their profit specified by the consumption of goods and leisure: **U = f (C, H)**

When maximizing their profit, the individual faces two restrictions that we have already mentioned above:

- a day has 24 hours (L + H = 24; therefore L = 24-H);
- consumption can be achieved only on the basis of their own labour. If the wage rate of that individual is equal to w, then we can write the second restriction as **C = w. L**

If we replace L with 24 – H , we get **C = w. (24 – H)**

C = 24 . w – w . H

0 = 24 . w – C – w . H

Utility maximization is conditioned by the fact that the individual maximizes their utility, if the marginal rate of substitution is equal to the ratio of prices (or, in other words, relative price). The two goods are substituted here by the variables of utility functions of consumption and free time. The relative price here is the wage rate. In this case, therefore, maximum utility condition is **w = MRS.**

In order for an individual to maximize their utility, they should work as many hours, so that the **marginal rate of substitution of free time by consumption is equal to wage rate (w).**

The optimal combination of leisure and work is graphically represented in Figure 84, which is reminiscent of a similar representation of the consumers' optimum. In theory of the consumer, we focused on how the default optimum of the consumer changes, if there occurs a change of price of one of the goods. We talked about the total effect of this price change, which we divided into two partial effects – the substitution and income effects. We will now proceed in similar fashion. We will explore how the change of the real wage rate (i.e. the price of labour), affects the layout of an individual's 24 hours between work and leisure time, i.e. this individual's optimum.

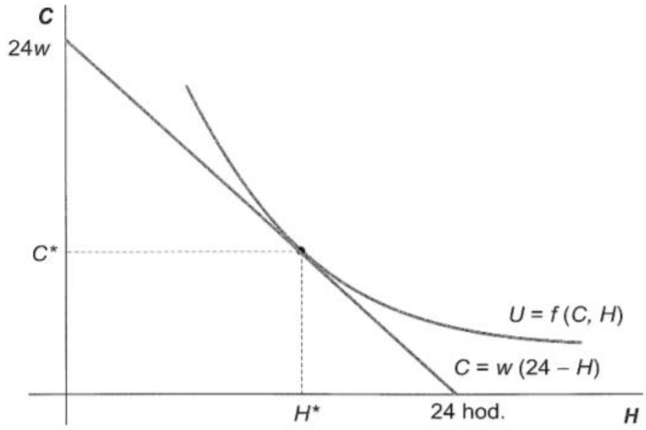

Fig. 85 Optimal distribution of 24 hours between labour and leisure time

Figure 85 presents point Q as the default optimum allocation of time between labour and leisure time. Point Q is specified by the combination H_1C_1. We will now assume that the wage rate grows from w_1 to w_2. This carries two consequences:

1. The simultaneous increase of the price of leisure time (with the increased wage rate, an individual who does not work is losing more money). This is reflected in the fact that the alternative cost of free time is the real wage rate.
2. This changes the constraint direction $C = 24 \cdot w - w \cdot H$. Its intersection with the horizontal axis of leisure time does not change (even when the individual does not work at all, they will be able to enjoy a maximum of 24 hours of leisure time per day). The intersection with the vertical axis, however, changes from 24*w1 to 24*w2. The growth of wage rate allows individuals to achieve a higher level of profit, which is represented by the indifference curve U2. Its optimum is now represented by the point S(H2C2).

Total effect (TE) of the growth of wage rate is represented by the shift of point Q to point S. This effect can be sub-analyzed into the substitution and income effects.

Substitution effect (SE) is the substitution of free time by labour. The increased real wage rate (meaning simultaneously the increased price of leisure time – see the 1st consequence) stimulates the individual to increase the number of working hours and reduce the number of hours of free time. Because of the fact that the increased price of leisure (free) time leads to a decrease in the number of hours of leisure time, the substitution effect is negative.

Income effect (IE) is related to the fact that the increased wage rate (meaning simultaneously an increased price of free time) leads to an increase of the individual's real income. The increasing real income allows the individual to increase their consumption of all goods, including their free time. Because of the fact that this increased price of leisure time leads to the growth of leisure, the income effect is positive. This is represented in Figure 85 by the shift from point R to point S.

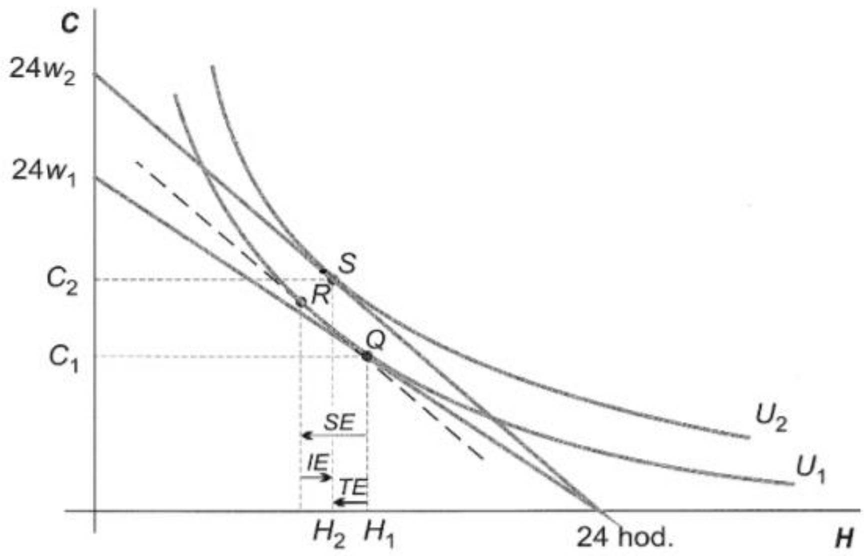

Fig. 86 Effect of changes in the real wage rate on the individual's distribution of time (SE > IE)

Since the substitution effect is negative and the income effect is positive, it cannot be determined whether the growth of the wage rate will lead to the growth of leisure time (and the simultaneous decline in the number of working hours) or to a decrease in free time (and the simultaneous growth of the number of working hours). In principle, there are two cases that may occur:

1. The substitution effect outweighs the income effect, so that the total effect will be negative. The increase of wage rate will in this case lead to a decrease in the number of hours of leisure time. It will also simultaneously lead to an increase in the working hours supplied by the individual. This situation is reflected in Figure 85.
2. The income effect outweighs the substitution effect, so that the total effect will be positive. This case is shown in Figure 86. The negative substitution effect (represented by the shift from point Q to point R) is outweighed by the positive income effect (the shift from point R to point S). The increased wage rate in this case leads to the growth of the number of hours of free time and the simultaneous decline in the number of hours of work that the individual is willing to supply.

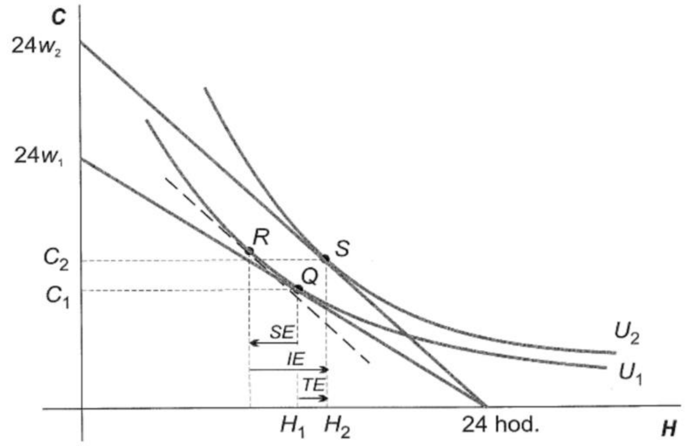

Fig. 87 Effect of changes in the real wage rate on the individual's distribution of time (IE > SE)

The specific derivation of the curves of individual labour supply is shown in pictures 87a and 87b. Figure 87a represents 3 different optima for an individual supplying their labour. These optima occur due to the alterations in the wage rate. The wage rate changes from w_1 = 100 CZK/hour, to w_2 = 200 CZK/hour, and to the w_3 = 300 CZK/hour. The optimal amount of free time is at w_1 equal to 18 hours, i.e. the optimal amount of labour or working time is 6 hours. Figure 87b shows 6 hours of supplied labour at the wage rate w_1. At w_2, the optimal amount of free time is 15 hours, so an individual maximizing the utility will offer 9 hours of labour (again, Figure 87b shows 9 hours of labour at the offered wage rate of w_2). At wage rate w_3, the optimal combination consists of 17 hours of free time and 7 hours of work (again shown in Figure 87b). By joining the points in Figure 87b we get a graphical representation of the individual labour supply. This backward-curved shape of the curve can be seen especially in the long-term (historically, the length of the working day shortens with the increasing level of real wage rate) and in certain groups of workers (according to the latest American studies, this can be observed in adult men, i.e. not women or adolescents). When analyzing the labour market in the short-run, we, as a rule, assume only the growing proportion of individual labour supply.

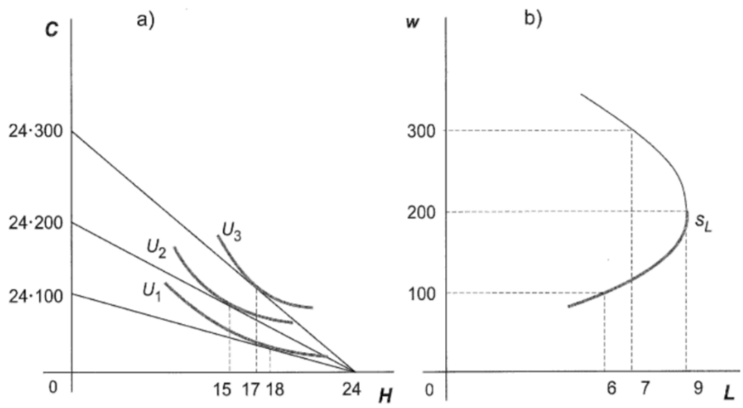

Fig. 88 Individual labour supply derivation

8.2.2 Market Labour Supply

The market labour supply represents the various amounts of labour that its owners are willing and able to supply to the market of the specific type of labour for a varying wage rate.

The market supply curve is created by a horizontal sum of all the individual labour supply curves. The above-mentioned effect of wage rate growth in the form of the “spillover" of labour from other walks of life is reflected in the fact that the market supply curve of labour is not backward curved, but is steadily growing. The intersection of the market labour supply and market labour demand determines the equilibrium wage rate and the equilibrium labour quantity.

8.3 Monopoly Power Enforcement in Labour Market

To illustrate a single supplier of specific labour in the given market, we can give the most often mentioned example of Labour Unions. Labour Unions aim to fulfill their own objectives, which may differ from the objectives of the firms.

The Labour Unions’ objectives may include

- maximizing the economic income realized by the Labour Union members,
- maximizing the total wages of the members of the Labour Union,
- maximizing employment.

The decisions of Labour Unions with regard to each of these given objectives are shown in Figure 88. The D_L curve represents the demand for the "product" of Labour Unions (i.e. for labour). For the derivation of the MRL curve, we must first express the total amount of the finances paid at the given wage rate to all units of labour involved in the production. This amount is referred to as the total wage:

$$\textbf{total wage} = \mathbf{w \cdot L}$$

The **marginal wage** (also known as Marginal Revenue of Labour Union, MR_L) represents the change of the total wage caused by the change in the volume of employment:

$$\mathbf{MR_L = \Delta(w*L)/\Delta L}$$

The marginal wage decreases faster than the demand for labour, as the decrease in the wage rate of an additional unit of work means that the lower wage rate will be paid to all already employed units of labour. Depending on the objective of the Labour Union, the marginal wage will be located at point A, B or C on the demand curve.

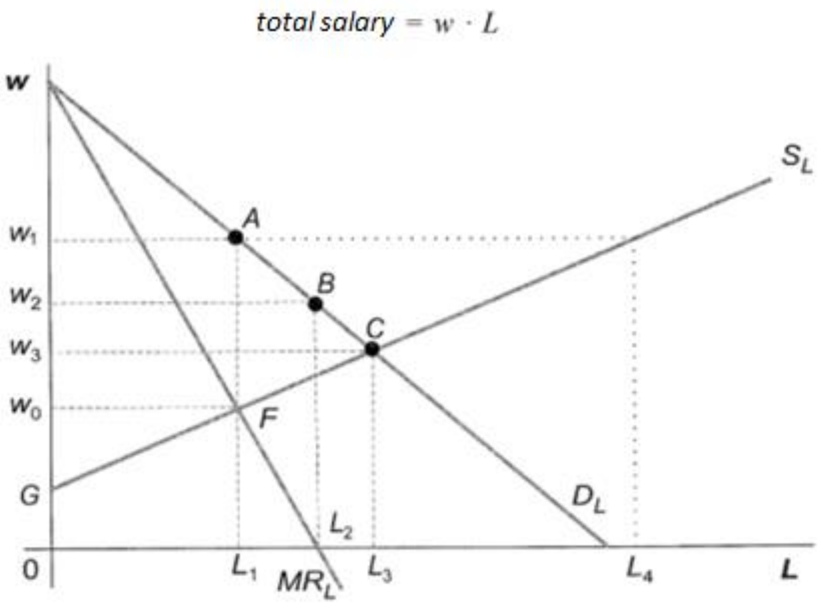

Fig. 89 Selection of the optimum amount of labour supplied by Labour Unions[2]

Point A represents a situation in which Labour Union should **maximize its total economic rent.** The economic rent can be defined as the difference between the real wage and the transfer price of labour. Transfer price of labour is the minimum wage rate, for which a unit of labour is prepared to enter the labour market. It is therefore identical to the alternative cost of labour. It is represented graphically by the labour supply curve S_L. The Labour Union would opt for such a volume of labour supplied that would equal the marginal wage and the transfer price of labour (in other words, the alternative labour cost).

If $MR_L > S_L$ (the given labour market has a higher wage rate than in other markets), the amount of work offered by a Labour Union grows. By analogy, if this market offers a lower wage rate than the other markets, the Labour Union reduces the quantity of labour supplied. The intersection of the MR_L and S_L determines the amount of labour supplied by the Labour Union as L_1. The monopoly power of the Labour Union is reflected in the fact that it does not only require the wage rate w_0 for its members, but rather forces the employers to pay the highest wage rate they are willing to pay, that is wage rate w_1. The maximum total economic rent would amount to the area between the

[2] Figure translation: „Celková mzda“: Total wage

wage rate w_1 for the amount of labour L_1 and part of the growing $S_L$curve between points 0 and L_1 (i.e. the area w_1AFG in Fig. 88).

At the wage rate w_1, the firms will demand the amount of labour L_1, while the individuals will supply the amount of labour L_4. The Labour Union policy aimed at maximizing total economic rent is thus linked to the predominance of the labour supplied over the labour demanded.

Point B is connected with **total wage maximizing.** The relation between the total and marginal value implies that the total wage will reach its maximum, if the marginal wage is equal to zero. The Labour Union would offer the amount of labour L_2 at the wage rate w_2. This pay rate also determines the excess of the amount of labour supplied over the amount of labour demanded. Graphically, the maximum total wage would be represented as the area, the size of which is given by the product of $w_2 \cdot L_2$ (i.e. area $0w_2BL_2$ in Fig. 88).

Point C is the point, in which the Labour Union would find itself if it **maximized the total employment of its members.** In a situation where labour supply and demand were in equilibirium, the volume of labor hired would be equal to L_3 at the wage rate of w_3. The same result would have been achieved, as if the labour market had a perfect competition.

8.4 Labour Market Bilateral Monopoly

If the two labour market subjects meet, one of which is in the position of a single buyer (monopsony) and the other is in the position of a single supplier (monopoly, here represented by the Labour Union), the situation is often described as a bilateral monopoly. Each of them posseses the strength that allows them to significantly affect the price of the purchased or sold labour. When specifying this price, however, it meets the strength of their partner that has – put simply – the opposite goal: while the firm in the position of a monopsony is seeking to pay the lowest wage rates to its employees, the Labour Unions are aiming to promote the highest possible wages of their members (if we assume that the Labour Unions aim to maximize their economic rent). The final result of a conflict of monopsony and monopoly cannot be determined; whether the resulting wage rate will be closer to the objectives of the monopsony or those of the monopoly, will depend on their respective bargaining power. Bilateral monopoly is shown in Figure 89.

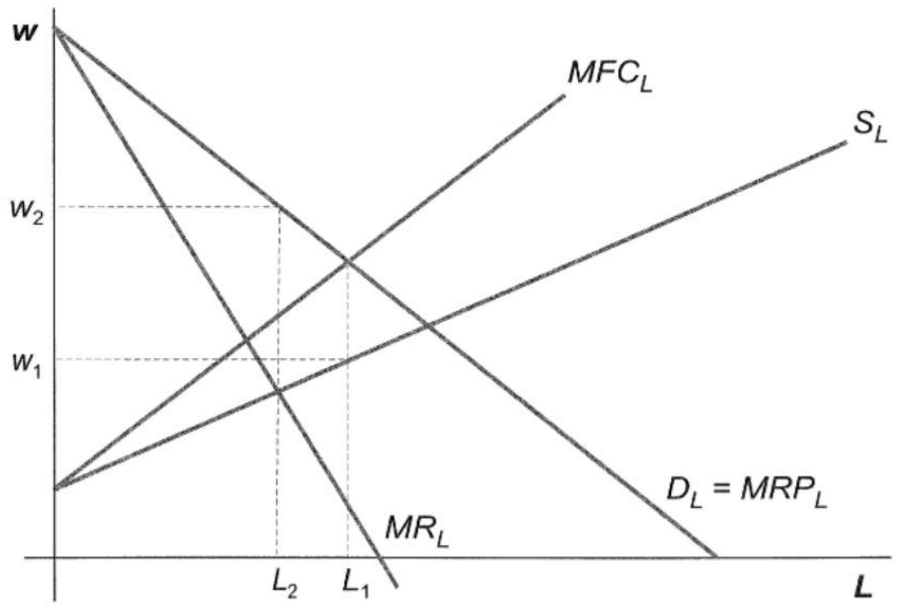

Fig. 90 Labour market bilateral monopoly

Monopsony will be based on the equilibrium of MRP_L a MFC_L: the optimal volume of employment represents the amount of labour units L_1. The monopsony will use its unique position to establish a lower pay rate of w_1. The amount of labour supplied to the monopoly will be determined by the equality of MR_L and S_L: the Labour Unions will pursue employment of L_2 labour units. Their privileged position is reflected in the amount of the required wage rate at the level of w_2. The final amount of the wage rate will, as has been said, depend on the negotiations between the two parties. Using the tools we have available, we cannot explicitly specify the final wage rate (one of the options would be the game theory with the use of a developed formal model of negotiation). The only thing we can say is that the wage rate will be moving in the interval between w_1 a w_2.

8.5 Labour Demand

8.5.1 Demand on the Perfectly Competitive Labour Market

If the labour market does have **perfect competition**, a firm entering into this market would be one of a large number of companies. Therefore the firm will not be able to affect the price of labour (wage rate). This is reflected in the fact that the labour supply curve is perceived by the firm as horizontal (it pays the same wage rate for each additional unit of labour). The marginal costs of labour are identical to the constant wage rate **($MFC_L = w$).**

If the labour market has **imperfect competition** only, the firm demanding labour is one of a small number of firms that enter into this market. Due to its favourable position, the firm will be able to influence the price of labour. In order to obtain an additional unit of labour, the firm will offer a higher wage, which will, however, be subsequently applied for all involved units of labour. It follows from this that the firm sees the labour supply as an increasing function and the marginal labour cost is higher than the wage rate **($MFC_L > w$).**

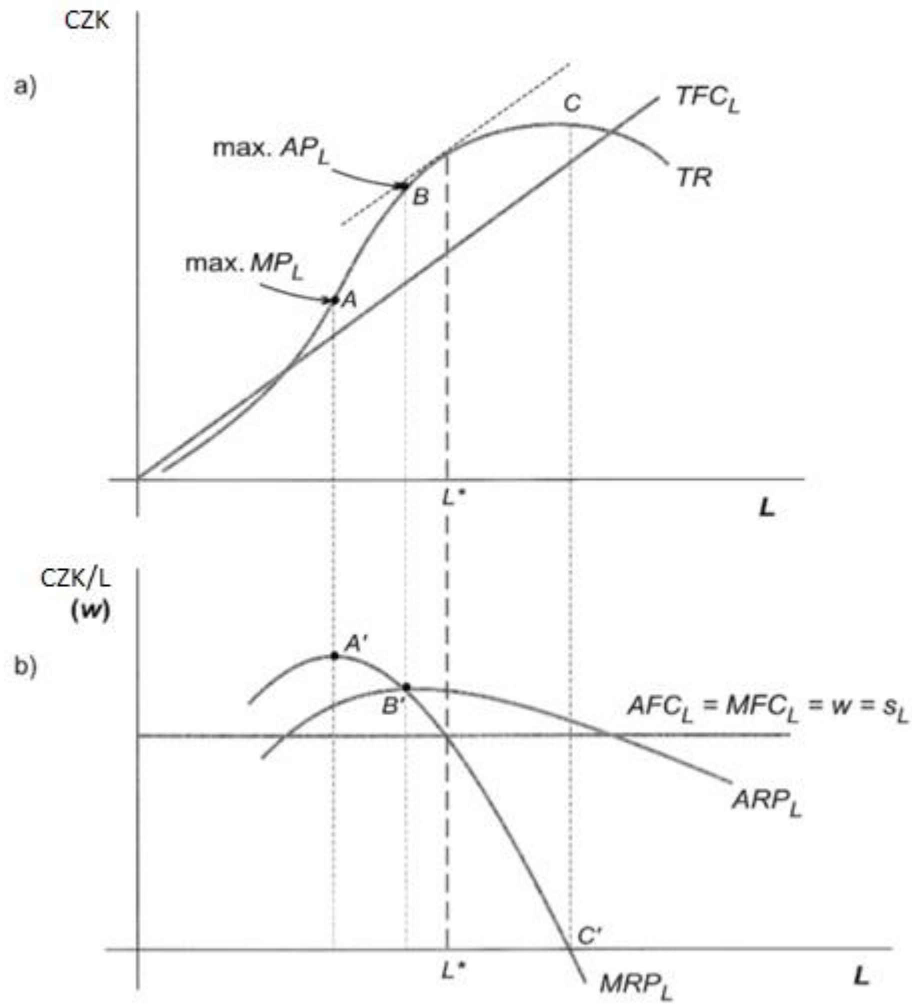

Fig. 91 Optimal amount of labour hired by the firm in the short- term

In point A in Figure 90, the increased revenue from the involvement of additional units of labour change into diminishing revenue; the MRP_L function thus reaches its maximum (point A). The total income is at its maximum in point C, which corresponds to a zero MRP_L in point C. The optimal amount of work is L*: just with the involvement of L* units of labour, the vertical distance between the features TR and TFC_L is at its greatest. Also the directives of both functions are identical and their additions are equal, therefore **$MRP_L = MFC_L$.** The firm maximizes its profit.

The labour demand curve of a perfectly competitive firm on a perfectly competitive market will consist of a declining part of the income function of the marginal labour product (which results from the optimization rule $MRP_L = MFC_L$), the upper limit of which is made up of the maximum level of income from the average labour product (result of respecting the relationship $ARP_L \geq w$).

8.5.1.1 Long-Term Labour Demand

The formation of a firm's labour demand in a situation, where all the inputs are variable (in our case labour and capital), is considerably more complex than it was in the previous case with only one variable input. The reason for this is the interdependence of the inputs: a change in the wage rate may not only lead to changes in the volume of labour, but also to changes in the volume of capital. For example, if the wage rates decline, the firm will be looking for a new optimum combination of these inputs, which may indicate the growth of the amount of labour and a decrease in the volume of capital. Reducing the amount of capital will affect, due to the cross-productivity effect, the marginal product of labour, which in view of the constant price of the product A must be reflected in the change of MRP_L ($P_A \cdot MP_L$). By default, a prerequisite for the formation of a firm's long-term labour demand is the change in wage rate. The change of the optimal combination of inputs represents the total effect of change of the wage rate. The total effect of change of wage rate can be divided into three sub-effects: the substitution effect, the production effect and the cost effect.

The substitution effect creates downward pressure on the MP_L curve, which only better expresses the fact that the additional unit of labour creates, due to lack of capital, a smaller output increment than it would if it had sufficient capital available.

The production effect shows the change in the optimum combination of inputs that is associated only with the change of the firm's output.

The cost effect is directly linked with the production effect, therefore, many economists do not distinguish the cost effect and integrate it with the production effect. Isolation of the cost effect emphasizes the fact that the change in wage rates, and thus the relative prices of labour and capital, will lead to a change in the quantity of hired inputs and output size only if the new output will be the output, at which the firm maximizes its profit. The point is that the decline in wage rates under normal conditions means a decline in marginal costs, allowing the firm to increase the optimal production volume.

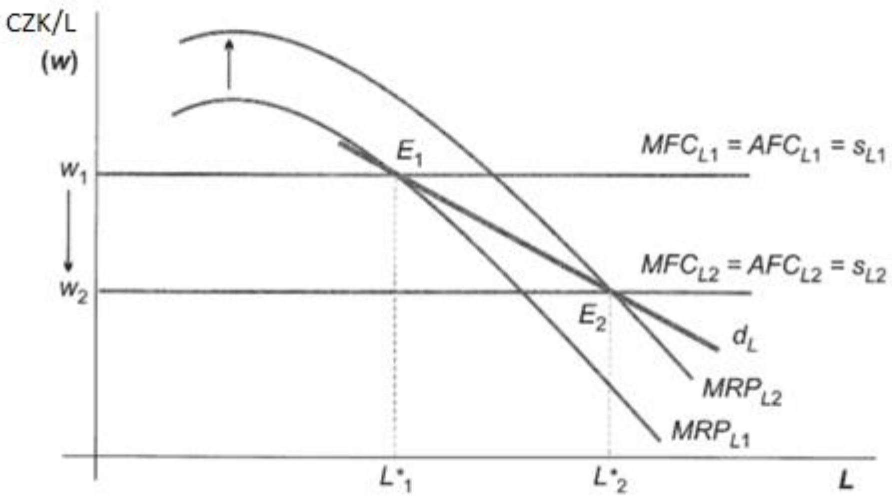

Fig. 92 Long-term demand of a perfectly competitive firm in a perfectly competitive labour market

At the initial wage rate w1, the optimal quantity of labour demanded by the firm is L1* (at point E_1, where $MRP_{L1} = MFC_{L1}$). When the wage rate falls to w2, due to the effects described above, the MRPL curve shifts upward, and a new intersection E2 is created where $MRP_{L2} = MFC_{L2}$. The firm will hire L_2* units of labour. By connecting the two optima, points E1 and E2, we obtain the curve D_L as the firm's long-run demand for labour. The long-run labour demand curve is flatter than the short-run labour demand curve. From this, we can infer that the firm's long-run demand for labour will be more elastic than its short-run demand. However, time is only one of the factors affecting the elasticity of the firm's demand for labour.

8.5.2 Factors Influencing a Firm's Demand for Labor

The demand for labor is primarily influenced by two groups of factors:

- the wage rate, changes in which cause movements along the given labor demand curve;
- other factors that shift the entire labor demand curve;
- Other factors that influence the change in revenue associated with a change in the quantity of labor, and that affect labor productivity:
    - The demand for the firm's product(s) and, indirectly, all factors that influence it;
    - The level of known technology and the directions of expected technological changes;
    - The number of complementary inputs, etc.

When we investigate the relationship between the percentage change in the quantity of labour demanded and the percentage change in its price, we derive the price elasticity of labour demand coefficient, denoted as e_{DL}.

The price elasticity and cross-elasticity of labour demand are formally analogous to the price elasticity and cross-elasticity of demand for goods. When explaining the first determinant of the elasticity of labour demand, it is important to remember that the demand for any input is a derived demand. It is therefore understandable that the elasticity of labour demand is influenced by the elasticity of demand for the product that is produced using labour: the more elastic the demand for the product, the more output the firm wants to produce, and the more labour it will hire.

The second significant factor affecting the elasticity of labour demand is the elasticity of substitution between inputs, or the degree of mutual substitutability between labour and capital. The third important factor affecting the elasticity of labour demand is the share of labour costs in the firm's total costs.

8.5.3 Long- and Short-Term Labour Demand

Short-run demand for labour reflects the firm's willingness to hire additional units of labour as long as the marginal cost of labour (MFC_L) equals its marginal revenue product (MRP_L). The firm takes the market wage rate as given, thus an increase in the wage rate shifts the intersection of the MRP_L and MFC_L curves along the MRP_L curve upward to the left, leading to a lower quantity of labour demanded. Conversely, a decrease in the wage rate has the opposite effect. The short-run labour demand curve is essentially the marginal revenue product of labour curve, which is bounded from above by the maximum average revenue product of labour ARP_L.

Firm's Long-term Labor Demand

Assuming the firm employs at least two variable inputs (labour and capital), we can use isoquant analysis to examine how changes in the wage rate affect the quantity of labour demanded. Unlike the short run, where a wage rate change induced three partial effects (substitution, output, and cost effects) that cumulatively shifted the MRP_L curve upward, the long-run demand for a non-perfectly competitive firm in a perfectly competitive labour market exhibits a fourth effect, known as the **revenue effect.** This effect stems from the firm's downward-sloping individual demand curve for output and the more rapidly decreasing marginal revenue (MR_A). If the wage rate decreases, ceteris paribus, the marginal cost curve shifts downward to the right, creating a new intersection of MR_A and MC, which determines a higher optimal output level, a lower price P_A, and lower marginal revenue MR_A. Since $MRP_L = MR_A \cdot MP_L$, the decrease in MR_A reduces MRP_L. The interaction of these four effects results in an upward shift of the MRP_L curve, but the fourth effect causes this shift to be smaller than in the short run. In Figure 91, this would mean that the MRP_{L2} curve representing this shift would lie closer to the MRP_{L1} curve.

8.5.4 Demand on the Imperfectly Competitive Labor Market

When a limited number of firms offer their products in the output market, it is known as imperfect competition (monopoly, oligopoly, and monopolistic competition). These firms capitalize on the advantages stemming from their privileged position as sellers. However, any market can also have a limited number of firms buying goods or services, leading to imperfect competition where buyers benefit from their privileged position. We are now interested in imperfect competition in the labour market, characterized by a limited number of firms entering this market as demanders of labour.

We can identify the following three cases:

- Monopsony: the labour market is entered by several companies,
- Oligopsony: a few firms enter the labor market to purchase labor;
- Monopsonistic competition: labour is bought by a large number of firms, each of which can influence the wage rate at least slightly.

The interaction of different market structures can lead to various situations: for example, a firm in an oligopolistic position in the output market may be one of many firms entering the labor market, a monopoly in the output market is simultaneously a monopsony in the labor market, etc. Under conditions of imperfect competition in the labor market, firms entering this market are in the position of price makers. The fundamental characteristic of an imperfectly competitive labor market is an upward-sloping individual labor supply curve (i.e., the labor supply to a single firm). For a firm to hire an additional unit of labor, it must pay a higher wage rate. The upward-sloping labor supply function is identical to the upward-sloping average factor cost of labor (AFC_L). function.

The fundamental characteristic of an imperfectly competitive labor market is the ability of the demanding firm to influence the price of labor. To hire an additional unit of labor, the firm offers a higher wage rate. Therefore, the individual labor supply curve is upward-sloping. The increasing nature of the individual labor supply curve implies a more rapidly increasing character of the marginal cost of labor: if the firm pays a higher wage rate for an additional unit of labor, it must pay this higher rate to all already employed units of labor. A monopsony exploits the advantage of its privileged position as the sole demander by paying a lower wage rate than would correspond to the intersection ofMRP_L and MFC_L. The labor demand curve of a monopsony cannot be constructed.

For any combination of a firm‘s position in the goods market and the labor market, the rule for maximizing profit is applicable in the form: $\mathbf{MRP_L/MFC_L. = MRP_K/MFC_K = 1}$

In the market for factors of production (FoP), individuals act as suppliers and firms act as demanders. The demand for an input, such as labor, is derived from the demand for the final good that is produced using that input. When maximizing profits, it is crucial for firms to purchase additional units of a given input only up to the point where the marginal cost of the input unit equals the marginal revenue that the unit generates. This principle is known as the condition for the optimal allocation of factors of production.

In the labor market, an individual seeking to maximize utility works a number of hours per day such that the marginal rate of substitution between leisure and consumption equals the wage rate. An increase in the wage rate affects the supply of labor through the substitution effect and the income effect. The substitution effect incentivizes working more hours because higher wages make work more attractive. Conversely, the income effect may lead individuals to work fewer hours, as higher wages increase their overall income, potentially leading to a preference for leisure over work.

These effects work against each other, and therefore the individual supply of labor is backward-bending. This phenomenon means that at a certain wage level, the supply of labor may, in some cases, be lower than expected. As a result, when wages increase, some individuals may choose to work fewer hours, causing the supply of labor to decrease.

The market supply of labor arises from the horizontal sum of all individual labor supply curves, representing the total volume of labor available in the market at various wage levels. This aggregate is used to analyze overall trends in the labor market and to assess the impact of various factors, such as changes in wages, employment, and working conditions.

A monopoly in the supply of labor typically refers to a situation where there is a single entity supplying labor in the market, often associated with labor unions. These unions can negotiate wages and working conditions on behalf of their members, giving them significant market power in the labor market.

On the other hand, a bilateral monopoly occurs when there is both a monopoly on the supply side (for example, a labor union) and a monopsony on the demand side, representing a situation where there is only one employer hiring workers in the market. In this arrangement, both parties strive to assert their advantages, which can lead to tension and negotiation between the unions and the employer.

A bilateral monopoly thus complicates the determination of the equilibrium wage, as the interests of both parties may be in conflict. In such an arrangement, firms aim to keep wages as low as possible, while unions seek to maximize wages and improve working conditions for their members. This dynamic can have implications not only for individuals but also for the entire economy, including the efficiency of resource allocation and overall labor productivity.

From the perspective of labor market theory, it is important to understand how different market structures, such as monopolies and monopsonies, influence the behavior of employees and employers. Knowledge of these relationships and market dynamics is key to accurately assessing economic conditions and the efficiency of the labor market as a whole. Firms and individuals must therefore adapt their decisions over time based on changing conditions, contributing to the overall development of the labor market and the economy.

Review questions

1. When the demand for a firm's product changes, does the firm's demand for factors of production also change?
2. Does market demand for the final product decrease due to the operation of the law of diminishing returns from the variable factor of production?
3. Explain why a firm will increase the quantity of the hired factor of production in the short run until the Average Revenue Product (ARP) is less than or equal to the price of the factor of production (P_F).
4. Explain what MR and MRP have in common and how they differ. Explain the relationship between them.
5. Explain the difference between MC and MFC.
6. Explain the backward-bending individual labor supply curve.

Test questions

1. What is the condition for the optimal allocation of factors of production?

a) Maximizing profit
b) Equating the marginal cost of an input unit with the marginal revenue
c) Minimizing costs

2. How does the substitution effect influence the supply of labor when the wage rate increases?

a) Reduces the supply of labor
b) Increases the supply of labor
c) Has no effect on the supply of labor

3. What does a backward-bending individual labor supply curve mean?

a) At a certain wage level, the supply of labor may be lower than expected
b) The supply of labor always increases with an increase in wages
c) The supply of labor is always constant

4. How is the market supply of labor formed?

a) By the vertical summation of all individual labor supply curves
b) By the horizontal summation of all individual labor supply curves
c) By the average of all individual labor supply curves

5. What is characteristic of a bilateral monopoly in the labor market?

a) The existence of only one employer
b) The existence of only one labor union
c) The existence of a monopoly on the supply side and a monopsony on the demand side

Answers:

1. b, 2. b, 3. a, 4. b, 5. c

Practical task

1. Analyze the impact of an increase in the minimum wage on the labor market in the context of the substitution and income effects. Explain how this change could affect the labor supply of different groups of workers (e.g., unskilled workers, part-time workers).

2. Prepare a case study of a bilateral monopoly using a real-world example (e.g., in the automotive industry). Describe how this market structure influences wage and working condition negotiations, and what the possible consequences are for efficiency and productivity in the industry.

Literature

[1] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005.

[2] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie*. 3. dopl. vydání. Praha: Management Press, 2003. s. 306–348. ISBN 80-7261-061-9. 548 s

[3] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[4] SOUKUPOVÁ, J., 2018. *Mikroekonomie*. 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 9

Capital Market

After studying this chapter you will be able to:

- characterize capital and its forms;
- explain the formation of demand and supply in the capital market of in various market conditions and in various time periods;
- distinguish between real and nominal interest rates and their influence on the decisions of investors.

Keywords:

capital, capital goods, interest rate, investment, current consumption, future consumption.

After studying this chapter, you will be able to

- Analyze the diminishing marginal product of capital and its impact on the demand for investment resources.
- Explain the relationship between the real and nominal interest rates and the influence of inflation on interest rates.
- Distinguish between the different components of the interest rate that compensate the lender for deferring consumption and for the depreciation of money due to inflation.
- Understand the equilibrium in the capital market in the short run and the long run, and the role of the interest rate in both cases.

Chapter preview

- The chapter focuses on capital as a key factor of production in the economy. Students will be introduced to various forms of capital and their significance for economic growth. The chapter will explain the processes that shape the demand and supply of capital in the market, with an emphasis on different market conditions and time periods. Special attention will be given to the concept of the interest rate, particularly the distinction between the real and nominal interest rates and their influence on investment decisions. Additionally, the relationship between current and future consumption in the context of investment and capital formation will be explained. The chapter will provide students with a comprehensive understanding of how the capital market functions and its importance for economic development. This includes an analysis of the factors influencing investment and capital accumulation in the economy.

Chapter objectives

- Understand how the diminishing marginal product of capital affects firms' demand for investment resources.
- Learn to distinguish between the nominal and real interest rates and understand their impact on the capital market.
- Analyze the equilibrium in the capital market in the short run and the long run, including the supply and demand for capital and the determination of the interest rate.
- Discuss the influence of interest rates on investment and savings decisions.

Estimated study time

- The estimated time required to study the chapter is 240 minutes.

9.1 Concept of Capital

Capital is most often identified as money. An important fact is that capital is the output that is produced in order to be further used as input for the production of goods and services. As a production input, capital is regarded as one of the three scarce factors of production and is seen as capital goods or real assets, which may take the form of:

- material (physical),
- non-material.

Material capital is represented by machinery, tools, buildings, vehicles, raw materials, stocks and various partially finished stages of the production process. A distinctive feature of the individual part of the material capital is that it remains in the production for a certain period of variable length. For this period, it is used for the production of goods and services.

Non-material capital refers in particular to the "human capital", which is composed of the technical and other knowledge and skills embodied in the labor force. Items such as technology, entrepreneurship, know-how, goodwill, trademarks, staff skills and experience can also be included in the category of non-material capital.

Kromě uIn addition to the categories mentioned above, there is also the financial capital, i.e. money and other financial assets.

Unlike labour and land, capital is not a primary factor, but is a secondary factor. It is the product of production, which once again enters into production. It is, therefore, as already stated, simultaneously an output as well as an input of the production. This means that:

- part of the factors of production must be released from the production of consumer goods and redirected to the production of capital goods;
- the production of capital goods is necessarily associated with limits of the present consumption.

The common feature of all the above forms of capital is the fact that an entity, which has currently spent certain resources to obtain these forms of capital, will get additional income from these forms of capital in the future. As soon as an individual enters the capital market, they consider the fact that they will have to solve these two basic questions:

- how much of their income are they supposed to use to purchase goods for their consumption = decisions on consumption over time,
- what form of capital should they purchase for their funds = investment decision-making.

A consumer's decisions about the level of current and future consumption depends on the following factors:

- the marginal rate of their time preferences that determine the shape of the indifference curves,
- size of the real interest rate, which specifies the directives of the lines of market oppor-tunities,
- the size of the current and future income of the consumer and the level of prices of a particular product in both of these periods.

The emergence of capital:

- at the beginning of the emergence of capital is the decision to postpone the current consumption to the future;
- the emergence of capital requires investment – i.e. the creation of capital goods able to increase labour productivity;
- the involvement of capital in production results in a growth of production volume and an associated possibility of greater consumption in the following periods.

This situation is referred to as an indirect method of manufacture using capital goods.

Definition

Capital is a factor of production that serves as an input in the production of goods and services. It can take the form of material assets (such as machines and buildings) and non-material assets (such as knowledge and technology). Unlike primary factors of production, such as labor and land, capital is a secondary factor whose creation requires the deferral of current consumption and investment in capital.

9.2 Capital Market

A perfectly competitive capital market is a market with no risk or inflation. An individual spends their income on the purchase of one type of goods in two periods: there is therefore the current consumption (C0) and future consumption (C1), consumer utility maximization: $\tau = r$, optimization of investment decision: $R = r$, optimization of investment decision does not depend on the preferences of the consumer

9.2.1 Investment Decision Making

The consumer has a current and future income, which they can use either for consumption, or they may use this income and invest it in production. The aim of the consumer is to continue to maximize their utility from the current and future consumption of goods C at a specific amount of current and future income.

Various investments can bring various returns for various periods of time. Therefore, the aim of the investment decision making is the determination of the most advantageous investment.

The general scheme of firms' investment decision making aims to reach a single goal – finding the most advantageous investment. This can be described as follows:

- Firstly, it is essential to make an estimate of the potential revenue from the individual investment projects. To do this, the firms must:

  (1) calculate the cost of purchasing the necessary capital goods;
  (2) estimate an annual net revenue;
  (3) compare the annual net revenue with the calculated expenses (divide "revenue" by "cost").

- From all the planned investment opportunities, the firms have to choose the one that is associated with the highest rate of return on capital.

In an economy where there is no risk and monopolies (we have so far not based any assumptions on this basis), it can be said that the rate of return on capital is equal to the market interest rate.

If we approach the reality of economic life and we leave these mentioned assumptions aside, we find that the return on capital can take various forms:

- for the firms that own capital goods and do not use these for their own production, but rent them for a certain period of time to other firms, this return represents the rent they receive for the leased capital goods;

- the most common form of return on capital is revenue.

The net revenue rate is calculated as the ratio of net revenue to capital stock. The calculated rate of net revenue may exceed (even in equilibrium conditions) the interest rate. The following are some of the reasons why this situation may occur:

- the existence of a monopoly;

- the introduction of technical innovations that result in a reduction;

- of unit costs, and thus cause the growth of rate of revenue in the firm;
- the existence of a risk.

9.2.1.1 Investment Decision Making Criteria

There are two main criteria for investment decision-making. The first of these is the current value of assets. It is necessary to choose such an investment, which brings the highest present value of total revenue. The second criterion is the future value of assets. It is necessary to choose such an investment that will yield the highest future value of assets.

The calculation of the rate of return on capital is also complicated by the fact that capital goods, as we have already explained, remain in the production process over a longer period of time. And yet the revenue in the current year is, as far as the firm is concerned, more important, than expected revenues in later years. Therefore, when making investment decisions, it is very important to find out the current value of the entire flow of future returns from the given capital good (the implemented investment).

The present value is calculated on the basis of discounting ("conversion") of the flow of future returns by the interest rate according to the following formula:

$$\mathbf{SH = N_1/(1+i_r) + N_2/(1+i_r)^2 + \ldots + N_n/(1+i_r)^n}$$

In this formula:

SH present value of the entire flow of expected future returns N1;

Ni net annual return from a given capital goods in the i-th year of its life, which is equal to n years;

Ir annual interest rate, which we assume to remain unchanged.

Present value of an annuity

Annuity, or in other words, a lifetime income, is a type of revenue that an investor receives for a given (specific) number of time periods. More on the issue in the next chapter.

9.2.2 Capital Return Rate

The rate of return on capital (r) is a ratio that indicates how much will the future consumption increase in relation to the reduction of current consumption. The higher the rate of return, the higher the portion of revenue that the market entities are willing to save. In the case of a temporary increase of future consumption, the return on capital rate will be determined by the formula:

$r = (X-S)/S = X/S - 1$

In the case of a steady (long-term) increase of future consumption, the return on capital rate will be determined by the formula: **$r = Y/S$**

From what we have discussed so far, it is clear that the investment is preceded by a deferred consumption, or in other words, savings. In a market economy with a developed division of labour, however, the individuals who save, and those who have investment opportunities, are often not the same individuals. Those who save, do not always have investment opportunities and, therefore, they lend their savings to those who have investment opportunities. This is done, for example, by giving their savings to banks and these banks then lend these savings to businesses with investment opportunities. It is therefore obvious that the "act of investment" and the "act of saving" are usually separated. This has lead to the development of lending funds market, that connects both of these "acts" of capital formation.

In this market:

- supply is made up of the savings of the people (and firms) that are saved in banks or that are used to buy securities – thereby creating lending funds;
- demand is made up of firms that want to invest. In order to do this, they use these funds by taking loans from banks or sell their securities.

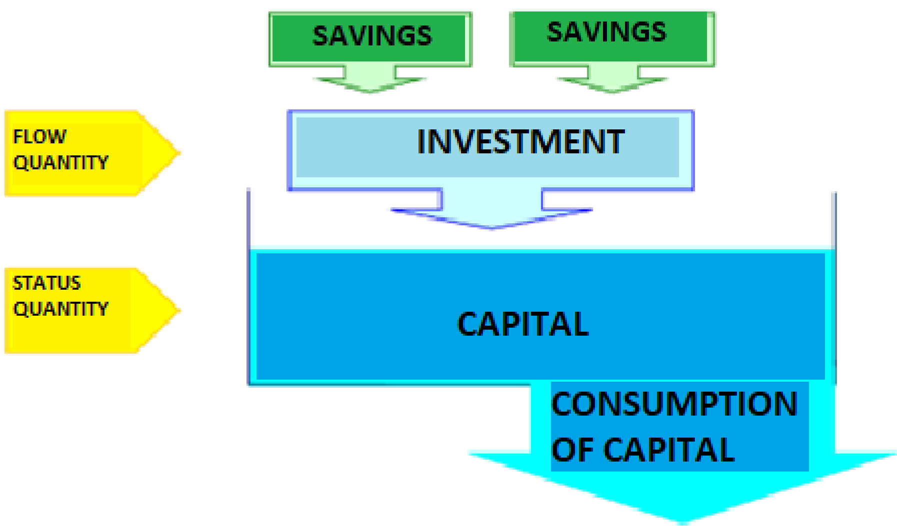

Fig. 92 Capital and investment

Figure 92 Illustrates the relationship between investment and capital. Capital is a "state" (buildings, machines, supplies, knowledge, etc.), investment is a "flow", which supplements, or where appropriate, increases the "state" of the capital.

The picture shows even a significant property of the capital – "the wear and tear of capital". Capital goods are not consumed in the production process at once, but remain here for a longer time. This means that they only wear out, so that their value is not transmitted into new products at once, but gradually. This transition of values is called amortization, which in accounting takes the form of depreciation of material capital goods.

We distinguish two types of wear and tear of capital goods:

- **physical wear and tear** – the use of capital goods in the production leads to a reduction of their accuracy, the individual components, parts, etc. wear out and this worsens their "quality"; in other words, this reduces the functionality of capital goods;
- **moral of the obsolescence** – the development of science and technology leads to new discoveries and to their implementation in production in the form of qualitatively new capital goods with higher productivity.

In connection with the wear and tear of capital goods, there arises the problem of determining a portion of their original price (value) that was actually transferred to the manufactured product. This problem is solved arbitrarily, i.e. the amount of the amortization (depreciation) is determined on the basis of the accounting regulations, the choice of which is a matter of economic policy.

The attention an economic policy devotes to determining the amount of depreciation is determined by the fact that the depreciation represents:

- a significant part of the production costs of each firm and, therefore, they significantly affect the size of the firm's revenue, which is the basis for the calculation of its taxes;
- a major source of funds for the purchase of new capital goods, which are to replace the original capital goods, that have already worn out, i.e. they play the role of the so-called investment restitution (IR).

9.2.3 Classic Capital Concept

When the firm decides to acquire new capital goods, it must have at its disposal a sufficient amount of financial funds. These can be resourced from the firm's own sources (amortization, retained revenue of public limited companies, etc.) or from other sources (i.e. borrowed).

The supply on the capital market consists of savings of economic entities (subjects). Household savings can take the form of:

- current and term accounts with various monetary institutions;
- different kinds of insurance premiums with insurance companies;
- pension fund deposits;
- purchase of shares or bonds.

On the capital market, all of the household savings take the form of capital supplied to firms.

The demand of capital market is formed by the firms by their need to finance the purchase of capital goods. The firms obtain the funds in the form of:

- bank loans;
- income from the sale of its own securities.

For simplicity, we will continue to abstract from the specific forms of financial instruments and we will assume that:

- households create savings in some abstract form,
- firms, in some abstract form, borrow these savings and convert them into new capital goods.

The capital market is a thus a platform, where the supply of savings, generated by households, meet the demand after these savings from firms or other entities. The price, which is constituted on the capital market, takes the form of the **interest rate**, which functions as a so-called mop-up price, i.e. it creates an accordance of the supply and demand.

Households will be willing to defer a portion of their consumption to the future and create savings only if they are sure to have an increase in future consumption. The reward for the waiting, associated with the deferring of the presentconsumption to the future is in the classical theory the amount they saved (SD). In the future this amount will return to them increased by the increment of the $\Delta S = S1 - SD$, which is generally referred to as the yield of thesavings. This yield most often takes the form of:

- the interest rate of the deposit or bonds;
- dividends from shares;
- capital yield resulting from the difference between the buying and the selling price of securities.

In the following section we will abstract from the specific forms of financial instruments and we will consider only one form of such yield – interest, which we can define as the rental price of loanable funds, paid by to those who want to invest, to those who want to save. Households, however, are usually not interested in the absolute amount of interest on the amount of money saved, but rather in the interest rate.

The households are motivated to save by size of the interest rate, from which they can derive the value of the future funds (i.e. future value S1, today's savings S0). If this future value will be paid after one year, it can be calculated as follows: $\mathbf{S1 = (1 + i)\,.\,S0}$

If today's amount will be paid after a certain time period (of n years), is the factor in parentheses (the so-called interest factor), multiplied to the power of the value of n: $\mathbf{S_1 = (1 + i)^n\,.\,S_0}$

9.2.3.1 Capital Market Supply - Creation of Savings

Economic entities do not have to spend all of their income immediately on their consumption. They may also keep a part of their income available as savings. Household income can be therefo-re split into:

- consumption spending (C),
- savings (S).

The amount and emergence of savings specify the supply of capital market, since the savings of economic subjects represent the portion of their disposable income, which has not been spent on consumption and was supplied on the capital market for firms to purchase capital goods.

From the perspective of time, the following capital market phenomena can be distinguished:

- short-term capital supply – in the short term, the exact amount of savings is clearly determined, i.e. this also means that the savings are constant and the supply curve is vertical,
- long-term capital supply – is the response of households to changes of interest rates by their varying amount of savings, i.e. the supply curve is growing.

Figure 93 represents a comparison of short-term and long-term capital market supply.

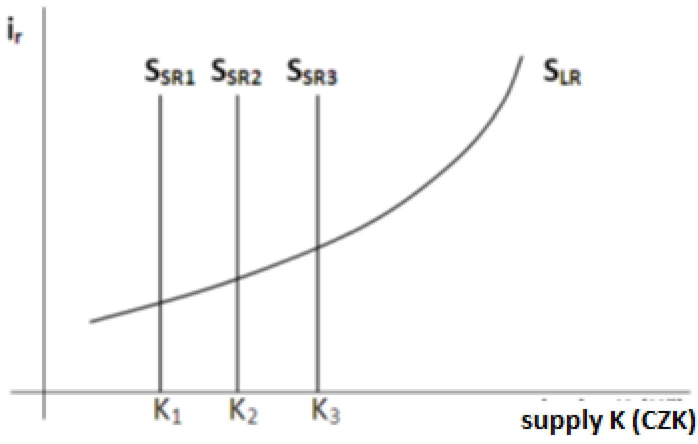

Fig. 93 Fig. 93 Capital market supply

9.2.3.2 Capital Market Demand

Entities (usually firms) that demand capital on the capital market, are lead by their effort to maximize profits.

The equilibrium of the firm occurs, if the marginal product revenue is equal to marginal cost (**MRP = MFC**). In the case of a perfectly competitive market, the marginal cost of capital is equal to the interest rate, therefore the cost is exactly equal to the price of capital.

The demand curve is again determined by the income of the marginal product curve, which is a multiple of marginal revenue (in perfect competition this is equal to interest rates) and the marginal physical product of capital.

The capital demand is therefore a decreasing function of the interest rate and is determined by the revenue from the marginal product. The capital demand is derived from the demand for the finished goods, that is produced using the given capital.

Due to the law of decreasing revenue, the marginal capital product function decreases. The downward progress is also reflected in the investment demand function (loan funds), because the more capital goods the firms use, the lower is their marginal product. Therefore, the firms demand more investment resources, only **at a lower interest rate**.

The demand for investment resources depends on **the real interest rate.**

In the case of financial loans, the interest rate consists of two parts:

- the first part is the compensation to the lender for postponing their consumption to the future,

- the second part compensates the lender for the depreciation caused by inflation.

Therefore, we distinguish between nominal and real interest rates. The nominal interest rate is the rate of interest paid on financial loans. The real interest rate is the nominal interest rate with a deducted inflation. The following applies:

nominal interest rate = real interest rate + expected inflation rate

9.2.3.3 Capital Market Equilibrium

As is clear from the different shape of the short- and long-term supply, we will on the capital market be distinguishing between a short-term and long-term equilibrium. In the short-term the capital supply is given, since savings are a state value and capital stock does not change in the short-term. The vertical shape of supply and the declining shape of demand are characterized for the short-term equilibrium, when the equilibrium interest rate is determined by their conflict, or to be more precise, only by the amount of demand.

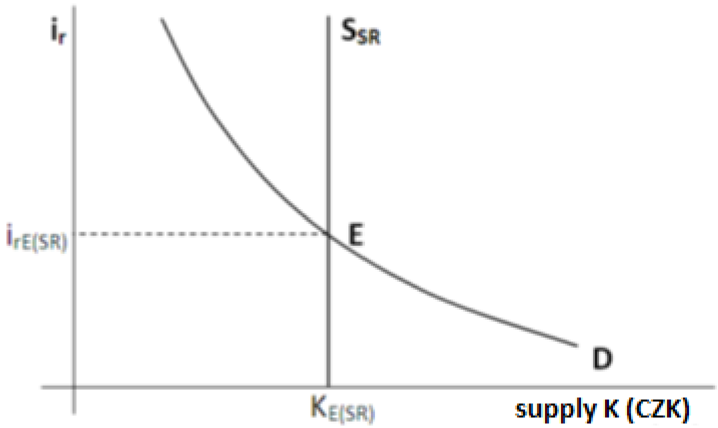

Fig. 94 Fig. 94 Short-term capital market balance

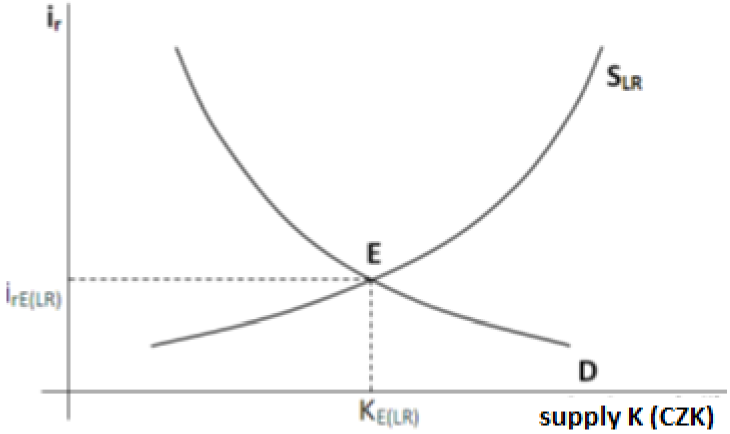

Fig. 95 Fi. 95 Long-term capital market balance

In the long-term, the capital supply is growing, since households have the possibility to decide whether and to what extent they will offer (supply) their savings on the capital market on the basis of the interest rate. Both equilibria are illustrated in Figures 94 and 95.

Remember

In the capital market, we distinguish between short-term and long-term equilibrium. In the short term, the supply of capital is fixed and the stock of capital does not change. During this phase, the supply curve is vertical, while the demand curve is downward-sloping, and the equilibrium interest rate is determined by their intersection. Conversely, in the long term, the supply of capital is increasing, allowing households to decide on the level of their savings based on the interest rate.

Σ

Capital represents an important factor of production in the economy, characterized by being a secondary form of productive input, i.e., a factor that has already been produced and is re-entered into production processes. Within this concept, capital contributes to production in various forms. Among the most well-known forms of capital are financial resources, but it is also important to distinguish between material and non-material capital.

Material capital, also known as capital goods, includes physical objects such as buildings, machinery, and inventories that are directly used in production processes. The quality and availability of these physical assets can significantly affect the efficiency and performance of businesses. On the other hand, non-material capital includes assets that do not have a physical form, such as software, licenses, skills, and expertise. In modern economies, intangible capital is playing an increasingly important role, as innovation and technological progress become key factors in competitive advantage.

The demand for capital is created by firms that need financing to expand their production capacities and invest in new technologies and innovations. On the supply side, there are households that have disposable financial resources and are looking for opportunities to increase the value of these resources. These savings of households are transformed into investments in the capital market, creating an interaction between demand and supply.

Equilibrium in the capital market occurs when the supply of capital equals its demand, which is reflected in the determination of the equilibrium interest rate. However, this equilibrium can have both short-term and long-term characteristics, and various factors such as changes in economic conditions, interest rates, and economic growth can influence this equilibrium and cause it to shift. As a result, understanding the functioning of the capital market is crucial for analyzing economic processes and making effective decisions in the areas of investment and production.

Review qusetions

1. Explain what happens in the capital market, if the real interest rate will be higher than its equilibrium level.
2. Explain the mutual relationship of investment and savings.
3. Explain the methods of return on capital determination.

Test questions

1. What is characteristic of capital as a factor of production? a

a) It is a primary form of productive input
b) It is a secondary form of productive input
c) It is not part of the production process

2. The forms of capital include:

a) Only financial resources
b) Material and non-material capital
c) Only physical capital

3. What creates the demand for capital?

a) Households seeking to increase the value of their resources
b) Firms needing financing to expand their capacities
c) Government institutions

4. When does equilibrium in the capital market occur?

a) When the supply of capital is higher than the demand
b) When the demand for capital is higher than the supply
c) When the supply of capital equals its demand

5. What can influence the equilibrium in the capital market?

a) Only changes in interest rates
b) Changes in economic conditions, interest rates, and economic growth
c) Only economic growth

Answers:

1.b, 2.b, 3.b, 4.c, 5.b

Practical task

1. Analyze the impact of technological progress on the significance of intangible capital in the modern economy. Provide specific examples of how the ratio between materiál and non-material capital is changing in various sectors (e.g., IT, manufacturing, services) and what consequences this has for firms' investment decisions.

2. Prepare a case study that illustrates how changes in interest rates affect the equilibrium in the capital market. Focus on a specific period in your country's history (e.g., the last financial crisis) and analyze how changes in interest rates influenced the investment behavior of firms and the savings of households. Discuss the short-term and long-term impacts of these changes on the economy.

Literature

[1] TULEJA, P. NEZVAL, P. MAJEROVÁ, I. *Základy mikroekonomie.* 1. vydání. Brno: CP Books, 2005, 262 stran. ISBN 80-251-0603-9

[2] JUREČKA, V. a kol. *Mikroekonomie.* Praha: Grada Publishing, 2010. 360 stran. ISBN 978-80-247-3259-6

[3] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[4] SOUKUPOVÁ, J., 2018. *Mikroekonomie.* 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 10

Investment Decision Making in Conditions of Risk and Insecurity

After studying this chapter you will be able to:

- characterize assets and their types;
- distinguish perpetuity and annuity;
- clarify the price of assets;
- explain the concepts of portfolio, risk, moral hazard.

Keywords:

Asset, real asset, financial asset, perpetuity, annuity, asset price, portfolio, risk, moral hazard.

After studying this chapter you will be able to:

- Characterize different types of assets and explain the difference between real and financial assets.
- Define the terms perpetuity and annuity and calculate their prices.
- Explain the relationship between the price of assets, expected return, and the interest rate.
- Describe the factors influencing the rate of return on an asset, including the impact of risk.

Chapter preview

- The following chapter focuses on the characteristics and types of assets, their valuation, and the factors influencing their profitability. It explains the concepts of real and financial assets, perpetuities and annuities, and their valuation. The chapter examines the relationship between the price of assets, expected return, and the interest rate. It also addresses the rate of return on assets and the impact of risk on investment decisions. Additionally, the chapter explains the concept of moral hazard and its effects on financial markets. Finally, it discusses the differences between investing and entrepreneurship, including specific aspects of entrepreneurial activity such as arbitrage and innovation. Overall, the chapter provides a comprehensive view of the issues related to investing and entrepreneurship.

Chapter objectives

- Provide students with basic knowledge about different types of assets and their valuation.
- Explain the relationship between risk and return in investing.
- Introduce students to the issue of moral hazard.
- Clarify the differences between investment and entrepreneurial activities.

Estimated study time

- Estimated time for study is 180 minutes.

10.1 Assets Types and Characteristics

The process of investment represents the handling (mainly purchases and sales) of assets. An asset is a source of revenue, therefore it is everything that brings its owner a particular valorization. We distinguish between real and financial assets.

Real assets – they bring their yield by their use in the production process. These are the real factors of production such as buildings, machinery, equipment, land, means of transport, etc. as well as non-material inputs, namely patents, licenses or trademarks.

Financial assets represent **the rights** to returns from the real assets through deposits in banks and investments in securities (stocks and bonds, in particular) that bring the revenue in the form of interest on deposits, bonds or dividends on shares.

10.1.1 Perpetuity and Annuity

Some of assets bring revenue permanently, such as land or shares. Such "eternal asset" is referred to as **perpetuity**. These are assets with a lasting value and greater certainty of return.

Other assets bring yield only after for certain time, such as material mines, patents or some types of bonds. Such an asset is called **annuity**.

The price of different types of assets is created by the market. However, we must distinguish between the rental prices and the purchase prices. For example, the price of land lease (rent) is the price at which this land is hired or rented, while the purchase price is the price at which this land can be sold and bought. In the latter case, the asset changes its owner.

10.1.1.1 Perpetuity and Annuity Price

The price of real asset – perpetuity is e.g. the price of land, because land has lasting value and infinite lifespan. The price is derived from the expected net yields from the assets and from the interest rate, which is the alternative (second best) opportunity to make use of the asset.

The expected net yields from the real assets – perpetuity is referred to by the symbol V and interest rate r. Then, the perpetuity market creates price p, for which the following applies: $\mathbf{P = V/r}$

The expected net proceeds from the assets can be found out when we compare the two situations: how much would we earn, if we didn't have the asset, and how much do we expect to earn, if we have this asset. The difference between the two results is the expected net yield from the asset.

It follows that the asset prices depend only on the expected income from assets and the interest rate. This knowledge is a very important and useful, especially when we are thinking about the market price of assets that have already been manufactured in the past and are now being bought and sold.

Another important fact is that the costs for which a real asset was produced, have no effect on its price. The price only depends on **the expected yield of the asset and the interest rate.**

The price of the asset has no relation to the cost of the past. These are referred to as sunk cost. The price of real assets depends only on the costs of the future, that is, what costs will the asset require in the future.

The price of financial perpetuity, such shares, is also derived from its expected yield and interest rate. The same formula applies: **P = V/r**, where P is the price, V is the expected net revenue and r is the interest rate that represents the second-best opportunity.

Annuity Price

Annuity is an asset that brings revenue just for some limited period of time. The price of an annuity, which will create a yield for just one year is: **P = V/(1+r)**

where P is the price of the asset, V is its expected net yield for that year, and r is the annual interest rate. (We assume that the income from the assets is coming during the year in the same time layout as an interest on a bank deposit.)

From this relation it follows that the investor converts the future expected return into its current value. This recalculation is described as discounting of future revenue. For example, if you get a yield of 100 CZK a year from now and the interest rate is 2 %, then today's value of the future 100 CZK is now 98 CZK. This is because if you saved the amount of 98 CZK in a bank at 2% interest, you will in one year have 100 CZK.

The investor is willing to pay for an asset at most the price, that is equal to the present value of the investor's expected future returns. If the asset will bring only one return, and that is a year from now, the maximum price that the investors will be willing to pay for it will be P = V/(1 + r), where V is the yield created in a year from now, and r is the interest rate. If the asset brings its yield not in one, but in two years from now, the investors once again recalculate the future yield into its present value that is V : $(1 + r)^2$. If they saved this amount for an interest, they would in two years receive exactly the amount V.

The market price of an annuity, that brings its yield for several years:

$$\mathbf{SH = V/(1+r) + V/(1+r)^2 + \ldots + V/(1+r)^n}$$

where P is the price of the asset, V is the expected annual yield of the asset, r is the annual interest rate, and n is the lifetime of the asset, i.e. the number of years during which the asset yields annual the yield V. It is therefore about an asset with a limited lifetime of n years.

An interesting and important finding of this chapter is the indirect (reciprocal) proportion between the cost of the asset and the interest rate. The decline in interest rates increases the market prices of assets — land, houses, machines, computers, even used cars, but also shares and other securities. The growth of interest rate decreases asset prices. Thus, if for some reasons, we expect the interest rate to decrease, we expect that it will increase prices of all real and financial assets.

10.2 Asset Return Rate

The rate of yield of assets is an important criterion for investment decision-making. In further explanation, we will limit ourselves solely to the return rate of perpetuities. Their return rate is defined as the proportion of yield and price of the asset: **v = V/P**

where v is the rate of yield, V is the yield and P is the price.

If we are counting the return rate for some asset in order to compare it with other investment opportunities, we have to relate the yield on asset to its current price, i.e. the price at which it is possible to buy or sell this asset on the market today. Investors compare the return rates of various assets and naturally buy the assets which yield the highest rate (for now, we are neglecting the asset risks).

The rates of return of different assets cannot vary in the long-term, because investors cannot try to always buy assets with higher rates of return. In this way, however, they raise the price of such assets, and as a result, their rate of return decreases. This means the emergence of the tendency towards convergence rates of returns from assets.

This also applies to financial assets. Let us imagine the shares of several different firms, which do not differ in their risk.

People sometimes think that the best investment is to buy stocks of well-known and prosperous firms. But the prosperity of the firm on the stock exchange will soon translate into high prices of its shares. When you buy shares of such well-known and prosperous firms, you find that you actually haven't done any good business – the rate of return from those shares (the ratio of their revenue to their market price) does not significantly differ from the yield provided by the shares of other firms. Does this mean that it is not worth to buy stocks of well-known and prosperous firms? No, it's worth it, but only when they are not yet known and prosperous. Then you can buy their shares at a low price. However, once they become known and prosperous, you will have to buy their shares at a high price.

The asset market promotes the tendency to the same rate of revenue. But one of these assets is also, a bank deposit, with the rate of return of the interest rate. It follows that **if the risk of bank deposits and other assets is the same, then the rate of return from assets tends to approach the interest rate of bank deposits.**

Remember

The rate of return on an asset, defined as the ratio of return to the price of the asset (v = V/P), is a key factor in investment decision-making, with investors preferring assets with the highest rate of return. There is a tendency for rates of return to converge because as investors start buying assets with higher returns, their prices increase and the rate of return decreases. If the risk of the assets is the same, the rates of return tend to approach the interest rate on bank deposits.

10.2.1.1 Risk and Rate of Return

In practice, the assets differ in their risk. The risk of the asset affects its return, but does not affect the rate of return, but rather the probability of achieving it. In this way, the risk affects the price of the asset. Assets with higher risk rate have a higher rate of return, because the higher level of risk reduces their price. Actually, we can observe that the lowest rates of return have the safest assets, such as government bonds, deposits in large and solid banks or shares of big and well-known firms. It is the different rate of risk, that explains why the rate of return of the various assets differs. The greater the risk, the higher the rate of return must be because only a high return rate will motivate investors to invest in risky assets. The rate of risk we can be expressed by the probability of expected yield.

Investing in risky assets is fairly similar to gambling. The investor's decision making is affected by two criteria:

- the rate of expected returns and

- the rate of risk expressed by the probability of expected yield

The investor would, course, want to maximize the rate of return while minimizing risk. However, this is difficult, because the assets with a higher yield are very risky.

Investment strategy that would reduce the risk as much as possible, without at the same time reducing the yield:

If the investor divides the total sum of investment between a large number of assets, they reduce their overall risk. An investor who wants to reduce their risk, should invest in a large number of different assets. Such a fan of assets into which the investor splits their money, is called **a portfolio.** Therefore, if you create such a portfolio, you will reduce the overall risk of your investment. It is unlikely that all the assets in your portfolio would fail and not provide any yield. Indeed, the various assets are subject to various kinds of risks. For example, the reduction in demand for office and retail areas will reduce return from real estate but will not reduce the return from shares of industrial companies or from bank deposits. The decline in interest rates, in turn, will reduce the return on bank deposits, but will increase the price of shares. The failure of one company does not mean that they all the other firms, the shares of which you have in your portfolio, will fail as well.

10.3 Moral Hazard

If a person invests their own money, he or she is governed by their self-interest and their own rational consideration, so that their investment decision making is efficient. Many people, however, do not have enough knowledge that is needed to invest. They are unaware of the existing investment opportunities and do not have experience with the implementation of purchase and sales transactions in asset markets — for example, on the real estate market or the stock market. Therefore, they use the services of different providers, who then invest for them. Professional investors can invest better than laymen, because they can better estimate the expected returns and risks. Their services are, therefore, useful, and generally well paid. But these people are investing other people's money. And it is precisely in such cases where sometimes a very undesirable phenomenon can occur. This phenomenon is called **moral hazard.**

Such moral hazard can occur when an investor is investing other people's money, and he or she participates in the profits of an investment but does not carry the risk of loss. In that case, he or she is motivated for the preference of riskier investments, even if they are (in terms of risk and yield comparison) less advantageous for the actual owners of the invested capital.

Causes of Moral Hazard

If the bank fully felt the risk of loss, the moral hazard would not exist. As soon as someone removes this risk however, the bank is tempted to hazard morally. Moral hazard of banks and other financial intermediaries can ultimately lead to their collective bankruptcy, when their losses from erroneous loans and investments grow to large sizes and the state (i.e. government) establishes that it is not able to cover these anymore. This is a very dangerous phenomenon, which may cause a financial crisis with very serious consequences for the entire economy.

A number of examples from practice shows how great the danger of moral hazard can be for the economy. The essence of entrepreneurship and investing is a realistic evaluation of risks. If the state provides guarantees to someone for their investment it takes away part of the risk, i.e. it artificially reduces the risk of loss. This can lead to inefficient investment – provision of bad loans and to excessively risky investments.

Remember

Investing one's own money typically stems from the interests and rational considerations of the individual. However, many people lack the necessary knowledge and experience, which forces them to rely on professional intermediaries. Moral hazard occurs when investors with other people's money favour riskier projects because they do not bear the risk of loss, and this can lead to inappropriate investments. This problem can have serious consequences for the economy, including inefficient allocation of resources and the potential collapse of financial institutions.

10.4 Business and Investment Activity

Business, as an activity, is full of uncertainty. An entrepreneur is not the same thing as an investor. The investor will be deciding between the purchase of various assets, appreciating their expected returns and risk. But an entrepreneur is doing something more – looking for gaps in the market. Entrepreneurs are always looking for new, still undiscovered and underutilized opportunities that promise (when someone discovers them and makes use of them) high economic profit. The market gaps are filled with two basic types of business activity: arbitrage and innovation.

Arbitrage is a widely spread kind of entrepreneurial activity and can produce a high rate of return. Arbitrage means finding cheap raw material markets or labour force and finding expensive markets of final products and services. The entrepreneur must have the ability to search and discover new markets and enter them (if possible, more quickly than the entrepreneur's competitors). This is what promises the entrepreneur big profits.

Another type of business activity is **innovation**. Innovation is the introduction of new products, services or production processes into the market. It is not so much the inventing itself (the inventor is not the entrepreneur) as more of the market search. The entrepreneur searches for a market in which people lack a specific product or service and are willing to pay well for it.

Therefore, entrepreneurial activity contains something more – it is the **discovery of previously unknown opportunities.** Entrepreneurs are pioneers on the markets by discovering yet unknown and undiscovered profit opportunities and exploiting them. And it is this discovery of new markets that opens the way to big profits. This entrepreneurial discovering has no end. Once the unknown becomes the known and discovered, as soon as the "gap" is taken up with the entry of new entrepreneurs, the starting high economic profit disappears. A real entrepreneur then continues discovering and looking at new and unexplored territory. Business responds to the fact that information is incomplete and markets are imperfect. Perfectly competitive markets offer no place for entrepreneurs. Only imperfectly competitive markets create opportunities for business. It is the incompleteness of information causes that there are gaps in the market and that filling these gaps brings profits. A business that finds and fills a gap in the market simultaneously reduces the imperfection of markets.

Σ

Assets are a key element of economics because they represent all property rights and resources that generate returns for their owners. These returns can take various forms, including income from investments, profits from sales, or dividends. Within assets, we distinguish between two main categories: real assets and financial assets.

Real assets include physical resources that are used in production processes. These resources can take the form of real estate, machinery, equipment, natural resources, and other physical commodities that are essential for the production of goods and services. Real assets are typically considered more stable and less volatile than financial assets, and therefore can provide steady returns over the long term.

On the other hand, financial assets include instruments such as stocks, bonds, derivatives, and other forms of investments that represent rights to earnings derived from real assets. Financial assets are often characterized by greater fluctuations in value and can offer higher potential returns, but also involve a higher degree of risk. Investors often use financial assets to diversify their portfolios and to maximize returns based on their risk tolerance.

Within the various types of assets, we encounter the terms perpetuity and annuity. A perpetuity is an asset that generates returns indefinitely, without a time limit. This form of asset can represent, for example, stocks that pay dividends without a specified end. In contrast, an annuity is an asset whose returns are paid out only for a certain period. A typical example of an annuity includes certain types of insurance products or pension plans that provide regular payments for a predetermined period.

The rate of return on an asset is an important indicator of investment efficiency. It is expressed as the ratio between the return generated by the asset and its market price. This rate is crucial for investors because it allows them to compare different assets and make decisions based on expected returns. The higher the rate of return relative to the market price, the more attractive the investment is for potential investors.

Investors' decision-making is primarily influenced by two main criteria: the expected rate of return and the level of risk. The expected return is the amount of money that an investor hopes to gain from an investment, while risk represents the uncertainty regarding whether and how returns will materialize. Investments with a higher level of risk typically offer higher potential returns, as investors expect to be compensated for taking on greater uncertainty.

In the realm of investments, the phenomenon of moral hazard can also occur. This happens when an investor invests other people's money (for example, through loans or investment funds) and gains a share of the profit from these investments but does not bear the full risk of loss. This situation can lead to irresponsible behavior by investors, who may have a tendency to take on greater risks than they otherwise would, which can disrupt the stability of the financial system. Moral hazard thus presents a challenge for regulators, who strive to ensure that investors act responsibly and that the market as a whole remains healthy.

Overall, it is essential for investors to carefully analyze and understand the relationship between risk and return. Knowledge of this relationship enables them to make informed decisions and avoid situations that could lead to unnecessary losses. In the modern investment environment, it is important to have not only knowledge about various types of assets but also the ability to effectively manage risks and maximize returns based on individual investment goals and needs.

Review questions

1. Explain the difference between financial and real assets.
2. What is the essence of a perpetuity?
3. What is an annuity?
4. How can we calculate the rate of return of the assets?
5. What is the position of an entrepreneur in the market of financial and real assets?

Test questions

1. Which category of assets is considered less volatile?

a) Real assets
b) Financial assets
c) Derivatives

2. What is characteristic of financial assets?

a) They always provide stable returns
b) They are less risky than real assets
c) They can offer higher potential returns, but with greater risk

3. What is the main difference between a perpetuity and an annuity?

a) Perpetuity has a finite duration, while an annuity has an infinite duration
b) Perpetuity generates returns indefinitely, while an annuity does so for a certain period
c) A perpetuity pertains only to real estate, while an annuity pertains only to securities

4. What influences the attractiveness of an investment for potential investors?

a) Only the market price of the asset
b) Only the expected return
c) The rate of return relative to the market price

5. When does moral hazard typically occur in the investment environment?

a) When the investor invests their own money and bears the full risk
b) When the investor invests other people's money and does not bear the full risk of loss
c) When an investor invests only in low-risk assets

Answers:

1.a, 2.c, 3.b, 4.c, 5.b

Practical task

1. Analyze the impact of technological innovations on the structure and functioning of financial markets over the past decade. Focus on the emergence and development of fintech companies, blockchain technologies, and cryptocurrencies. Discuss how these innovations are changing traditional models of investing, asset management, and risk management. Consider the potential benefits and risks associated with these new technologies for individual investors and the stability of the financial system as a whole. Propose how regulators and traditional financial institutions could respond to these changes to ensure the effective functioning of markets and the protection of investors. Reflect on the future development of financial markets and possible scenarios for what they might look like in the next ten years.

2. Develop a strategy for creating and managing a long-term investment portfolio for a young professional who is just starting their career. Consider their current financial situation, potential future income growth, and long-term financial goals, such as purchasing real estate and retirement planning. Propose an appropriate allocation of assets among different investment classes, including stocks, bonds, real estate, and alternative investments. Discuss the importance of diversification and regular portfolio rebalancing. Consider the use of passive investment strategies, such as index funds, and actively managed funds. Propose a plan for regular review and adjustment of the investment strategy over time, taking into account changing life circumstances and economic conditions. Include an educational aspect in your strategy to help the young investor understand the principles of investing and financial planning.

Literature

[1] FUCHS, K., TULEJA, P. *Základy ekonomie.* 2. rozš. vydání. Praha: Ekopress 2005.

[2] HOLMAN, R. *Ekonomie.* 5. vyd. V Praze: C. H. Beck, 2011. Beckovy ekonomické učebnice. ISBN 978-80-7400-006-5

[3] HOLMAN, R. *Mikroekonomie. Středně pokročilý kurz.* 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[4] MACÁKOVÁ, L., SOUKUPOVÁ, J. *Mikroekonomie – repetitorium.* Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[5] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie.* 5. aktualizované vydání. Praha: Management Press, 2010. s. 507-533. ISBN 978-80-7261-218-5. 574 s.

[6] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[7] SOUKUPOVÁ, J., 2018. *Mikroekonomie.* 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

Chapter 11

General Equilibrium and Efficiency

After studying this chapter you will be able to:

- distinguish between the results of partial and overall (general) equilibrium analysis;
- explain the nature and practically solve problems of efficiency in a broader concept than just economic efficiency;
- define production efficiency, efficiency in production and consumption efficiency;
- know the basic principles of the formation of general equilibrium.

Keywords:

general equilibrium, overall equilibrium analysis, partial equilibrium analysis, exchange efficiency, box scheme of production, efficient allocation of resources,

allocation rules, contractual exchange curve, boundaries of attainable utility, marginal rate of product transformation, social well-being, overall efficiency.

After studying this chapter you will be able to:

- Explain the concept of general economic equilibrium.
- Describe the conditions for achieving general equilibrium.
- Distinguish between economic efficiency and fairness.
- Define social well-being and the factors that influence it.

Chapter preview

- The following chapter addresses the concept of general economic equilibrium and the conditions for achieving it. It explains the difference between economic efficiency and fairness in the context of resource allocation and wealth distribution. Furthermore, it delves into the theory of social well-being, including its definition and the factors that influence it. The chapter presents various criteria for evaluating social well-being, such as the egalitarian standard, the "social conscience" standard, and the "below the poverty line" standard. It also analyzes the conflict between economic efficiency and fair distribution in the economy. The chapter includes a graphical representation of the social well-being function and the utility possibility frontier. In conclusion, it discusses the possibilities for achieving optimal social welfare and the compromises between efficiency and fairness in economic decision-making.

Chapter objectives

- Distinguish between partial and general (overall) equilibrium analysis and explain their significance in economic theory.
- Define and apply the concepts of productive efficiency, exchange efficiency, and productive-consumptive efficiency in a broader context than just economic efficiency.
- Explain the basic principles of forming general equilibrium and their importance for understanding the functioning of the economic system as a whole.
- Analyze and interpret key terms related to general equilibrium, including efficient resource allocation, contract curve of exchange, utility possibility frontier, and marginal rate of product transformation.

Estimated study time

- Estimated time to study is 150 minutes.

In the previous study chapters, we paid attention to the analysis of consumer behavior and the analysis of company behavior. The basis of this analysis was the individual markets in their isolated form. In other words, in previous lectures we focused our attention on Partial ***Equilibrium Analysis***, that is, ***each market was examined separately as an independent system.***

Now we begin ***the analysis of the general (overall) equilibrium*** (General Equilibrium Analysis), which means that we will see the economy as a system ***of interconnected individual markets*** (it is not yet a macroeconomic approach, but it is the "first step" in the direction of macroeconomic analysis). The aim of this analysis will be to determine the conditions of equilibrium on all markets simultaneously, thereby defining the conditions for the effective functioning of the market economy. Knowledge of these conditions makes it possible to analyze the causes of the so-called ***market failure*** and subsequently to justify ***the task of the state. General equilibrium and efficiency, market failure and the state's microeconomic policy*** represent three areas of problems that will be addressed in turn in the next three lectures.

11.1 Assumptions of the General Equilibrium Model and the Definition of Efficiency

The economy represents a very complex system of individual entities and the relationships between them. Therefore, a simplified model of the economy is used to analyze the general equilibrium and to derive the efficiency conditions. We will use this model to derive and formulate the ***general equilibrium conditions*** of a simple economy, representing an economic system consisting of interconnected markets. The construction of this model is based on the following assumptions.

Assumptions of the 2x2x2x2 model

1. There are only ***two people (A and B)*** who make up the company.
2. There are ***only two goods (X and Y)*** and consumers spend all of their income to obtain them.
3. There are only ***two factors of production*** – labor (L) and capital (K), which consumers own and rent to earn income. The total amount of resources is ***constant*** and there are ***perfect substitutes for them.***
4. There are only ***two companies*** that produce the given goods.
5. All markets are perfectly competitive.
6. There is only natural exchange.
7. The economy is closed, there is no foreign trade.

The first four assumptions are the basis of why the model is called the **2 x 2 x 2 x 2 *model*** . Already in this simple (very unrealistic) model we can identify six markets:

- the labor market in production X and the labor market in production Y,
- the capital market in the production of X and the capital market in the production of Y,
- the market for final good X and the market for final good Y.

General equilibrium assumes a state of equilibrium in production and consumption at the same time . To achieve it, the following prerequisites must be met:

- Efficiency in production
- Efficiency in shift
- Production - consumption efficiency

11.1.1 Definition of the Concept of Efficiency

Efficiency occurs in a situation where the resources of the economy are allocated (distributed) in an optimal way. And that's why we need to know how economic theory understands the optimum. Defining the efficient allocation of resources is very difficult because there is not only one quantity to be maximized. For example, it is not possible to talk about the output of the economic system as a whole, because there is no clear way of connecting (aggregating) all produced goods. Similarly, we do not know how to measure the utility of the whole society, and so we can hardly decide whether the overall welfare is as high as the existing resources allow. For these reasons, it is clear that optimality is a multidimensional matter and thus the problem of aggregation arises. However, these facts do not prevent us from introducing the concept of efficiency in its abstract form.

Definition of efficiency

If there are multiple beneficial activities, we can call the situation efficient if one of these activities cannot be increased without simultaneously reducing another activity.

The opposite case – the situation is inefficient if the conditions can be improved.

Assuming that we have a fixed amount of various goods (consumer) and a fixed amount of production factors (production), we can, in accordance with the above definition, define the effective allocation as follows:

1. ***production efficiency -*** the allocation must achieve a situation where it is not possible to redistribute the available production factors in such a way that more of one good is produced while the amount of other goods remains unchanged;
2. ***exchange efficiency*** – it must not be possible to redistribute a given stock of goods in such a way that the utility of one consumer increases and the utility of other consumers does not change;
3. ***efficiency in the consumption of production factors*** (i.e. efficient combination of produced goods) – it must not be possible to change the structure of production in such a way that the utility of one consumer increases without reducing the utility of another consumer.

11.2 Production Efficiency, Exchange Efficiency and Production Consumption Efficiency

11.2.1 Production Efficiency

Production efficiency can be defined as follows: A fixed amount of resources will be efficiently distributed in the economy if it is not possible to produce more of one good without having to reduce the production of another good.

Determining the conditions that enable efficiency in production is complicated by the large number of firms that use different resources and may produce many products. It is therefore necessary to monitor not only the allocation of resources within individual companies, but also how resources are distributed between companies. We obtain the resulting information through three successive steps aimed at deciding on:

- allocation of inputs within the firm,
- allocation of inputs between firms,
- structure of the company's output.

If we summarize the conditions that must be observed in all these three situations, we obtain "allocation rules", the validity of which is a prerequisite for the efficient course of production.

11.2.1.1 Optimal Selection of Inputs by One Company

The starting point of the analysis: one firm, two inputs (labor and capital), production of two goods X and Y.

A firm will produce efficiently if it is not possible to relocate resources in such a way that the production of one good cannot be increased without limiting the production of the other good. We can find out the assumptions of such a situation using ***the marginal rate of technical substitution (MRTS),*** which is expressed by the isoquant guideline.

Allocation rule 1

The first condition of production efficiency is the allocation of a fixed amount of labor and capital within the firm, in which the marginal rate of technical substitution of both production factors is the same for both produced goods, and both production factors are fully utilized. We can graphically express the 1st allocation rule graphically using a box diagram and a production possibilities frontier curve.

The production box diagram shows all possible ways of allocating *K* and *L* between the production of two goods. It is created by joining two isoquant maps (graphical representation of production functions ***X*** and ***Y***).

The quantity of inputs that is maximally available for the company determines the dimensions of the box: the quantity *L* is given by the distance $\boldsymbol{0_X L^X}$**, the quantity *K* by the distance** $\boldsymbol{0_X K}$ **,** *or* $\boldsymbol{0_Y L^Y}$**, and** $\boldsymbol{0_Y K^Y}$**,**

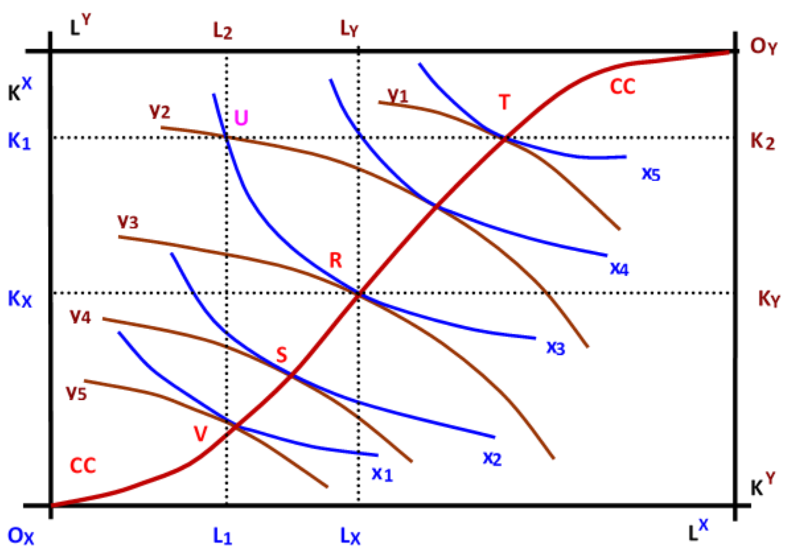

Fig. 96 Production box diagram

E.g. the U point represents a situation where the economy spends:

- L1 units of labor and K1 units of capital to produce x3 units of X;
- L2 units of labor and K2 units of capital to produce y3 units of Y.

Point U is a possible production point, but it is not a point of efficient production, since it is possible to move along the isoquant x3 to point R on the contract curve CC without limiting the production of good X, while achieving with the same volume of resources an increase in the production of good Y (shift from isoquants y3 to y4).

It is clear from the figure that all points lying on ***CC*** *(Contract Curve)* show the efficient allocation of resources (labor and capital), i.e. at these points it is not possible to increase the production of one good without reducing the production of the other good. For points lying outside the contract curve, this conclusion does not apply – they are points of technically inefficient allocation of resources.

Since we already know that the isoquant line expresses *the marginal rate of technical substitution* of capital for labor ($MRTS_{K,L}$), we can define the condition of efficient production by the relation **:**
$(MRTS_{K,L})_X = (MRTS_{K,L})_Y$

The same information provided by the production box diagram can be represented by the Production Possibility Frontier (PPF) ***curve, showing the alternative combinations of two products that can be efficiently produced with a certain fixed range of resources.***

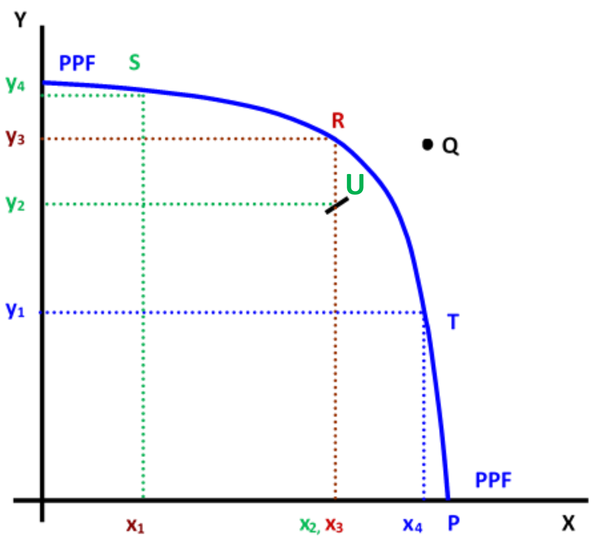

Fig. 97 Production Possibility Frontier (PPF)

For a given range of production resources (labour and capital):

- it is impossible to reach a point lying outside the PPF curve (eg point Q); producing a product corresponding to point Q requires more resources;
- it is possible to produce in the range of all points below the PPF curve (eg point U), but it would be inefficient production because it would not use all resources;
- only the points on the PPF curve represent an efficient allocation of resources and fully correspond to the same points on the contract curve. This means that, for example, an increase in the production of good X is possible only at the expense of a decrease in the production of good Y (and vice versa).

The shape of the production possibilities frontier curve

Properties of the *PPF curve* :

- **PPF is *declining*** $\rightarrow$ as the production of *X increases, the production of Y* must decrease (with a given range of inputs and a given technology);
- **The PPF is *concave*** $\rightarrow$ as the production *of X increases, the production of Y* decreases at an increasing rate. The slope of *the PPF* curve expresses the marginal rate of product transformation.

The line of the PPF curve is called ***the Marginal Rate of Product Transformation*** (MRPT).

Marginal rate of product transformation expressing the rate at which one good can be transformed into another, or by how many units must the production of one good be reduced in order to produce an additional unit of another good.

Formally, it is possible to express the marginal rate of transformation as the ratio of the change in the product of two goods, or by the inverse ratio of MC:

$$MRPT = -\frac{dY}{dX} = \frac{MP_Y}{MP_X} = \frac{MC_X}{MC_Y}$$

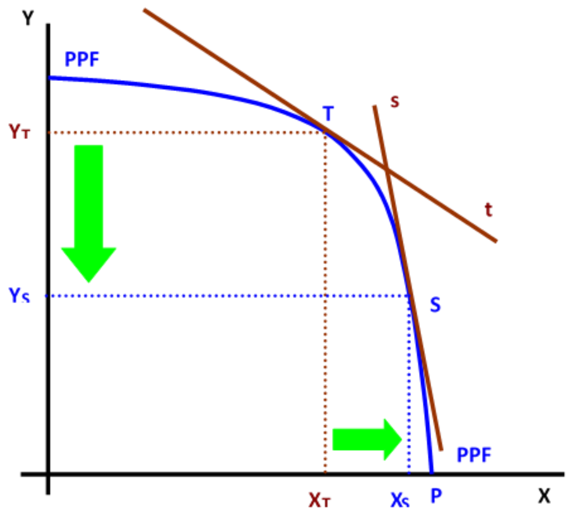

Fig. 98 Marginal Rate of Product Transformation (MRPT)

As the production of *X* increases and the production of *Y* decreases, the productivity of inputs in the production of *X* decreases (***MC_X*** increases*)* and the productivity in the production of *Y increases* (*MC_Y* decreases*).* An equally large increase in the production *of X* requires an ever-greater reduction in the production *of Y* . Approaching the x-axis, *the MRPT value* increases as the ***MC_X /MC_Y*** ratio increases. MRPT growth is therefore affected by the same factors that affect the MC_X /MC_Y ratio:

- ***diminishing returns*** – if both goods are produced under conditions of diminishing returns, increasing output of good X will increase MC_X, while a decrease in product Y will decrease MC;
- ***specialized inputs –*** the concave shape of the PPF can also be explained by assuming that some inputs are better suited to the production of one good than to the production of another good. In this case, the growth of X would require the use of additional (less appropriate) inputs in the production of that good. Therefore, the MC of $_X$ would increase and the MC of $_Y$ would decrease as the inferior product Y could only be produced with more appropriate inputs. Such an argument can apply, for example, to a farmer who has varyingly fertile land for growing a certain crop – in an effort to increase its production, he must use less and less favorable land.
- ***differences in the fund intensity of production*** – even in the case of homogeneous factors and constant returns to scale, the PPF curve will be concave if both products are produced with different proportions of inputs. If, for example, the production of good X is more capital-intensive than the production of good Y, then at every point of the contract production curve (Fig. 1) the ratio K to L in production X exceeds the ratio K to L in production Y. The contract curve must be above the diagonal of the box scheme. In the opposite case, when the production of good X is less capital intensive than the production of Y, the contract curve is below the diagonal.

Opportunity cost - the law of increasing opportunity cost

The concept of opportunity cost is used to express the fact that, assuming fixed inputs, in order to produce a larger quantity of one good, it is necessary to limit the production of another good. The cost of producing more of product X can be more easily measured by the reduction in output Y caused by this increase. The cost of each additional unit of X is best measured as the MRPT (X per Y) at the corresponding point on the PPF curve.

The law of increasing opportunity costs expresses the fact that the more a certain good is produced, the higher are the opportunity costs of these additionally produced units of goods (as a result of decreasing returns, specialized inputs or the existence of different capital requirements of production). In the case of constant opportunity costs, the production possibilities frontier would be represented by a straight line.

11.2.1.2 Effective Distribution of Resources among Firms

If the economy has a fixed supply of resources, then another condition for the overall efficiency of production must be the effective distribution of these resources between individual firms.

Allocation rule 2

The second condition of production efficiency is the allocation of a fixed amount of labor and capital between the two firms, in which the marginal product of both factors of production is the same for both produced goods. Graphically, the 2nd allocation rule can be represented using MP_L curves as we can see in the following Fig. 99.

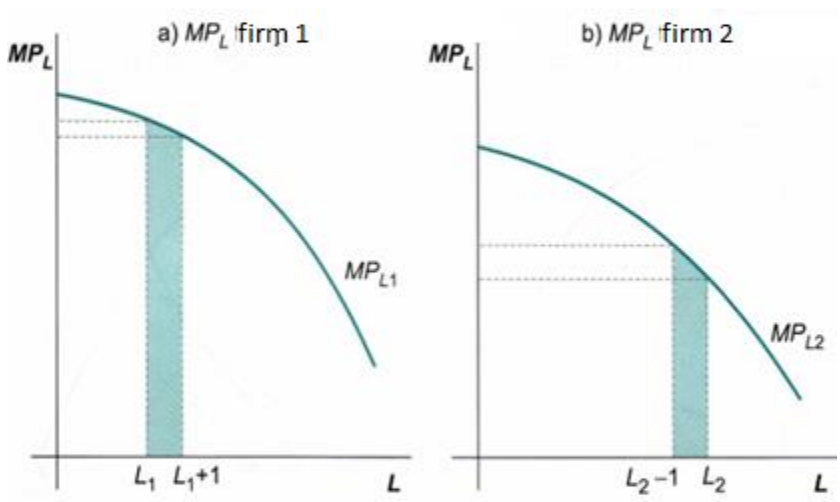

Fig. 99 Marginal Product of Labour of Two Companies

Consider the initial situation as ***the allocation of L_1 workers in firm 1*** and ***L_2 workers in firm 2***. The figure shows that ***the MP_L*** of the first firm ***is greater*** than ***the MP_L*** of the second firm. Using the allocation rule, work was transferred from the second company to the first. If we transferred one worker, the change would be as follows. Figure a) shows the expansion of production and one worker, and figure b) firm B's production is limited due to the loss of a worker. The transfer helped to equalize the MP_L of both firms (in the first firm MP_L decreased and in the second MP_L increased. The transfer of workers will continue until MP, in this case labor, equalizes).

11.2.1.3 Effective Choice of Product Structure by a Firm

Assuming that the previous two conditions of production efficiency are met, another condition must be met – firms must produce an efficient combination of products. Prerequisites are summarized by the 3rd allocation rule.

Allocation rule 3

The third condition of production efficiency is the structure of the production of both goods, in which the marginal rate of product transformation is the same for both firms.

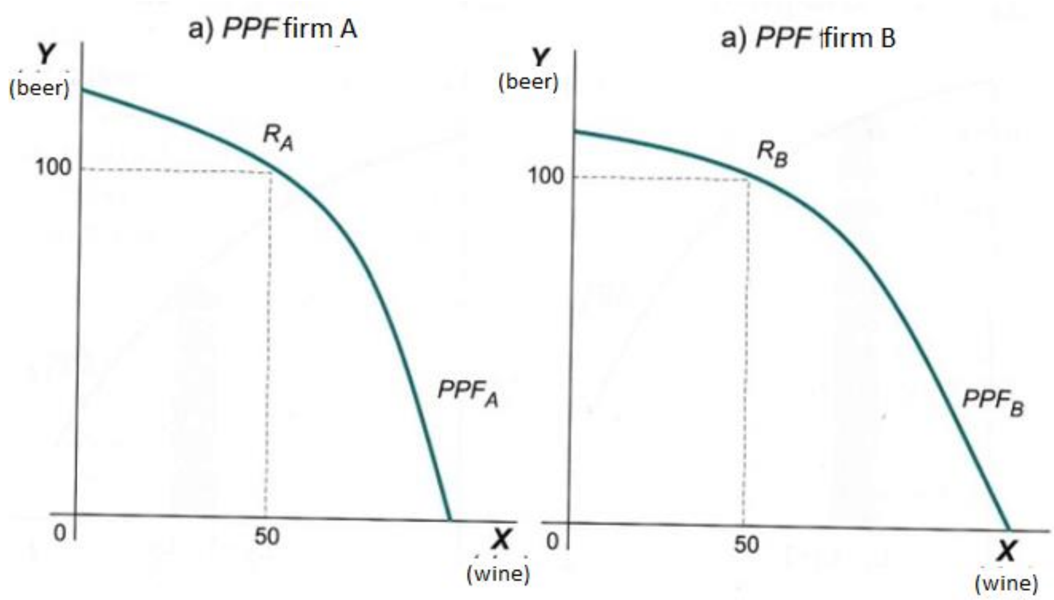

Fig. 100 Production Possibility Frontier Curves of Two Firms

Assume 2 companies A and B, both producing beer and wine. Their PPF curves are shown in Fig. 100. Firm A chooses the production combination at point R_A, where the MRPT is 2/1. In this situation, the company has to give up the production of 2 bottles of beer if it wants to produce 1 more bottle of beer.

Firm B chooses the production combination at point R_B and its MRPT at this point is 1/1. Under such an assumption, production can be increased if firm A produces more beer and firm B produces more wine. The reorganization of production will make it possible to increase the total beer output without decreasing the wine output. The default choice of companies A and B was therefore inefficient.

11.2.2 Exchange Efficiency

The distribution of a fixed quantity of goods is (Paretically) efficient if its redistribution cannot make any consumer better off without simultaneously harming another consumer. The condition of such an allocation is that the MRS_C between any pairs of goods must be the same for all consumers $\mathbf{MRS_A = MRS_B}$

If this condition is not met, it is possible to redistribute both products between both consumers in such a way that the utility of at least one consumer increases without the utility of the other consumer decreasing.

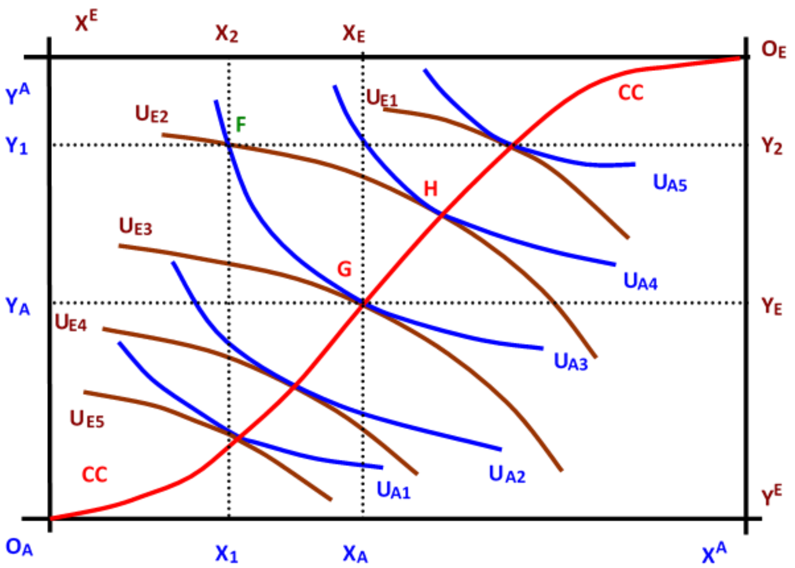

Fig. 101 Box Diagram and Contract Exchange Curve

Box diagram of exchange depicts all possible ways of dividing two products between two consumers is formed by the indifference maps of two consumers (A and B). ***The contract curve of exchange*** **(CC)** = the set of points that represent the effective allocation of two different goods between two consumers is formed by the points of contact of the indifference curves of both consumers in which the marginal measures of substitution in the consumption of both goods are identical.

For an effective exchange, the following must apply:

$$MRS_{C(A)} = MRS_{C(E)}$$

The consumer's indifference map with A is related to the origin O_A and the consumer's indifference map E starts from the origin O_E . The dimensions of the box diagram are given by the quantities of goods X and Y that were produced. Each point in the box plot represents a certain allocation of two products between the two consumers. For example, suppose point F in graph 6. At this point, consumer A consumes X_1 unit of X and Y_1 unit of Y, and E consumes X_2 units of X and Y_2 units of Y. Point F represents an inefficient allocation of products because other allocations are possible that will bring more benefit to both (for example, represented by point G, which means an shift to a higher indifference curve for both consumers) or at least make one better off and not worse off for the other (as, for example, at point H). As in the case of the box-diagram of production, here too, all the equilibrium points form a contract curve.

11.2.3 Production and Consumption Efficiency

The goal of an economic system is to satisfy human needs – efficiency in production may not be desirable at all if the wrong combination of goods is produced in terms of consumer demand. In order to achieve overall efficiency, the conditions for achieving efficiency in production and efficiency in the shift must be met at the same time.

The prerequisite for achieving overall efficiency is the conformity of consumer preferences with production possibilities. In other words, the ratio in which two goods are substitutable in consumption must match the ratio in which they are substitutable in production: $\mathbf{MRS_C = MRPT}$

If a consumer wishes to exchange 2 L of beer for 1 L of wine, but the inputs to production have been allocated so that 1 L of beer could be exchanged for 1 L of wine, then the situation is inefficient – too little wine is produced and consumers give it a higher value than their opportunity costs in production. To produce one additional liter of wine, it would be necessary to reduce beer production by 1 liter - but consumers want 1 liter of wine for 2 liters of beer. As always, when the extent to which consumers wish to exchange goods differs from the extent to which they technically can, there is the possibility of reallocation to achieve a more advantageous situation.

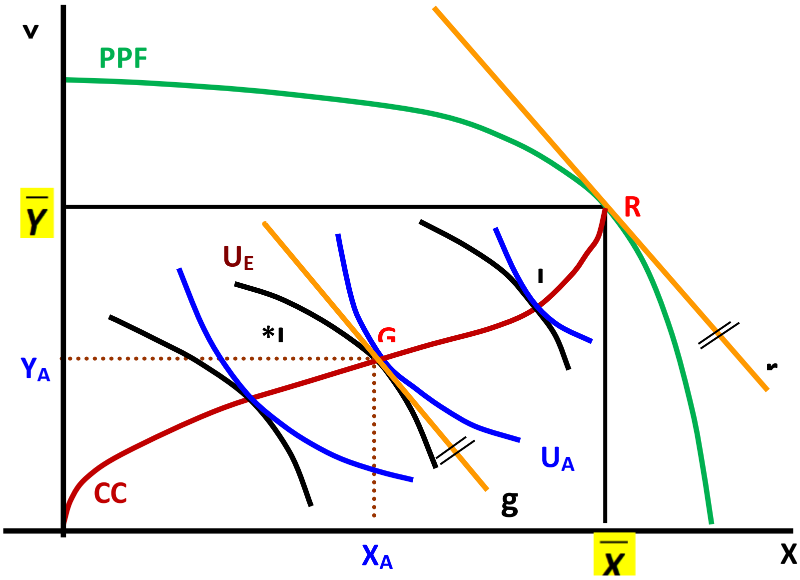

Fig. 102 Production and Consumption Efficiency – Overall Balance

Figure 102 was created by connecting (combining) Figure 2 and Figure 6. A total of $\bar{Y}$ Y units and $\bar{X}$ units will be produced in the economy; $\bar{Y}$ is simultaneously the total quantity of Y and $\bar{X}$ is the total quantity of X that will be divided between both consumers. We know that if the product is to be divided the product is to be divided in such a way as to maximize the satisfaction of the needs of both consumers, the marginal rate of transformation must be the same as the marginal rate of substitution. In the figure, this means that the optimal distribution of both goods between consumers should be represented by the point on the contract curve at which the direction of their indifference curves coincides with the direction of the production possibilities frontier curve at point R (the rag tangents are parallel). This condition is satisfied at point G (and R):

- consumer A receives X_A units of X and Y_A units of Y;
- consumer B receives $(X_S - X_A)$ units of X and $(Y_S - Y_A)$ units of Y.

It follows from the above that in order to achieve an equilibrium situation in the 2x2x2x2 model, the following conditions must be met:

1. the marginal rate of technical substitution of one production factor for another should be the same for both goods;
2. the marginal rate of substitution in the consumption of one good for another should be the same for both consumers;
3. the common marginal rate of substitution in consumption should equal the common marginal rate of transformation.

Conditions for achieving overall equilibrium (many production factors, many consumers and goods):

1. the marginal rate of technical substitution of each factor of production for every other factor of production should be the same for all goods;
2. the marginal rate of substitution of each good for every other good should be the same for all consumers;
3. the common marginal rate of substitution should equal the common marginal rate of transformation for all pairs of goods.

11.3 Achieving General Equilibrium

Until now, in the simplified model of the economy (2x2x2x2), we have assumed the existence of only natural exchange, which cannot be used for the analysis of a complex economy. In such an economy (assuming perfect competition), general equilibrium is achieved through ***the price mechanism.***

11.3.1 Price System and Production Efficiency

We are looking for an answer to the question, how does the price system ensure the full use of all production factors and at the same time their optimal distribution between sectors?

The condition of maximizing the company's profit (on the labor market ***)***

$$MP_{LX} \,.\, P_X = w = MP_{LY} \,.\, P_Y \rightarrow MP_{LY}/MP_{LX} = P_X/P_Y$$

$$P_X / P_Y = \textit{relative price of X to Y}$$

We know that the ratio of the marginal products of labor in both industries is equal to the marginal rate of product transformation between the two products ($MP_{LY}/MP_{LX} = MRPT = P_X/P_Y$) and determines ***the direction of the PPF curve.***

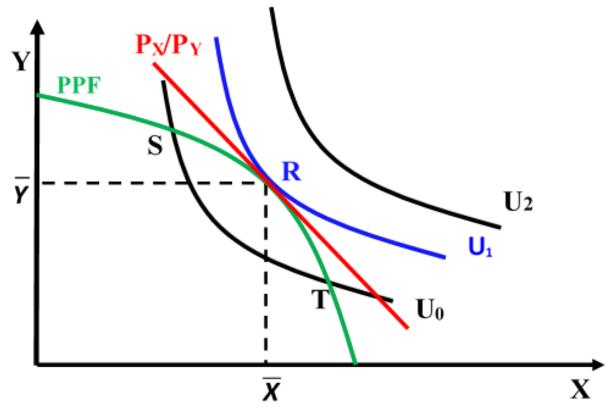

Fig. 103 Relative Price and Production Structure

From the theory of the firm, we know that the ratio of labor and capital used in the production of a certain amount of product depends on the relative price of inputs, and the firm chooses a combination of inputs that corresponds to the point where the isocost touches the isoquant, i.e., their slopes are equal. Since the isocost line expresses the relative price of inputs and the isoquant line the marginal rate of technical substitution, ***MRTS = w/r applies at the considered point.***

Graphically, it is possible to show the equilibrium and non-equilibrium distribution of labor and capital between the production of two products in a box diagram of production (see Fig. 104 and 105).

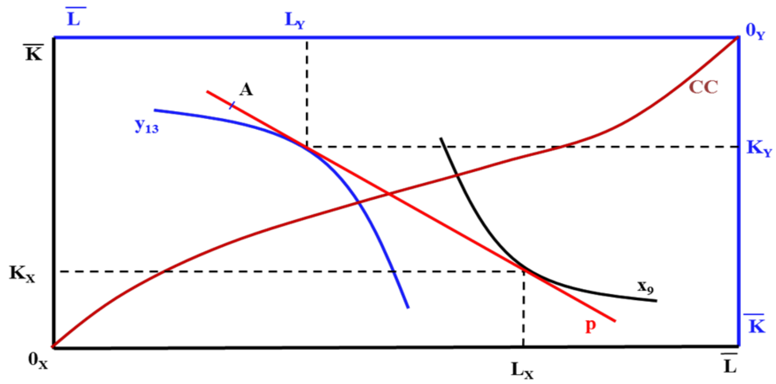

Fig. 104 Relative Prices and Inefficient Allocation of Production Resources

Let us consider the initial distribution of production factors as the distribution shown by point ***A*** in Fig. 104. at the ratio of prices of production factors **w/r** (expressed by the straight line p). In this situation, there will be an excess of capital supply and labor demand: $K_X + K_Y < K$ **and** $L_X + L_Y < L$.

The situation shown in Figure 105 will lead to an increase in the price ratio w/r (line p') and a subsequent redistribution of labor and capital in the production of products.

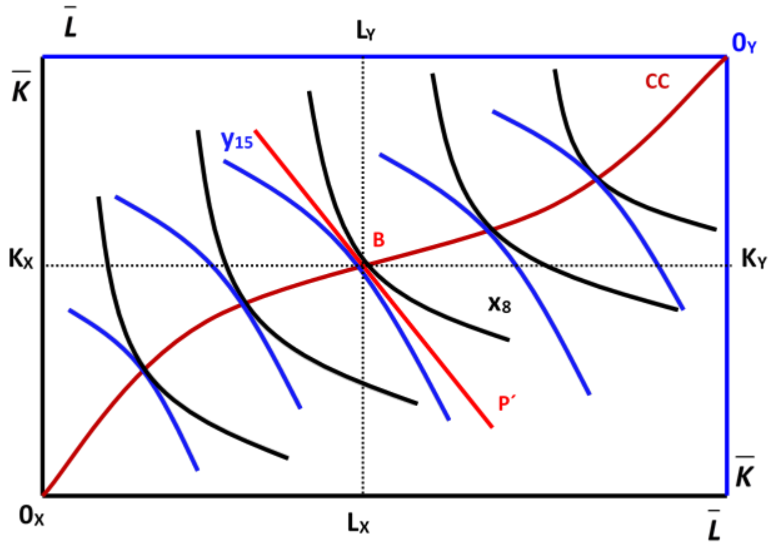

Fig. 105 Relative Prices and Efficient Allocation of Production Factors

11.3.2 Price System and Efficiency in Exchange

From the theory of consumer behaviour, we know that the consumer chooses a combination of goods that corresponds to the point where the budget line touches the highest possible indifference curve; the slope of the budget line (MRS_E, reflecting the relative price of the two products) and the slope of the indifference curve (indicating the marginal rate of substitution in consumption) coincide at this point, and at the considered point it must be true that $\mathbf{MRS_E = P_X/P_Y}$.

The following figures 106 and 107 show a comparison of the equilibrium and non-equilibrium situation in the shift. The non-equilibrium situation shown in Fig. 106 will lead to a change in the price ratio of both products, and the perfectly competitive market will move towards an equilibrium price ratio, such as the ratio expressed by the line p* in Fig. 107, when the indifference curves touch the price line at the same point (D) and the point of contact is the point of the contractual exchange curve.

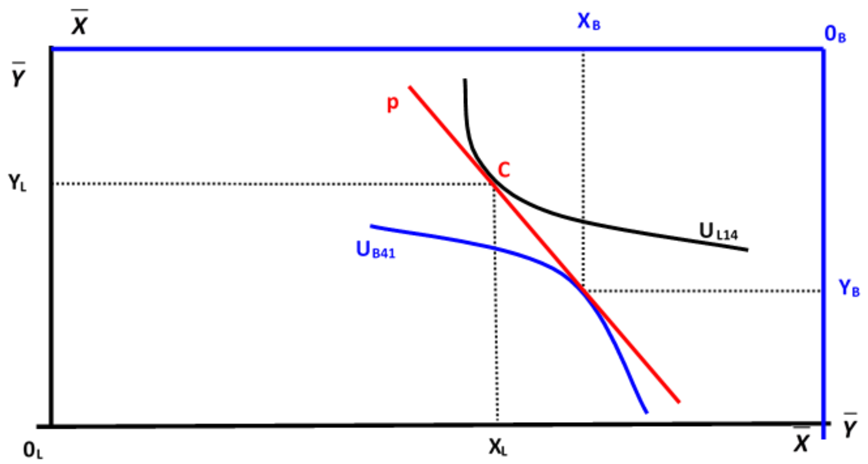

Fig. 106 Relative Prices and Inefficient Product Allocation

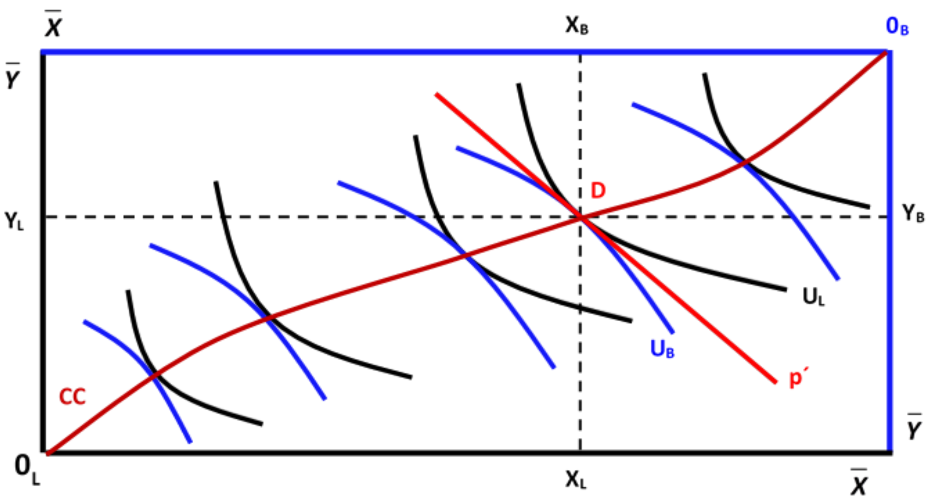

Fig. 107 Relative Prices and Effective Product Allocation

11.3.3 Price System and General Equilibrium

General equilibrium occurs when both consumers and producers encounter and accept the same prices. Therefore, at the equilibrium point it must be true that the ***relative MC of production = P_X/P_Y = relative MU → MRPT = P_X/P_Y = MRS_C***

The general balance is conditioned by the simultaneous fulfillment of 3 conditions:

1. The MRTS of one factor of production for another must be the same for both goods;
2. The MRS of one good for another must be the same for both consumers;
3. the joint MRS must equal the joint MRPT.

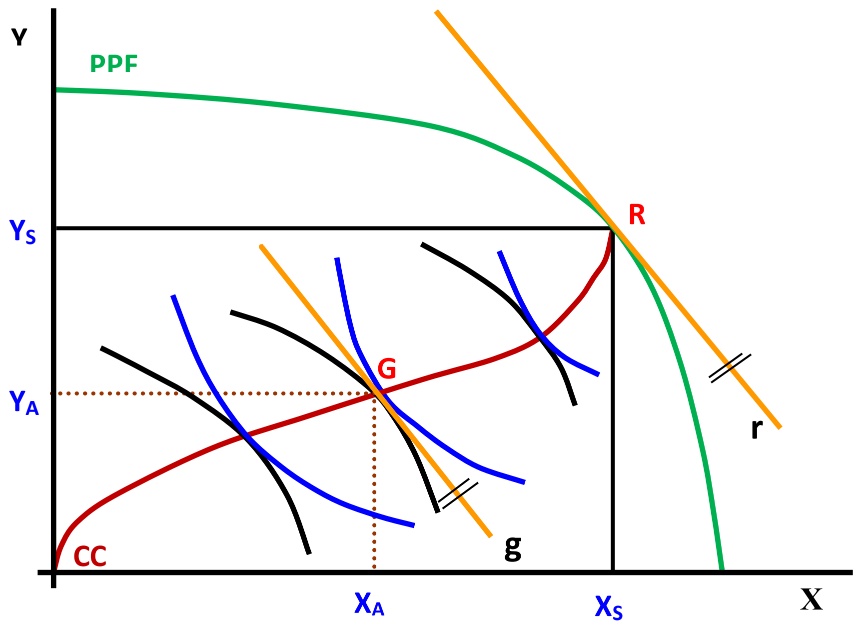

Fig. 108 General Equilibrium

11.3.4 Achieving General Equilibrium

General equilibrium is achieved through the price system. In perfect competition, labor moves between industries until the same wage rate is reached in all industries. In the equilibrium situation, it holds

$$MP_{L(X)} \cdot P_X = w = MP_{L(Y)} \cdot P_Y,$$

The ratio of ***$MP_{L(X)}$* to *$MP_{L(Y)}$* is equal to MRPT** and determines the slope of the ***PPF curve*** *,* so it is expressed as follows:

$$MRPT = -\frac{dY}{dX} = \frac{MP_{L(Y)}}{MP_{L(X)}} = \frac{P_X}{P_Y}$$

When all the labor is employed, the economy is on the *PPF curve* - where, depends on the relative price of the products. The relative prices of goods determine the production structure of the economy.

The relative price of inputs determines ***the ratio of L and K*** used in production, respectively. ***MRTS = P_L/P_K applies .*** In conditions of perfect competition, producers are led to produce on the contractual production curve, or For efficient production through a pricing system.

At the consumer's optimum point, the degree of substitutability of goods in consumption equals the degree of substitutability given by the financial possibilities of the consumer = *the degree of substitutability in exchange* : $MRS_C = P_X/P_Y$. Relative prices determine *the consumption structure of the economy* .

Recall once again that general equilibrium occurs when both consumers and producers encounter and accept the same prices. Firms compare relative prices with relative marginal costs of production, and consumers compare relative prices with relative marginal benefits. Therefore, at the equilibrium point, when aligning the interests of both economic entities, it must be true that

$$\textit{relative MC} = P_X/P_Y = \textit{relative MU}$$

$$MRPT = P_X/P_Y = MRS_C$$

11.4 Efficiency and Fairness

Problem: an economically efficient situation may not be socially desirable, fair and vice versa **.**

To find those allocations that are *economically efficient* (it is not possible to redistribute that would make someone better without harming anyone else) ***and at the same time* fair** (resulting in a socially desirable distribution of income and wealth)**, the theory of social welfare is aimed at.**

In reality, there are several types of obstacles that prevent perfectly efficient distribution of resources by the market. These are mainly ***information barriers, monopoly power, externalities and public goods***. Even in the case of a perfectly competitive model, an efficient allocation may not be socially desirable, as some consumers may prefer one allocation of goods while others prefer another allocation.

For example, in Figure 109 when choosing between efficient allocations E_1 or E_4 (other combinations can also be selected) some consumers may prefer situation E_1 while others prefer situation E_4 .

The division of consumer goods is:

- ***economically efficient*** →when no redistribution is possible that would make someone better off without harming anyone else;

- **fair** → when it results in a socially desirable distribution of income and wealth.

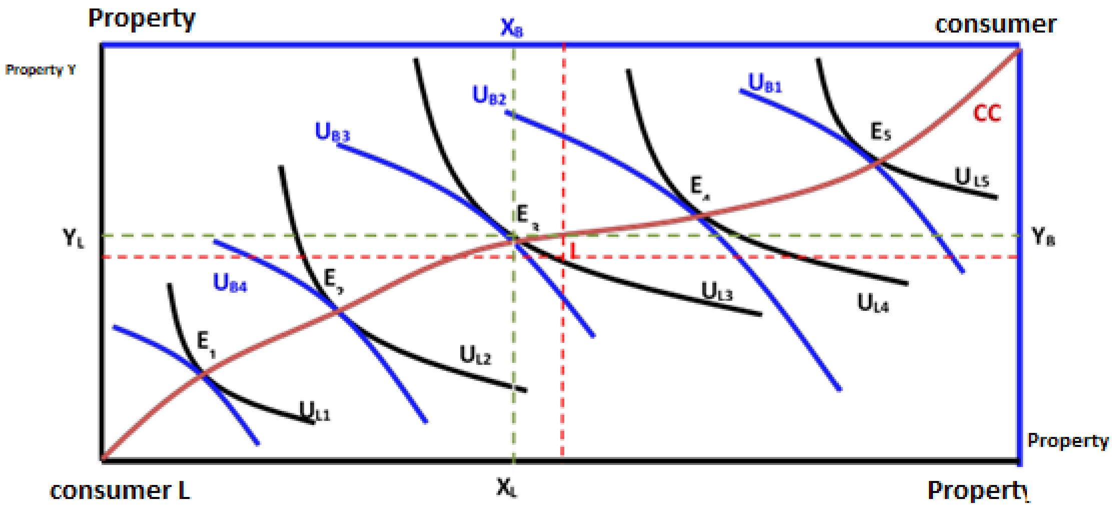

Fig. 109 Box Diagram of Exchange and Fair Allocation

Social welfare theory is focused on finding those allocations that are both efficient and fair***.***

Social welfare (Social Welfare – *SW)* = synonymous with the level of satisfaction of members of society and depends on well-being, or *benefit of individual consumers* . If we denote the utility of the n^th consumer by U_n, then $\boldsymbol{SW = f(U_1, U_2, \dots, U_n)}$.

The factors determining social well-being are mainly:

- total quantity of goods,
- method of distribution of goods;
- but also the health of society, the amount of free time, the degree of environmental pollution, political stability, but also the amount of water precipitation, etc.

Social welfare function: **SW = f (Q, D, H, L, P, S, R, Z)** or $\mathbf{SW = f(U_1, U_2, \dots, U_n)}$.

11.4.1 Social Welfare Criteria

The criterion of economic efficiency states that any change in production and distribution that benefits someone without harming anyone else will increase social welfare. ***The criteria of justice*** offer some basis for social choice between these efficient allocations, i.e., for cases where one member of society can be better off only at the simultaneous worsening of someone else's situation.

Thus, a **utility possibility frontier (UPF) curve can be constructed**, which is shown in Figure 110 and shows different combinations of the utility of two consumers that are achievable assuming a fixed quantity of consumed products.

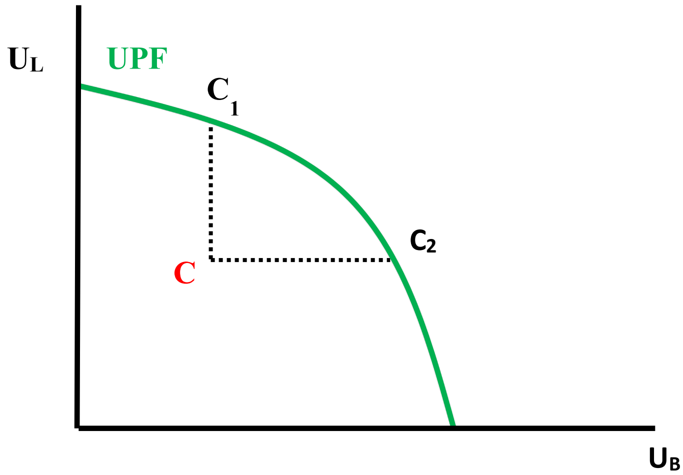

Fig. 110 Utility Possibility Frontier

There are two problems to solve:

- the problem of efficiency criteria and social welfare;
- the problem of criteria of justice and social welfare.
  - ***Egalitarian standard -*** distribution requires absolute equality of all members of society. In our model, both consumers should enjoy the same level of wealth.
  - ***The standard of "social conscience" –*** distribution is based on the idea of mutual dependence of consumer utility in the sense that the well-being of one consumer depends not only on the amount of goods falling to him, but also on the amount of goods available to other consumers. In other words, a rich person has a social conscience in the sense that when he has reached a certain level of well-being, he can increase it further by leaving some of his possessions to the poor.
  - ***The "underdog" standard –*** requires an increase in the share of lower income groups in the company's total income. This perspective requires defining the poverty line and raising the incomes of those below that line.

The social welfare function can be represented graphically by means of indifference curves, or social welfare curve = *W curve* in Figure 111 (simplifying assumption: society = 2 consumers). ***The social welfare curve*** (Welfare Curve – W) depicts all combinations of the utility of both consumers that represent the same level of social welfare.

Assuming the utility is measurable, it is possible to use different combinations of the utility of both consumers to construct the curve of the frontier of achievable utility (see Fig. 111). ***Achievable utility frontier (UPF) curve*** depicts the different combinations of two consumers' utility achievable with a fixed amount of the good produced by an efficient combination of inputs. At each *UPF* point, all three efficiency boundary conditions are met.

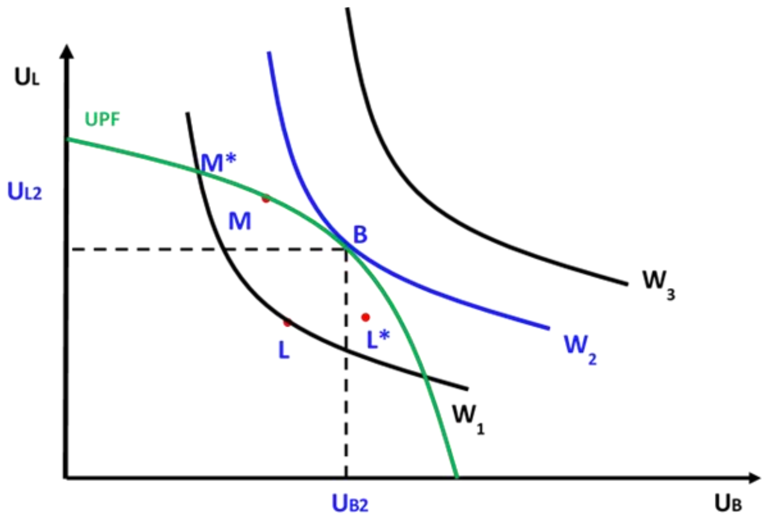

Fig. 111 Social Welfare

The social optimum is the point of contact of the UPF curve and the W curve = the bliss point (Bliss Point) is point B, →the society is on the highest attainable social indifference curve with the efficient use of inputs.

11.4.2 Conflict between Efficiency and Fairness

The conflict between efficiency and justice (or society's value system) makes the point of bliss unattainable. In the initial situation M (Fig. 16), income redistribution in an attempt to reach point B can weaken the motivation to work and lead to a decrease in efficiency = reaching point L, which is located below the UPF curve, or on the lower social welfare curve W_1.

To achieve a certain level of social welfare, it may make sense to accept a certain degree of inefficiency. E.g. in the case of a significantly asymmetric initial distribution, or in a very unfair situation M*, the pension transfer could lead to reaching the point L*. In such a case, what was sacrificed from the point of view of efficiency would be compensated (in terms of social welfare) by an increase in fairness. →The point L* is located below the UPF curve, but on a higher social welfare curve than the W_1 curve.

Σ

Economic analysis is a key tool for understanding the functioning of markets and the processes that affect the prices and quantities of goods. It generally focuses on partial equilibrium problems, which involves the study of determining the price and quantity of a good that is sold in a particular submarket. This approach assumes that a given market is not influenced by other markets and analyzes its functioning in isolation. On the other hand, general equilibrium analysis focuses on all markets simultaneously, taking into account their interconnectedness and conditionality. This more comprehensive approach allows for a better understanding of the interactions between different markets and how changes in one market can affect others. To express the allocation of resources and products within this analysis, box diagrams of production and exchange are often used, which visually depict the relationships between different production processes and exchanges.

The production possibilities frontier (PPF) curve is one of the main tools used in this analysis. This production box plot curve integrates all general equilibrium assumptions into a simple model. The PPF shows the alternative combinations of two products that can be efficiently produced with a given range of available resources. This graphical model helps illustrate the concept of production efficiency, a key assumption of general equilibrium. Achieving production efficiency is, however, dependent on the fulfillment of three allocation rules. The equality of the marginal rate of technical substitution of individual resources for different products ensures that production processes are optimized in such a way that the available combination of resources is used as effectively as possible. The equality of the marginal productivity of different firms ensures that firms produce based on the same level of efficiency, meaning that no firm can increase its productivity without adversely affecting the others. Equality of product transformation across firms refers to the ability of firms to transform resources into final products with similar efficiency.

It is important to emphasize that an economically efficient allocation of resources may not always be socially desirable. This means that even if resources are allocated effectively from the point of view of production efficiency, it does not necessarily mean that the results of this allocation lead to the maximum satisfaction of the needs and wants of society as a whole. This issue is the focus of the theory of social welfare, which tries to find such an arrangement of resources and production that would best meet the needs of society. Social welfare is defined as the level of satisfaction or utility achieved by members of society. This theory seeks to find a balance between efficiency and equity in the allocation of resources to ensure that economic outcomes lead to the general benefit and well-being of society as a whole.

Review questions

1. Explain the difference between general equilibrium analysis and partial equilibrium analysis.
2. Using a box diagram, explain the terms of efficiency in production.
3. Using a box diagram, explain the terms of shift efficiency.
4. Explain production-consumption efficiency and define the conditions of general equilibrium.
5. Explain how the price mechanism can ensure the efficiency of production, exchange, and general equilibrium.
6. Discuss the interrelationship and criteria of efficiency and equity.

Test questions:

1. What is the main objective of general equilibrium analysis?

a) Study of individual markets in isolation
b) Analysis of all markets simultaneously and their mutual interactions
c) Focus only on the labor market

2. What does the production possibilities frontier (PPF) curve show?

a) Total demand for goods
b) An alternative combination of two products that can be efficiently produced with the given resources
c) Money supply in the economy

3. Which of the following is not an allocation rule for achieving production efficiency?

a) Equality of the marginal rate of technical substitution of resources for different products
b) Equality of marginal productivity of different firms
c) Profit maximization of all firms

4. What does it mean when we say that an economically efficient allocation of resources may not always be socially desirable?

a) Efficient allocation always leads to maximum social welfare
b) Efficient allocation in terms of production does not necessarily maximize the satisfaction of society's needs
c) Social welfare is not important in economic analysis

5. What is the main goal of social welfare theory?

a) Profit maximization of companies
b) Striking a balance between efficiency and equity in resource allocation
c) Minimizing government spending

Answers:

1.b, 2.b, 3.c, 4.b, 5.b

Practical task

1. Develop an analysis of the mutual influence of the labor market and the goods market in the context of general equilibrium. Focus on how changes in one market (e.g. raising the minimum wage) might affect the equilibrium in the other market. Discuss the potential effects on prices, employment and overall economic efficiency. Suggest possible measures that could mitigate possible negative impacts and promote the achievement of a new general equilibrium.

2. Create a model that demonstrates the conflict between economic efficiency and social justice in the allocation of resources. Use a specific example, such as the distribution of health care or education. Analyze how different approaches to allocating these resources can affect the overall efficiency of the economy and societal well-being. Suggest a possible compromise solution that would balance the demands of efficiency and fairness and discuss its potential advantages and disadvantages.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie*. 5th updated edition. Praha: Management Press, 2010. s. 467-506. ISBN 978-80-7261-218-5. 574 pp.

[2] MACÁKOVÁ, L.; SOUKUPOVÁ, J. *Mikroekonomie – repetitorium.* Slaný: Melandrium, 1998. 205 pp. ISBN 80-86175-01-4.

[3] HOLMAN, Robert. *Ekonomie*. Praha: C. H. Beck, 1999. 726 pp. ISBN 80-7179-255-1.

[4] HOLMAN, Robert. *Mikroekonomie. Středně pokročilý kurz.* 2nd edition. Praha: C. H. Beck, 2007. 592 pp. ISBN 978-80-7179-862-0.

[5] JUREČKA, V. a kol., 2018. *Mikroekonomie.* 3rd updated edition. Praha: Grada, 360 pp. ISBN978-80-271-0146-7.

[6] SOUKUPOVÁ, J., 2018. *Mikroekonomie.* 6th updated edition. Praha: Management Press, 576 pp. ISBN 978-80-7261-538-4.

Chapter 12

Market Failure and the Macroeconomic Policy of the State

After studying the chapter, you will be able to:

- characterise the main causes of market failure and obstacles to achieving efficiency;
- explain the essence and significance of the state's microeconomic policy and its measures aimed at eliminating the negative effects of market price distortions, the existence of imperfect competition, externalities, public goods and asymmetric information on economic efficiency.

Keywords:

externalities, public good, mixed good, irreducibility, non-excludability, moral hazard, adverse selection, "black passenger", antitrust policy, state regulation, price regulation, internationalization of externalities, Pigou's tax.

After studying the chapter, you will be able to

- Explain the effect of fiscal policy on the market, especially the effects of the sales tax.
- Describe the issue of providing public goods and regulating asymmetric information.
- Discuss the conflict between efficient resource allocation and a company's value system.
- Identify the causes and manifestations of state failure.

Chapter preview

- The chapter deals with the role of the state in the market economy. It discusses the instruments of fiscal policy, especially the sales tax and its effects on the market. He also deals with the issue of public goods and asymmetric information, including methods of their provision and regulation. It also examines the conflict between economic efficiency and social values that can arise during state interventions in the economy. In the conclusion, the causes and manifestations of state failure are described, such as imperfect decisions by state authorities, time delays in the implementation of measures, or the influence of lobby groups. The chapter thus provides a comprehensive view of the role of the state in the economy, its instruments and potential problems associated with its operation.

Chapter objectives

- Understand the effects of fiscal policy on the market.
- Understand the issue of public goods and asymmetric information.
- Be aware of possible conflicts between economic efficiency and company values.
- Get to know the causes of state failure

Estimated study time

- The estimated time required to study the chapter is 180 minutes.

The content of this chapter of the study text immediately follows on from the analysis of the way the price system ensures the efficient use of all production factors and at the same time their optimal allocation between industries and producers in conditions of perfect competition, which we discussed in the previous text. Attention is now given to analyzing the obstacles that prevent the price mechanism from allocating resources efficiently. We refer to this fact as "***market failure***".

It is characteristic of a perfectly functioning market mechanism that it provides unbiased (objective) information about the market situation to both the demand side and the supply side. The bearer of this information is ***the price*** that is formed on the market. In the real economic world, there are a large number of obstacles to perfect competition, or ***market failure.*** The most important of these can be classified into four groups: monopoly power, externalities, public goods, and imperfect information. The mentioned four groups of "obstacles" are characterized by one common feature, which is "***price distortion***", which does not provide market participants with adequate, correct information. As a result, ***market failure and inefficient allocation of resources occur***.

12.1 Imperfect Competition (Monopoly Power)

It describes a situation where economic entities can exercise a certain monopoly power in setting the price. The basic feature of such markets is that ***the price is not the same as the MR, or P > MR***.

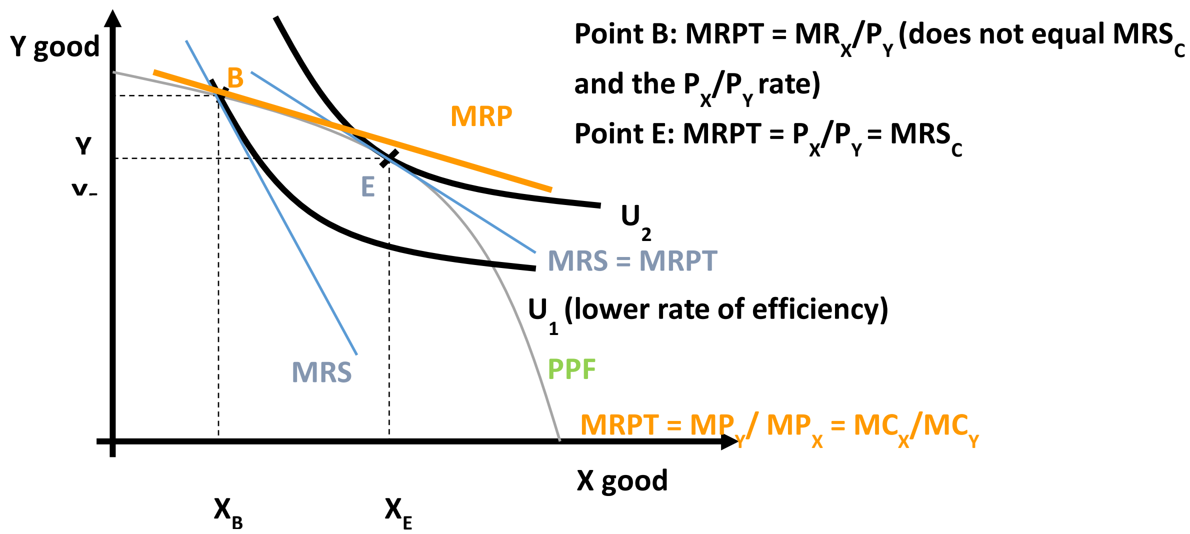

Fig. 112 Monopoly Power and Efficiency

Let us describe the situation using a simple model. The economy consists of only 2 producers producing 2 goods X and Y: X is produced by a monopoly firm and Y by a price-receiving firm. The goal of both producers is profit maximization, i.e. they choose the output at which MR = MC. When producing Y , this means that P = MC; when producing X , this means that P > MR and P > MC. In this case, MRPT is not equal to the share of goods prices, but to the share of MR and P (see Fig. 112).

It is clear from the figure that point E is the point of efficient allocation in the case where both *X* and *Y* are produced in perfect competition: ***MRS = P_X/P_Y = MRPT(X for Y).*** Point B: X is produced under monopoly conditions and Y under perfectly competitive conditions. For the profit-maximizing output, it holds

$$MRPT = \frac{MR_X}{P_Y} < \frac{P_X}{P_Y} = MRS \quad \text{i.e. } MRPT < MRS$$

MRPT* and *MRS inequality is given by the fact that $MR_X < P_X$. *As a result, less product X* (the monopolized product) and more product *Y* are produced than is optimal given existing preferences and technology. Producers respond to a different set of relative prices than consumers.

Monopoly power eliminates the unified criterion for consumer and producer decision-making: firms' decision-making is governed by MR, consumer's decision-making by the price of goods (while MR and P do not coincide). Only the equality P = MC directly links consumer demand to firm decisions; if not met, resources are not deployed efficiently.

12.2 Externalities

An externality occurs when the production or consumption of one entity causes unintended costs or benefits to other entities without those who caused the costs or gained the income paying for them.

Externalities can be of two types:

- ***positive externalities*** – the activity of one entity ***benefits*** another and the latter does not have to cover the costs associated with it, e.g. a guard dog in a residential area;
- ***negative externalities*** – the activity of one entity ***brings costs*** to another entity that are not reimbursed to him, and he does not receive any benefit from them at the same time, e.g. discharge of chemical waste into a fish pond.

Points to remember

An externality arises when the activity of one entity unintentionally affects the costs or benefits of another entity without compensation being provided for them. There are positive externalities where the action of one benefits others without costing them, for example an investment in public transport that makes travel easier for the whole community. Conversely, negative externalities occur when one activity causes costs to another entity without compensation, such as pollutant emissions from a plant that negatively affect the health of nearby residents.

12.2.1 Externalities and Efficiency Conditions

- Negative externalities

If production produces negative externalities, the producer compares price (P) and marginal cost (MC) when choosing the amount of output, taking into account only his private MC, not the additional costs arising as a result of his production to the company, or other subjects. The formation of the price of this product is not based on all the costs associated with its production.

Effectiveness conditions

Total Marginal Cost (Social Marginal Cost – SMC) = Private Marginal Cost of Production (MC) + External Marginal Cost (External Marginal Cost – EMC): ***SMC = MC + EMC***

Total Marginal Rate of Product Transformation (Social Rate of Product Transformation - SMRPT) = rate at which society can transform one good into another. It is determined by the ratio of total marginal costs in the production of Y and X: ***SMRPT =*** SMC_Y/SMC_X

- Positive externalities

The originator of positive externalities does not take into account the additional benefit that his activity brings to another economic entity. As a result, the total marginal utility (Social Marginal Utility - SMU) consists of the private marginal utility (MU), which the given activity brings to its originator, and the external marginal utility (External Marginal Utility - EMU), which is obtained by other subjects: ***SMU = MU + EMU***

The total marginal rate of substitution (Social Marginal Rate of Substitution - SMRS) = the rate at which consumers wish to exchange one good for another. Efficient allocation requires: ***SMRPT = SMRS***

Economic entities make decisions based on the private marginal rate of product transformation and the marginal rate of substitution. When externalities arise, total and private rates differ and the price system is unable to induce efficient allocation.

12.2.2 Negative Externalities and Inefficiency

To explain the problem of the consequences of negative externalities, we will again use a simple model of the economy in which two types of goods X and Y are produced, while the production of X is linked with a negative externality and this means that the SMC_X relation holds $> MC_X$. Production of Y is without externality – **SMCY =** MC_Y.

A necessary condition for social efficiency is the equality of ***SMRPT = SMRS*** . However, the price system coupled with the private motivation of firms leads to an allocation based on equality:

$$MRPT = \frac{P_Y}{P_X} = SMRS$$

so ***SMRPT > SMRS*** because

$$SMRPT = \frac{SMC_Y}{SMC_X} > MRPT = \frac{MC_Y}{MC_X}$$

This means that the rate at which society can transform the production *of X* and *Y* exceeds the rate at which the production of these goods can be transformed privately. Since MRPT and SMRPT do not match, the conditions for optimal resource allocation are not created. The private sector tends to produce too many goods produced with negative externalities because it makes decisions based on underestimated costs.

12.2.3 Social Costs of Inefficiency

The socially efficient amount of output is determined by the equality of total marginal benefit and total marginal cost, i.e. ***SMU = SMC.*** In graphical terms, the total marginal utility can be expressed by a demand curve. The SMC curve *is given* ***by the sum of*** *the MC and EMC curves.*

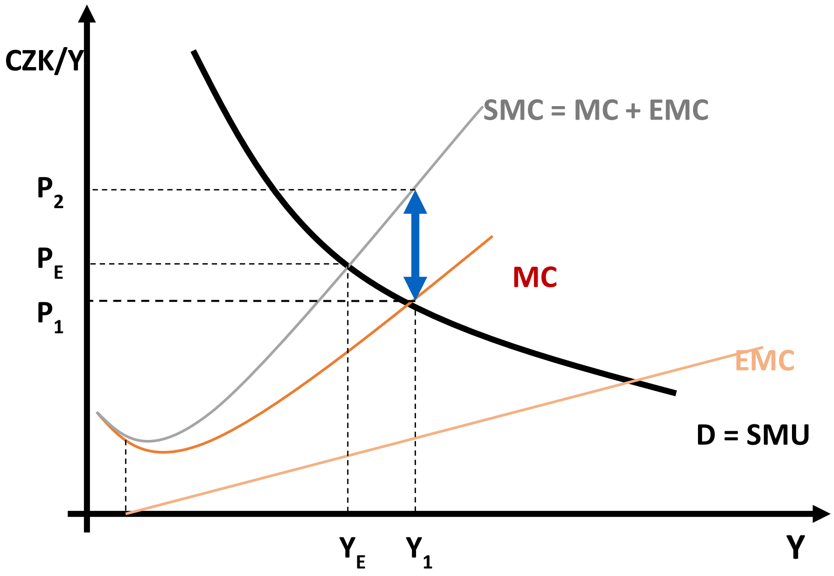

Fig. 113 Social Marginal Utility

The inefficiency is that the price is set based on private marginal cost and not social, so it is set at P_1 at a product higher than the efficient one. Therefore, the price does not cover the external marginal costs that are related to production. The total cost of this inefficiency is represented by the yellow shaded triangle. The source of inefficiency is the inadequate price of production. The market price P_1 is too low to cover total marginal cost. The total marginal cost of production Y_1 would be covered by the price P_2. At the same time, ***P_1 is too high*** for private firms to produce an output of only the amount ***Y_E*** and its height motivates the production of an excess range of output Y_1. For any output greater than the equilibrium output Y_E, the marginal total cost of inefficiency is given by the difference between the total marginal cost and the total marginal benefit.

12.2.4 Positive Externalities and Inefficiency

Positive externalities lead to too low production (see Fig. 114). Curve D expresses the private MU that the given activity brings to its originator. He chooses the Y_I range its activity, corresponding to the intersection of the D and S curves of the MC. Its activity simultaneously produces an external benefit to other entities, see the EMU curve. The total social marginal utility is expressed by the sum of the D and EMU curves = SMU curve.

The socially efficient level of output Y_E at which the SMU of an additional unit of activity coincides with the SMC of this activity, determines the intersection of the SMU and SMC curves.

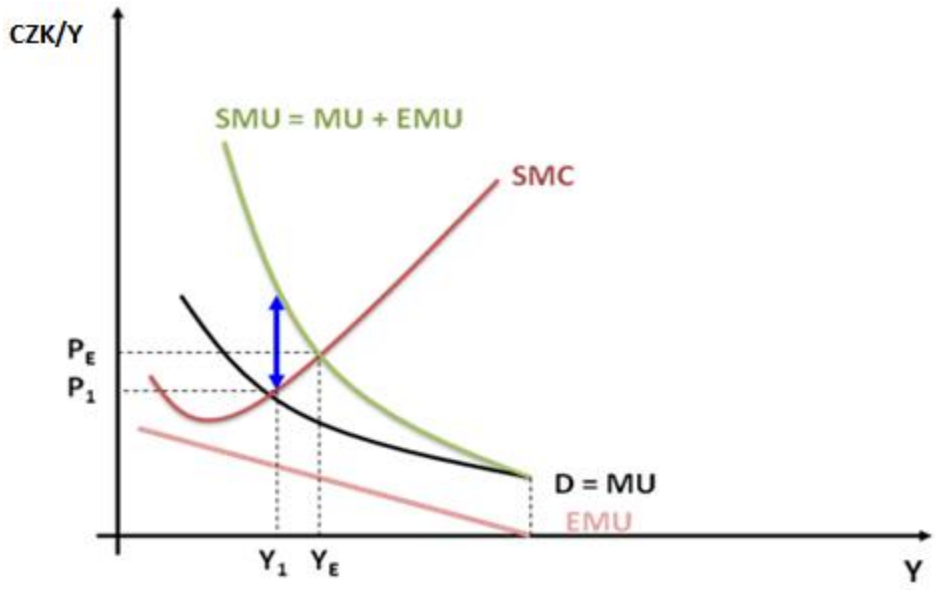

Fig. 114 Social Costs of Inefficiency

Inefficiency arises because the operator of the activity does not obtain all the benefits that flow from its activity. Therefore, the price is P_I too low to cause it to output at the socially desirable level = Y_E (it will only provide output Y_I).

The social costs of inefficiency arising from the existence of positive externalities are represented in the figure by the ocher shaded area, which represents the difference between SMU and SMC for an output level between Y_I and Y_E . For any output less than Y_E is the social cost of inefficiency given by the difference between SMU and SMC. The total social cost of inefficiency is given by the sum of the differences between SMU and SMC for all levels of production that are below the efficient level.

12.3 Public Goods

Public goods generally mean goods or services that are characterized by two characteristics:

- ***Non-rival consumption (irreducibility)*** →whoever consumes this good, his consumption has no bearing on how much of the good others may consume; as a consequence, the MCs of providing the public good to an additional consumer are zero (for most private goods, *the MCs* of production are increasing).

- ***Inevitability*** →it is not possible (or prohibitively expensive) to exclude non-paying consumers from the consumption of the good; the consequence of this is that public goods can be enjoyed without being directly responsible for them was valid.

Indivisibility and non-excludability leads to a tendency to take the position of a *"**black passenger**",* i.e. to avoid paying for the consumption of this good in the certainty that it is not possible to exclude anyone from the consumption of this good. Indivisibility and non-excludability lead to the provision of inaccurate information about the desirability of the good, and thus to the tendency to spend insufficient resources on its production.

Estates can be:

- **indivisible and non-excludable = public goods (Public Goods);**
- **divisible and excludable = private goods (Private Goods);**
- **indivisible but excludable,** or divisible but not excludable = Mixed Goods .

Typical public goods are, for example, national defense and security, a network of highways and roads, but also mandatory vaccination against infectious diseases, etc.

Points to remember

Public goods are characterized by two key properties: non-rivalrous consumption, which means that one individual's consumption does not limit the amount of the good available to others, and non-excludability, where it is not possible to exclude non-paying users from its consumption. These characteristics lead to the "black passenger" phenomenon, where people profit from the good without paying for it, which can cause underinvestment in their production. Examples of public goods include national defense, public transportation, and mandatory vaccinations, while private goods such as food or clothing are divisible and excludable.

12.3.1 Public Goods and Efficiency

The effective level of private goods is determined by the equality ***MU = MC***. The same principle applies to public goods, with some differences:

- While for private goods, *MU is* measured as the utility that the consumer receives, in the case of a public good, it expresses how each entity values an additional unit of output. The optimal quantity of the public good is determined by equality: total MU = MC; the total marginal utility of an additional unit of a public good is given by the sum of the marginal utilities of all its members:

$$SMU_V = MU_{V1} + MU_{V2} + \dots MU_{Vn}$$ (at 1, 2, ...,n – members of the community).

- Everyone can consume the same amount of this good, although individuals assign different values to it. (In private goods markets, everyone can consume different amounts depending on the price of the good and their income.)

The benefit from the increase in the quantity of the public good provided will fall on all members of the community, the benefit from the increase in the production of the private good by one unit will accrue only to the consumer who acquires it. Most public goods are provided by the government, exceptionally by profit-maximizing private companies. However, the question always arises as to what extent (quantity) the given public good should be provided.

12.3.1.1 Optimal Quantity of a Public Good

As with private goods, effective output can be determined based on the equality of MU and MC, with the difference that:

- For a private good, MU is measured as the additional benefit that the consumer receives. In the case of a public good, MU expresses the evaluation of an additional unit of output by individual entities. The optimal quantity of a public good determines the equality of the total marginal utility (given by the sum of the valuation of the good by all entities that use it) with the marginal costs.
- Everyone can consume the same amount of a public good, even though individuals assign different values to it.

a) Market willingness-to-pay curve ***b) Optimal amount of public good***

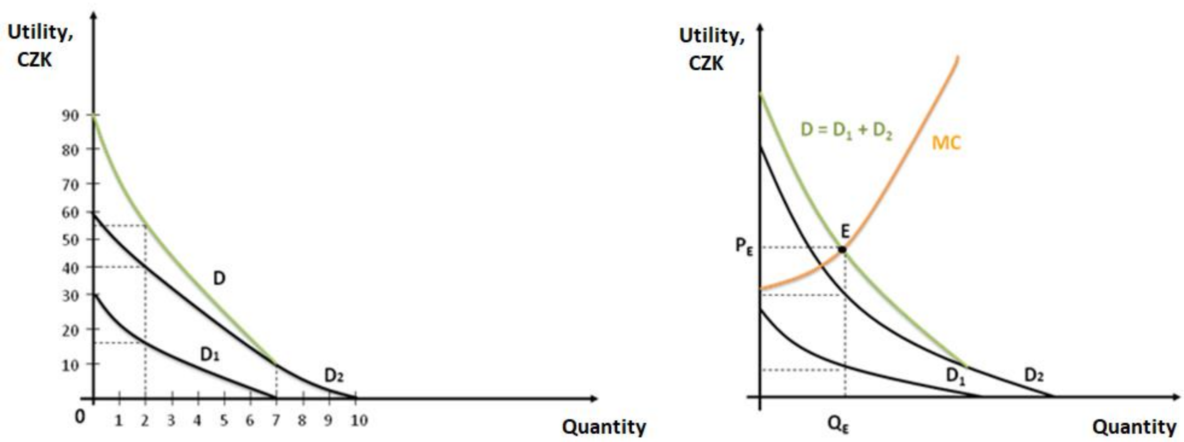

Fig. 115 Market Willingness-to-Pay Curve and the Optimal Quantity of a Public Good

The premise of the graphical model is that a public good is provided by the government to only two citizens, each of whom assigns a different value to it. In Figure 115, the quantity of a public good (output) is expressed on the horizontal axis and its value on the y-axis. Curve D_1 represents the demand of the first, curve D_2 the demand of the second consumer. Each of them expresses the MU that the consumer obtains when consuming each level of output.

Aggregate demand curve (D), the so-called willingness-to-pay curve = vertical sum of curves D_1 and D_2 = the sum of *MU* of both consumers for each level of output. Because it is derived on the basis of the unrealistic assumption of the possibility of finding out the real demand, the market demand curve for the public good (***D***) is considered ***a pseudo demand curve.***

Supply curve of a public good is given by *the MC* of its production. The *MC* curve is horizontal because it is assumed for simplicity that providing an additional unit of a public good does not require an increase in additional costs.

The optimal quantity of a public good is determined by the intersection of the market demand curve with the supply curve, or by the MC curve , i.e. at point E in Fig. 115 b. The sums that both consumers would be willing to pay together for an additional unit of the good exactly cover the costs of providing this additional unit. ***A public good is provided in an efficient quantity only when the marginal benefit equals the marginal cost.***

The claim that $\boldsymbol{Q_E}$ in Fig. 115b represents the optimal level of the public good, valid only if the total cost does not exceed the amount the public would be willing to pay for that quantity. It is a similar condition as in the case of a firm producing a private good to the extent corresponding to MR = MC, provided TR covers TC (or VC in the short run). If the government is to ensure the production of the QE amount of the public good, it must have large enough revenues to cover the total cost of producing that amount.

12.4 Asymmetric Information

Asymmetric information represents situations where one side of the market knows more than the other. It arises as a result of classified activity or classified information:

- ***covert activities*** are those that cannot be observed accurately and without significant additional costs by other entities;
- ***insider information*** corresponds to situations where one side of the market has more expertise than the other.

Two subproblems are related to the existence of asymmetric information: ***moral hazard*** and ***adverse selection.***

12.4.1 Moral Hazard

Moral hazard is defined as the activity of one (informed) economic subject who, while maximizing his benefit, ***reduces the benefit of other*** (uninformed) participants in a market transaction. As a rule, he does not bear the full consequences for this activity, because this effect of his is not perfectly observable and verifiable (example: property insurance - the insured person probably pays less attention to the insured matter, because he knows that he will be compensated for any loss. This creates a higher probability of an event against which is the client insured). Moral hazard arises when the object of insurance is insured for a higher price than the cost of restoration. A moral hazard is therefore a systemic error enabling profit to be made by acting at the expense of other subjects to whom the risk was unilaterally transferred.

The most common example ***of moral hazard*** are situations where a principal-agent ***relationship arises between economic entities***. These are situations where someone hires another person to perform a task that affects their well-being. The person who executes the lease = ***principal (P)*** and the one who is hired = ***agent (A)*** →acts for or on behalf of the principal.

Examples: in a company: the manager is the agent and the owner of the company (shareholder) is the principal; in healthcare, the doctor is the agent and the patient is the principal. In this sense, even ***the government,*** which acts on behalf of the population, acts as an agent.

The following applies to the relationship between the principal and the agent:

- the principal usually delegates a certain part of the decision-making powers to the agent;
- the agent performs certain tasks or decides certain matters for the principal and shares the consequences of this action with him. Therefore, there is a mutual connection between the benefit of the agent and the principal.
- decision-making takes place in a context of uncertainty. The exact result of the agent's activity cannot be determined in advance.
- there is asymmetric information and the problem of moral hazard associated with it.

Importantly, both the agent and the principal are motivated by self-interest, both wanting to maximize their utility. There is no reason to believe that their interests coincide. The agent tends to maximize his own utility rather than the principal's utility.

The agent can be motivated to act as much as possible in the interest of the principal in basically two ways:

- the inclusion of ***the share ownership rule***, e.g. in the method of dividing the income between the agent and the principal;
- sometimes it is possible to compel *the agent* to act for *the principal* by observing his activity, but this requires additional costs.

Agent incentive and reward system

The remuneration system is a means by which the principal tries to interest the agent in acting in his interest. In determining the optimal compensation system, the relationship of the principal and the agent to risk plays an important role. If we do not assume the exceptional possibility of risk preference, then the following alternatives are possible:

- both are risk averse;
- both have a neutral relationship to risk;
- one is risk averse and the other is indifferent.

Based on the risk approach assumed above, several possible reward systems come into consideration, all involving the sharing of revenue between the principal and the agent.

a) **The agent assumes all the risk** – the principal receives a fixed sum paid from the revenue achieved and the agent receives the remaining part of the revenue. Since the return depends on random factors as well as on the agent's efforts, the agent assumes all the risk associated with random factors (the state of the economy, the behavior of competitors, etc.). This contract *is not optimal in terms of risk sharing* .

b) **The principal bears all the risk** – the agent receives a fixed reward because he rejects the risk and the principal bears all the risk. This system is optimal in terms of risk sharing, but *not optimal in terms of motives* . An agent who is guaranteed a fixed reward will have no incentive to increase effort, which reduces his utility. Subsequently, there will be a higher probability that the principal will get a lower yield.

If the agent is protected from risk by a fixed reward system, then he is likely to behave as if he were indifferent to risk in his decision-making.

There is a problematic relationship between optimal risk sharing and motives. This relationship is referred to as ***the fundamental stress***. The following reward systems propose a solution to the problem of motivation:

- the remuneration system reflects the agent's efforts – it is based on the requirement that the risk-averse agent is not guaranteed a fixed income;
- the so-called coercive contract or linear contract - a similar idea linking the agent's reward to his effort.

Definition

Moral hazard is a situation where an informed economic entity maximizes its utility at the expense of uninformed market participants, while not bearing the full consequences of its actions, which are difficult to observe. It often arises in principal-agent relationships, where the agent performing tasks for the principal may tend to act for his own benefit rather than the benefit of the principal. In practice, moral hazard manifests itself, for example, in insurance, where the insured person may take less care of his property because he believes that any losses will be compensated by the insurance company.

12.4.2 Adverse Selection

Adverse selection is a process that results in "less desirable" market subjects (buyers or sellers) participating in voluntary exchange rather than others.

Asymmetric information leads to the displacement of higher-quality goods from the market by lower-quality goods. Among other things, this is because in the real world (as opposed to perfect competition) consumers cannot easily discern the quality of goods until they have purchased and used them for some time (existence of asymmetric information). If the seller has better information about the product than the buyer, a "Black Jack" market develops in which low-quality goods displace high-quality goods.

A classic example that vividly explains the nature of "adverse selection" is the development of the used car market. Dealers in car dealerships are better informed about the quality of each car than buyers. Buyers are of course aware that they are taking on the risk of buying "Black Jack" - a low-quality product that the seller is trying to get rid of. Buyers do not know which car is "Black Jack", but they know the probability of getting it and therefore the average quality of the cars. They will therefore be prepared to pay only for average quality, so some cars will be underpriced (cars of above average quality) and others overpriced (cars in poor condition).

The consequence of such a situation will be that those who sell better quality cars will leave the market. The average quality of the cars that remain on the market will decline, and buyers will be willing to pay less for the cars that remain on the market. This will lead to further crowding out of better quality cars, a further drop in average quality and a subsequent drop in prices, etc. The market will narrow to products of ever lower quality. Figure 116 illustrates this development.

a) High quality car market ***b) low quality car market***

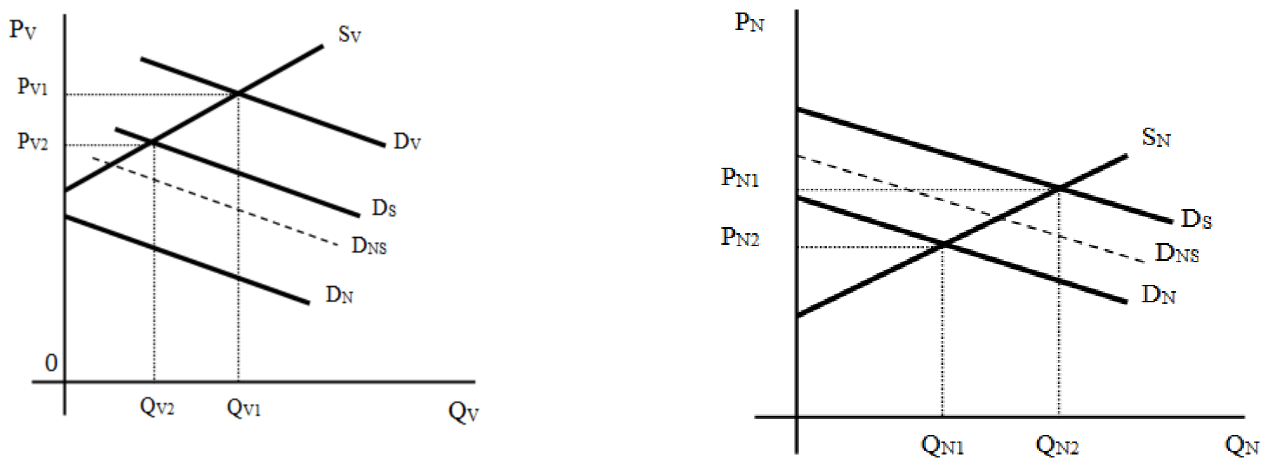

Fig. 116 Adverse Selection

The high quality car market (Fig. 116a) is initially characterized by the range of S_V and demands D_V; the market for low quality cars (Fig. 116b) is initially characterized by the supply S_N and demand D_N.

A low-quality good is displaced by a high-quality good →a shift in the demand curve from level D_V on D_S in Fig. 116ab and as a result of the deterioration of the perception of the average quality of cars on the market + a shift of the demand curve for low-quality cars from D_N on D_S in Fig. 5b. Sales of high quality cars will decline from Q_{Vl} on Q_{V2} and sales of low-quality cars will increase from Q_{N1} on Q_{N2} . Fewer and more low-quality cars will be sold. When consumers discover this circumstance, their demand will change, →the demand curve will shift to level D_{NS} , which reflects the fact that the quality of cars sold is on average between medium and low.

The shift would continue until only inferior cars remained on the market. In that case, the price was too low to attract better quality cars to the market, so buyers correctly assumed that any car they bought would be of poor quality. In that case, the demand would be represented by the curve D_N.

Signaling behavior

The negative consequences of "adverse selection" can be eliminated by means of so-called signaling behavior, i.e. through signals (observable activities) the informed party transmits credible information to the uninformed party.

An important prerequisite for effective signaling is that, in order to ensure it, the signaling costs associated with the sale of high-quality goods must be lower than the costs associated with the sale of lower-quality goods.

Examples of signals:

- setting **a high price**;
- provision of **longer manufacturing guarantees**;
- on the labor market, **the education achieved** or **the completion of professional courses, etc.**

An example of a generally accepted signal in the above sense is education on the labor market. Education directly or indirectly increases a person's productivity because it provides him with information, skill and general knowledge that are useful in work. Education is a signal of higher productivity also in the sense that more productive people more easily achieve a higher level of education - they are usually more intelligent, more energetic and more interested in the work they do = characteristics that are also beneficial in school. More productive people are therefore more likely to attain higher levels of education in an attempt to signal their productivity to firms in order to obtain better paying jobs.

12.5 Microeconomic Policy of the State

In the economy, in addition to the economic subjects of households and companies, the state also figures not only on the supply and demand side, but can also influence the market through its economic policy. In general terms, economic policy is associated with **the macroeconomic activity of the state**, i.e. with the action of the state in such areas of economic policy as, for example, fiscal or monetary policy, and with solving problems such as high inflation or high unemployment.

However, with its measures, the state also affects the behavior **of microeconomic subjects** – consumers and companies, by changing the conditions under which decisions are made by these subjects and thus affects the balance in sub-markets and the general market balance.

12.5.1 Causes and Tools of the Microeconomic Policy of the State

There are three basic causes, or groups of initiatives for state regulation:

1. Inefficient allocation of production factors and final output due to imperfect competition, externalities, public goods and imperfect information;
2. Social unacceptability of the distribution of pensions, the effort to change the distribution of pensions;
3. Conflict between efficiency and the value system of society.

Microeconomic policy tools:

I. Determining the rules of conduct of individual market entities:

- defining ownership rights in the company;
- anti-monopoly legislation or state regulation of prices, quality of goods, etc.

II. Fiscal policy of the state: revenues and expenditures of the state budget (in conjunction with taxes) change the distribution of income in society against the situation created by the market and affect the process of creating equilibrium.

III. The state as a market entity in the markets of final production and in the markets of production factors can influence the situation existing in these markets.

12.5.2 Rules of Conduct of Market Entities

Rules of conduct of market entities, or The "rules of the game" of individual market participants form the so-called ***institutional framework*** of the economy. The rules develop evolutionarily and their formulation and enforceability is taken over by the state. The basic rules should apply to all market entities, and therefore the rules promulgated by the state must apply primarily to the state. An important part of the institutional framework is ***the definition of property rights,*** which the state should precisely define and ensure their effective protection.

12.5.2.1 Definition of Property Rights for a Market Economy

From the point of view of property rights, we can encounter two basic cases = factors of production (F) are free and available to everyone or they are privately owned.

1. **The factor of production is free** (every interested party can use the factor)

As an example, let's use fishing on the high seas, where marine resources can be used by anyone interested. Assume a perfectly competitive market, decreasing MR and decreasing AR (the causes are not on the demand side, but on the supply side) → actually the marginal and average "product" of the free production factor decreases (each boat catches less fish than the previous boat, and each new boat reduces the catch of all previous ships). However, this situation is different from the situation in the general case of the development of MP and AP in the market of production factors.

Since the factor of production is a free good, TC = 0 (AC = 0, MC = 0). Since operating costs are also zero (or we do not take them into account), the number of producers will increase as long as they make a positive profit, or Until AR = 0 (since AR = AC, so also until AC = 0). Under this condition, a normal profit will be achieved and it is also determined that L_Y will be fully utilized production factors.

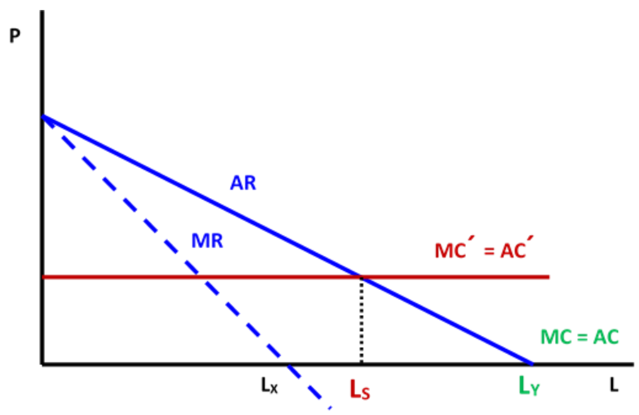

Fig. 117 Rate of Utilisation of Free Resources

2. **The factor of production is privately owned**

If the factor of production is privately owned, then TC > 0 and assuming constant returns to scale, the graphical representation of MC = AC will be a line parallel to the X-axis (in the recommended figure 117. The number of producers will increase until AR = AC', so using L_S production factors.

Under conditions of perfect competition and free factors of production, more activities are achieved than under conditions of private ownership (where the use of a factor of production must be paid for). The task of the government is therefore to strive for the clearest possible distribution of scarce production resources to individual owners and to prevent their waste.

However, it is a fact that even in market economies ***there are free production resources*** . In many cases, this is because ***the costs associated with collecting revenue from their ownership*** can ***be higher than the revenue thus obtained*** (eg collecting tolls from city streets, as a tollbooth would have to be at every intersection). In addition, ***there may not be an owner of the factor of production*** (fishing on the high seas), or ***the owner may not be interested in collecting income*** from the factor of production (e.g. it may be decided that driving on highways will be free).

Chicago economist RH Coase demonstrated that an economy will arrive at an efficient allocation of resources if property rights are well defined and effectively enforceable, i.e. if owners can resolve disputes among themselves with very low bargaining costs.

Coase's theorem: regardless of the initial distribution of property rights, the final outcome in a market economy will be efficient, provided that the initial legal distribution of property is well defined, and transactions involving trade in property rights have zero costs.

According to Coase's theorem, the possibility of agreement exists, but the existence of the possibility does not imply the automatic realization of agreements:

- one of the subjects may expect a better result from a conflict than from an agreement;
- the necessary prerequisites may not be met (e.g. high prices for legal services or ill-defined property rights).

The state must then intervene and help to resolve any disputes.

12.5.3 State Regulation

The efficiency of a market-driven economy decreases if there is imperfect competition, externalities occur, it may fail in the allocation of public goods and in unequal access of market subjects to information. Economic policy can thus assist the market in the efficient allocation of production factors and final production. The state's microeconomic measures are aimed at weakening the consequences of the four forms of market failure.

State regulation is usually used ***against firms in a situation approaching a monopoly*** . In the case of regulation, the company remains privately owned, but the state regulates its prices, the level of services provided, safety standards, etc.

Price regulation by the government is motivated by an effort to exclude the element of monopoly profit from the price, to reduce the monopolistically high price, and to increase the quantity of goods produced.

- Price regulation at MC level ***(P = MC)*** – can lead (very likely in case of monopoly) to ***AC > AR = loss*** of the firm **→** ***its exit from the industry*** or ***pressure on the state*** **→** the state has to compensate for the loss (by increasing the price or subsidies).
- Price regulation at the AC level ***(P = AC)*** **–** then ***AC = AR*** **→** the firm achieves ***only*** normal profit.

Price regulation by the state slows down companies' response to market signals → it can therefore contribute to inefficient allocation of production resources and final production.

12.5.4 Anti-monopoly Legislation

Anti-monopoly legislation is intended to control the growth of monopolies and to protect competitors from undesirable practices of monopolistic firms, which enable the reduction of the number of competitors (lower prices and in certain areas, discounts for larger purchases, mergers of companies with the same program, discounts for privileged customers).

The problem of its effectiveness is mainly linked to the factor that these measures are mostly implemented on the basis of court decisions → an entity must be found to invoke the judicial process.

12.5.5 Fiscal Policy

Fiscal policy affects the market situation mainly through sales tax, subsidies and fines.

Sales tax

Sales tax is included in the price of goods and is paid to the state budget by sellers of goods. Sales tax affects the behavior of economic entities by:

- affecting market equilibrium;
- affecting individual subjects unevenly;
- reducing the allocative efficiency of the market.

A sales tax usually lowers the equilibrium quantity (Q_E) and increases it price (P). The situation in the market where the tax is included in the price = it is paid by the sellers of the goods is shown in the picture. The second picture shows the situation where the tax would be paid by consumers. In both cases, less goods would be sold, but in the first case at a higher price.

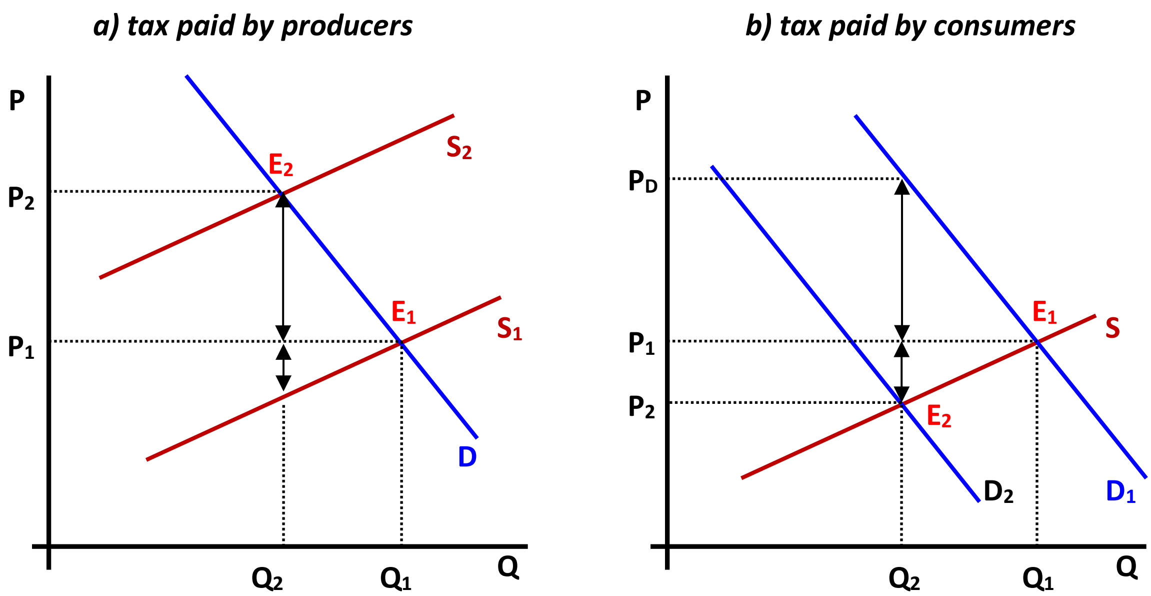

Fig. 118 Impact of Sales Tax on the Submarket

Exception: if supply or demand is ***completely inelastic (demand or supply curve is vertical), then only price*** changes , not quantity.

Uneven impact of sales tax

Note: for explanation, use the image "the impact of the sales tax at different slopes of the supply and demand curves"

The shortcoming of sales tax is the fact that its effect does not have to be limited only to the entity on which it is imposed. The entity that pays the tax usually tries to shift the tax burden to other market entities. The move can go towards:

- ***forward*** (ie to a downstream producer or consumer);
- ***backwards*** (i.e. to the supplying company or the owner of the production factor).

The effect (impact) of the tax expresses the allocation of the tax burden between individual entities. The tax burden is spread between producers and consumers – to some extent by producers and to some extent by consumers.

Graphically, different situations of the impact of the tax burden occur due to the different slope of the supply and demand curves. The impact of the tax on the volume of production is ***the smallest*** if the demand or supply curve has ***a large slope***, i.e. when demand or supply is ***inelastic*** → ***the tax is reflected in the amount of the price and its effect on the quantity offered, or the quantity demanded is negligible.***

Sales tax and market allocative inefficiency

Taxes usually reduce the efficiency of the allocation of production factors as well as final production. In most cases, the tax burden is greater than the amount paid in tax → additional tax impact = ***additional tax burden.***

Total tax burden = sum of own tax revenue and additional tax burden. The graphical representation of the additional tax burden is ***the deadweight cost*** of the company, which = the additional tax burden borne by the company as a result of the introduction of the sales tax = area ***Z*** in the graph used in Fig. 119.

The consumer's surplus does not remain entirely with the consumer - part of it is received by the state in the form of a part of the tax revenue. In the same way, the producer does not retain the entire market value of the production - part of it is received by the state in the form of the second part of the tax revenue.

State subsidies and fees

State subsidies are a tool to eliminate ***negative externalities,*** as well as fines for their production. The choice between subsidies and fees (penalties) depends on their impact on the volume of externalities and the after-tax deadweight costs.

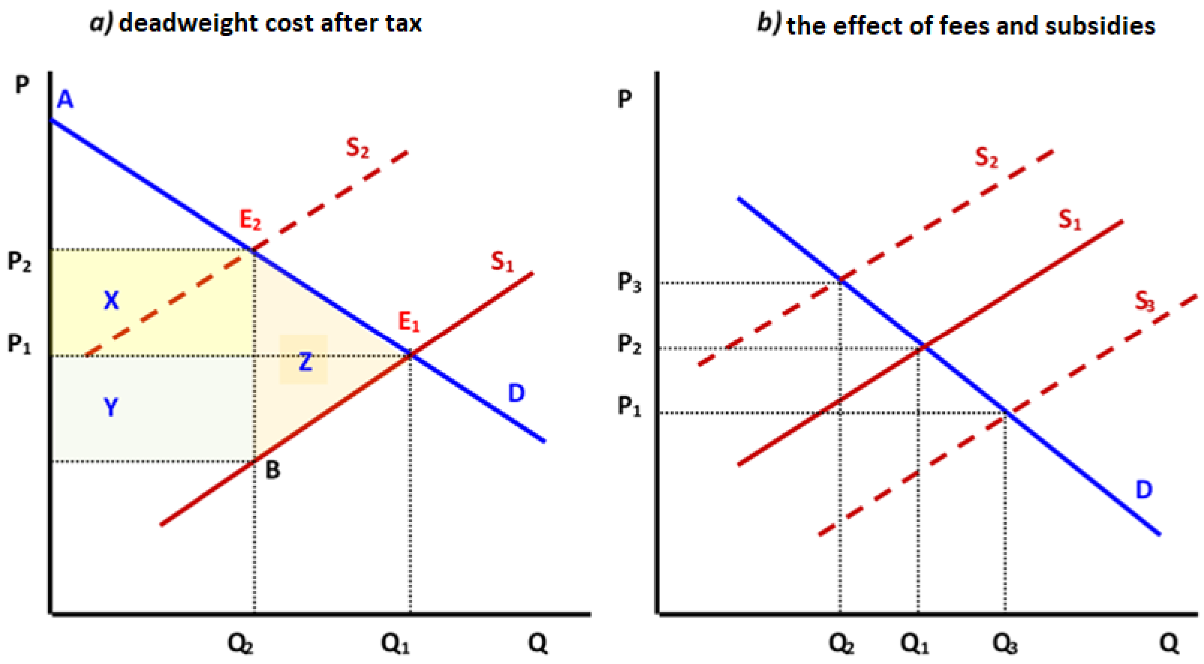

Fig. 119 Effect of Taxes, Subsidies and Fees on Market Equilibrium

The fee increases the firm's costs and thus leads to a decrease in supply → the supply curve shifts = the volume of externalities is reduced, but the price increases and the volume of production offered decreases.

Subsidizing part of the costs of removing externalities increases the firm's supply → shift of the supply curve = increase in the quantity offered and decrease in price + the risk that a higher volume of production will ultimately produce a higher volume of negative externalities.

12.5.6 Providing Public Goods and Regulating Asymmetric Information

The following problems are connected with the existence and provision ***of public goods***:

- determination of their optimal volume;
- obtaining sufficient funds for their production (exclusion of the "black passenger") - taxation of all potential consumers ⟶ the necessary amount of money is collected;
- choosing an effective way (mechanism) of their provision ⟶ (a) they will be provided by the state (defense, security...), (b) private companies.

Information (or at least a certain part of it) is very close to public goods in its features**:**

- It is very costly to obtain information, but once certain information is available, the MC for providing it to other interested parties is ***zero,*** i.e. the information is characterized ***by irreducibility.***
- It is difficult and often even impossible to transfer certain information only to those who are willing to pay for it, i.e. information is characterized ***by non-excludability.***

But as soon as information acquires the character of a public good, ***the market will tend to produce it in insufficient quantities, or not to offer it at all***.

State regulation of the information market:

- by requiring the production of goods in accordance with health or safety regulations;
- requiring sellers to disclose key information about the products offered;
- ensure public availability of necessary information (state agencies);
- to subsidize private entities engaged in providing information to the public.

12.5.7 Conflict between Efficient Resource Allocation and Value System

Efficient allocation of resources may not always be the goal pursued by the company. The market can, for example, offer goods that are harmful to consumers. The state is trying to strengthen the position of consumers (e.g. by limiting certain types of advertising), or to correct the preferences of some consumers (drugs). The justifications for these state measures are mostly ethical or moral, not economic. Therefore, ***the efficient allocation of resources can come into conflict with other goals of society*** (e.g. a classic example is prohibition, which leads to the reduction of highly efficient production of alcoholic beverages).

Redistribution of income

Society (represented by the state) may not find the distribution of income as created in markets to be **socially acceptable.** The state therefore carries out **redistribution processes.**

The default situation can be changed:

a) ***by changing the structure of ownership, i.e. redistribution of property*** – for example, there may be a redistribution of ownership of asset X in favor of consumer A or asset Y in favor of consumer A;
b) **redistribution of income** – state authorities do not change the ownership structure of production factors, but drain part of the income from one consumer and provide it to other consumers.

Redistribution changes the initial distribution into **a final distribution**.

12.6 State Failure

State failure has causes in two areas (the result is always the same – a reduction in economic efficiency, a negative influence on the distribution of income, etc.).

The first area is ***the imperfect decision-making of state authorities*** - the decisions of state officials can be wrong. The reason for these failures is usually time delays. ***Information lag*** is the difference between the moment when the need for state intervention arose and the moment when the state becomes aware of it and intervenes. It is determined both by the time during which the problem is processed by the state apparatus and by the time of the decision-making process (the so-called ***political time lag***). It then takes a certain amount of time for the effects to become apparent from the moment a certain measure is taken – ***implementation lag*** .

In addition, the state may be subject to lobbying groups - lobbying pressure may result in the phenomenon of so-called **"*rent seeking*"** - the unproductive expenditure of resources by a market entity that ***brings it an economic profit higher than the profit would be in conditions of perfect competition*** (attempts to appropriate a monopoly through state authorities profit) – an administrative decision that makes it e.g. the sole supplier in the public sector, the imposition of import quotas, a high tariff – all companies strive for monopoly profit.

Points to remember

State failure results from imperfect decision-making by state authorities and time delays that affect the effectiveness of state interventions. Among the main causes are the information delay, when the state reacts to the need for intervention only after the time when the problem arose, and the implementation delay, when it takes time for the effects of the measures taken to be felt. Moreover, *the state may be subject to lobbying pressures, leading to unproductive rent-seeking, where market actors seek to obtain monopoly profits through administrative decisions.*

In the real economic world, markets face many obstacles that disrupt perfect competition and lead to market failure. These barriers can be systematically divided into four main groups: monopoly power, externalities, public goods, and imperfect information. Each of these groups presents specific challenges that affect the efficiency of market functioning and the allocation of resources.

Monopoly power occurs when individuals or firms have the ability to set prices, which distorts the competitive environment and leads to situations where prices do not reflect the true costs or values of goods and services. Externalities occur when the actions of one person or firm affect others without this interaction being reflected in prices. For example, environmental pollution is a negative externality that can cause the cost of production to be lower than it actually is for society as a whole.

Public goods such as national defense or clean air have characteristics that make them difficult to provide on the basis of market competition. Since no one can be excluded from their use, there is no incentive for individuals to pay for them, which can lead to an undersupply of these goods. Imperfect information that market entities do not have, or only partially have, also impairs effective decision-making. For example, consumers may not have access to information about product quality or price, which may lead to suboptimal decisions.

These four groups of barriers are characterized by a common feature, which is "price distortion". This price distortion prevents market participants from correctly evaluating and receiving information that would allow them to effectively respond to supply and demand. As a result, market failure occurs, which manifests itself in the inefficient allocation of resources and the loss of the overall welfare of society.

Microeconomic activity of the state, or of the government, is an integral part of economic policy and its main goal is to strengthen tendencies towards the effective functioning of the market mechanism. The existence of market failures raises the need for an active state policy aimed at eliminating the negative effects of these failures on efficiency. This policy may include regulating monopolies, ensuring a fairer allocation of public goods, implementing measures to internalize externalities, and improving the information environment for market entities.

The aim of these interventions is not only to improve the functioning of markets, but also to ensure that economic activities benefit the wider society. A properly designed economic policy can lead to greater efficiency and equity, thereby contributing to the achievement of greater social well-being.

?

Review questions

1. Explain the nature of market failure and its causes.
2. Discuss externalities (both positive and negative) as one cause of market failure and explain their effect on inefficiency.
3. Explain public goods as a cause of market failure and clarify the optimal choice of the extent of the public good to be provided.
4. Discuss the problem of asymmetric information, moral hazard, and adverse selection.
5. Explain the reasons for the existence and instruments of microeconomic policy of the state.
6. Discuss the importance of the precise definition of property rights for the efficient functioning of the market.
7. Explain the objectives of state regulations, their possibilities and negatives.
8. Explain the content of fiscal measures in the state's microeconomic policy.

?

Test questions:

1. Which of the following is not considered a major group of barriers to perfect competition?

a) Monopoly power
b) Externalities
c) Inflation

2. What is a characteristic of public goods?

a) High price
b) The impossibility of excluding someone from their use
c) Easy production by the private sector

3. What does "price distortion" mean in the context of market failures?

a) The prices are always too high
b) Prices do not reflect the actual costs or values of goods and services
c) Prices are set by the government

4. What is the main goal of microeconomic activity of the state?

a) Increase tax revenue
b) To strengthen the tendencies towards the effective functioning of the market mechanism
c) Increase inflation

5. What is one consequence of market failures?

a) Efficient allocation of resources
b) Profit maximization of all firms
c) Loss of the overall welfare of society

Answers:

1.c, 2.b, 3.b, 4.b, 5.c

Practical task:

1. Analyze a specific case of an externality in your neighborhood (e.g. air pollution by a local factory). Describe how this externality affects different groups in society and what its economic effects are. Suggest possible solutions that could help internalize this externality and discuss the potential advantages and disadvantages of each proposal. Consider the role of the state in solving this problem and propose specific policy measures.

2. Choose one public good (eg street lighting, national defense) and analyze its provision in your city or country. Identify the problems associated with its financing and distribution. Suggest an innovative way that this public good could be provided more efficiently using market mechanisms or new technologies. Discuss the potential advantages and disadvantages of your proposal, including its effects on different groups in society and overall economic efficiency.

Literature:

[1] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie*. 5th updated edition. Praha: Management Press, 2010. s. 507-533. ISBN 978-80-7261-218-5. 574 s.

[2] FUCHS, K., TULEJA, P. *Základy ekonomie*. 2. rozš. vydání. Praha: Ekopress 2005.

[3] MACÁKOVÁ, Libuše; SOUKUPOVÁ, Jana. *Mikroekonomie – repetitorium*. Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[4] HOLMAN, Robert. *Ekonomie*. Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[5] HOLMAN, Robert. *Mikroekonomie.* Středně pokročilý kurz. 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[6] JUREČKA, V. a kol., 2018. Mikroekonomie. 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[7] SOUKUPOVÁ, J., 2018. Mikroekonomie. 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

List of Literature and Sources Used

[1] FUCHS, K., TULEJA, P. *Základy ekonomie.* 2. rozš. vydání. Praha: Ekopress 2005.

[2] MACÁKOVÁ, Libuše; SOUKUPOVÁ, Jana*. Mikroekonomie – repetitorium.* Slaný: Melandrium, 1998. 205 s. ISBN 80-86175-01-4.

[3] HOLMAN, Robert. *Ekonomie.* Praha: C. H. Beck, 1999. 726 s. ISBN 80-7179-255-1.

[4] HOLMAN, Robert. *Mikroekonomie. Středně pokročilý kurz.* 2. vydání. Praha: C. H. Beck, 2007. 592 s. ISBN 978-80-7179-862-0.

[5] JUREČKA, V. a kol., 2018. *Mikroekonomie*. 3. akt. vyd. Praha: Grada, 360 s. ISBN978-80-271-0146-7.

[6] JUREČKA, V. a kol. *Mikroekonomie.* Praha: Grada Publishing, 2010. 360 stran. ISBN 978-80-247-3259-6

[7] SAMUELSON, P. A. a W. D. NORDHAUS. *Ekonomie.* 18. vydání. Praha: NS Svoboda, 2007. ISBN 978-80-205-0590-3

[8] SOUKUPOVÁ, J., 2018. *Mikroekonomie.* 6. akt. vyd. Praha: Management Press, 576 s. ISBN 978-80-7261-538-4.

[9] SOUKUPOVÁ, J., HOŘEJŠÍ, B., MACÁKOVÁ, L., SOUKUP, J. *Mikroekonomie.* 5. aktualizované vydání. Praha: Management Press, 2010. s. 507-533. ISBN 978-80-7261-218-5. 574 s.

List of Figures

List of Tables