Romanian Statistical Review
Revista Română de Statistică

THE JOURNAL OF NATIONAL INSTITUTE OF STATISTICS

4 /2022

www.revistadestatistica.ro

CONTENTS 4/2022

Statistical Coordinates Of Youth's Living Standards. A Comparative Analysis in EU-27

Ciprian IFTIMOAEI[1]
(Ciprian.Iftimoaei@iasi.insse.ro)
National Institute of Statistics, "Al. I. Cuza" University of Iasi

Ionuț-Cristian BACIU
(Ionut.Baciu@iasi.insse.ro)
Iasi County Directorate of Statistics

Vicențiu-Robert GABOR
(Vicentiu.Gabor@iasi.insse.ro)
"Al. I. Cuza" University of Iasi, Iasi County Directorate of Statistics

ABSTRACT

After a health crisis that kept them at home, physically distant from friends, colleagues, and teachers, young people gathered in classrooms, offices, and their favourite places to spend their free time. As declared by the authorities, the end of the pandemic coincided with the start of a geopolitical and security crisis: the war started by Putin's Russia that impacted the quality of life of European citizens. European economies face the restriction of gas supply from Russia, the scourge of ever-increasing energy prices, and galloping inflation. The fear of war replaced the fear of disease during the COVID-19 pandemic, followed by the fear of poverty. The year 2022 is declared by the European Union as the "European year of youth". In this respect, the European Commission proposes as a priority objective the inclusion of young people and their problems in the development of future policies, as well as the organization of activities dedicated to young people throughout the EU. From the perspective of the European Commission, the most pressing problem related to young people is the transition from school to work. Young people are the most affected by economic inactivity, underemployment, unemployment, and poverty. Many European countries, including Romania, have problems with the lack of labour force, with young people leaving school early, with young people who do not study or follow any form of professional training. The completion of studies by young people is only sometimes followed by the search for a job to engage and carry out an activity in one of the economic sectors. Postponing youth employment also leads to leaving the parental home late, which represents a form of social self-protection. To analyze the aspects mentioned above, which define the standard of living of young people, we use descriptive statistical analysis, Bravais-

1. Corresponding author

Pearson correlation, simple regression and ascending hierarchical classification. The data used in this study come from the Eurostat database.

Keywords*: youth, young people, living standard, quality of life, poverty, unemployment, education*

JEL Classification*: H520, I310, J130*

1. INTRODUCTION

The year 2022 has been declared by the European Union as the *European Year of Youth*, with the European Commission proposing as a priority objective "to include young people and their priorities in future policy-making and to organise youth activities throughout the European Union (EU)". According to Eurostat, one in six Europeans is aged 15-29. At EU level, 2005 was the year when the share of the young population under 15 (16.3%) overtook the share of the elderly population aged 65 and over (16.6%) in the total population.

Europe's population is undergoing a process of demographic ageing. In Romania, 2009 was the year of change in the ratio of young to old people as a share of the total population. Demographic ageing is a firm and long-lasting process; in Romania, it is determined by the falling birth rate, external migration and the increase in life expectancy after 1990. According to demographic projections by the National Institute of Statistics (2022), the young population aged 0-14 years will decrease significantly between 2021-2100 by about 1.2 million people.

In recent decades, young Europeans' values, concepts and lifestyles have changed in the context of economic, social and cultural developments. The age at first marriage has risen continuously. Young people decide to marry after graduation and enter the labour market in order to have the financial resources needed to make a decent living as a couple. More and more young people prefer common-law unions or civil partnerships to marriage. The increase in the average age at first marriage also increases the average age at which mothers decide to have a child. More and more children are born out of wedlock, the decision to have a child no longer being dependent on married status. In Romania over the last decade, the number of live births to mothers with higher levels of education has increased and the proportion of those whose mothers have low levels of education (secondary school at most) has decreased. This phenomenon is also linked to the increase in the share of women with higher education.

Sociological research at European level reveals that one in three young people faced difficulties during the COVID-19 health crisis, 40% of young Europeans feared for their health, youth unemployment in the EU increased from 11.9% in 2019 to 13.3% in 2020. Eurostat shows that over the

period 2015-2019, the rate of severe material deprivation for young people aged 15-29 in the EU decreased from 8.4% to 5.4%, but increased to 6.5% in the pandemic year 2020.

Even though they have gone through periods of social and economic problems, young people in the EU remain attached to the European idea, with 75% of them holding a positive view of the European Union and democracy. From the European Commission's perspective, the most pressing issue for young people is the transition from school/education to work. Young people are most affected by economic inactivity, unemployment, poverty. They are increasingly falling prey to drug and alcohol abuse, prostitution, human trafficking, anxiety and depression.

Young people are affected by unemployment to a greater extent than adults because there is not always a link between their education and job vacancies or because they lack the work experience required by employers. Young people are often employed on part-time, fixed-term contracts and are easily dismissed when the economy is weak. But there are also young people who find work while still in education and training.

Eurostat has recently disseminated data on the labour market participation of young students and apprentices. In Romania, only 2% of students and apprentices were participating in the labour market in 2021 (including part-time, weekend jobs, etc.), the lowest level in the European Union and more than ten times below the EU-27 average of 23%. The Netherlands recorded the highest labour market participation of students and apprentices aged 15-29 at 70%. Countries in the north of the EU score best on this indicator, while the southern Mediterranean remains problematic.

In Romania, employers complain of labour shortages due to external migration after the EU accession. External migration is a determinant of poverty, with around 14% of the employed population at risk of relative poverty. More than one million employment contracts are concluded at the gross minimum wage.

While Romania is facing a labour shortage, the education system is facing the problem of early school leaving. According to the National Institute of Statistics (NIS), the early school leaving rate was 15.6% in 2020, above the limit of 11.3% set under the Europe 2020 Strategy. According to Eurostat, the EU-27 average value in 2020 for the early school leaving rate of young people (18-24 years) was 9.9%. Spain has the highest value of early school leaving rate (16%) and Croatia the lowest (2.2%). Romania ranks second last in the EU-27 ranking for this indicator.

From an educational and vocational point of view, young people can be in the following situations: (1) young people who are in education and/or

training but not in employment; (2) young people who are in education but in employment; (3) young people who are not in education but in employment; (4) young people who are neither in education nor in employment.

Regarding the NEET issues, we should not only consider unemployed, young people registered by employment agencies as job seekers. NEETs also include young people who are unwilling to work or return to the education system because they are looking for job opportunities in line with their aspirations (working conditions/wage package), because they have disabilities and cannot find jobs adapted to their special needs, young people who are sick or caring for sick people, or simply young people who are taking a gap year to travel, read, volunteer (voluntary NEETs).

The COVID-19 pandemic has amplified the problems of NEETs, especially for young people in rural and underdeveloped urban areas, those who have only completed compulsory education or who left school early, without a qualification required by the labour market, excluded from the digital society. They become victims of discrimination, marginalisation and social exclusion, with reduced chances of socio-professional (re)integration, at risk of involvement in crime and organised crime.

According to Eurostat, in 2021, the share of young people aged 15 to 29 in Romania who neither work nor study was 32.7%. In 2021, NEET rates in EU Member States for people aged 15-29 with low educational attainment ranged from 6.4% in Sweden to 32.7% in Romania. Looking more closely at these figures, six countries had higher NEET rates than the EU average: Slovakia (16.6%), Spain (18.4%), Malta (20.3%), Italy (23.0%), Bulgaria (24.4%) and Romania (32.7%).

Completion of undergraduate studies does not mean that all young people will face the labour market. Postgraduate studies (masters and PhD) have long been a transition period from school to work. Delaying employment is also reflected in leaving home late. Eurostat data show that in 2021, young Europeans will have left home at an average age of 26.5 years. Young people in Romania leave home on average at the age of 28, men at 30 and women earlier at 26. In most northern and western EU countries, young people left home on average in their early to mid-twenties, while in southern and eastern countries the average age was in their early thirties.

Leaving home late is also a form of social self-protection for young people. Eurostat data show that in 2020, at EU level, young people aged 15-29 who were not living with their parents were more at risk of poverty (25.9%) than those who lived with their parents (17.8%). In Romania, the at-risk-of-poverty and social exclusion rate among young people was over 36% in 2020. Due to poverty, more and more young people in Romania will consider

emigration as a solution to their living standards problems. There is a demand for labour in Romania, but many vacancies are poorly paid, especially in the so-called "hospitality industry" (HORECA). Wages in construction, industry and agriculture have become somewhat more attractive in 2018-2022, but work in these economic sectors involves more effort and wages are lower compared to developed European countries.

2. LITERATURE REVIEW

Stiglitz, Sen and Fitoussi (2009) consider that quality of life is a concept that includes, in addition to productive activity, the standard of living and the whole range of non-material factors (leisure time, quality of social interactions, general life experience).

An analysis by Brzinsky-Fay (2017) of the relationship between educational supply and labour market demand, focusing on youth unemployment in 30 OECD countries, reveals that no single institution can influence relative youth unemployment. Institutions interact with each other in the sense that the results (effects) of their activities are always conjunctural.

In a study of 24 European countries, Jongbloed and Giret (2020) examine the quality of life of NEET according to the context of each country analysed. The study's hypothesis is that the well-being of young NEETs will be higher when there are more comprehensive social protection measures, but that inequalities in well-being between NEET and non-NEET groups will be minimised if the transition period from education and training to actual integration into work is reduced.

To assess the quality of life of young people, Tvaronaviciene et al. (2021) propose the creation of an index composed of factors belonging to the economic environment, the socio-political environment, the social environment and the natural environment. The analysis was conducted on a sample of 384 respondents aged 17-25 years in Ukraine (Rivne region). The most important indicators of young people's well-being are social environment factors, while sociopolitical factors have the least influence. Young people consider family relationships and health the most important indicators of the quality of life.

To explain differences between different European countries regarding the age at which young people leave home, Van den Berg et al. (2021) used panel analysis and found that cultural context forms the most important explanation for cross-national variation (80% of cross-national variation in leaving home). Other important predictors explaining the age at which young people leave the parental home are the following: religiosity, individualistic family values, changing youth unemployment, GDP and net replacement rate.

In Romania, research on standard of living for youth is a constant concern of the Research Institute for Quality of Life (ICCV) of the Romanian Academy, established at the beginning of the post-communist transition (1990). One of the first collective works on the quality of life of young people – *Tineretul Deceniului Unu. Provocările anilor '90 (Youth of the First Decade. The Challenges of the 90s)* – was coordinated by the sociologist Ioan Mărginean (1994) of the ICCV together with professors and researchers from the Department of Sociology and Social Work of the University of Bucharest, the Research Centre for Youth Problems under the Ministry of Youth and Sport, as well as other institutions and departments of the Ministry of Education, the Ministry of Labour and Social Protection, the Ministry of Public Works and Territorial Planning and the Ministry of Internal Affairs. This work brings together contributions from sociology, demography, statistics, social anthropology, psychology, multidisciplinary approaches to youth issues in contemporary society: youth in the macro-social transition in Romania, theoretical and methodological guidelines in youth research, the economic standard of living of young people, young family values, attitudes and behaviours, health status of youth, youth education, socio-professional integration of young people, anomic behaviours among young people, risk groups among young people, analysis of youth policies.

In a study focusing on the perception of local public authorities regarding young people and their participation in the decision-making process, Popescu and Preoteasa (2004) conducted a series of interviews with local authorities/people from local councils (mayors, deputy mayors, councillors, spokespersons), county youth and sports directorates, school management, etc. The main themes were youth problems, youth participation in decision-making in the respective cities, obstacles encountered and solutions for increasing youth participation. The findings of the study show that the most important problems of young people are related to their low standard of living: lack of jobs, low-paid jobs that do not offer prospects for professional advancement and personal development, lack of housing, limited leisure opportunities. All these problems lead to a low participation of young people in social life and, by extension, in the decision-making process in public authorities and in the affairs of their communities.

Sociologist Sorin Mitulescu (2009) conducted a study in which he sought to record the reaction of young people (pupils and students) who are in the process of shaping their own educational style, depending on whether they accept or ignore innovations. His paper concludes that, at the moment, only a small proportion of young people in Romania are ready to take advantage of educational innovations. Most are still in a state of passivity and conservatism.

A qualitative research on young people in Romania is conducted by Constantinescu (2013), focusing on the following dimensions of the quality of life of young people in Romania: health, material wealth and work. The author also believes that the concept of quality of life needs to be better promoted among young people.

A study by Sandu et al. (2014) combined both quantitative and qualitative aspects and focused on young people aged 15-29 in Romania (a sample of 1302 respondents). Ten focus groups were conducted, both in urban and rural areas. The main issues covered were: socio-economic context; family and society; youth education and labour market; preferences, lifestyles and leisure time; religion and spirituality. The findings of the study show that more than 50% of young people over 18 live with their parents due to a lack of economic or social opportunities. Almost half of these young people consider their parents to be middle class and one in three see them as working class. In terms of better education access, there are gaps between young people in poor rural regions and urban areas, reducing the chances of young people from rural backgrounds to pursue higher education.

In an article based on a secondary analysis of statistical data, Leovaridis and Antimiu (2017) highlight the ways in which young people spend their leisure time, the factors that influence leisure time, classification of leisure time functions. The study focuses on three categories of young people: single young people, married young people and married young people with children. All three categories have in common watching films and TV series. Young singles also prefer two other passive ways of spending leisure time, reading and surfing the Internet. Young married couples are more inclined to go out with friends and go on trips outside the capital, thus active ways of spending leisure time; and young married couples with children are also more active, preferring to go to the park with their young children.

Sociologist Ion Ionescu's article (2017) on young people at social exclusion risk focuses on presenting some of the factors that hinder the social inclusion of young people, as well as the social protection measures offered to this vulnerable group. The social inclusion of young people refers to a process that seeks to provide the support needed to realise the potential of any person, regardless of their experience or circumstances. In order to support the active participation of young people in socio-economic, cultural and political life, opportunities for access to education, employment and decent living conditions must be provided. Social exclusion of young people could have long-term consequences for their living conditions, social and economic participation, emotional life and health. At EU level, the main determinants of social exclusion are rooted in social inequalities: barriers to accessing quality

education and training, securing suitable jobs, encountering discriminatory practices and attitudes, etc.

The relation between education and occupation in the youth population is analysed by Gabriela Neagu (2015) from a perspective based on the most known theories in sociology and economics: human capital theory, unequal access to education theory, and signal theory. In the empirical analysis, the author of the study used data from qualitative research conducted in Romania on the education-employment relation, as well as data from the National Institute of Statistics, the Romanian Institute for Evaluation and Strategies (IRES) and Eurostat. The ICCV researcher concludes that people who succeed in multiple levels of education are more likely to get a job.

Petrescu, Neguț and Mihalache's (2021) paper on the implementation of the Youth Guarantee Programme (YGP) in Romania during 2014-2020, identifies the main barriers to the implementation of this programme in our country: lack of coordination of measures between institutions, lack of flexibility in registering NEET, low levels of partnership with local authorities, companies and NGOs, delays in funding measures and lack of centralised monitoring data to provide a picture of progress and therefore of the necessary improvement measures.

3. CONCEPTS, INDICATORS, AND DATA SOURCES

In this paper, the central concept is that of "standard of living" which is defined as the set of elements that make up the well-being in terms of physical, social, cultural, health, etc. state in which people – particularly young people – live, the nature of their activities, social interactions, access to goods and services, lifestyles, leisure patterns, dissatisfaction/satisfaction with personal, family and community life. The concept of standard of living is an evaluative one, which can be measured using quantitative (statistical) and qualitative (sociological) methods.

From a socio-economic point of view, the standard of living is a concept with a high degree of complexity and relativity that underlies the measurement of poverty. The complexity of this concept derives from comprising a multitude of material, educational, health and cultural elements. The relativity of this concept results from the dynamics or volatility of these elements from one historical period to another, from one society to another. For example, what is considered a decent standard of living in a developing country can be interpreted as a minimum standard of living in a socio-economically developed country. A synonymous concept for standard of living is quality of life.

Looking at it from the European level, the standard of living is measured by the ratio of price of certain goods and services and the income in the country concerned. For this purpose, a common national unit called the purchasing power standard (PPS) is used. Comparing gross domestic product (GDP) per capita expressed in PPS gives an overview of living standards in the EU.

To assess the standard of living, the concept is operationalised in two dimensions: the subsistence minimum and the decent living minimum. A subsistence minimum is defined as a level of income that allows a person or household to survive physically and function in optimal physical and mental health. The calculation of the subsistence minimum includes those expenses related to food consumption, clothing, and household maintenance. A decent minimum subsistence level means that the expenses related to the survival/ subsistence of the person or household are added to the expenditure on participation in social, cultural and political life, access to education and vocational training, and labour market orientation. For the measurement of the two dimensions (subsistence minimum and decent minimum), the "minimum monthly subsistence basket" and the "minimum monthly decent living basket" are used.

To measure the standard of living, Eurostat and the statistical institutes of the member states use indicators constructed from data derived from the Quality of Life Survey (QLS), the Household Budget Survey (HBS) and the Household Labour Force Survey (HFLS). All these statistical surveys are in fact selective (sample-based) surveys whose data are representative at European, national and regional level. The system of statistical indicators for measuring the living standards of young people in Romania, in the European context, is composed of:

- relative poverty rate;
- at-risk-of-poverty or social exclusion rate (AROPE index);
- rate of material and social deprivation;
- the share of NEETs in the young population;
- youth mortality (the proportion of young people who die before their 15th birthday);
- the rate of early leaving from education and training;
- share of the youth population with tertiary education in the total youth population;
- the estimated average age of young people leaving the parental household (years);
- youth employment rate;
- youth unemployment rate;
- Gini coefficient;
- the share of government expenditure on education in GDP

The **relative poverty rate** (%) is defined as the share of poor people (by the relative method of estimation) in the total population. Persons in households with a disposable income per adult-equivalent (including or excluding the consumption from own resources) below the poverty line are considered poor. This indicator is currently determined for the threshold of 60% of median disposable income per adult-equivalent.

The **at-risk-of-poverty or social exclusion rate (AROPE)** (%) is a composite indicator adopted at European Union level under the Europe 2020 Programme to promote social inclusion and reduce poverty, representing the proportion of the total population at risk of poverty or social exclusion. People subject to the AROPE indicator are those in at least one of the following situations: have disposable income below the poverty line, are in a state of severe material deprivation, live in a household with very low work intensity household.

The **material and social deprivation rate** (%) is defined as the proportion of people living in ordinary dwellings who are unable to meet the costs of at least five out of thirteen basic necessities of life considered desirable or necessary for an acceptable standard of living.

NEET (%) refers to young people who are neither in employment nor in education or training. This statistical indicator is used to monitor and evaluate European policies on the socio-professional integration of young people aged 15-24, an age group later extended to 15-29. The category of young adults aged 20-34 is also used in Eurostat reports.

Early school leaving rate (18-24 years) (%) refers to the share of young people aged 18-24 who have completed at most lower secondary education and no further education or training in the four weeks preceding the Labour Force Survey (LFS).

Population with tertiary education (levels 5-8) (%). The **educational attainment level** of an individual is the highest ISCED (International Standard Classification of Education) level successfully completed, where the successful completion of an education programme is validated by a recognised qualification, i.e. a qualification officially recognised by the relevant national education authorities or recognised as equivalent to another formal education qualification.

Youth employment rate (%) is the share of the employed population of age group x in the total population of the same age group x. From 2021 the methodology of the Household Labour Force Survey has been revised to meet the requirements of the new European regulations that came into force on 1 January 2021. The most important change concerns the exclusion from employment of persons who produce agricultural goods intended exclusively

or mainly for self-consumption and has a noticeable impact on the survey estimates, particularly in terms of: - a reduction in the number of employed and active population (and in the employment and activity rates); - an increase in the unemployment rate (mainly due to the decrease in the active population which is the denominator of the indicator).

Youth unemployment rate (%) is the share of unemployed people in the labour force. National and regional analyses use the indicator based on data from administrative sources (National Agency for Employment), and comparative analysis uses the indicator based on the methodology of the International Labour Organisation (AMIGO Research).

The **Gini coefficient** is an indicator of the severity of poverty that describes the inequality of income/resource distribution among members of society. The size of the indicator represents the proportion of total income/ resources that would have to be redistributed to members of society to achieve a perfectly equal distribution. The value of the indicator ranges from 0 to 100%, with extreme values having the following hypothetical meanings: 0% means perfect equality and 100% means that all income is held by one person. Reality has shown that inequality never goes away, and that there will always be rich and poor. A relatively low value of the Gini coefficient describes a society with an acceptable standard of living for its members, low poverty, low social and economic inequality.

The **share of public expenditure on education in Gross Domestic Product** (GDP) in a given financial year shows the proportion of annual national financial achievement allocated by the government to education development. To calculate this indicator, the sum of total public expenditure on education is divided by GDP in a given financial year and multiplied by 100. The data source is total public expenditure on education and Gross Domestic Product for a given financial year.

The data used in this study are taken from the Eurostat database.

4. RESEARCH METHODOLOGY, DATA ANALYSIS, AND RESULTS

In this study, we analyze the standard of living of young people at the level of the European Union in 2021. We use descriptive statistics, simple regression analysis, and building a model based on the Ascending Hierarchical Classification (CIA). In the first part of the analysis, we will descriptively present the key indicators considered in this paper.

Descriptive statistics

Summary statistics: *Table 1*

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Relative poverty rate for the 15-19 age group	27	0	27	10,000	30,600	19,974	6,260
AROPE index for the 15-19 age group	27	0	27	11,900	44,500	23,789	7,785
Material and social deprivation for the 15-19 age group	27	0	27	3,000	41,200	11,785	9,360
NEETS age group 15-19 years	27	0	27	2,500	13,200	6,211	2,657
Share of the population aged between 15 and 19 out of the total population	27	0	27	4,100	6,500	5,104	0,556
Share of newborns who die before reaching the age of 15 (youth mortality)	27	0	27	0,290	0,850	0,472	0,146
Early leavers from education and training (18-24 age group)	27	0	27	2,400	15,300	8,226	3,393
Share of the population by level of education (tertiary education)	27	0	27	23,300	62,600	44,533	9,907
Estimated average age of young people leaving the parental household	27	0	27	19,000	33,600	26,989	3,674
Youth employment rate for age group 15-19	27	0	27	2,000	64,900	13,815	14,544
Youth unemployment rate for age group 15-19	27	0	27	8,600	50,800	28,823	12,120
Real GDP per capita in 2021	27	0	27	6690,000	84490,000	28067,407	18414,014
Gini coefficient of equivalised disposable income in 2021	27	0	27	20,900	39,700	29,378	4,310
Total general government expenditure on education, 2020	27	0	27	3,100	7,000	5,222	0,915

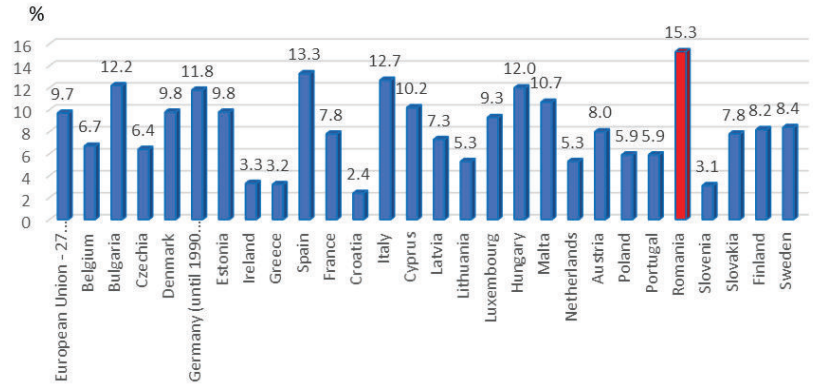
Source: Eurostat

An important indicator reflecting participation in education is the early school leaving rate of young people (18-24 years). In 2021, the rate in Romania was 15.3%, the highest in the European Union. This high rate is driven by factors such as poverty, low level of support from parents and family, low accessibility to education services, lack of infrastructure (especially in rural areas).

In countries such as Croatia, Slovenia, Ireland, Greece, the indicator shows the lowest values (about 3% of young people aged 18-24 left education early).

Early leavers from education and training (%) (aged 18-24 years)

Table 2



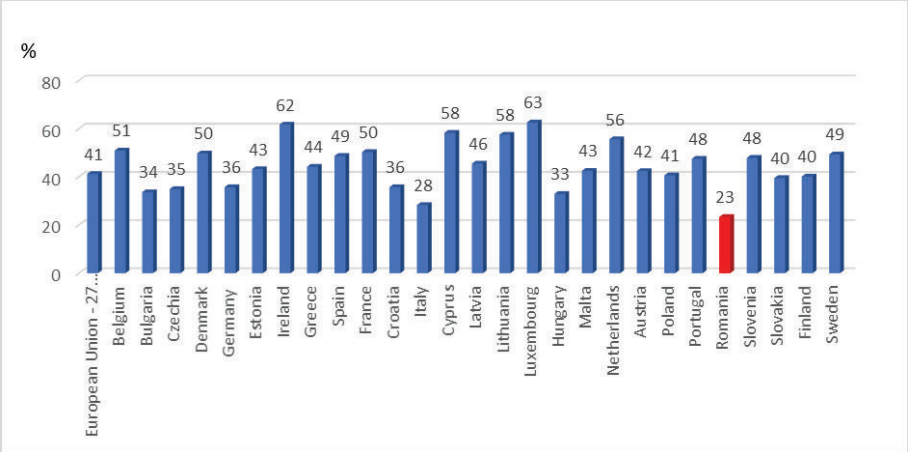
Source: Eurostat

Factors leading to early school leaving are family poverty, seasonal work – especially for rural children, poor parental education, high costs of childcare at school, low accessibility of rural children to prestigious urban schools.

Romanian schools are facing dropouts and functional illiteracy. As interest in school declines, so does the phenomenon of young people delaying employment. More and more young people prefer to delay adulthood, defer actively looking for a job, and are supported by parents at home or by those who have left to work abroad. The category of young people who drop out of school because they are neither interested in a particular qualification/training nor working has also expanded as a result of the changing values, principles and lifestyles characteristic of an 'open society'.

Share of people aged 25-34 with tertiary education(levels 5-8) (%)

Table 3



Source: Eurostat

The share of people aged 25-34 with higher education in Romania is 23.3%, well below the EU average of 41.2%. Luxembourg has the highest share at 62.6%. By 2030, EU countries aim to increase the share of the population aged 25-34 completing tertiary education to 45%, due to its importance for a more developed labour market. In 2021, 14 countries have not yet reached 45%.

The youth unemployment rate in Romania in 2021 in the 25-29 age group is 7.5%, below the EU average. In the 20-24 age group, the unemployment rate is 17.4%, 2.2 percentage points above the EU average.

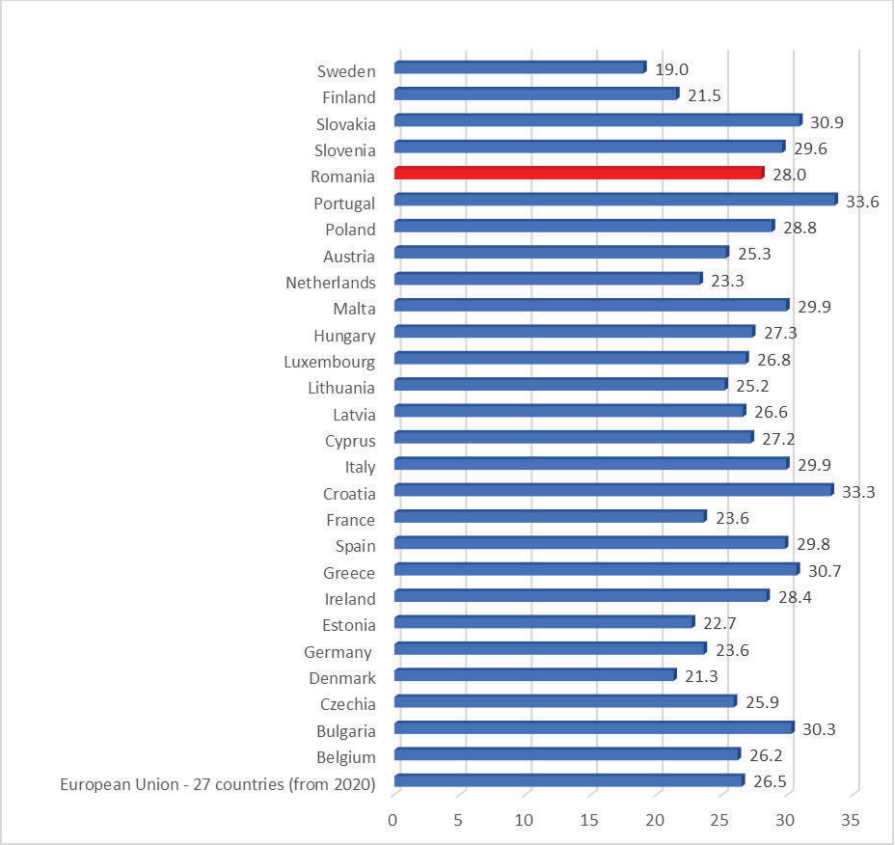
High values are recorded in countries such as Greece and Spain, where more than 30% of the 20-24 age segment are unemployed. Unemployment is highest among graduates of lower (medium and low) education levels.

In terms of the employment rate among young people aged 20-24, Romania records a value of 38.6% for 2021, below the EU average of 50.6%. The Netherlands stands out with an employment rate of 77.9%.

For the 25-29 age group, the employment rate in Romania is 71.3%, 3 percentage points below the EU average. In this age group, Malta has the highest employment rate (88.9%).

Estimated average age of young people leaving the parental household (age)

Table 4



Source: Eurostat

According to Eurostat, in 2021 the average age at which young people no longer live with their parents is 26.5 years at the EU level. Young people in the Nordic countries (Sweden, Finland) move alone the earliest. At the other end of the spectrum are young people in Portugal and Croatia who leave home after the age of 33. In Romania, the average age is 28. It should also be noted that in most EU countries, men decide to leave their parents' home later than women, with a significant gap in Romania (25.7 years for women compared to 30.3 years for men).

The decision to live alone is closely linked to the labour market. Countries where young people leave home at an older age are more likely to have a lower employment rate for the 15-29 age group. In Romania, among the causes leading to an increase in the average age of leaving the parental home are: high rents and apartment prices in relation to purchasing power; level of financial dependency; traditional considerations.

Romania is above the EU average in terms of the percentage of the population both under 15 (15.8% compared to 15.15% EU average) and for the 15-19 age group (5.3% compared to 5.2% EU average) in relation to the total population. In the 20-24 and 25-29 age segments, their share is 5.2% of the total population.

The Bravais-Pearson correlation is one of the most useful methods to validate hypotheses and to identify new avenues of research based on the identified links.

Bravais-Pearson correlation and p-values

Table 5

Correlation matrix (Pearson):

Variables		Var. 1	Var. 2	Var. 3	Var. 4	Var. 5	Var. 6	Var. 7	Var. 8	Var. 9	Var. 10	Var. 11	Var. 12	Var. 13	Var. 14
Var. 1	Relative poverty rate for the 15-19 age group (2021)	**1**	**0,933**	**0,583**	0,370	-0,157	0,354	0,316	-0,148	0,340	**-0,399**	**0,529**	-0,138	**0,766**	-0,352
Var. 2	AROPE index for the 15-19 age group (2021)	**0,933**	**1**	**0,793**	**0,399**	-0,030	**0,475**	**0,384**	-0,223	0,325	**-0,406**	**0,526**	-0,171	**0,741**	**-0,495**
Var. 3	Material and social deprivation for the 15-19 age group (2021)	**0,583**	**0,793**	**1**	0,319	-0,058	**0,625**	0,310	-0,334	0,317	**-0,433**	**0,397**	**-0,425**	**0,562**	**-0,519**
Var. 4	NEETS age group 15-19 years (2021)	0,370	**0,399**	0,319	**1**	-0,283	**0,441**	**0,518**	**-0,448**	0,339	-0,164	0,243	-0,219	0,357	**-0,400**
Var. 5	Share of the population aged between 15 and 19 out of the total population (2021)	-0,157	-0,030	-0,058	-0,283	**1**	-0,247	-0,165	**0,483**	-0,365	**0,427**	-0,027	**0,588**	-0,249	-0,026
Var. 6	Share of newborns who die before reaching the age of 15, 2021 (youth mortality)	0,354	**0,475**	**0,625**	**0,441**	-0,247	**1**	0,297	**-0,467**	0,379	-0,175	0,090	**-0,487**	0,324	-0,334
Var. 7	Early leavers from education and training, 2021 (18-24 age group)	0,316	**0,384**	0,310	**0,518**	-0,165	0,297	**1**	**-0,437**	-0,150	-0,100	0,239	-0,108	**0,385**	-0,142
Var. 8	Share of the population by level of education, 2021 (tertiary education)	-0,148	-0,223	-0,334	**-0,448**	**0,483**	**-0,467**	**-0,437**	**1**	-0,229	0,312	-0,284	**0,619**	-0,177	0,264
Var. 9	Estimated average age of young people leaving the parental household (2021)	0,340	0,325	0,317	0,339	-0,365	0,379	-0,150	-0,229	**1**	**-0,559**	**0,406**	**-0,396**	0,234	**-0,526**
Var. 10	Youth employment rate for age group 15-19 (2021)	**-0,399**	**-0,406**	**-0,433**	-0,164	**0,427**	-0,175	-0,100	0,312	**-0,559**	**1**	**-0,610**	**0,515**	-0,303	0,212
Var. 11	Youth unemployment rate for age group 15-19 (2021)	**0,529**	**0,526**	**0,397**	0,243	-0,027	0,090	0,239	-0,284	**0,406**	**-0,610**	**1**	-0,258	0,246	-0,235
Var. 12	Real GDP per capita in 2021	-0,138	-0,171	**-0,425**	-0,219	**0,588**	**-0,487**	-0,108	**0,619**	**-0,396**	**0,515**	-0,258	**1**	-0,322	0,030
Var. 13	Gini coefficient of equivalised disposable income in 2021	**0,766**	**0,741**	**0,562**	0,357	-0,249	0,324	**0,385**	-0,177	0,234	-0,303	0,246	-0,322	**1**	-0,306
Var. 14	Total general government expenditure on education, 2020	-0,352	**-0,495**	**-0,519**	**-0,400**	-0,026	-0,334	-0,142	0,264	**-0,526**	0,212	-0,235	0,030	-0,306	**1**

Values in bold are different from 0 with a significance level alpha=0,05

p-values:

Variables	Var. 1	Var. 2	Var. 3	Var. 4	Var. 5	Var. 6	Var. 7	Var. 8	Var. 9	Var. 10	Var. 11	Var. 12	Var. 13	Var. 14
Var. 1 Relative poverty rate for the 15-19 age group (2021)	**0**	**0,000**	**0,001**	0,058	0,434	0,070	0,109	0,461	0,083	**0,039**	**0,005**	0,491	**0,000**	0,072
Var. 2 AROPE index for the 15-19 age group (2021)	**< 0,0001**	**0**	**0,0001**	**0,039**	0,881	**0,012**	**0,048**	0,264	0,098	**0,036**	**0,005**	0,394	**0,0001**	**0,009**
Var. 3 Material and social deprivation for the 15-19 age group (2021)	**0,001**	**0,0001**	**0**	0,105	0,773	**0,000**	0,116	0,089	0,108	**0,024**	**0,040**	**0,027**	**0,002**	**0,006**
Var. 4 NEETS age group 15-19 years (2021)	0,058	**0,039**	0,105	**0**	0,153	**0,021**	**0,006**	**0,019**	0,084	0,415	0,222	0,273	0,068	**0,039**
Var. 5 Share of the population aged between 15 and 19 out of the total population (2021)	0,434	0,881	0,773	0,153	**0**	0,215	0,411	**0,011**	0,061	**0,026**	0,892	**0,001**	0,211	0,898
Var. 6 Share of newborns who die before reaching the age of 15, 2021 (youth mortality)	0,070	**0,012**	**0,000**	**0,021**	0,215	**0**	0,132	**0,014**	0,051	0,383	0,655	**0,010**	0,099	0,089
Var. 7 Early leavers from education and training, 2021 (18-24 age group)	0,109	**0,048**	0,116	**0,006**	0,411	0,132	**0**	**0,023**	0,457	0,619	0,230	0,592	**0,048**	0,479
Var. 8 Share of the population by level of education, 2021 (tertiary education)	0,461	0,264	0,089	**0,019**	**0,011**	**0,014**	**0,023**	**0**	0,250	0,113	0,152	**0,001**	0,376	0,184
Var. 9 Estimated average age of young people leaving the parental household (2021)	0,083	0,098	0,108	0,084	0,061	0,051	0,457	0,250	**0**	**0,002**	**0,036**	**0,041**	0,241	**0,005**
Var. 10 Youth employment rate for age group 15-19 (2021)	**0,039**	**0,036**	**0,024**	0,415	**0,026**	0,383	0,619	0,113	**0,002**	**0**	**0,001**	**0,006**	0,125	0,289
Var. 11 Youth unemployment rate for age group 15-19 (2021)	**0,005**	**0,005**	**0,040**	0,222	0,892	0,655	0,230	0,152	**0,036**	**0,001**	**0**	0,193	0,217	0,238
Var. 12 Real GDP per capita in 2021	0,491	0,394	**0,027**	0,273	**0,001**	**0,010**	0,592	**0,001**	**0,041**	**0,006**	0,193	**0**	0,101	0,883
Var. 13 Gini coefficient of equivalised disposable income in 2021	**< 0,0001**	**0,0001**	**0,002**	0,068	0,211	0,099	**0,048**	0,376	0,241	0,125	0,217	0,101	**0**	0,120
Var. 14 Total general government expenditure on education, 2020	0,072	**0,009**	**0,006**	**0,039**	0,898	0,089	0,479	0,184	**0,005**	0,289	0,238	0,883	0,120	**0**

Values in bold are different from 0 with a significance level alpha=0,05

Source: *own processing*

At EU level, the level of material and social deprivation among young people is higher in countries with social class inequalities, while more socially balanced countries are more equal and offer more opportunities for young people to integrate into society.

According to the analysis, countries that invest a higher percentage of GDP in education are less at risk of poverty among the young population, and according to the correlation results, nations in this category have a smaller young population to manage in relation to the total population and are currently affected by the demographic ageing process.

The share of the NEET (Not in Education, Employment, or Training) population is lower in countries that invest in the education sector. States with a higher GDP per capita have (as expected) a more favourable framework for the empowerment of the young population and at the same time a framework is built for the early integration of the mature and financially independent population. The level of material and social deprivation is lower in countries with a dynamic economy, the proportion of the population with higher education is higher than in EU countries (Romania, Bulgaria, Italy) with problematic economies, and the employment rate among young people is high, which implicitly leads to a lower risk of poverty.

All the characteristics listed contribute significantly to creating a favourable framework for increasing the level of independence of the young population, which ultimately reflects an earlier exit from the parental household.

The countries of the European Union that build an adequate framework for the integration of young people into the economy provide important

support in reducing social groups where economic problems are a constant feature of everyday life.

Using simple linear regression analysis, we analyze the statistical relationships between the following variables:

- NEET (depedent variable) and total general expenditure on education (independent variable);
- AROPE (depedent variable) and early leavers from education and training (independent variable);
- employment rate (depedent variable) and total general expenditure on education (independent variable);
- employment rate (depedent variable) and estimated average age of young people leaving the parental household (independent variable).

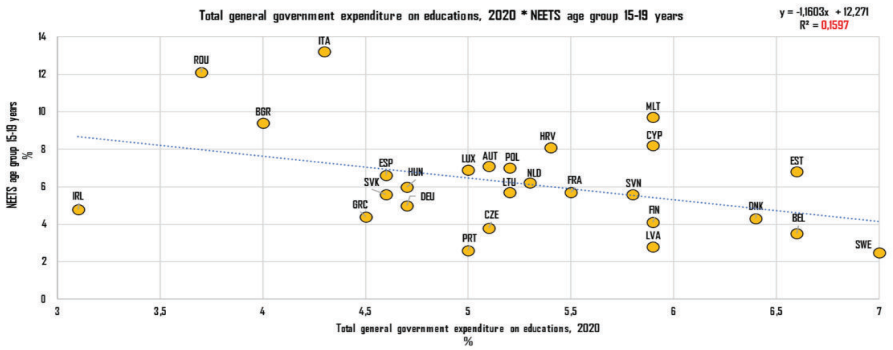
In the simple regression model, the evolution of the dependent variable is defined in terms of an independent variable. In general, the statistical model of simple linear regression is determined by the relation:

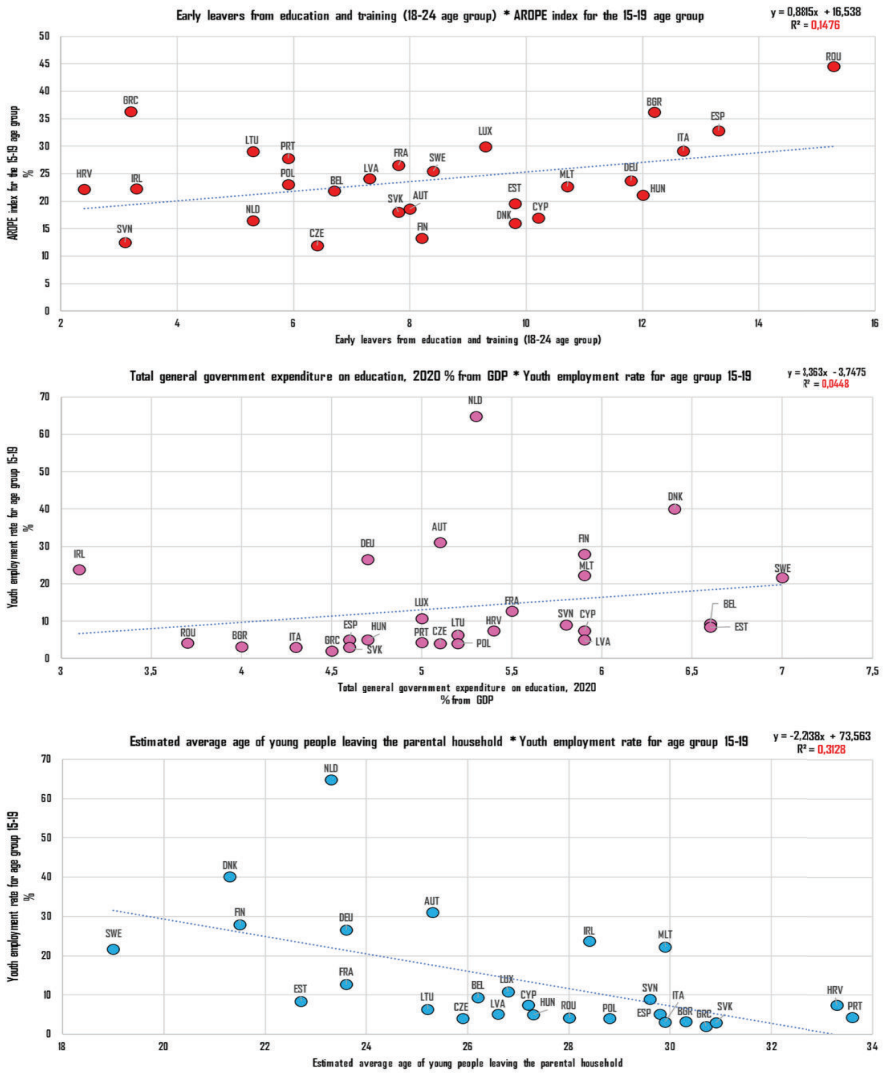
$$Y = \alpha + \beta X_1 + \varepsilon$$

where: Y is the dependent variable
X_1 the independent variable
α, β are the regression coefficients

Simple linear regression models

Graph 1





The estimated equation of the regression model between NEETS (age group 15-19 years) and total general government expenditure on education is:

NEETS = 12.271 - 1.1603*Education_expenditure

A 1% increase in total general government expenditure on education results in a 1.16% decrease in the percentage of NEETS (age group 15-19 years).

The estimated equation of the regression model between AROPE (dependent variable) and early leavers from education and training (independent variable) is of the form:

AROPE=16.538+0.8815*early_levers_from_education

It can be noticed that there is a direct link between the 2 variables, a 1% increase in early leavers from education and training causes a 0.88% increase in the rate of poverty or social exclusion (age group 15-19 years).
Regarding the relationship between employment rate (dependent variable) and total general expenditure on education (independent variable), for the 15-19 age group, the regression equation is:

Employment_Rate = -3.7475+3.383* Education_Expenditure

It is found that an increase in education expenses will increase the employment rate among young people (age group 15-19 years).

The statistical link between the estimated average age of young people leaving the parental household (independent variable) and the employment rate (dependent variable) is of the form:

Employment_rate = 73,563 - 2,213* age_leaving_parental_house

Thus, an increase in the age of leaving the parental home by young people leads to a decrease in the employment rate among them.

Building a CIA (Hierarchical Ascending Classification) model is quite straightforward: spatial grouping of units based on similarity criteria into alternating groups until all units are integrated into a single class is done. Spatial units are grouped based on the principle of similarity between the datasets describing each spatial unit. The principle used in this analysis is that of the Euclidean distance, a distance that facilitates the measurement of the difference between the centre of gravity of a class and the centre of gravity of the cloud of points described by all the spatial units described by all the variables inserted in the analysis (Groza, 2001).

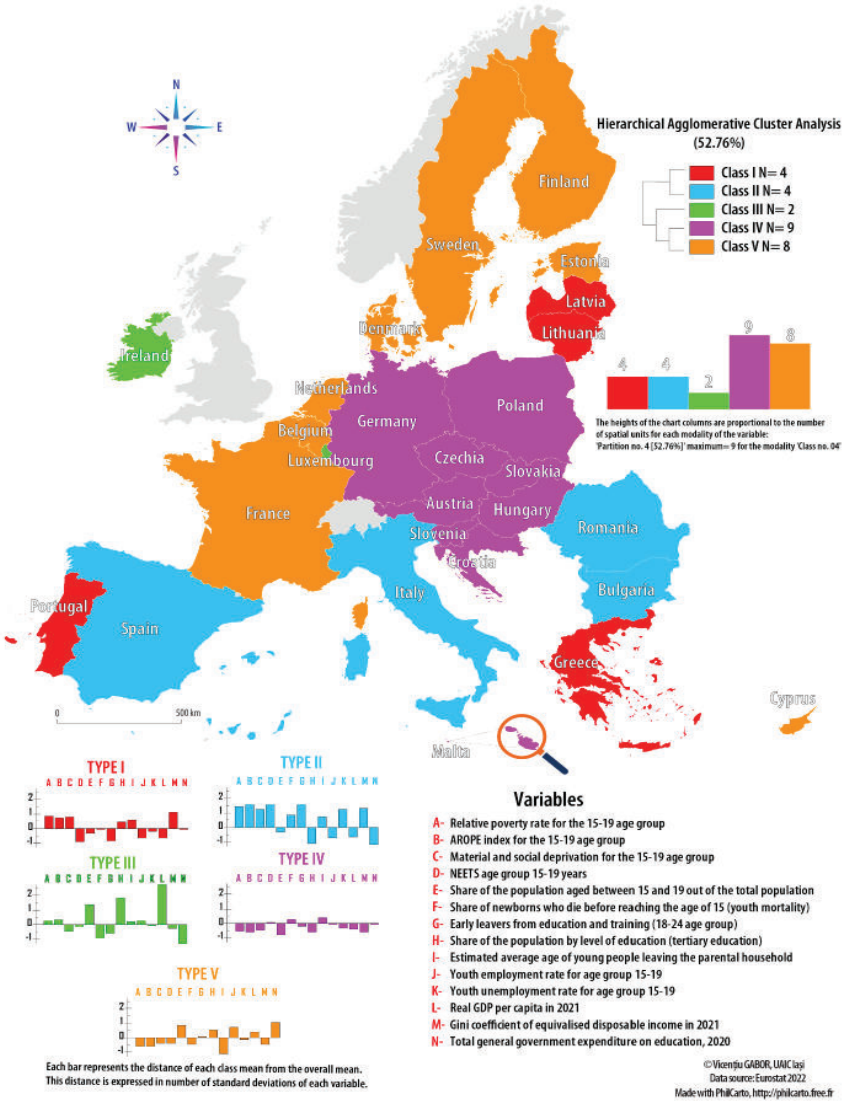
Each house generated will be located in the vicinity or further away from the barycenter of the cloud; the dispersion of the centres of gravity of all classes forms the inter-class variance. The value of the inter-class variance is the parameter that shows how much the classes are separated by the Hierarchical Ascending Classification. The higher its values, the more different the classes are from each other (Groza, 2001).

The dispersion of the values of the variables of each class around the barycenter of that class forms the intra-class variance. The intra-class variance value is the parameter that shows how similar states in the same class are to each other and how different they are from states of other classes/types. The smaller the coefficient, the more representative the division by class, as states are increasingly similar to those in their class and increasingly dissimilar to

those in neighbouring classes. The sum of the two variances constitutes the total inertia (variance) of the cloud of points (respectively 100%).

The results of the multivariate analysis and mapping of the types of spatial units within the European Union according to variables describing the socio-economic framework of young people

Graph 2



Source: own processing

At the general level, the degree of intraclass similarity is 52.76%, the generation of a level of similarity over 50% of the spatial entries was also the main reason for performing the analysis with 5 analysis classes.

The five typologies present the following common features that led to their integration into a known and generic group under the name of typology:

Type 1 is a typology of states with high values regarding the share of the poor young population with high existing inequalities expressed from the perspective of the Gini coefficient. Four states from the European Union, from the north of Europe: Latvia and Latvia and two countries from the Mediterranean area, Greece and Portugal, were integrated into this class.

Type 2, the second typology, integrated the European Union states with the most significant problems from the perspective of the quality of life of the young population on the territory of the European Union. These states have the highest rates of relative poverty, the highest values for the AROPE index, the highest values for the material deprivation of the young population and the most significant shares in the presence of NEETS groups. Also, in the states of this class, young people postpone leaving the parental household the most and encounter the biggest problems with finding a job, a situation reflected in the high unemployment rates among young people in the states of the outlined typology. Four countries from Eastern Europe (Romania, Bulgaria) and two states from Southern Europe (Italy and Spain) were integrated into this typology.

Type 3 is a typology consisting of only territorial entities, Ireland and Luxembourg, and this typology was defined based on variables with values that far exceed the European Union average. High values of the share of the young population in the total population, high shares of the population with higher education (tertiary level) and the highest values for the reported GDP per capita also represent the main peculiarity of the class.

Type 4 is the most representative class in the analysis with the most states included, 9 nations. The typology is one of balance; most of the values of the variables used are close to the averages recorded at the European level. From a geographical point of view, the states of this typology are grouped like a nucleus, integrating countries from central and eastern Europe (Germany, Poland, the Czech Republic, Slovakia, Austria, Hungary and Croatia. The only state that is outside this well-represented territorial grouping geographically is Malta.

Type 5, the last typology of the analysis, was defined by the presence of high values of the analysis variables. The states in this class invest the most in the education sector, young people in these nations leave their parents' residence at an early age, and the employment rate of the young population

(15-19) is also high. In the last typology, the Scandinavian states (Sweden, Finland, Denmark), a Baltic state (Estonia) and part of the states in western Europe (the Netherlands, Belgium and France) were included. Also, in the last typology, the Mediterranean state of Cyprus was integrated.

5. CONCLUSIONS

Europe's population is affected by demographic ageing. The share of young people in the total population decreases every year due to the decrease in the birth rate and fertility, but also the increase in life expectancy. Some eastern countries, such as Romania, the demographic ageing is du to the external migration of young people and young adults.

Young people are fewer and fewer and more affected by poverty. From the perspective of education and employment in the labour market, young Europeans represent a group in a situation of risk or vulnerability. Official statistics show that young people are affected by poverty more than adults or older people.

The developed countries of Western Europe partially solved their labour demand through migration from the eastern countries of the continent. Meanwhile, less economically developed countries facing labour shortages have problems with how the education and training system works. In 2021, Romania's early school leaving rate was 15.3%, the highest in the European Union. The indicator shows the lowest values in Croatia, Slovenia, Ireland, and Greece (about 3% of young people aged 18-24 left education early).

The share of people aged 25-34 with higher education in Romania is 23.3%, well below the EU average of 41.2%, although graduates of tertiary education in Romania find work relatively easily, even if not always in the field in which they trained. Luxembourg has the highest share at 62.6%, with tertiary education in the total population.

The youth unemployment rate in Romania in 2021 in the 25-29 age group is 7.5%, below the EU average. In the 20-24 age group, the unemployment rate is 17.4%, 2.2 percentage points above the EU average. High values are recorded in countries such as Greece and Spain, where more than 30% of the 20-24 age segment are unemployed. Unemployment is highest among graduates of lower (medium and low) education levels.

Regression analysis showed that a 1% increase in total general government expenditure on education results in a 1.16% decrease in the percentage of NEETS (age group 15-19 years). Also, a 1% increase in early leavers from education and training causes a 0.88% increase in the rate of poverty or social exclusion (age group 15-19 years). It was found that an

increase in education expenses will increase the employment rate among young people (age group 15-19 years).

Young people postpone leaving the parental household because encounter a big problem with finding a job, a situation reflected in the high unemployment rates among young people in some European states: two countries form Eastern Europe (Romania, Bulgaria) and two states from Southern Europe (Italy and Spain).

In Romania, Bulgaria, Italy and Spain, young people face the most significant problems related to their quality of life. In these countries that form a particular category, as it was resulted from the Hierarchical Ascending Classification, young people are in risk of poverty, affected by early leaving from education and training, do not work and do not follow professional training, and leave the parental home late.

Taking into account the problems faced by young Europeans, revealed by the statistical analysis, the strategies and public policies for youth must be oriented towards increasing the interest of this age segment in education and professional training, the correlation of professional activity with the job offer, active measures to increase youth employment, developing digital and entrepreneurial skills, combating precarious youth work, poverty or social exclusion.

REFERENCES

1. **Brzinsky-Fay, C**., 2017, *The interplay of educational and labour market institutions and links to relative youth unemployment*, Journal of European Social Policy, Vol.27(4), pp.346-359, https://doi.org/10.1177/0958928717719198
2. **Constantinescu, M.**, 2013, *Educating young people for quality of life improvement*, Procedia–Social and Behavioral Sciences, Vol.93(1), pp.395-399, https://doi.org/10.1016/j.sbspro.2013.09.210
3. **Groza, O.**, 2001, *Reprezentarea şi comunicarea datelor statistice teritoriale. Cazul analizelor multidimensionale,* Analele Ştiinţifice ale Universităţii « Al. I. Cuza » din Iaşi » (serie nouă), tomul XLVII, s. II c. Geografie, suplimentul Lucrările Simpozionului "Sisteme Informaţionale Geografice", nr. 7, Editura Universităţii « Al. I. Cuza » Iaşi, ISSN 1223-5334, pp. 97-109
4. **Ionescu, I.**, 2017, *Tinerii-grup expus riscului de excluziune socială:analizarea factorilor care le îngreunează situația pe piața muncii şi în educație*, Revista Calitatea Vieții, Vol.28, No.1, pp.75-104
5. **Jongbloed, J., Giret, J.-F.**, 2020, *Quality of life of NEET youth in comparative perspective:subjective well-being during the transition to adulthood*, Journal of Youth Studies, Vol25(3), pp.321-343, https://doi.org/10.1080/13676261.2020.1869196
6. **Leovaridis, C., Antimiu, R.**, 2017, *Modalități de petrecere a timpului liber pentru tinerii din țara noastră*, Revista Calitatea Vieții, Vol.28, No.2, pp.283-312
7. **Mărginean, I.**, 1996, *Tineretul deceniului unu. Provocările anilor 90*. Editura Expert, Bucureşti
8. **Mitulescu, S.**, 2009, *Răspunsul tinerilor la inovația educațională*, Revista Calitatea Vieții, Vol.20, No.1-2, pp.141-150

9. **Petrescu, C., Neguț A., Mihalache, F.**, 2021, *Implementation of the youth guarantee programme in Romania*, Revista Calitatea Vieții, Vol.32, No.4, pp.449-467
10. **Popescu, R., Preoteasa**, A.M., 2004, *Percepția autorităților publice locale cu privire la tineri şi participarea lor la procesul de decizie*, Revista Calitatea Vieții, Vol. XV, No.3-4, pp.255-269
11. **Sandu, D., Stoica, C., Umbres, R.**, 2014, *Romanian youth: worries, aspirations, values and life style*, Research report by the Center for Urban and Regional Sociology for Friedrich-Ebert-Stiftung Romania, DOI:10.13140/2.1.4543.3125
12. **Stiglitz, J.E., Sen, A., Fitoussi, J.-P.**, 2009, *Report of the Commission on the measurement of economic performance and social progress*, http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report
13. **Tvaronaviciene, M., Mazur, N., Mischchuk, H., Bilan, Y.**, 2021, *Quality of life of the youth:assessment methodology development and empirical study in human capital management*, Economic Research, Vol.35 (1), pp.1088-1105
14. **Van den Berg, L., Kalmijn, M., Leopold, T.**, 2021, *Explaining cross-national differences in leaving home*, Population, Space and Place, Vol.27(3), pp.1-18, https://doi.org/10.1002/psp.2476
15. https://ec.europa.eu/eurostat/data/database
16. https://ourworldindata.org/

Evaluation of the Effect of the Covid-19 Pandemic on the Financial Performances of Information Technology Sector Companies Traded on Borsa Istanbul with Data Envelopment Analysis

Merve ERYİĞİT
Department of Statistics, Hacettepe University

Esra POLAT (espolat@hacettepe.edu.tr)
Department of Statistics, Faculty of Science, Hacettepe University, 06800, Ankara, Turkey

ABSTRACT

Data Envelopment Analysis is a non-parametric analysis method based on linear modeling. In this method, it is aimed to measure the relative efficiency of decision making units that produce homogeneous outputs by using homogeneous inputs. Many sectors affected positively or negatively by Covid-19 pandemic. In this study, the impact of the Covid-19 pandemic on the financial performance based efficiencies of companies in the information technology (IT) sector traded in Istanbul Stock Exchange (Borsa Istanbul/BIST) examined. In addition, the effect of the Covid-19 pandemic on the periodical financial performance efficiency by determining the periods with the highest and lowest average efficiency for the quarterly periods between 2019:03 and 2021:06 is investigated. As a result, due to inevitable changing situations such as pandemics affect the countries; this study will shed light on both the comparison and evaluation studies for the IT sector and similar studies for other sectors to be made in Turkey or in other countries in the future.

Keywords: *Borsa Istanbul (BIST), Data Envelopment Analysis, Covid-19, Efficiency Values of Companies, Performance*

JEL Classification: *C61, C81, D53, D57, L25*

1. INTRODUCTION

The coronavirus emerged at the end of December 2019 in Wuhan, China's seventh largest city, and it spread from China to many countries of the world in a short time. As a result of the worldwide spreading of the virus and the increasing number of cases and deaths, a pandemic was declared by the World Health Organization (WHO) on 11 March 2020 (Kılıç,2020). In Turkey, the first case was seen on 11 March 2020.

The pandemic has brought with it regulations for people' social and economic life such as not to be together in groups. In many countries, distance education system has been adopted, many companies started to home working system in business life, domestic and international travels have been cancelled, shopping centers and restaurants have been closed. In addition, the obligation to use masks and curfew restrictions were applied from time to time during the process (Kılıç, 2020).

Since the structures and dynamics of the sectors are different, the situations and degrees of being affected by the Covid-19 pandemic also differ. In this respect, it is different and important to examine the impact of the Covid-19 pandemic in terms of sectoral and businesses operating in the sector.

Stock exchange institutions bring investors together in order to conduct their trading activities in a reliable, fair and competitive environment in every country with a stock market. Borsa İstanbul is the only stock exchange institution operating in this field in Turkey. It started its operations under the name of Istanbul Stock Exchange in 1986 and was renamed Borsa Istanbul (BIST) on April 5, 2013. Investors who buy and sell stocks on the stock exchange in Turkey continue their activities within Borsa Istanbul. Borsa Istanbul is supervised by the Capital Markets Board (CMB) of Turkey within the framework of the Capital Markets Law (https://www.matriksdata.com/website/borsa-istanbul-bist; Eryiğit, 2022).

Information technology (IT) sector shows a growth trend due to the technological developments and changes brought by the age. The aim of the study is to examine the change in the efficiency status of companies operating in the field of informatics in Borsa Istanbul, with the first shock impact of the March 2020 period, when the Covid-19 pandemic was first reported in Turkey. This study is a part of master thesis of Eryiğit (2022). For this purpose, it is aimed to investigate the effect of the pandemic period on the financial ratios and efficiencies of 13 companies traded in Borsa Istanbul using Data Envelopment Analysis (DEA). In the research as the data of periods between 2019:03 and 2021:06 is used, it makes possible to interpret the efficiency changes resulting from the financial performance of the companies on a

quarterly basis. Moreover, studying on these periods enables making comments on the companies' efficiency changes due to pandemic by comparing the average efficiency of the 5 periods before Covid-19 (2019:03-2020:03) with the average efficiency of the 5 periods after Covid-19 (2020:06-2021:03). At the same time, the highest and lowest periodical average efficiencies of the sector are determined by the periodical average efficiency values of the CCR and BCC DEA models. The remarkable aspects of this study are working with current data of including the Covid-19 pandemic period for companies operating in the ever-growing information technology sector that traded in BIST and examining the efficiency changes on a company basis (Eryiğit, 2022).

2. **LITERATURE REVIEW**

There are many studies that have been carried out in Turkey and all around the world about the efficiency evaluations of companies in the IT sector using DEA. Some of the studies in the literature are as follows.

Thore et al. (1996) ranked the companies in the US computer industry in terms of efficiency over a 10-year period. R&D expenditures and real capital investments were taken as input variables, profits, market capitalization and sales revenues were taken as output variables. In this study, it was concluded that enterprises that provide long-term and continuity are consistent in terms of efficiencies (Thore et al., 1996). Tektüfekçi (2010) used the data set between 2007 and 2009 for evaluating the financial performance of publicly traded technology companies traded on the Istanbul Stock Exchange (ISE). Receivables turnover, total debt/equity ratio and current ratio are input variables, profit/sales and earnings per share are output variables. As a result of the study, only one company has found as efficient (Tektüfekçi, 2010). Grupta et al. (2013) evaluated the relative efficiencies of 139 companies in the Indian IT sector with DEA method when marketing factors differed. Input variables are selling and distribution expenses, communication expenses, and output variables are total income and total sales. As a result of the study, efficient companies were determined (Grupta et al., 2013). Atilla and Kabataş (2015) analyzed the data between 2010 and 2014 in order to evaluate the financial performances of 11 IT sector companies traded in BIST (Atilla and Kabataş, 2015). Özdağoğlu (2015) analyzed the year 2014 data of 12 companies operating in BIST informatics index. Input variables are stocks, liquid assets, equity capital and tangible fixed assets, and output variables are profit/loss before tax and gross sales. As a result of the study, compared to the sector average the companies with high gross sales are actually in inefficient, while

companies with less gross sales are efficient (Özdağoğlu, 2015). Yeniay (2017) made an efficiency analysis based on the financial performances of 15 IT sector companies traded in BIST, using the data between 2013 and 2015. DEA analysis, Window Analysis and Malmquist TFP Index methods were used. Current ratio, average collection period and stock turnover are used as input variables, return on equity and return on assets are used as output variables. As a result of the study, during the years the financial ratios of the companies and their efficiency values were determined (Yeniay, 2017). Özcan (2020), analyzed the financial performances of 15 BIST technology companies with DEA using quarterly data of 2017:03 and 2019:03. Input variables; total debt/equity, receivables turnover, current ratio, output variables are net profit margin and return on equity. As a result of the study, 4 companies were found efficient in 2019 (Özcan, 2020). Gedik (2020) used DEA continuous Malmquist TFP Index method in order to determine the efficiencies of 7 companies in the IT sector between the years 2014-2016. Input variables; inventories, liquid assets, tangible assets, equity, output variables are revenue and profit/loss before tax. As a result of the study, improvement rates are given for inefficient companies to become efficient (Gedik, 2020). Özkan (2021) analyzed the efficiencies of 18 IT sector companies traded in BIST using the 2019 data with DEA method. Input-oriented CCR model was used and Super Efficiency analysis was also applied in order to determine the efficiency ranking between companies. As a result, target values were calculated for the companies that were not efficient (Özkan, 2021). Uygurtürk and Yıldız (2021) analyzed the information technology sector in BIST in order to determine the relationship between the efficiency and financial performance of companies. The data between 2014 and 2018 were analyzed and total assets, operating expenses, sales costs were used as input variables, sales revenue and pre-tax income from continuing operations were used as output variables. DEA method was used to determine the efficiency of the companies and the Gray Relational Analysis method was used to determine the financial performance rankings. As a result of the study, it was concluded that there is a consistency between the two methods (Uygurtürk and Yıldız, 2021).

3. RESEARCH METHODOLOGY

Efficiency and productivity are used in many texts in the same way or instead of each other, which is not true. The two terms are not exactly the same. Because every point on the production frontier represents the maximum efficiency, but this does not mean maximum productivity. Only at a certain point in the production frontier is productivity at its maximum. For this

reason, it can be said that efficiency is a part of productivity (Coelli et al., 2005). The DEA method is widely used for measuring the relative efficiencies of homogeneous DMUs in various production and service areas and for developing suggestions for inefficient DMUs.

Charnes, Cooper and Rhodes extended Farrell's technical efficiency measurement method (1957), which was calculated through a single input/output as a multi-input/output measurement of relative efficiency and this method entered the literature in 1978 as a CCR model (Charnes et al., 1994; Yeşilyurt, 2007). DEA can be addressed through two different approaches. Input-oriented approach is preferred when there is a possibility of intervention on input variables that how the best available inputs are obtained. The output-oriented approach is preferred if the possibility of intervention on output variables is high and the maximum amount of output to be produced using the available inputs is concerned (Charnes et al., 1994; Sarı, 2015). In summary, the input-oriented approach seeks the optimal combination of inputs to produce a particular output efficiently, while the output-oriented approach seeks the maximum amount of output to be produced with a particular input (Düzgün, 2011). The CCR (Charnes, Cooper, Rhodes) and BCC (Banker, Charnes, Choper) models are the basic models for DEA. CCR model measures efficiency under the assumption of constant returns to scale and measures total efficiency (Kuşkonmaz, 2014). Cooper et al. (2002) describe that CCR model is built on the assumption of constant returns to scale of activities. That is, if an activity (x,y) is feasible, then, for every positive scalar t, the activity (tx, ty) is also feasible (Cooper et al., 2007). Banker, Charnes and Cooper (1984) proposed BCC Model with the addition of a constraint to the CCR model. The difference of the BCC model, which is basically similar to the CCR model, is the addition of the convexity constraint to the model (Çağlar, 2003). The BCC model is a model that provides the measurement of the technical efficiency of DMUs with the assumption of variable returns to scale. Therefore, for any DMU having an efficiency value of 1 for the CCR model, the relevant DMU must be efficient both in terms of technical and scale efficiency. For the BCC model, a DMU having an efficiency value of 1 could be not efficient in terms of scale efficiency but it is technically efficient (Kuşkonmaz, 2014).

3.1. Charnes, Cooper and Rhodes (CCR) Model

Input Oriented CCR Model

Input Oriented CCR Primal Model:

The Input Oriented CCR Primal model form is expressed as follows (Tarım, 2001):

$$\max w_0 = \sum_{r=1}^{s} u_r y_{r0}$$

Constraints are defined as (Aktaş, 2014):

$$\sum_{i=1}^{m} v_i x_{i0} = 1 \qquad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 \qquad u_r,\ v_i \ge 0 \qquad \begin{matrix} i = 1,2,\cdots,m \\ r = 1,2,\cdots,s \\ j = 1,2,\cdots,n \end{matrix}$$

In the formula, n: number of DMUs, s: number of outputs, m: number of inputs, v_i is the input weight and u_r is the output weight, $\sum u_r y_{rj}$ is the sum of the outputs, and $\sum v_i x_{ij}$ is the sum of the inputs. In the input-oriented CCR model, the weighted average of the outputs of all DMUs, respectively, is maximized. For the weighted average of the inputs to be 1 in all DMUs, the weighted average of the inputs of the relevant DMU is equalized to 1 in the constraints. The next constraint ensures that the weighted average of the outputs is smaller than the weighted average of their inputs. Thus, it ensured that the output/input ratio is at most 1 for all DMUs. Therefore, the optimal output mean for any efficient DMU is at most equal to 1. The DMUs that are below the efficiency limit are inefficient. These inefficient DMUs' average of the weight outputs, in other words, the efficiency value is less than 1 (Örkcü, 2004; Aktaş, 2014).

Input Oriented CCR Dual Model:

$$\min z_0 = \theta$$

Constraints are defined as (Aktaş, 2014):

$$\theta_0 x_{i0} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- = 0 \qquad \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = 0 \qquad \begin{matrix} \lambda_j \ge 0 \\ s_i^- \ge 0 \\ s_r^+ \ge 0 \end{matrix} \qquad \begin{matrix} i = 1,2,\cdots,m \\ r = 1,2,\cdots,s \\ j = 1,2,\cdots,n \end{matrix} \qquad [1]$$

The solution results of Dual and Primal models give equal results according to linear programming theory ($\min z_0 = \theta_0^* = \max w_0 = w^*$). Decision units with $\sum u_r y_{rj} / \sum v_i x_{rj} = 1$ are reference points for efficiency. w_0 is the performance coefficient assigned to DMU_0, it is calculated with the performances of all other DMUs, and the weights that maximize this value are the weights v_i^* and u_r^*. Other v_i^*, u_r^* weights do not make this result better. When $w_0^* = 1$ for DMU_0, DMU is completely efficient compared to others.

Efficiency constraints are given as: $\theta_0^* = 1,\ \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ = 0$ [2]

DMU_0 is efficient provided that the conditions in eqn [2] are met, and it is not possible to improve the others without changing the specific input and output amount for the relevant DMU_0. Conversely, DMU_0, is said to be inefficient if one or both of the conditions are not met. Constraints on θ_0^* and slack variables indicate poor performance and its origin. If any s_i^{+*} value of a DMU is non-zero, it must increase the corresponding output amount for DMU to become efficient; if s_i^{-*} value is non-zero, it must reduce the amount of the relevant input for DMU to become efficient (Örkcü, 2004). The constraints in eqn [2] can be interpreted as follows; in the input oriented CCR dual model eqn. [1] form, $\theta_0^* < 1$ indicates the performance of other DMUs. In other words, DMU_0 can reduce the amount of input in the ratio of $\left(1-\theta_0^*\right)$. The condition on slack variables is fulfilled if all $\left(s_i^-, s_i^+\right)$ values are zero. Regarding slack variables, it is said that the input x_{i0} can be decreased if $s_i^{-*} > 0$ and the output y_{i0} can be increased if $s_i^{+*} > 0$ and the output y_{i0} can be increased if $s_i^{+*} > 0$ without decreasing or increasing other input and output variables (Aktaş, 2014).

3.2. Banker, Charnes and Cooper (BCC) Model

Input Oriented BCC Model

Input Oriented BCC Primal Model:

$$\min z_0 = \theta$$

Constraints;

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{r0} \qquad r = 1,2,\cdots,s \qquad \theta x_{i0} - \sum_{j=1}^{n} \lambda_j x_{ij} - s_i^- = 0 \quad i = 1,2,\cdots,m \qquad [3]$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad \lambda, s^+, s^- \geq 0$$

Input Oriented BCC Dual Model:

$$\max q_0 = \sum_{r=1}^{s} \mu_r y_{r0} + u_0$$

Constraints; $\sum_{i=1}^{m} v_i x_{i0} = 1 \quad \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + u_0 \leq 0 \quad j = 1,2,\cdots,n$

$$\mu, v \geq 0, u_0 \text{ independent} \qquad [4]$$

When the eqn. [3] and eqn. [4] forms are examined, it is seen that the model is quite similar to the CCR model structure. The difference from the CCR model is that the sum of λs is equal to 1 in the primal model, and the u_0

variable is added in the dual model. These changes caused the efficiency limit to change. In the CCR model, the efficiency line passes through the origin, while in the BCC model, it does not have to pass through the origin. With this feature, the BCC model differs from the CCR model. There are no differences in the interpretation of the models in terms of other variables (Örkcü, 2004; Aktaş, 2014).

4. RESULTS

The efficiencies of 13 BIST IT companies traded in BIST in quarterly periods between 2019:03 and 2021:06 are examined. Since the Covid-19 pandemic started on March 11, 2020 in Turkey, 2020:06, 2020:09, 2020:12, 2021:03 and 2021:06 periods are considered as the post-Covid-19 pandemic periods. The data is obtained from the financial reports of the companies from the website of the Public Disclosure Platform (PDP) (www.kap.gov.tr). The commonly used variables in the literature, which are thought to be important in terms of evaluation of the financial performance of the companies, are selected as input and output variables.

The input variables are Current Ratio, Accounts Receivables Turnover Ratio, Cash Ratio and output variables are Return on Assets, Return on Equity. The information about for each input and output variables are given in below (Ilgaz, 2021; Eryiğit, 2022).

Current Ratio: It gives information about the net working capital adequacy of the businesses and their capacity to pay their short-term debts. The current ratio is found by dividing the current assets of the business by the short-term liabilities of the business.

Current Ratio=Current Assets/ Short-term Liabilities

Accounts Receivables Turnover Ratio: It is the ratio that measures the ability of companies to collect their trade receivables. A high receivables turnover rate is an indication that the company collects its receivables quickly, and in this case, the company's liquidity situation is assumed to be high. In this way, the company does not suffer from cash shortages and can evaluate receivables in different economic areas without melting over time. The collection period of the company's receivables may change due to seasonal fluctuations and inflation. In the interpretation of the ratio, comparisons should be made with the previous periods of the company and other companies in the same sector or industry. The low turnover ratio is an indication of low competitiveness, difficulty in debt collection, ineffectiveness of the company's

collection method and insufficient selectivity in credit sales customers. The high turnover ratio reveals the efficiency of the companies in the collection of receivables and ensures that the company is in a comfortable position to repay its debts even if the current and liquidity ratio of the relevant company is low.

Accounts Receivables Turnover Ratio=Net Credit Sales/ Average Trade Receivables

Cash Ratio: It measures the liquidity situation of the company in a narrow frame. Because trade receivables and non-trade receivables are not included in this ratio and it indicates a company's capacity to pay off its short-term debts by current assets that can be converted into cash in a short time. Therefore, it gives information about the extent to which the most liquid assets of the business meet its debts in case of economic difficulties.

Cash Ratio= Liquid Assets/ Short-term Liabilities

Return on Assets: It is used to measure the size the company has reached and the efficiency it has achieved from this size. It expresses the extent to which all the assets owned by the companies are used effectively. It is calculated by dividing net profit by total assets. Since the net profit will be the amount resulting from the interest rates cuts, the ratio varies according to the financial method of the company. For example, it is expected that the net profit to total asset ratio of a company with a high interest rate burden and the high usage of foreign funds will be low when compared to a company that finances itself significantly with its own equities.

Return on Assets= Net Profit /Total Assets

Return on Equity: It measures the profitability per unit of the amount of funds left by the shareholders to the company. In this respect, success in company management is an important indicator in the analysis of profitability status. In addition, it is important in terms of measuring the return of the company's funds left to use. The point that distinguishes this ratio from the earnings per share, which takes into account the paid-in capital, is that it includes all the funds provided by the shareholders to the company, namely the equity. In this respect, the ratio is also used to measure the return on participation of investors who are shareholders in the company, apart from measuring management success.

Return on Equity= Net Profit /Shareholders' Equity

In this study, Efficiency Measurement System Version 1.3 (EMS 1.3.0) package program, developed by Holger Scheel and based on Excel software, was used for the analysis of the DEA model. Since in the data set the presence of negative values for companies would violate the assumption of the DEA to be positive, for the relevant variables negative values are converted to positive values with the formula $\frac{X_{rj} - X_{jmin}}{X_{jmax} - X_{jmin}}$. Here, X_{rj}: r_{th} ouput value of j_{th} decision unit, X_{jmin}: Minimum r value and X_{jmax}: Maximum r value (Yıldız, 2007).

4.1. Data Envelopment Analysis Results

CCR model measures the total efficiency of DMUs, the BCC model measures the technical efficiency of DMUs. Since it is desired to have more detailed information about whether the total inefficiency of DMUs is due to technical efficiency or scale efficiency; both the CCR and BCC models are used. In this study, input-oriented CCR and BCC models are used, and scale efficiency values are calculated using the results of these two models. When input-oriented approach is selected, the efficiency values of the decision units on the efficiency limit is 1 and the efficiency values of the decision units that are not included in the efficiency limit are less than 1. Efficiency values are determined by the distances of decision units from the efficiency boundary (Cooper, 2007; Eryiğit, 2022).

The DEA results obtained for the input-oriented CCR model are shown in Table 1. The companies with an efficiency value of 1 are found to be efficient in terms of performance. When Table 1 is evaluated on a company basis, ALCTL and ARMDA are not efficient in any periods, while DESPC company remains efficient for all periods. On the other hand, PKART company is efficient only for the period of 2021:06, while it is inefficient in other periods. According to the results of the input-oriented CCR model, the graph of the average efficiency values of periods and the number of efficient companies for the periods are given in Figure 1 below.

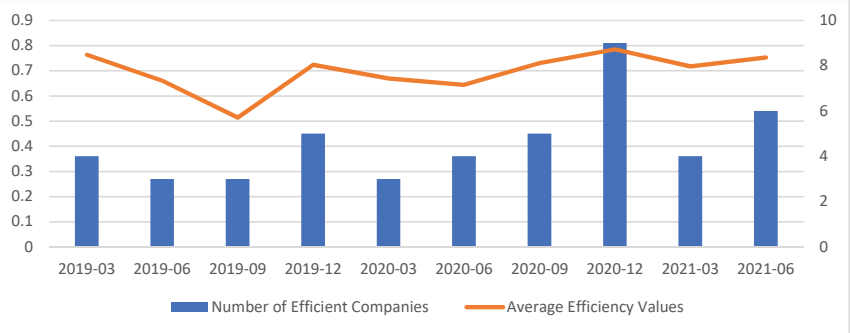
Input Oriented CCR Model Efficiency Scores

Table 1

Companies	2019:03	2019:06	2019:09	2019:12	2020:03	2020:06	2020:09	2020:12	2021:03	2021:06
ALCTL	0.000	0.000	0.058	0.000	0.484	0.953	0.586	0.176	0.064	0.506
ARENA	**1.000**	0.684	0.116	0.488	0.284	0.369	0.311	0.479	0.329	0.312
ARMDA	0.560	0.415	0.179	0.405	0.151	0.136	0.041	0.226	0.800	0.485
DGATE	0.812	0.742	0.588	0.912	**1.000**	**1.000**	0.914	**1.000**	0.700	0.867
DESPC	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
FONET	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.795	0.993
INDES	0.897	0.814	0.695	**1.000**	0.965	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
KFEIN	0.834	0.511	0.436	0.526	0.508	0.132	0.228	**1.000**	0.458	0.000
KAREL	**1.000**	**1.000**	**1.000**	**1.000**	0.244	0.590	0.727	**1.000**	0.697	**1.000**
KRONT	0.882	0.833	0.379	**1.000**	0.914	0.361	**1.000**	**1.000**	**1.000**	0.625
LINK	0.913	0.516	0.437	0.759	0.973	0.551	**1.000**	**1.000**	0.810	**1.000**
LOGO	0.626	0.582	0.519	0.834	0.834	0.984	0.804	**1.000**	**1.000**	**1.000**
PKART	0.414	0.488	0.281	0.484	0.338	0.293	0.181	0.329	0.672	**1.000**
Average Efficiency	**0.764**	**0.660**	**0.514**	**0.724**	**0.669**	**0.644**	**0.730**	**0.785**	**0.717**	**0.753**
Number of Efficient Companies	**4**	**3**	**3**	**5**	**3**	**4**	**5**	**9**	**4**	**6**

The average efficiency values of periods and the number of efficient companies for periods with Input Oriented CCR Model

Figure 1



According to Figure 1, the period of 2019:09 is the period with the lowest average efficiency of all companies with a value of 0.514. The highest average efficiency for all companies is obtained in the period of 2020:12 and 9 companies out of 13 companies are efficient for this period. The number of efficient companies and their average efficiency fluctuating. When post-Covid-19 periods are evaluated, the average efficiency of all companies and the number of efficient companies increased from the period of 2020:06 until

the period of 2020:12, decreased in the period of 2021:03 and increased again in the period of 2021:06.

The input-oriented BCC model results are presented in Table 2 that companies with an efficiency value equal to 1 are found to be efficient in terms of performance.

Input Oriented BCC Model Efficiency Scores

Table 2

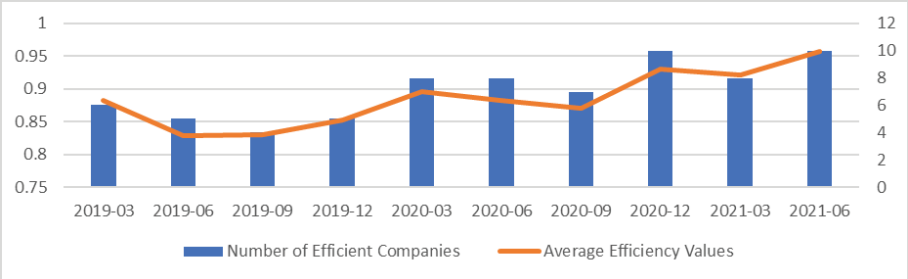
Companies	2019:03	2019:06	2019:09	2019:12	2020:03	2020:06	2020:09	2020:12	2021:03	2021:06
ALCTL	**1.000**	**1.000**	0.966	0.631	0.688	**1.000**	**1.000**	0.767	0.737	0.627
ARENA	**1.000**	**1.000**	0.703	0.808	0.845	0.780	0.726	0.909	0.915	0.891
ARMDA	0.720	0.655	0.752	0.868	**1.000**	**1.000**	0.933	**1.000**	**1.000**	**1.000**
DGATE	0.883	0.808	0.766	0.934	**1.000**	**1.000**	**1.000**	**1.000**	0.808	0.920
DESPC	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
FONET	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
INDES	0.913	0.915	0.978	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
KFEIN	0.874	0.579	0.617	0.622	0.515	0.445	0.298	**1.000**	**1.000**	**1.000**
KAREL	**1.000**	**1.000**	**1.000**	**1.000**	0.994	0.910	0.931	**1.000**	0.812	**1.000**
KRONT	**1.000**	0.981	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
LINK	0.962	0.610	0.878	0.761	**1.000**	0.646	**1.000**	**1.000**	**1.000**	**1.000**
LOGO	0.687	0.611	0.624	0.846	**1.000**	**1.000**	0.976	**1.000**	**1.000**	**1.000**
PKART	0.441	0.605	0.508	0.624	0.604	0.697	0.461	0.423	0.707	**1.000**
Average Efficiency	**0.883**	**0.828**	**0.830**	**0.853**	**0.896**	**0.883**	**0.871**	**0.930**	**0.921**	**0.957**
Number of Efficient Companies	**6**	**5**	**4**	**5**	**8**	**8**	**7**	**10**	**8**	**10**

According to Table 2, while the DESPC company maintains its efficiency for the BCC model for all periods, as in the CCR model, the FONET company is also efficient in all periods. PKART company is efficient only in the period of 2021:06, as in the CCR model. INDES company has increased its efficiency starting from the period of 2019:03 and maintained its efficient position during the Covid-19 period. The efficiency value of KRONT company in the period of 2019:06 is 0.981 and the company is efficient in all other periods. According to the results of input-oriented BCC model, the

graph of the average efficiency values of periods and the number of efficient companies for the periods are given in Figure 2 below.

The average efficiency values of periods and the number of efficient companies for periods with Input Oriented BCC Model

Figure 2



According to Figure 2, the period of 2019:06 is the period with the lowest average efficiency of all companies with a value of 0.828. The highest average efficiency for all companies is obtained in the period of 2021:06 with a value of 0.957 and 10 companies out of 13 companies are efficient in this period. The number of efficient companies and average efficiencies are fluctuating for the pre-Covid-19 and post-Covid-19 periods.

Detailed comments on the basis of companies for both CCR and BCC DEA models will be mentioned later. Scale efficiency values are found in the form of CCR/BCC, and the scale efficiency of DMU equals 1 in case for both scales it has an efficiency value of 1. If a technically efficient DMU is inefficient due to scale, the relevant DMU can't be efficient in total. Therefore, in studies where CCR and BCC models are used together, scale efficiency values for DMUs also can be calculated (Eryiğit, 2022). Hence, the scale efficiency values of the companies for the period between 2019:03 and 2021:06 are given in Table 3.

Scale Efficiency Values

Table 3

Companies	2019:03	2019:06	2019:09	2019:12	2020:03	2020:06	2020:09	2020:12	2021:03	2021:06
ALCTL	0.000	0.000	0.060	0.000	0.703	0.953	0.586	0.229	0.087	0.807
ARENA	**1.000**	0.684	0.165	0.604	0.336	0.473	0.428	0.527	0.360	0.350
ARMDA	0.778	0.634	0.238	0.467	0.151	0.136	0.044	0.226	0.800	0.485
DGATE	0.920	0.918	0.768	0.976	**1.000**	**1.000**	0.914	**1.000**	0.866	0.942
DESPC	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
FONET	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.795	0.993
INDES	0.982	0.890	0.711	**1.000**	0.965	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
KFEIN	0.954	0.883	0.707	0.846	0.986	0.297	0.765	**1.000**	0.458	0.000
KAREL	**1.000**	**1.000**	**1.000**	**1.000**	0.245	0.648	0.781	**1.000**	0.858	**1.000**
KRONT	0.882	0.849	0.379	**1.000**	0.914	0.361	**1.000**	**1.000**	**1.000**	0.625
LINK	0.949	0.846	0.498	0.997	0.973	0.853	**1.000**	**1.000**	0.810	**1.000**
LOGO	0.911	0.953	0.832	0.986	0.834	0.984	0.824	**1.000**	**1.000**	**1.000**
PKART	0.939	0.807	0.553	0.776	0.560	0.420	0.393	0.778	0.950	**1.000**
Scale Average Efficiency	**0.870**	**0.805**	**0.609**	**0.819**	**0.744**	**0.702**	**0.749**	**0.828**	**0.768**	**0.785**

From Table 3 it is seen that, for post-Covid-19 periods 2020:06 and 2020:09, 4 and 5 companies are efficient, respectively. For the period of 2020:12, the scale efficiency of 9 companies equal 1 that is the highest number, this number decreased to 4 companies for the period of 2021:03 and increased to 6 companies for the period of 2021:06. When the scale efficiency averages are examined, it is seen that the highest scale efficiency average with a value of 0.870 belongs to the period of 2019:03 and as of this period, the average scale efficiencies are fluctuating. When the CCR, BCC models and scale efficiency results are examined together, especially for the post-Covid-19 period of 2020:12, it can be said that the total efficiency of the ARMDA company is low due to its low scale efficiency. The companies, with a scale efficiency value of 1, have the most appropriate input-output combinations.

When the CCR, BCC models are evaluated together, the comments made for each company are as follows.

According to the CCR DEA model result, ALCTL company is not efficient in any periods, and the average efficiency of the company's 5 periods before the Covid-19 pandemic is 0.108, and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.457. For BCC model, the company has achieved efficiency in 4 periods, 2 of which are the 2020:06 and 2020:09

periods, considered as the post-Covid-19 periods. The average efficiency of the 5 periods before the Covid-19 pandemic is 0.857 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.826. The company experienced an increment in average efficiency in the CCR model and decrease in the BCC model for the post-Covid-19 period. Since for none of the periods it is efficient in the CCR model, it could not reach the value of 1 for scale efficiency.

According to the CCR DEA model result, ARENA company is found to be efficient only in the period of 2019:03, the company's average efficiency for the 5 periods before the Covid-19 pandemic is 0.514 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.360. According to the BCC model, the company is found to be efficient in the periods of 2019:03 and 2019:06, the average efficiency of 5 periods before Covid-19 is 0.871 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.844. The company's average efficiencies decreased in the post-Covid-19 periods compared to the 5 pre-Covid-19 periods for both models. Generally, it can be said that the efficiency performance of the company has been negatively affected by the pandemic process. In scale efficiency, it reached the value of 1 only in the period of 2019:03.

According to the CCR DEA model result, ARMDA company is not efficient in any periods and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.342 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.338. According to the BCC model, it is found to be efficient for the periods 2020:03, 2020:06, 2020:12, 2021:03, 2021:06, the average efficiency of the 5 periods before the Covid-19 pandemic is 0.799 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.987. There is an increase in the average efficiency in the post-Covid-19 periods in the BCC model, and a slight decrease in the CCR model. Since the company in none of periods is found to be efficient in the CCR model, it could not reach the value of 1 in scale efficiency.

According to the CCR DEA model result, DGATE company is found to be efficient for the periods 2020:03, 2020:06, 2020:12, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.811 and the average efficiency of the 5 periods after the Covid-19 is 0.896. For the BCC model, it is found to be efficient for the periods 2020:03, 2020:06, 2020:09, 2020:12, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.878 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.946. The company has experienced an increase in the average efficiency of the 5 post-Covid-19 periods for both models compared to the average efficiency of the 5 pre-Covid-19 periods. Generally, it can be mentioned that efficiency performance of the company has been positively

affected by the pandemic process. Scale efficiency values is equal to 1 for 3 periods.

Since DESPC company is efficient for all periods according to the results of CCR and BCC DEA models, it can't be mentioned that the Covid-19 outbreak has a positive or negative effect on the company's efficiency performance. As a result of this, its scale efficiency is equal to 1 for all periods.

According to the CCR model result, the FONET company maintains its efficiency until 2021:03, a decrease in its efficiency in the period of 2021:03 and an increase in the period of 2021:06. The average efficiency of the 5 periods before the Covid-19 pandemic is 1, the average efficiency of the 5 periods after the Covid-19 pandemic is 0.958. According to the results of the BCC model, it is found to be efficient for all periods.

According to the CCR model result, INDES company follows an efficient process after the period of 2020:03, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.874, the average efficiency of the 5 periods after the Covid-19 pandemic is 1. For the BCC model result, its efficiency increases from the period of 2019:03 to the period of 2019:12 and becomes an efficient company as of the period of 2019:12 and maintains its efficiency. For the BCC model, the average efficiency of the 5 periods before the Covid-19 pandemic is 0.961, and the average efficiency of the 5 periods after the Covid-19 pandemic is 1. The company experiences an increase in the post-Covid-19 periods in terms of average efficiency for both models compared to the 5 pre-Covid-19 periods. Generally, it can be mentioned that efficiency performance of the company has been positively affected by the pandemic process.

According to the CCR model result, KFEIN company is found to be efficient only in the period of 2020:12, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.563, and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.364. Since company's return on equity and return on assets are negative in the period of 2021:06, there is a decrease in its efficiency in this period. According to the BCC model result, it is found to be efficient for the periods 2020:12, 2021:03 and 2021:06, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.641, and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.749. According to the BCC model result, it is found to be efficient for the periods 2020:12, 2021:03 and 2021:06, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.641, and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.749. The company experiences a decrease in the average efficiency of post-Covid-19 periods compared to the average efficiency of the 5 pre-Covid-19 periods for CCR model, and

conversely an increase in the BCC model. The company's scale efficiency is adversely affected in the first 2 quarters of 2021 after the pandemic, but its performance maintained in terms of technical efficiency.

According to the CCR model result, KAREL company is found to be efficient for the periods 2019:03, 2019:06, 2019:09, 2019:12, 2020:12 and 2021:06, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.849, the average of efficiency for 5 periods after the Covid-19 pandemic is 0.803. According to the BCC model, for the periods of 2020:03, 2020:06, 2020:09 efficiencies are close to 1, and the average efficiency of the 5 periods before the Covid-19 pandemic is 0.999, the average efficiency of the 5 periods after the Covid-19 pandemic is 0.931. According to both models, the company experienced a slight decrease in its average efficiency of post-Covid-19 periods compared to the average efficiency of the pre-Covid-19 periods.

According to the CCR model result, KRONT company is found to be efficient for 4 periods, that 3 of them are the post-Covid-19 periods, and the average efficiency for 5 periods after Covid-19 is 0.797. For the BCC model result, it is found efficient for all periods except for the period of 2019:06, and the efficiency of the relevant period is very close to 1. According to the BCC model result, it can't be mentioned that the Covid-19 pandemic has a negative impact on the company's efficiency performance.

According to the CCR model result, the LINK company is found to be efficient only for the periods 2020:09, 2020:12 and 2021:06, which are considered as post-Covid-19 periods. The average efficiency of the 5 periods before the Covid-19 pandemic is 0.720 and the average efficiency for the 5 periods after the Covid-19 pandemic is 0.872. As a result of the BCC model, the company is also efficient for the periods of 2020:03, 2021:03, and the average efficiency of the company for the 5 periods before the Covid-19 pandemic is 0.842 and the average efficiency for the 5 periods after the Covid-19 pandemic is 0.929. For both models, there is a decrease in efficiency for the period of 2020:06, which is considered as the first Covid-19 period, compared to the period of 2020:03. However, the company has experienced an increase in the average efficiency for the post-Covid-19 periods for both models compared to the average efficiency of the pre-Covid-19 periods 5 periods, and it can be said that generally the efficiency performance of the company has been positively affected by the pandemic process.

According to the CCR model result, LOGO company is efficient for the periods 2020:12, 2021:03 and 2021:06. The average efficiency of the 5 periods before the Covid-19 pandemic is 0.679 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.958. For the BCC model

it is efficient for same periods as in CCR model, moreover, it is efficient for the 2020:03 and 2020:06 periods, and its' efficiency nearly close to 1 for the period of 2020:09. The company's average efficiency of the 5 periods before the Covid-19 pandemic is 0.754 and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.995 for BCC model. The company has experienced an increase in the average efficiency of the post-Covid-19 periods compared to pre-Covid-19 periods for both models. Generally, it can be mentioned that efficiency performance of the company has been positively affected by the pandemic process.

The PKART company is only efficient for the period of 2021:06 according to both models. According to the CCR model, the average efficiency of the 5 periods before the Covid-19 pandemic is 0.401 and the average efficiency of the 5 periods after the Covid-19 is 0.495. According to the BCC model, the average efficiency of the 5 periods before the Covid-19 pandemic is 0.556, and the average efficiency of the 5 periods after the Covid-19 pandemic is 0.658. The company has experienced an increase in the average efficiency for the post-Covid-19 periods for both models compared to the 5 pre-Covid-19 periods. It can be said that generally the efficiency performance of the company has been positively affected by the pandemic process.

5. CONCLUSION

The Covid-19 outbreak has caused different effects on the sectors due to the fact that its dynamics and factors affecting its structure are different. Although some sectors are negatively affected due to the measures and restrictions taken during the pandemic period, it had a positive effect on some sectors. In this study, the effect of the pandemic on the financial performance of the IT sector companies traded in the BIST in Turkey is analyzed with the DEA method. Since the first case in Turkey was reported on March 11, 2020, periodical data of between 2019:03 and 2021:06 is included in the analysis.

The periods 2020:06, 2020:09, 2020:12, 2021:03 and 2021:06 are considered as post-Covid-19 periods. In the DEA method, both the CCR model, which measures the total efficiency and the BCC model, which measures the technical efficiency, are used. Since it is thought that the possibility of intervention in the inputs of the companies is more, input-oriented approach is used. The scale efficiencies of the companies are found by the dividing CCR model efficiency scores to BCC model efficiency scores. In this way, it is seen in which period the companies operate at an appropriate scale.

According to the CCR DEA input-oriented model results, the average efficiency of all companies and the number of efficient companies increase or

decrease in some of pre-Covid-19 periods. At the beginning of the Covid-19 process, with the first shock, the number of efficient companies increased, although there is a slight decrease in average efficiency in the period of 2020:06. When the post-Covid-19 periods are evaluated within themselves, the average efficiency of all companies and the number of efficient companies increased in the period of 2020:06 until the period of 2020:12, decreased in the period of 2021:03 and increased again in the period of 2021:06. In terms of the observed periods, the period of 2020:12, which is under the influence of the Covid-19 pandemic, is determined as the period with the highest average efficiency value of 0.785 and the highest number of efficient companies of 9. The period of 2020:12 is followed by 2019:03, the period in which the average efficiency is 0.764 and the number of efficient companies is 4. 2019:09 is the period with the lowest average efficiency value of 0.514 and it has the least efficient number of companies of 3. According to the BCC input-oriented model result, the average efficiency of all companies and the number of efficient companies are fluctuating as increasing or decreasing in some periods of pre-Covid-19. At the beginning of the Covid-19 process, with the first shock, there is no significant change in the average efficiency in the period of 2020:06, but a positive effect is observed in the average efficiency for the period of 2020:12. For the observed periods, the period of 2021:06, which is under the influence of the Covid-19 pandemic, is determined as the period with the highest average efficiency value of 0.957 and the highest number of efficient companies of 10. The period of 2021:06 is followed by the period of 2020:12, in which average efficiency is 0.930 and the number of efficient companies are 10. The period with the lowest average efficiency with a value of 0.828 is the period of 2019:06. When the results of the two models are examined together, DESPC company is found to be efficient for all periods and PKART company is found to be efficient only for the period of 2021:06. Although the rankings of the 2020:12 and 2021:06 periods in terms of average efficiency change for the models, they are the periods with the highest number of efficient companies and high average efficiency values for both models. It is concluded that the negative impact of the Covid-19 pandemic on the average efficiency of all companies can't be mentioned, since the periods of post-Covid-19 pandemic are not the periods with the lowest average efficiency for both models. When the scale efficiency results are examined, since DESPC company is efficient for all periods according to the CCR and BCC models, the scale efficiency value is 1 for all periods and the company has the most appropriate input and output combination. The period in which the companies has the most appropriate input and output combination with a value of 0.870 is the period of 2019:03, and this period is followed by the period of 2020:12

with a value of 0.828. Since 2019:09 is the period with the lowest average efficiency in the CCR model, it is also the period with the lowest average scale efficiency.

The post-Covid-19 pandemic periods are the periods that there are precautions about preventing people from being together in mass and there are also curfews at various time intervals due to the contagiousness of the pandemic and its spreading effect. For this reason, distance education activities were carried out intensively in Turkey during these periods, and the use of electronic commerce (e-commerce) channels increased in this process. As a result, considering the impact of all these situations, though the Covid-19 pandemic has different effects on the financial performance efficiencies of the companies, when the average efficiencies of the periods concerned in this study are examined, it has been concluded that generally there is not a negative impact of the pandemic on the IT sector.

REFERENCES

1. **Aktaş, E.,** 2014, *"Service, Construction and Manufacturing Sectors Performance Analysis with Data Envolopment Analysis"*, Master Dissertation in Turkish, Gazi University Graduate School of Natural and Applied Sciences Department of Statistics, Turkey, Ankara.
2. **Atilla, İ., Kabataş, Y.**, 2015, *"Türkiye'de Bilişim Sektöründe Faaliyet Gösteren İşletmelerin Finansal Performanslarının Değerlendirilmesi: Veri Zarflama Analizi Uygulaması"*, Vergi Sorunları Dergisi, 38(327), 166-175.
3. **Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M.,** 1994, *"Data Envelopment Analysis: Theory, Methodology and Applications"*, Kluwer Academic Publishers.
4. **Coelli, T.J., Rao, D.S.P., O'Donnell, C. J., Battese, G.E.**, 2005, *"An introduction to efficiency and productivity analysis",* 2nd ed. New York, NY, United States: Springer. https://doi.org/10.1007/b136381
5. **Cooper, W.W., Seiford, L.M, Tone, K.,** 2007, *"Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software"*, Second Edition, Springer.
6. **Çağlar, A.,** 2003, *"Efficiency Measurement of Municipalities with Data Envelopment Analysis"*, Phd Dissertation in Turkish, Hacettepe University Graduate School of Natural and Applied Sciences Department of Statistics, Turkey, Ankara.
7. **Düzgün, M.**, 2011, *"Veri Zarflama Analiziyle Elektrik Dağıtım Şirketlerinin Etkinlik ve Verimlilik Analizi"*, Master Dissertation in Turkish, Ankara University Graduate School of Social Sciences, Turkey, Ankara.
8. **Eryiğit, M.,** 2022, *"Evaluation of the Effect of the Covid-19 Pandemic on the Financial Performances of Information Technology Sector Companies Traded on Borsa Istanbul with Data Envelopment Analysis"*, Master Dissertation in Turkish, Hacettepe University Department of Statistics, Turkey, Ankara.
9. **Gedik, A.**, 2020, *"Ekonomik Kalkınmada Bilişim Sektörünün Etkinliği Üzerine Bir Uygulama"*, Journal of Academic Perspective on Social Studies, 1, 67-83.
10. **Grupta, O., Kavidayal P.C., Mishra R.C.**, 2013, *"Evaluating Significance of Marketing in Indian IT Companies Using DEA"*, Global Journal of Management and Business Research Marketing, 13(5).

11. **Ilgaz, B.,** *Oran Analizleri (Ratio Analysis)*, https://www.academia.edu/26025403/Oran_Analizleri_SM_Bar%C4%B1%C5%9F_ILGAZ_www_bilgaz_net_1 . Access date: 26.09.2021.
12. **Kılıç, Y.**, 2020, *"Borsa İstanbul'da COVID-19 (Koronavirüs) Etkisi"*, Journal of Emerging Economies and Policy, 5(1), 66-77.
13. **Kuşkonmaz, S.,** 2014, *"Determining the Effects of Financial Crises in Turkey, on the Fiscal Performance of Companies Traded in Borsa Istanbul by Using Data Envelopment Analysis"*, Master Dissertation in Turkish, Marmara University Institute of Social Sciences, Turkey, İstanbul.
14. **Örkcü, H.H.,** 2004, *"A Multi Criteria Approach to Weight Distribution Problem in Data Envelopment Analysis"*, Master Dissertation in Turkish, Ankara University Graduate School of Social Sciences, Turkey, Ankara.
15. **Özcan, M.,** 2020, *"BIST Teknoloji Firmalarının Finansal Performanslarının Veri Zarflama Analiziyle Ölçülmesi"*, Akademik Sosyal Araştırmalar Dergisi, 8(102), 55-65.
16. **Özdağoğlu, A.,** 2015, *"Bilişim Sektöründeki Şirketlerin Etkinliklerinin Veri Zarflama Analizi ile Değerlendirilmesi"*, Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 22(2), 331-340.
17. **Özkan, T.**, 2021, *"Borsa İstanbul Bilişim Sektöründe İşlem Gören İşletmelerin Etkinliklerinin Değerlendirilmesi"*, Maliye ve Finans Yazıları, 115, 253-268.
18. **Sarı, Z.,** 2015, *"Data Envelopment Analysis and Application"*, Master Turkish Dissertation, Hacettepe University Graduate School of Natural and Applied Sciences Department of Statistics, Turkey, Ankara.
19. **Tarım, A.,** 2021, *"Veri Zarflama Analizi: Matematiksel Programlama Tabanlı Göreli Etkinlik Ölçüm Yaklaşımı"*, T.C. Turkish Court of Accounts Publications, Turkey, Ankara.
20. **Tektüfekçi, F.,** 2010, *"İMKB'ye Kayıtlı Halka Açık Teknoloji Şirketlerinde Finansal Etkinliğin Veri Zarflama Analizi (VZA) ile Değerlendirilmesi"*, Organizasyonel ve Yönetim Bilimleri Dergisi, 2(2), 69-77.
21. **Thore, S., Fred, P., Ruefli, T.W., Yue, P.**, 1996, *"DEA and the Management of the Product Cycle: The U.S. Computer Industry"*, Computers & Operations Research, 23(4), 341-356.
22. **Uygurtürk, H., Yıldız İ.**, 2021, *"İşletmelerin Etkinlikleri ile Finansal Performansları Arasındaki İlişki: Bilişim Sektörü Üzerine Bir Araştırma"*, Verimlilik Dergisi/ Journal of Productivity, 1: 3-15.
23. **Yeniay, B.,** 2017, *"Analysis of Dynamic Effıciency Based on Financial Performances of Information Sector Companies Which Are Treated in Istanbul Stock Exchange (ISE)"*, Master Dissertation in Turkish, Muğla Sıtkı Koçman University Graduate School of Natural and Applied Sciences Department of Statistics, Turkey, Muğla.
24. **Yeşilyurt, C.,** *"Türkiye'deki İktisat Bölümlerinin Göreceli Performanslarının Veri Zarflama Analizi Yöntemiyle Ölçülmesi: KPSS 2007 Verilerine Dayalı Bir Uygulama"*, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 23(4), 138-140.
25. https://www.matriksdata.com/website/borsa-istanbul-bist . Access date: 18.01.2022.
26. **Public Disclosure Platform (PDP)**, www.kap.gov.tr. Access date: 10.09.2021.

Typology of urban agglomerations from the perspective of environmental effects: a quantitative approach using multivariate statistical techniques

Giani Ionel Grădinaru (giani.gradinaru@csie.ase.ro)
The Bucharest University of Economics Studies,
Institute of National Economy - Romanian Academy

Bianca Conțolencu (biancacontolencu@gmail.com)
The Bucharest University of Economics Studies

ABSTRACT

In recent decades, urban agglomerations have experienced a substantial development from a demographic point of view and has resulted in the emergence of urban agglomerations. The purpose of this study is to identify some characteristics of urban agglomerations in terms of environmental effects, to find typologies of groups of urban agglomerations. To carry out quantitative research, the database created by Data-Driven EnviroLab was analyzed, consisting of several indicators relevant to the study. To achieve the first goal, principal component analysis was used to highlight the types of agglomerations. The results showed that public transport is an important source of pollution due to the high values for pollution indicators analyzed in the study, and the level of development, including the economic growth of urban areas is a topic to be discussed and brought to everyone's attention. For the second goal, clustering was used to highlight similarities between the analyzed urban agglomerations.

Keywords *: urban agglomeration, environmental effects, urban ecosystem, principal component analysis, cluster analysis*

1. INTRODUCTION

Industrialization is one of the elements that have contributed to the transformation of human society into a more stable place for society as a whole (Ashton, 1997). However, this process took place due to industrial revolutions, which impacted society in an overwhelming and highly evolving way (Bell,

1976). At the same time, industrialization has boosted the economy, transport, health and, of course, the evolution of urban settlements (Groumpos, 2021). The emergence of industrial revolutions and the development of branches once appeared, for example factories, technology, mechanization, have resulted in industrial agglomerations and gathering of people in urban areas called rural-urban migration (Ahmad et al., 2021, Chien et al., 2021). The general rate of urbanization has increased from about 36% in 1970 to about 55% in 2019, but also in particular cases, for example in China, the same phenomenon occurs, even the rate of urbanization has increased even more significantly, from only 17% in 1970 to more than 60% in 2019 (Ahmad et al., 2021), much higher than the overall urbanization rate. According to studies developed by the United Nations (UN DASE, 2018), this conscious action is expected to continue. By 2050, almost 7 out of 10 people globally will live in cities, i.e., 68% of the world's population will come from urban areas. Over 80% of global GDP comes from cities and, if well managed through increased productivity, innovation and new ideas, urbanization can stimulate sustainable growth. At the same time, cities are largely responsible for environmental issues, such as air and water pollution and the high use of non-renewable energy sources that contribute to climate change (Grimmond, 2007; Guerra et al., 2016).

Urban agglomeration is a developing urban spatial form, determined by factors such as industry and population concentration, strongly connected transportation network, perfect central city, and favorable regional incentive policies (Fang and Yu, 2017; Geddes et al., 1998; McLoughlin, 1969). Although most researchers believe that cities, respectively groups of cities and urban agglomerations have evolved during industrialization (Gottman, 1957; Kunzmann and Wegener, 1991; Chaofan et al., 2020; Hashmi et al., 2020; Tian et al. al., 2022), the scientist Cui, in 1992, argued that urban agglomeration cannot be associated with the term grouping of cities, in other words densely populated urban areas tend not to be associated with industrialization; interactions between different cities are often spontaneous and disorganized. In the context of the relationship between city and environment, population growth and concentration and diversification of technological processes in industry and agriculture, especially chemical processes, cause relatively sudden ecological imbalances due to an interdependence between urbanization and natural space (Alpopi, 2008; Voukkali and Zorpas, 2022). . In the absence of management based on an adequate urban plan, the inefficient use of resources, urban infrastructure and overcrowding in urban areas lead to increased air pollution, implicitly to environmental damage; especially in developing countries leading to environmental degradation (Hashmi et al., 2020). Also, according to a recent report by the World Health Organization (WHO, 2022),

growth in the urban area makes the quality of the environment vulnerable, and in a recent study conducted in 2022, in 6,743 human settlements, of which 75% are presented of urban centers in developing countries, about 97% of them did not meet the air quality standards set by the World Health Organization.

The paper aims to highlight characteristics of urban agglomerations in terms of environmental effects, finding typologies of groups of urban agglomerations in which pollution and environmental effects are similar. So as to achieve the proposed goal, the database created by Data-Driven EnviroLab made up of several indicators relevant to the study was used.

By using the principal component analysis and the cluster analysis, the relevant information will be obtained in order to identify significant feature groups of urban agglomerations. At the level of each group, the aim is to find typologies of urban agglomerations where the level of pollution and environmental effects are treated similarly in terms of urban agglomerations on all continents of the world.

2. **MATERIALS AND METHODS**

The database used includes 162 cities around the world and includes five areas focused on environmental assessment, namely the urban ecosystem, transport, water and air quality and climate change, economic and development indicators. The data set was created by Data-Driven EnviroLab (2020) in order to develop an index of urban environment and social inclusion, and as a starting point were the Sustainable Development Goals and the Habitat III conference, where we wanted to define an agenda so that sustainability initiatives and the efforts of city governments can be evaluated and compared with each other. This could be analyzed by making it possible for programs that are efficient and beneficial to be used by all cities on all continents. Also, the EnviroLab Data-Driven study (2020) was conducted for 32 pilot cities based on the indexes created for the primary variables. The data were collected via remote sensing satellites, geospatial data sets and official statistics measured and officially reported by governments.

Description of variables

Table no. 1

Short Name	Description	United	Observation
CITY PRODUCT PER CAPITA	The measure of the economic production of an urban agglomeration, where the number of inhabitants is considered	Euro	The result of dividing the region's gross domestic product by the total population
CO_2	The trend in carbon dioxide emissions from fossil fuels	Unitless	A positive value indicates an increase in emissions and a negative value shows a declining trend
NO_2	Average Exposure of Nitrogen Dioxide	$\mu g/m^3$	
$PM_{2.5}$	Average Exposure to Fine Particulate Matter	$\mu g/m^3$	
PUBTRANS	Distance to Public Transit	Percent	
TRANSCOV	Access to Public Transit	Percentage of population	Percentage of the population in a neighborhood with access to public transport, walking distance is defined as a radius of 420 meters
TREECAP	Tree cover per capita	Square meters per person	
TREEPROP	Tree cover proportion of neighborhood	Percent	
UHI	Urban heat island effect	Degrees C	
WATSTRESS	Annual water withdrawal relative to water availability	Ratio	
NDBI	Normalized difference built-up index	Percent	A measure of how much an urban area is built up
NDVI	Normalized difference vegetation index	Percent	Satellite-derived. A measure of surface greenness
INC_GINI	Gini coefficient of income inequality between neighborhoods	Ratio	

Source: Own processing

In this article, only 74 cities have been selected for use, to highlight the character of urban agglomerations, namely the high number of inhabitants.

Urban Agglomerations

Table no. 2

Continent	Urban Agglomerations
Africa	Addis Ababa, Algiers, Casablanca, Lagos, Maputo, Nairobi
Asia	Bangalore, Bangkok, Beijing, Chengdu, Chongqing, Guangzhou, Hangzhou, Hongkong, Jakarta, Kolkata, Manila, Nanjing, Qingdao, Seoul, Shanghai, Shenzhen, Singapore, Tel Aviv, Tianjin, Tokyo, Wuhan
Australia	Melbourne, Sydney
Europe	Amsterdam, Athens, Barcelona, Berlin, Bratislava, Brussels, Bucharest, Budapest, Dublin, Hamburg, Istanbul, Kyiv, London, Madrid, Marseille, Milan, Moscow, Munich, Paris, Saint Petersburg, Seville, Vienna, Warsaw
North America	Atlanta, Boston, Chicago, Las Vegas, Los Angeles, Montreal, New York, Ottawa, Seattle, Toronto
South America	Bogota, Buenos Aires, Lima, Manaus, Medellin, Mexico City, Montevideo, Quito, Rio de Janeiro, San Francisco, Santiago, São Paulo

Source: Own processing

The impact that urban regions have on the natural environment can be perceived from many points of view, as an urban ecosystem or by individually analyzing each element of nature, water, air, and soil. In this paper the factors will be analyzed together, so that the techniques used will be multivariate. (Gorunescu, 2006).

The first multivariate technique used in the paper is the principal component analysis. The principle on which this analysis is based is to minimize the number of initial variables, building new artificial variables and graphical representations to visualize the relationships between variables and possible individuals or groups of variables (Smith, 2002). With its help we identified the characteristics of urban agglomerations by reducing the number of variables, thus obtaining a small number of principal components. The resulting tables, namely the one specific to the descriptive analysis, the covariance matrix, to identify the correlations between the variables and the table of eigenvalues will be interpreted to obtain the principal components (Appendix 1). Also, the scree plot chart helps us to choose the number of components to be retained by the sudden interruption of the slope of the chart (Appendix 1).

A first step in the analysis is to observe the relationships between variables through the covariance matrix. Before applying the principal component analysis, the data will be standardized because the descriptive analysis performed above showed that the variables do not have the same unit of measurement and do not have a normal distribution. The method used in this case is the Range method.

The second technique applied in the study is the analysis of the hierarchical cluster. The goal is to produce a unique set of clusters by pairing the variables that will result in the creation of classes with similar characteristics (Girolami, 2002). Clustering-specific algorithms process data and find groups that can be formed. One such method is the Ward method. The algorithm is based on the creation of a correlation matrix, where all clusters and unclustered variables are brought together in pairs, and the pair that produces the highest average intercorrelation within the cluster is chosen as a cluster (Appendix 2). Through the graph, called the dendrogram (Appendix 2), one can observe the class inclusion relationships between groups and the value of the grouping criterion associated with each (Bridges, 1966). The cluster analysis was useful in this paper because it offers the possibility to identify typologies of urban agglomerations on all continents in which the environmental effects and actions of the inhabitants are similar. The first step to follow is to observe the number of clusters of urban agglomerations that are formed according to each component analyzed using the principal component analysis because it was wanted to highlight the clusters within each principal component. The hierarchical K-Means clustering method was used in the present study.

3. RESULTS

3.1 Statistical analysis on the environmental effects of urban agglomerations

All urban agglomerations involved in the analysis show poor performance in terms of pollution levels, highlighting the agglomerations on the Asian continent, where the average indicator $PM_{2.5}$ is 36 $\mu g/m^3$ being in the fourth position out of six in the field of concentrations for particles in suspension in urban agglomerations in Europe, where the average of this indicator was 15.1 $\mu g/m^3$. Looking individually, on each urban agglomeration, it is observed that in the top of the ranking there are urban agglomerations from Asia, in the order of Tianjin (73.1 $\mu g/m^3$) , Nanjing (56.71 $\mu g/m^3$) , Beijing (56.24 $\mu g/m^3$) , Shanghai (52.15 $\mu g/m^3$) , Qingdao (50.55 $\mu g/m^3$), Wuhan (49.32 $\mu g/m^3$) Among the cities analyzed in Europe, in first place is Milan (27.11 $\mu g/m^3$), followed by Budapest (21.52 $\mu g/m^3$),

Warsaw (2083 $\mu g/m^3$), Moscow (20.55 $\mu g/m^3$), Bucharest (19.34 $\mu g/m^3$). The most unaffected cities are Melbourne (3.61 $\mu g/m^3$), Sydney (4.1 $\mu g/m^3$) and Ottawa (5.82 $\mu g/m^3$).

The population of urban areas in Asia, Africa, South America have to walk, on average, the most to reach a public transport station, for example the people of Chongqing have to travel, on average , about 17,000 meters, those in Hangzhou about 8,500 meters, Qingdao about 7,400 meters, those in Lagos 4,534 meters, and those in Mexico about 3,200 meters. We can see that although these are urban agglomerations, they are developing. At the opposite pole are the citizens of Europe and the USA. Thus, public transport is easy in Paris (116 meters), Athens (135 meters), Brussels (151 meters), Boston (175 meters), Chicago (232 meters). Exceptions to this ranking are Tokyo (250 meters) and Seoul (330 meters). In the case of the percentage of the population in a neighborhood with access to public transport, the walking distance is defined as a radius of 420 meters. The trend is maintained, namely the most inaccessible neighborhoods are in cities in Africa, South America, and Asia. , with the exception of the agglomerations of San Francisco and Buenos Aires, Tokyo and Seoul. Athens, Brussels, Paris, Barcelona, Dublin, Amsterdam, Vienna are among the most accessible in terms of public transport. The situation in Bucharest is favorable, namely 86.41% of the population of a neighborhood has access to public transport.

The existence of trees around homes is useful from many perspectives, such as it can improve the social, physical and economic health of a community (Data-Driven EnviroLab, 2020). In this situation, Singapore is in the top of the studied agglomerations, with 28,389 m^2/ person, followed by the capital of Ethiopia, Addis Baba with 18,810 m^2/ person and Brussels with 7312 m^2/ person. Lima, Buenos Aires, Tel Aviv, Athens represents the urban agglomerations where the surface per person covered by trees is at the lowest level, namely 0.25 m^2/ person, 0.93 m^2/ person, 1.3 m^2/ person, respectively 1.8 m^2/ person. In the case of Bucharest, the area covered by trees is 11.2 m^2/ person. The most forested neighborhoods are in cities with developing economic growth, which may be due to the low level of development of industrialization in these cities. Thus, the most forested neighborhoods are in Lagos, Maputo, Nairobi with proportions greater than 70%. The situation in Europe is favorable for Kyiv (58%), Berlin (55%), Hamburg (42%). On the other hand, Lima Bangalore, Qingdao, and Buenos Aires are among the cities where the percentage of afforestation of the neighborhoods is the lowest. The percentage achieved by Bucharest is 8%.

Heat islands are urbanized areas that experience higher temperatures than peripheral areas. Structures such as buildings, roads and other infrastructure

absorb and transfer heat from the sun more than natural landscapes, such as forests and water bodies do. In the presented situation, most of the urban agglomerations in South America register the highest values, such as Medellin (10.33 ℃), Bogota (7.66 ℃), Mexico (7.27 ℃), Santiago (6.67 ℃), Sao Paulo (6.61 ℃). Residents of these cities are affected by heat stress in urban areas, which can lead to prejudicing human health. Barcelona (3.63 ℃), Istanbul (3.37 ℃) and Marseille (2.23 ℃) are the agglomerations in Europe that meet the same phenomenon, but on a smaller scale. At the opposite pole are the cities of Africa, Asia and Europe. In the case of Africa, this phenomenon may have such values due to the equatorial climate. In the case of the Bucharest agglomeration, the situation is not very accentuated, the value being 1.04 ℃).

In terms of the level of built-up area, the lowest performance is found in Lima, Seville, Sao Paulo, but also Barcelona, Madrid or Marseille and Athens. In contrast, most urban areas in Europe and the USA have a high level of built-up area, for example Ottawa, Saint Petersburg, Montreal, Warsaw, Moscow, Kyiv, Bratislava, Hamburg. It can be noticed that many of the most built agglomerations are in the countries where the communist regime existed.

3.2 The link between economic growth and pollution levels

It was studied whether the level of pollutants analyzed, CO_2 and NO_2 the $PM_{2.5}$ proportion of built-up area are influenced by the economic status of urban agglomerations, in other words, whether economic growth accompanied by gas emissions, air pollution and the level of physical development is a turning point. developing countries, where the process of industrialization takes place. In this case, we will apply the Student's test together with the Pearson correlation coefficient.

The analysis shows that between economic production and the carbon dioxide pollution index there is a moderate to strong negative correlation, i.e., when the level of production will increase the level of pollution caused by road traffic will decrease. The same phenomenon happens in the case of the other two indices. Also, analyzing the correlation between the economic production of urban agglomerations and the proportion of built area, we will be able to justify what has been demonstrated by researchers, namely that developing cities face this problem related to physical development, but once reached the evolved level from the economical point of view this problem is diminishing.

3.3 Common characteristics of urban agglomerations

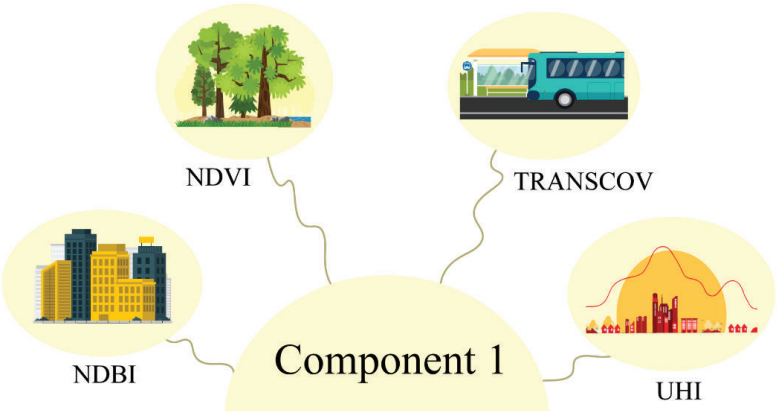
The correlation table shows that all variables are positively correlated. There is a very strong and positive correlation between pollution indicators

(over 80%), and the strongest is 86.3% between the normalized difference built-up index and urban heat island effect which shows that with an increase of the proportion of built-up areas in the urban environment the difference between atmospheric degrees also increases. The variables CO_2(carbon dioxide), NO_2 (nitrogen dioxide) and $PM_{2.5}$(suspended dust) have strong correlations with the insufficiency of quality water and the distance to public transport, over 60%. The correlation between most variables is average, about 50%. At the same time, the two indicators of sustainable public transport are positively correlated with the indicators of air pollution, suggesting that cities that have had good results in providing efficient public transport have had lower levels of air pollution. This model makes sense because public transport reduces the number of private cars, i.e., dependence on them reduces congestion, which will increase local urban air pollution if public transport is not ecofriendly.

According to the table of eigenvalues, the first principal factor retains about 63% of the variation of the initial variables. To draw relevant conclusions, the cumulative percentage of information retained by the components must exceed a threshold of 70%. Analyzing eigenvalues table, shows that the first two components retain approximately 73% of the variation of the variables. In other words, the projection on the plan given by the first two principal axes is explained by 73.12% of the total variability. At the same time, the slope of the scree plot graph changes suddenly to the value of 2, thus suggesting the selection of the first two components.

Component variables 1

Figure. 1

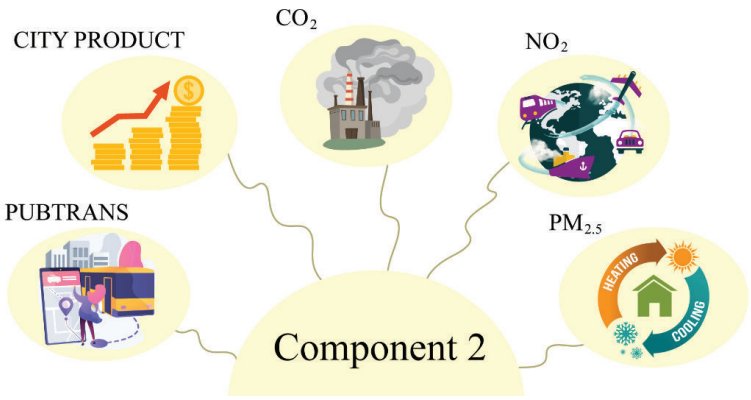


Source: own processing

The table of eigenvectors shows which are the representative variables for the formed components. The first principal component is explained in a positive way by the indicators: the normalized difference built-up index, the normalized difference vegetation index, the urban heat island effect, and the access to public transit. It is found that this component is made up of those indicators related to **the urban agglomeration ecosystem**. Representative indicators show that this variable component will suggest that urban agglomerations can be characterized by the level of buildings built in urban areas, the proportion of vegetation in urban neighborhoods, existing temperature levels because of certain factors and citizens access to public transport within a radius of 420 meters.

Component variables 2

Figure 2



Source: Own processing

With regard to the second component, it is characterized in a negative way by the variables that show **the level of pollution**, namely CO_2, NO_2 and $PM_{2.5}$ and distance to public transit. This time it is noted that urban agglomerations are negatively affected by pollution indicators from car exhaust, combustion process, industrial activities, and electricity production. In the same sense, the variable that considers the proximity of a public transport station to the area of residence of people in an urban neighborhood. At the same time, the second component is determined in a positive way by the indicator that refers to the city's production per capita, which suggests that pollution indicators increase when the level of production relative to the citizen of the city increases. This conclusion can be proved by referring to urban agglomerations on the Asian continent, such as Beijing or Tokyo, which

although they are among the most developed cities in the world, they are also placed in a high ranking when it comes to environmental pollution.

In the graphical representation of the scores of urban agglomerations, the cities outside the existing ellipse on the Asian continent are distinguished, namely Tianjin, Chongqing, Qingdao. They have very high values in terms of indicators that characterize pollution, so they have a very high level of pollution, also the level of development is medium to high. So, these urban agglomerations are highly polluted and developing agglomerations. Also noteworthy is the place of the city of Singapore, on the ellipse line, at the top. It is known that in Singapore the air quality is better than that of many cities in Asia and is comparable to that of urban areas in the United States and Europe. The outliers noted in the graph suggest that urban agglomerations in Asia may be treated differently due to the exceptions that apply to them. They also register a high degree of pollution compared to all other urban agglomerations studied in the analysis.

3.4 Typology of urban agglomerations

In the case of the first principal component called **the urban agglomeration ecosystem**, six clusters were obtained. The first cluster is characterized by urban agglomerations that have a high proportion of buildings with an urban and overcrowded infrastructure, suggesting that they are developing cities that do not have a development and maintenance plan for the environment, implicitly for air quality. Cluster one includes urban agglomerations on the Latin American continent, cities such as Mexico, Sao Paolo, and Bogota. The typology of urban agglomerations in the second cluster is represented by those cities that do not have the most accessible public transport, although the built-up area index has the highest value among clusters, and urban areas are not warmer than their surroundings. The agglomerations of Lima and Casablanca represent this class. The unfavorable quality of the cities of Lima and Casablanca is caused by a fleet of cars and buses for over 20 years (Industrial Economics, 2021, Anas et al., 2018).

The third cluster is composed of urban agglomerations that have a well-implemented and efficient access to public transport, highlighting the attention of the management on the efficiency of fuel consumption compared to other types of transport leading to a sustainable urban form. At the same time, these cities have a smaller area than other urban agglomerations. Most are found in countries with a high level of economic development, and the long-term policy that limits air pollutants is present. The component cities include Bucharest, Berlin, Bratislava, Warsaw, Kyiv, Saint Petersburg, Moscow, Hamburg, Seattle, Chicago, and Montreal. The fifth cluster consists

of urban agglomerations economic centers that have grown rapidly in the last 30 years. It is possible that due to this increase and the increase in economic production, the cities in this cluster will have among the lowest values in terms of access to public transport among all clusters. Until recently, the main objective of these cities was industrialization, so that the governments of these urban agglomerations invest less in public transport. At the same time, the level of the built-up surface is an average one supporting the idea presented above. The agglomerations presented in this cluster are Beijing, Bangkok, Algiers, Atlanta, Hong Kong, Istanbul, Rio de Janeiro, Sydney. The level of the built-up urban area in cluster six has an average to high performance, which states that the present agglomerations are mostly capital. This cluster includes Barcelona, Bangalore, Boston, Buenos Aires, Kolkata, London, New York, Dublin, Paris, San Francisco, Seoul, Tokyo. Also, in these agglomerations, investment was made in efficient public transport, reaching the measure of changing the modes of transport from vehicles to public transport, observing the strongest performance in this cluster for the variable access to public transport.

For **the pollution level component** , taking into account the selection criteria, three clusters were obtained. The first cluster represents urban agglomerations where the level of the economy is developing compared to the other areas processed in the analysis and represents the most polluted cities in terms of carbon dioxide, nitrogen dioxide and suspended particles; it is made up of those urban agglomerations in Asia that have a very high level of pollution, such as Beijing, Shanghai, Tianjin or Hangzhou. The urban agglomerations that formed the second cluster are part of the underdeveloped countries, and the pollution indices have a relatively high performance. This phenomenon may be caused by the lack or non-compliance with environmental policy and the failure to address pollution issues. In the situation presented, that of urban agglomerations in Africa, Latin America, and a few cities in Asia, such as those in India and Indonesia, it is known that the level of the city from the economic point of view is very low, in addition they are polluted as a consequence. the standard of living of the citizens is affected. Depending on the third cluster, those urban agglomerations are highlighted where the production of the city per capita is the highest, and the pollution indicators have the lowest values. Thus, it can be suggested that urban agglomerations with a high level of economy address issues such as the efficiency of public transport to the detriment of private transport and the decrease in the values of pollution indicators. Cluster three brought together urban agglomerations from Europe, the United States and Australia, along with the Romanian capital.

CONCLUSIONS

The urban population is constantly growing, the emphasis is also on the sustainable development of urban agglomerations through policies developed because cities can contribute to mitigating climate change and developing opposition to their impact. The decisions of future city leaders need to meet development goals and promote successful environmental management. Many of the developing cities in the southern part of the world have the least financial resources per capita (Bear et al., 2016). Urban leaders in these regions often face a test of balance between addressing the urgent and growing need for essential services and making long-term decisions and investments that will influence living standards in urban areas and its inhabitants in the years to come. In developed cities, officials have to deal with previous land use and infrastructure decisions that have led to unsustainable resource consumption practices (Hower, 2016). Common global challenges, such as air pollution and climate change, endanger millions of people.

The paper highlighted the impact that urban agglomerations have on the environment. Analyzing the accumulated information, it is discovered that the development of an urban area brings continuous consequences on the environment, and without practices that require respect for the environment, our planet will not perform by itself.

The research identified 2 main components that illustrate common features of urban agglomerations. For the first component, the urban ecosystem reflects the impact that development has on urban areas and how economic growth, the level of built area, the level of vegetation in neighborhoods and the access of inhabitants to urban transport have characterized urban agglomerations. The second component, the level of pollution in relation to public transport is supported by the literature. This analysis highlights an outlier, Singapore, who has performed in terms of air quality through the policies and practices applied in recent years.

Regarding the division of urban agglomerations into clusters according to each component, it is found that while some agglomerations perform well on environmental indicators, almost all face a poor distribution of air pollution, surface covered by trees, public transit, and urban heat. It is observed that revenues still play an significant role in terms of pollution levels in many urban agglomerations, and cities tend to be grouped with other cities with similar economic development. However, higher levels of economic production do not lead to favorable results, as the results show that both developed and underdeveloped and developing urban agglomerations have not placed concern for the quality of the environment in city plans.

References

1. **Ashton, TS**, 1997, *The industrial revolution 1760-1830* . OUP Catalog
2. **Bell, D**., 1976, *The coming of the post-industrial society. In The Educational Forum*. Taylor & Francis Group
3. **Groumpos, PP**, 2021, *A Critical Historical and Scientific Overview of all Industrial Revolutions* . IFAC-PapersOnLine, DOI: https://doi.org/10.1016/j.ifacol.2021.10.492
4. **Rees, J**., 2016, *Industrialization and urbanization in the United States, 1880–1929.* In Oxford Research Encyclopedia of American History
5. **Chien, CF, Tseng, ML, Tan, RG, Tan, K., & Velek, O.**, 2021, *Industry 3.5 for sustainable migration and total resource management* . Resources, Conservation and Recycling
6. **Ahmad, M., Khan Z., Khalid, Answer, M., Jabeen, G**., 2021, *Do rural-urban migration and industrial agglomeration mitigate the environmental degradation across China's regional development levels?* , Sustainable Production and Consumption
7. **United Nations**, 2018, *World Urbanization Prospects: 193*
8. **Grimmond, S.**, 2007, *Urbanization and global environmental change: local effects of urban warming*
9. **Osório de Andrade Guerra, JB, Pereira Ribeiro, JM, Fernandez, F., Bailey, C., Barbosa, SB, Neiva, S.**, 2016, *The adoption of strategies for sustainable cities: a comparative study between Newcastle and Florianópolis focused on urban mobility*, Journal of Cleaner Production
10. **Fang, C., Yu, D.**, 2017, *Urban agglomeration: An evolving concept of an emerging phenomenon. Landscape and Urban Planning*
11. **Geddes, P., LeGates, R., Stout Eds. F**., 1998, *Cities in Evolution*
12. **McLoughlin, JB**, 1969, *Urban & Regional Planning. A Systems Approach*
13. **Gottmann, J.**, 1957, *Megalopolis or the Urbanization of the Northeastern Seaboard*, Economic Geography
14. **Kunzmann, KR, Wegener, M.**, 1991, *The pattern of urbanization in Western Europe*
15. **Chaofan, C., Yawen, S., Qingxin L., Feng J.**, 2020, *Impacts of industrial agglomeration on pollution and ecological efficiency-A spatial econometric analysis based on a big panel dataset of China's 259 cities.* Journal of Cleaner Production
16. **Hashmi, SH, Fan, H., Fareed, Z., Shahzad, F.**, 2020, *Asymmetric nexus between urban agglomerations and environmental pollution in top ten urban agglomerated countries using quantile methods* . Environmental Science and Pollution Research
17. **Tian, G., Xia, Q., Wu, Z., Fu, T.**, 2022, *Ecological network analysis of industrial wastes metabolism based on input-output model for Jiangsu, China* . Waste Management
18. **Alpopi, C.**, 2008. *The effects of urban agglomeration on the environment*
19. **Voukkali, Antonis A. Zorpas**, 2022, *Evaluation of urban metabolism assessment methods through SWOT analysis and analytical hierocracy process* , Science of The Total Environment
20. **World Health Organization**, 2022, *Ambient air quality database*
21. **EnviroLab Data-Driven**, 2020, *The Urban Environment and Social Inclusion Index*. Singapore: Yale-NUS College
22. **Gorunescu, F.**, 2006, *Data mining. Concepts, Models and Techniques* . Cluj-Napoca, Romania
23. **Smith, L.**, 2002, *A tutorial on Principal Components Analysis*
24. **Girolami, M**., 2002, *Mercer kernel-based clustering in feature space.* IEEE Transactions on Neural Networks
25. **Bridges, CC**, 1966, *Hierarchical cluster analysis* . Psychological Reports
26. **Industrial Economics, Incorporated**, 2021, *Assessing the Mortality Burden of Air Pollution in Lima-Callao*

27. **Anas, A., Sarkar, S., Timilsina, G.**, 2018, *The Optimal Mix of Pricing and Infrastructure Expansions to Alleviate Traffic Congestion and In-Bus Crowding in Grand Casablanca*
28. **Bear, VA, Mahendra, A., & Westphal, MI**, 2016, *Towards a More Equal City: Framing the Challenges and Opportunities.* Working Paper
29. **Hower, M.**, 2016, *The 5 Toughest Challenges Cities Face* . GreenBiz

Appendix A

Matrix correlation

Fig.A. 1

Uncorrected Correlation Matrix

		INC_GINI	TREELOSS	WATTSTRESS	City Product per capita euro	CO2	NO2	PM25	PUBTRANS	TRANSCOV	UHI	TREECAP	TREEPROP	NDBI	NDVI
INC_GINI	INC_GINI	1.0000	0.6884	0.6419	0.5206	0.6932	0.5890	0.5963	0.4108	0.6865	0.7235	0.1423	0.6306	0.7677	0.7126
TREELOSS	TREELOSS	0.6884	1.0000	0.5097	0.5480	0.6037	0.5085	0.5162	0.3249	0.6070	0.6673	0.3574	0.4916	0.6975	0.6568
WATTSTRESS	WATTSTRESS	0.6419	0.5097	1.0000	0.6014	0.5956	0.6871	0.6923	0.3722	0.6772	0.6209	0.2238	0.4546	0.6575	0.6005
City Product per capita euro	City Product per capita euro	0.5206	0.5480	0.6014	1.0000	0.4311	0.5310	0.5341	0.2841	0.8469	0.7444	0.3366	0.6671	0.7245	0.7941
CO2	CO2	0.6932	0.6037	0.5956	0.4311	1.0000	0.8595	0.8589	0.7670	0.4919	0.6663	0.1835	0.5956	0.7063	0.6405
NO2	NO2	0.5890	0.5085	0.6871	0.5310	0.8595	1.0000	0.9993	0.6746	0.5805	0.6611	0.2114	0.5122	0.6927	0.6117
PM25	PM25	0.5963	0.5162	0.6923	0.5341	0.8589	0.9993	1.0000	0.6716	0.5856	0.6659	0.2090	0.5116	0.7005	0.6146
PUBTRANS	PUBTRANS	0.4108	0.3249	0.3722	0.2841	0.7670	0.6746	0.6716	1.0000	0.2349	0.4734	0.1459	0.5407	0.4840	0.5022
TRANSCOV	TRANSCOV	0.6865	0.6070	0.6772	0.8469	0.4919	0.5805	0.5856	0.2349	1.0000	0.8149	0.2408	0.6642	0.8212	0.7775
UHI	UHI	0.7235	0.6673	0.6209	0.7444	0.6663	0.6611	0.6659	0.4734	0.8149	1.0000	0.2280	0.8090	0.8629	0.8092
TREECAP	TREECAP	0.1423	0.3574	0.2238	0.3366	0.1835	0.2114	0.2090	0.1459	0.2408	0.2280	1.0000	0.2937	0.2180	0.2345
TREEPROP	TREEPROP	0.6306	0.4916	0.4546	0.6671	0.5956	0.5122	0.5116	0.5407	0.6642	0.8090	0.2937	1.0000	0.6663	0.8072
NDBI	NDBI	0.7677	0.6975	0.6575	0.7245	0.7063	0.6927	0.7005	0.4840	0.8212	0.8629	0.2180	0.6663	1.0000	0.8304
NDVI	NDVI	0.7126	0.6568	0.6005	0.7941	0.6405	0.6117	0.6146	0.5022	0.7775	0.8092	0.2345	0.8072	0.8304	1.0000

Eigenvalues of main components

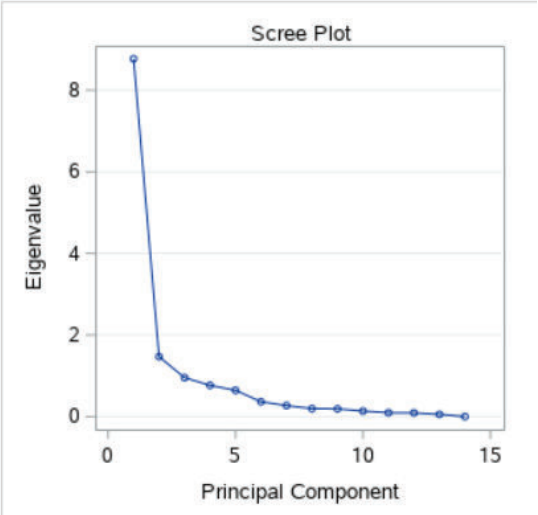
Fig.A. 2.

Eigenvalues of the Uncorrected Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	8.76934466	7.30192058	0.6264	0.6264
2	1.46742408	0.51261763	0.1048	0.7312
3	0.95480645	0.18876339	0.0682	0.7994
4	0.76604306	0.12211477	0.0547	0.8541
5	0.64392829	0.28001721	0.0460	0.9001
6	0.36391108	0.09397405	0.0260	0.9261
7	0.26993703	0.07483218	0.0193	0.9454
8	0.19510485	0.00441019	0.0139	0.9593
9	0.19069466	0.05498886	0.0136	0.9729
10	0.13570580	0.03962022	0.0097	0.9826
11	0.09608558	0.00480430	0.0069	0.9895
12	0.09128129	0.03606150	0.0065	0.9960
13	0.05521979	0.05470639	0.0039	1.0000
14	0.00051340		0.0000	1.0000

Scree plot

Fig.A. 3



Appendix B

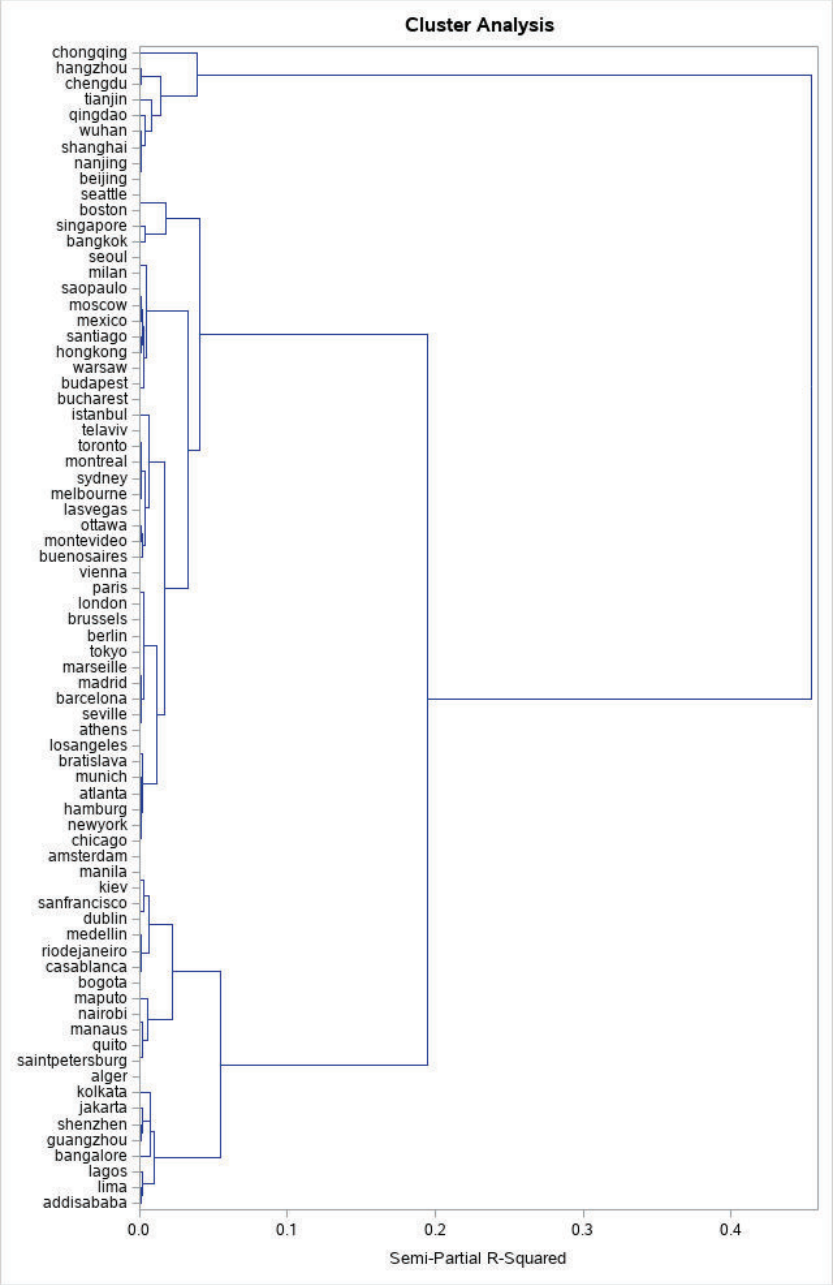
Cluster History

Fig.B.1

Cluster History						
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Tie
7	CL18	CL16	14	0.0221	.815	
6	CL9	CL19	39	0.0325	.783	
5	CL10	chongqing	9	0.0384	.744	
4	CL6	CL8	43	0.0402	.704	
3	CL12	CL7	22	0.0551	.649	
2	CL3	CL4	65	0.1946	.454	
1	CL2	CL5	74	0.4542	.000	

Dendogram

Fig.B.2



Statistics for Variables

Fig.B.3

Statistics for Variables				
Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
City Product per capita euro	0.22772	0.16582	0.484272	0.939007
CO2	0.28345	0.13198	0.789152	3.742764
NO2	0.20593	0.12070	0.665845	1.992625
PM25	0.20565	0.12108	0.662828	1.965845
PUBTRANS	0.15712	0.10732	0.546235	1.203783
OVER-ALL	0.21981	0.13089	0.655127	1.899620

Cluster Means

Fig.B.4

Cluster Means					
Cluster	City Product per capita euro	CO2	NO2	PM25	PUBTRANS
1	0.1991970645	0.8225927178	0.6489701685	0.6502207842	0.3759266067
2	0.1172913584	0.4585896137	0.1952489476	0.1991714418	0.1031275298
3	0.4640677310	0.1223363329	0.1507112649	0.1535794765	0.0293555718

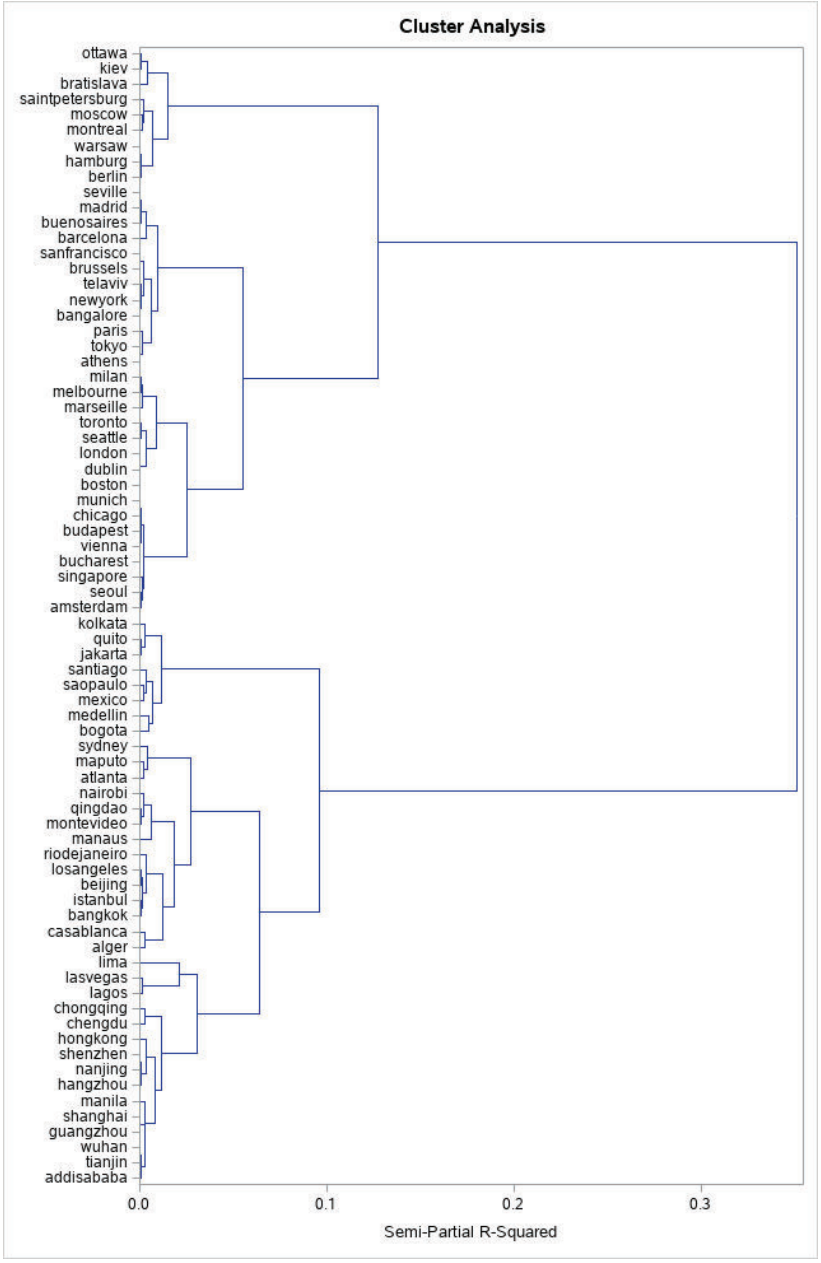
Cluster History

Fig.B.5

Cluster History						
Number of Clusters	Clusters Joined		Freq	Semipartial R-Square	R-Square	Tie
7	CL10	CL24	14	0.0270	.723	
6	CL13	CL9	15	0.0309	.692	
5	CL8	CL15	28	0.0549	.638	
4	CL6	CL7	29	0.0639	.574	
3	CL4	CL14	37	0.0957	.478	
2	CL5	CL11	37	0.1270	.351	
1	CL3	CL2	74	0.3510	.000	

Dendogram

Fig.B.6



Statistics for Variables

Fig.B.7

Statistics for Variables				
Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
TRANSCOV	0.28867	0.12696	0.819815	4.549849
UHI	0.15767	0.09472	0.663861	1.974956
NDBI	0.21983	0.12920	0.678215	2.107669
NDVI	0.20269	0.16458	0.385832	0.628220
OVER-ALL	0.22226	0.13122	0.675332	2.080073

Cluster Means

Fig.B.8.

Cluster Means				
Cluster	TRANSCOV	UHI	NDBI	NDVI
1	0.5535039231	0.7435226276	0.6691740458	0.2813042473
2	0.4211734188	0.0698136168	0.9198970541	0.2127963952
3	0.7931132895	0.3504582904	0.1903796906	0.3584308496
4	0.2688939696	0.3972955353	0.5586067158	0.3306854784
5	0.3925695805	0.3533537902	0.6126906085	0.6557164034
6	0.9134225809	0.3291151635	0.6318942221	0.3539860732

Estimation Of Some Aspects Of Gender Equality on the Labour Market Using Methods Of Multiple Linear Regression and ANOVA[1]

Tatiana COLESNICOVA, PhD, Associate Professor (ctania@gmail.com)
National Institute for Economic Research, Moldova

Elvira NAVAL, PhD, Associate Professor (elvira.naval@gmail.com)
"Vladimir Andrunachievici" Institute of Mathematics and Computer Science, Moldova

Ghenadie CIOBANU, PhD, scientific researcher (gciobanu01@gmail.com)
University ARTIFEX, Romania

ABSTRACT

Gender inequality is one of the social and economic challenges of society. Women actively contribute to social and economic development as employees, entrepreneurs and service providers. However, they are increasingly faced with unequal labour market conditions compared to men. Thus, the purpose of this research is to assess certain aspects of gender inequality in the labour market based on data selected from a gender questionnaire developed by the STCU project team using multiple linear regression and ANOVA methods necessary to monitor this situation. Of the twenty-seven indicators of the questionnaire, five were selected: perceptions of the existence of inequality in the workplace, gender, age, level of education and the economic activity sector. The results of the regression analysis and quantitative assessment of factors revealed that in spite of the significantly low level of education of men, compared to women, they occupy the most qualified jobs with significantly higher wages; women with even higher qualifications are less in demand in the private sector, both in leadership positions and in higher paid ones; for both men and women, age has the same impact in the workplace.

1. This paper has been developed within the framework of the Scientific Project for the period 2020-2023, registered in the State Register of projects in the field of science and innovation of the Republic of Moldova with the code 20.80009.0807.29 Proiect Program de Stat "Perfecționarea mecanismelor de aplicare a instrumentelor inovaționale orientate spre creșterea durabilă a bunăstării populației Republicii Moldova" / Project State Program "Improving the mechanisms for applying innovative instruments aimed at sustainably increasing the welfare of the population of the Republic of Moldova".

Key words: *ANOVA method, gender inequality, labour market, gender questionnaire, least squares method, multiple linear regression.*
JEL: C*12, J16, J7.*

INTRODUCTION.

Gender-related problems are one of the key issues widely argued over, the tackling of which is the objective of policies in most of the modern democracies. There is a variety of entities that deal with these issues from social, economic, ideological and other perspectives from feminist NGOs to international organizations, as well as scientific institutions.

Gender equality is one of the sustainable development goals of United Nations. Gender equality, as well as the empowerment of women and girls, have been recognized as key drivers for achieving all sustainable development goals. Within the framework of the Millennium Development Goals, certain successes have been achieved in the field of gender equality. However, throughout the world, women and girls continue to suffer from discrimination and violence, including in the labour market. Over the past years, noticeable changes have taken place for the better: the number of women in elected positions in government bodies and state institutions has increased; the gender gap was reduced among primary school students, to a lesser extent in secondary schools; the presence of women in the labour market and in international labour migration has increased. These changes in women's lives are associated with the social transformations that accompany economic development. The problems of men and women, including in the labour market, have become one of the central issues in the global and national debates. In many cases, changes in the status of women are caused or accelerated by state reforms and social movements.

To conduct a competent state policy in the field of gender equality, including for equal chances for men and women in the labour market, regular monitoring of the situation is necessary, which consists in applying the correct methodology and includes qualitative and quantitative analyses of the current situation. In this regard, the development of tools and methods for analyzing various aspects of this issue is the task of experts and researchers economists dealing with this problem. The main task of scientists in this matter is to develop the correct methodology, given the fact that the methods of qualitative and quantitative analysis must be applied together, otherwise the result will be incomplete or completely unreliable.

Qualitative researches, in contrast to quantitative, do not focus on statistical measurements, but rely on the understanding, explanation, and

interpretation of empirical data and are the source of the formation of hypotheses and productive ideas. If quantitative research does answer the question "How much?", then qualitative research answers the questions "What?" "How?" and "Why?". In qualitative research, projective and stimulating techniques are widely used - unstructured, non-prescriptive ways to ask questions that help the researcher to reveal motives, circumstances, attitudes, preferences, values, degree of satisfaction, respondents' problems, etc., regarding the object under research. Projective techniques help to overcome communication difficulties, such as verbalization of feelings, relationships, etc., as well as the identification of latent motives, implicit attitudes, repressed feelings, etc.

This research is based on data from a survey done in the framework of the STCU project for further quantitative analysis of the situation of gender equality in the labour market. Thus, the unique data from this questionnaire were used for our calculations.

Quantitative research is the main tool for obtaining the necessary information for planning and decision-making in the case when the necessary hypotheses regarding the main gender problems in the labour market are already formed or formulated. The methods of quantitative research are always based on clear mathematical and statistical models, which allows the result to be based not on opinions and assumptions, but on exact quantitative or numerical values of the studied indicators. The main merit of quantitative research is that they reduce the risk of making wrong decisions and choosing inaccurate planning parameters. Confidence that even without research everything is known about the problem, often turns into insufficiently thought out and insufficiently effective actions and resembles the trial and error method. Thus, quantitative studies are the most appropriate way to numerically assess or examine the depth of the problem.

The gender studies often offer a secondary role to quantitative aspects of gender inequality, to which factors contribute to a positive or negative perception of the gender issues. That's why, the authors are attempting to cover this gap and focus on these aspects by researching the degree of influence of these factors. The research is focused on the Moldovan data and presents an unique econometric model that quantifies the degree of influence of socioeconomic characteristics on the perception of gender inequality in the Republic of Moldova. Thus, the purpose of this research is the estimation of some aspects of gender equality on the labour market by applying a quantitative scientific approach, such as regressional analysis of the results of the questionnaire, elaborated by the authors, on gender problems on the labour market, necessary for monitoring the situation in the country.

THE DEGREE OF SCIENTIFIC APPROACH AND REFLECTION IN THE SPECIALIZED LITERATURE

The problem of gender equality, especially on the labour market, has been researched for many years by scientists from different fields and countries. At the initial stage of research on gender equality, mainly only qualitative research methods were used. With the development of this area many researchers began to use the mathematical apparatus to measure the depth of the problem being studied and one of the types of its intellectual tools is statistical models and methods. Since, gender studies contain different types of regression analysis for the estimations of different aspects of the investigated object. For example, such econometric models as: probit, logit, multiple linear regression models, multiple non-linear regression models, etc. were applied by scientists for quantitative analysis in gender studies.

Applying the apparatus of production functions and estimating its coefficients began with the works of Gary Becker [4-5]. Subsequently, his results were supplemented and developed in practical works of J. Mincer [9-10, 19], R. Anker [2], O. Duncan [17], R. Oaxaca [21], A. Blinder [6], B. Chiswick [5, 7-10], etc., as well as in the theoretical works of T. Shultz [25], L. Thurow [27], etc. Nowadays, the multiple linear regression by the type of extended Mincer earnings equation for gender groups was traditionally used for the estimation of the influence of the human capital and other socioeconomic factors on payment of men and women and was applied by the scientists from the developed and developing countries. For example, this method was applied by: M. Baker and N. Fortin using Canada's data [3], F. Pastore and A. Vereshchagina using Belarusian data [23-24], D. Andren and Th. Andren using Romanian data [1], C. Ogloblin using Russian data [22], S. Roshchin and O. Gorelkina using Russian data [34], A. Oshchepkov using Russian data [32], T. Stuken using Russian data [36], D. Jolliffe using Bulgarian data [18], T. Colesnicova using Moldovan data [14] and scientists from other countries.

Also, the econometric regression analysis as probit-model traditionally used to reveal and estimate the influence of various factors on the probability to be employed for men and women was applied by the scientists from many countries. For example, this method has been applied by: F. Pastore and A. Vereshchagina [23]; I. Mal'tseva and S. Roshchin [33]; S. Roshchin [35]; K. Abazieva [28], T. Bandjukova [29], T. Colesnicova [14].

Also, the examples of applying in gender researches the quantitative methods can serve as the calculations of various indices by international organizations. For example, to quantify the gender inequality the United Nations Development Programme in their yearly Human Development Reports present a composite index and a comparison between world countries.

Another indicator of gender inequality is presented and analyzed by the international organization World Economic Forum in their annual Gender Gap reports. The non-governmental organization World Economic Forum has developed a quantitative method for assessing gender equality (*Gender Gap Index*). Since 2006, analysts at the World Economic Forum have been calculating this Index for most countries in the world; another report was published at the end of 2018. The index takes into account the gap between men and women in the economic, political and other fields, as well as the tendency of its change over time. According to the authors of the research, the world is moving very slowly towards the full realization of the potential of women. According to World Economic Forum estimates, at the current rate of change in the Index, global gender equality will be achieved in 108 years [26].

As well, International Monetary Fund in many of their publications analyze various gender-related indicators that focus on inclusion, employment and technology.

Aspects of gender inequality are quantified by many others indicators – quantitative indicators, like: Duncan Segregation Index, Gender Parity Index, Gender Equity Index, Gender Empowerment Measure and many others more or less specific indicators.

The issues of women's involvement in the production process, the participation of women and men in the labour market, discrimination of women in the field of employment, analysis of gender problems in Moldovan SMEs are reflected in autochthonous scientific researches with the participation of the authors of this paper and are edited in scientific journals recognized in the country and in the materials of national and international conferences [11-15, 20, 30-31].

Scientific novelty of this research is in the applying methods of multiple linear regression and analysis of variance – ANOVA for the estimation of some aspects of gender equality on the labour market. Also, the originality of this work is in the using of the new information sources – dates from the authors' questionnaire on gender problems, elaborated by the team of the STCU project. The feature of this research is that it is based not on the time series data, but on the binary variables obtained as a result of the survey, which could not be drawn up as time series data.

DATA SOURCES AND USED METHODS

In the research the multiple linear regression is applied. Multiple is the linear regression, in the model of which the number of independent variables is two or more. The multiple linear regression equation has the form (***Figure 1***):

The multiple linear regression equation

Figure 1

$$Y = b_0 + \sum_{i=1}^{n} b_i x_i + \varepsilon,$$

where ***Y*** *– estimated dependent variable,* b_0 *– intercept,* ***n*** *– number of independent variables* $(n \geq 2)$, b_i *– coefficients of the independent variables,* x_i *– independent variables,* ε *– error term*

As in the simple linear regression, the b_i parameters of the model are calculated using the least squares method. The difference between simple and multiple linear regression is that instead of a regression line, it uses a hyperplane. The advantage of multiple linear regression over simple is that the use of several input variables in the model allows to increase the share of the explained variance of the output variable, and thus improve the model's fit to the data, in other words, by each new variable added to the model, the coefficient of determination increases.

ANOVA (ANalysis Of VAriance) is a method in mathematical statistics aimed at finding dependencies in experimental data by examining the significance of differences in average values. The essence of analysis of variance is to study the effect of one or more independent variables, usually referred to as factors, on the dependent variable. Analysis of variance was developed by R. Fisher to analyze the results of experimental studies.

For the elaboration of the econometric models, 245 questionnaires of respondents were selected from the Republic of Moldova, that's why 245 Moldovan people participated in the survey. The results were divided on gender basis: out of which 46.6% were male respondents, and 53.4%- female respondents; by age group: under 35 years – 44.3%, and persons over 35 years – 55.7%; by education level with higher education: licentiate/master degree – 57.3%, doctorate/post doctorate – 42.7%; by economic activity sector: government sector – 59.2% of people, from private sector – 40.8% of respondents.

Based on the responses presented on 27 positions, five indicators were selected and they received binary values: one and zero.

The data of the elaborated questionnaire were adapted to the binary data structure (0, 1) by combining insignificant indicators, the values of which surpassed two variants of response. For example, the age indicator: under 35 years = 1; 35-50 years = 2; over 50 years = 3 was replaced by: under 35 years = 1; over 35 years = 0, given the insignificant number of those surveyed over 50 years. The gender indicator was noted as follows: female gender - 1;

male gender - 0. At the level of education, the indicator of the respondents with licentiate / master's studies was 1 and in the case of those with doctoral / postdoctoral studies - 0. The private sector was labeled by binary value 0 and the state sector - by value 1. In the case of the dependent variable (the perception of gender inequality), the perception of its existence was denoted by the value 0, and the perception of the absence - by the value 1. In the regression was used: ordinary least squares method; multiple linear regression method; ANOVA method, Data Analysis/Regression application in Microsoft Excel©.

THE FORMULATION OF THE PROBLEM AND THE OBTAINED RESULTS

The dependent variable selected from the gender questionnaire for our regression function is *the existence of inequality at the workplace* (ineq_ labour) in relation to independent variables: *age* (age), *education level* (educ_ level), *gender* (gender), *economic activity sector* (sector).

The regression obtained for the respondents of both sexes is indicated in the ***Figure 2***.

Equation of regression for respondents of both sexes

Figure 2

Men and women

ineq_labour =	**0.205*gender +**	**0.2994*age**	**+0.358*educ_level**	**+0.294*sector**
σ	(0.0827)	(0.0766)	(0.0688)	(0.0688)
t	[2.4789]	[3.9068]	[5.2049]	[3.4028]

R^2 =0.7479; R^2_{adj} =0.7342; F =94.943

Source: Elaborated by authors based on questionnaire data

The coefficients of the estimated regression function are significant, which is confirmed by the values of Student statistics (t-statistics), greater than 2.81 (except the first one from the gender variable, which shows the significance level of 0.01, therefore does not fall within the significance level of 0.005). At the same time, the standard error values - σ - are small enough.

The coefficients of determination R^2 and adjusted R^2_{adj} indicate the fact that the selected regression function is a good one, expressing the degree of correspondence of the model of statistical dependence between the dependent variable *labour inequality* and independent variables: *gender, age, education level* and *the economic activity sector*.

The values of the estimated coefficients themselves determine the impact of each of the coefficients on the dependent variable, namely: about 35.8% is due to the education level, about 29.9% - to the age, with almost the same percentage – 29.4% influences the sector of activity, and on the last position with only about 20.5% is the influence of the gender variable. It should be noted that the examined sample includes both women and men.

Shows interest the evaluation of the separate opinion of women and men. In order to perform this assessment, it is necessary to divide the questionnaire responses separately for women and men.

Two behavioural functions dependent on the same independent variables will be estimated. The results of the separate assessment of the gender inequality in the workplace of women and men are presented in the ***Figure 3*** and ***Figure 4***.

Equation of regression for male respondents

Figure 3

Men

ineq_labour =	**0.3292*age**	**+0.3599*educ_level**	**+0.4395*sector**
σ	(0.1582)	(0.1298)	(0.1642)
t	[2.0869]	[4.0917]	[1.9076]

R^2=0.8310; R^2_{adj}=0.6897; F =28.19

Source: Elaborated by authors based on questionnaire data

Estimations of the coefficients of the behavioural function for men is dependent on three factors: age, education level and sector of activity and show us the influence of each factor on the issue of men's perceived inequality at work. Approximately 33% are due to age, almost 36% are the influence of the education level and approximately 44% is the impact of the activity sector. Student statistics are between the significance level 0.05 and the significance level 0.005 and standard deviations are moderate. The selected behavioural function shows a high degree of compliance to the statistical data.

Equation of regression for female respondents

Figure 4

Women

ineq_labour = 0.33*age +0.531*educ_level+0.313*sector

σ (0.1582) (0.1298) (0.1642)

t [2.0869] [4.0917] [1.9076]

R^2=0.8310; R^2_{adj}=0.6897; F=28.19

Source: Elaborated by authors based on questionnaire data

Estimations of coefficients of the behavioural function for women dependent on the same three factors as for men, i.e. age, education level and sector of activity show us the influence of each factor on the issue of inequality at work, as perceived by women. About 33% is due to age, 53.1% is the influence of the education level and about 31.3% is the impact of the sector of the activity. Student statistics are between 0.05 - the significance level (for the sector) and the significance level of 0.005 for the remaining variables. Standard deviations are moderate. The selected behavioural function shows a high degree of compliance to statistical data.

Let us calculate a few ratios using the estimated coefficients for women and men of the proposed behavioural functions (***Figure 5***):

Ratios based on estimated female and male coefficients of the proposed behavioural functions

Figure 5

***age women/men*=0.33/0.3292=1.002**

***educ_level women/men*=0.531/0.3599=1.475**

***sector women/men*=0.313/0.4395=0.712**

Source: Elaborated by authors based on questionnaire data

These coefficients show that women and men, involved in the economic activity, are approximately of the same age, but women's level of education is much higher than that of men, and fewer women are involved in the private and mixed (private and state) sectors.

In the ***Figure 6*** a synthetic indicator, showing the ratio between women and men is calculated according to the three indicators calculated above.

The ratio of women to men according to the three calculated indicators

Figure 6

Synthetic indicator = (age women/men + educ_level women/men + sector women/men) / 3 = 1.063

Source: Elaborated by authors

The synthetic indicator 1.063 calculated as average upon three ratios demonstrate the fact that on an average gender inequality on the labour market is at the level of 106.3%. A synthetic indicator reveals how many times the result is more than 1. As consequence, gender problem on the labour market is not so bad if we examine average indicators while at the sector and education level situation is not so good. The conclusion is that the empirical studies are needed to be done at the specific level, not at the average level.

CONCLUSIONS

As a result of the research, according to the elaborated econometric models, calculations were made that determine the degree or depth of gender inequality in the labour market in the Republic of Moldova on the basis of the identified relationship between some different socioeconomic factors in the gender aspect. The same apparatus applied in the research was used for the regression analysis. The novelty of this approach consists in using data from a questionnaire survey, presented in the form of logical variables.

The results of the research showed a high level approximation of the inequality indicator between women and men on the labour market of the data by the behavioural functions depending on four logical variables. According to the results of this research, a group of indicators was calculated, on the basis of which a synthetic indicator was compiled. The issue of gender equality in the employment is a matter of major significance. It is in this segment where the gender disparities are the most pronounced. Despite the fact that men's level of education has been much below that of women, the ratio being equal to 147.5%, they occupy more qualified jobs with higher wages. Women, even with a higher level of qualification, are less requested in the private sector both for senior positions and for better paid positions.

The sector of activity is a decisive factor, particularly the private sector, which for many reasons offer the predilection for men (43.95%), but only 31.3% - for women employed in the private sector, the ratio between women and men being equal to 71.2%. The elaborated econometric models have shown that, for men, the sector of activity (with 43.95%) is a factor with a higher influence on the perceived gender inequality in the workplace than the age, the influence of which is nearly the same (33.0% for women and

32.9% for men), the education level having a nearly similar impact for men (35.99%) as the age. On the contrary, for women the sector (with 31.3%) has the lowest impact on the perceived gender inequality (a little lower than the impact of the age (33.0%)), while education level had the highest influence (53.1%). For both men and women, age has the same impact.

In conclusion, it is necessary for women to be protected in one way or another when they participate in various competitions for valuable functions and positions. In turn, women have to attend interviews and contests on the same footing as men. For the future, it will better also to include in estimated behavioural functions such an indicator as human perception of the women and men on professional abilities when engaging on work position. Also, further research is needed for a quantitative assessment of the role of the analyzed factors by types of economic activities and by size of the companies and, also by the hierarchical positions inside the enterprises or other criterion.

References

1. **ANDREN, D., ANDREN**, 2019, *Th. Occupational gender composition and wages in Romania: from planned equality to market inequality*. In: Working papers in economics, 2007, no. 261, August, Goteborg University – School of Business, Economics and Law. 53 p., [Accessed on 10.06.2019]. Disponible at: http://gupea.ub.gu.se/bitstream/2077/4734/1/gunwpe0261.pdf
2. **ANKER, R., & KNOWLES, J. C**., 1978, *A micro-analysis of female labour force participation in Africa: case study no. 9*. 1978
3. **BAKER, M., FORTIN, N**., 2019, *Occupational gender composition and wages in Canada: 1987-1988*. In: NBER Working Paper, 1999, no. 7371. 41 p., [Accesed on 13.07.2019]. Disponible at: http://www.nber.org/papers/w7371
4. **BECKER, G**., 1967, *Human Capital and the Personal Distribution of Income*, W.S. Woytinsky Lecture No. 1, Ann Arbor, University of Michigan. 1967
5. **BECKER, G. & CHISWICK, B. R**., 1966, *Education and the Distribution of Earnings*. In: American Economic Review, 56(2), pp.358-369. 1966
6. **BLINDER, A. S**., 1973, *Wage Discrimination: Reduced Form and Structural Estimates*. In: The Journal of Human Resources, Vol. 8, No. 4 (Autumn, 1973), pp. 436-455
7. **CHISWICK, B. R**., 1967, **Human Capital and the Distribution of Personal Income**, Unpublished Ph.D. Dissertation, Columbia University. 1967
8. **CHISWICK, B. R**., 1974, *Income Inequality: Regional Analyses Within a Human Capital Framework*, New York: National Bureau of Economic Research. 1974
9. **CHISWICK, B. R. & MINCER, J.**, 2003, Experience and the Distribution of Earnings. University of Illinois at Chicago and IZA Bonn. Discussion Paper No. 847, pp.1-38. 2003
10. **CHISWICK, B. R. & MINCER, J.**, 1972, *Time Series Changes in Personal Income Inequality*. In: Journal of Political Economy, 80 (3) Part 2, S34-S66. 1972
11. **COLESNICOVA, T.**, 2010 *Analiza problemelor de gen în IMM-urile Moldovei*. In: Analele ale Institutului Umanistic Contemporan. Ch.: IUC, ed. 6-7, 2010, - 226 p. ISSN 1857-1271, ISBN 978-9975-9935-6-2, pp.187-191
12. **COLESNICOVA, T.**, 2008, Approaches to estimation of the value for working women's labor at the household in Moldova. In: Revista economică. Sibiu-Chişinău: IEFS, 2008, nr.5-6 (42-43), vol. II, - 188 p. ISSN 1582-6260, pp.136–145

13. **COLESNICOVA, T.**, 2010, *Empirical analysis of social-economic factors which influence the probability to be employed for gender groups in the Republic of Moldova*, In: Progrese în teoria deciziilor economice în condiţii de risc şi incertitudine: Conferinţa XXV ştiinţifică naţională cu participare internaţională din 17 septembrie 2010. România: Institutul de Cercetări Economice şi Sociale „Gh. Zane", Academia Română – Filiala Iaşi. Iaşi: Tehnopress, 2010, vol. XI, -234 p., ISBN 973-702-795-7, pp.117–126
14. **COLESNICOVA, T.**, 2011, *Evaluation of the human capital influence on men and women salaries in the Republic of Moldova*. In: Economie şi sociologie nr.1/2011. Ch.: IEFS, 2011, - 216 p. ISSN 1857-4130, pp.204-213.
15. **COLESNICOVA, T. & NAVAL, E.**, 2018, *Impactul diminuării discriminării de gen asupra creşterii sustenabile în Republica Moldova*. In: Materialele Conferinţei Internaţonale "Economic Growth in Conditions of Globalization: competitiveness, innovation, sustainability, the 13th edition, October 11-12, 2018, Vol. II. pp. 232-236. Chisinau: INCE – ISBN 978-9975-3202-9-0
16. **DUMOUCHEL, W. H., & DUNCAN, G. J.**, 1983, *Using sample survey weights in multiple regression analyses of stratified samples*. In: Journal of the American Statistical Association, 78(383), pp. 535-543. 1983
17. **DUNCAN, O. D., & DUNCAN, B.**, 1955, *A methodological analysis of segregation indexes*. In: American sociological review, 20(2), pp. 210-217. 1955
18. **JOLLIFFE, D.**, 2002, *The gender wage gap in Bulgaria: A semiparametric estimation of discrimination*. In: Journal of Comparative Economics, June 2002, vol. 30, Issue 2, pp. 275-295
19. **MINCER, J., & POLACHEK, S.**, 1974, *Family investments in human capital: Earnings of women*. In: Journal of Political Economy, 82(2, Part 2), S76-S108. 1974
20. **NAVAL, E.**, 2018, *Earnings evaluation of the Moldovan budget employees*. In: Proceedings of the International Symposium "Experience. Knowledge. Contemporary Challenges 3rd Edition "Romania in the Year of the Centenary. The European and Global Socio-Economic Context", December 13th-14th, 2018, Bucharest, pp.106-115
21. **OAXACA, R.L., & RANSOM, M.R.**, 1994, *On discrimination and the decomposition of wage differentials*. 1994
22. **OGLOBLIN, C.**, 1999, *The gender earnings differential in the Russian transition economy*. In: Industrial and Labor Relations Review, July 1999, vol. 52, no. 4
23. **PASTORE, F., VERESHCHAGINA, A.**, 2004, Private returns to human capital over transition: A case study of Belarus. In: IZA DP, November 2004, no. 1409. 40 p., [Accessed on 13.07.2019]. Disponible at: http://ftp.iza.org/dp1409.pdf
24. **PASTORE, F., VERESHCHAGINA, A.**, 2019, *The determinants of female labour supply in Belarus*. In: IZA DP, April 2008, no. 3457. 26 p., [Accessed on 01.07.2019]. Disponible at: http://ftp.iza.org/dp3457.pdf
25. **SCHULTZ, T. W.**, 1961, *Investment in human capital*. In: The American economic review, pp. 1-17. 1961
26. THE GLOBAL GENDER GAP REPORT 2018. World Economic Forum. [Accessed on 13.07.2019]. Disponible at: https://www.weforum.org/reports/the-global-gender-gap-report-2018
27. **THUROW, L. C.**, 1975, Generating inequality. Basic books. 1975
28. **АБАЗИЕВА, К**., 2009, *Работать или не работать после пенсии: гендерные аспекты выбора*. In: TerraEconomicus - Экономический вестник Ростовского государственного университета, 2009, том 7, №1, с. 67-74
29. **БАНДЮКОВА, Т.**, 2019, *Случайная занятость в России: численность, состав, мобильность. Центр трудовых исследований ГУ ВШЭ*. [Accessed on 13.07.2019]. Disponible at: http://www.budgetrf.ru/Publications/Analysis/arett/arett200702081014/arett200702081014_p_007.htm

30. **КОЛЕСНИКОВА, Т.**, 2009, *Гендерное вертикальное неравенство на рынке труда Молдовы*. In: Economie şi sociologie nr.2, 2009. Ch.: IEFS, 2009, -180 p. ISSN 1857-4130, pp.151–159
31. **КОЛЕСНИКОВА, Т**. Опыт регулирования гендерного равенства в сфере занятости в Республике Молдова. In: Analele Institutului de Economie, Finanţe şi Statistică. Ch.: IEFS, 2011, -184 p. ISSN 1857-3630, ISBN 978-9975-4176-4-8, pp.176-179
32. **ОЩЕПКОВ, А.**, 2006, Гендерные различия в оплате труда в России. In: Экономический журнал ВШЭ, 2006, №4, с. 590-619
33. **МАЛЬЦЕВА, И., РОЩИН,** С., 2007, Гендерная сегрегация и трудовая мобильность на российском рынке труда. 2-е изд., Москва: ГУ ВШЭ, 2007. 295 с.
34. **РОЩИН, С., ГОРЕЛКИНА, О.**, 2004, Гендерные различия в заработной плате: микроэкономический анализ факторов и тенденций. In: Гендерное неравенство в современной России сквозь призму статистики. М.: Едиториал, УРСС, 2004, с.130-146
35. **РОЩИН, С.**, 2019, *Предложение труда в России: микроэкономический анализ экономической активности населения*. Препринт WP3/2003/02/. М.: ГУВШЭ, 2003. 58 с., [Accessed on 12.07.2019]. Disponible at: http://demoscope.ru/weekly/2003/0109/biblio06.php
36. **СТУКЕН, Т.**, 2007, Дискриминация в заработной плате на Российском рынке труда. In: Известия ИГЭА, 2007, №6 (56), с.127-130

Decomposing Income Inequality by Education in Romania

Ionela-Roxana PETCU – Assistant Lecturer, PhD. (roxana.glavan10@yahoo.com)
Bucharest University of Economic Studies, Bucharest, Romania

ABSTRACT

As long as education is not consolidated a sustainable society no longer grows. All components of education are of most importance when analysing their contributions to wage inequalities in Romania. Decomposition analysis is an essential tool that provides insights into the social dimension of income inequality. The current study aims to explore the role of education and its influence in the decomposition of income inequality in Romania, by using both multivariate regression-based and Theil decomposition. First, to measure the inequality in the distribution of individual income, Gini index, two Theil indices and Variance of log are used. Then, a multivariate regression-based decomposition function is applied on the dataset. As a well-known econometric measure of investigating economic inequality, due to its properties, the Theil index is employed to decompose both overall income inequality in Romania into the between and within components and furthermore by Education. The data subset relies on the UK Data Service collection - European Working Conditions Survey for 2015. The analysis reveals consistent results for the computed indices. The Theil decomposition analysis results regarding the income inequality in Romania indicate that by the attained level of education the main contributing groups to the inequality are people with medium education and high education. The group of people with high education contributed with 40% to the inequality while for Theil L. decomposition the same group contributed with 28.72%. These findings are also consistent with the fact that a higher level of education is better rewarded.

Keywords: *Romania, Theil index, entropy, group inequality, decomposition*
JEL Classification: *C1, J7*

1. INTRODUCTION

Access to equal income opportunities is an essential dimension of residential differentiation in the context of population dynamic and social change, presenting new global challenges. The interplay between population growth, family structure and education will create potential income imbalances for the Europe population over the coming decades (Guerin, 2013).

Many research studies conclude that the rise in income inequality is mainly due to the increase of two driving factors: earnings inequality and the number of single-mother families (Cholezas and Panos, 2007; Kollmeyer, 2013). Earnings inequality increases as a reward for educational certifications, work experience and digital skills, required for the new jobs

where new technologies, automatization, robotics and communication tend to replace traditional working (Cholezas and Panos, 2007; Castellano, 2017). In respect to the second factor, the observed increase in single-mother families has altered the income distribution (Kollmeyer, 2013). These social changes in family structure heightens income inequality in the long run. Findings across countries reveal that single-parent households are at greater risk of poverty than households with regular families (Brady and Burroway, 2012; Maldonado and Nieuwenhuis, 2015). Moreover, Maldonado and Nieuwenhuis (2015) point out that single mothers are more likely to be poor than single fathers. In the past decade, European economies experienced changes in labour market conditions (Huber, 2007) and in population patterns (Botev, 2012; Gerőházi et al., 2011) due to the accelerating economic requirements and due to the global Great Recession started in autumn 2007 (Cho, and Newhouse, 2013). According to Matysiak et al. (2021), it is seen a fertility decline strongly corelated to unemployment increase. Therefore, as the crisis impacts the economic performance, youth unemployment rate increases (Cho, and Newhouse, 2013; OECD, 2013), with mixed results for young adult workforce in Europe (O'Higgins, 2012). Crises in the past millennia had a periodic occurrence affecting independently different parts of the world. With the increase in the interdependence between the major global regions the resiliency to crises shifted to the lower end, requiring increased efforts. Furthermore, the emergency measures and lockdowns based on social distancing to contain the spread of COVID-19, contributed to the rise in the wage inequality and poverty in all European countries (Palomino et al., 2020) with grater increases in Eastern and Southern Europe. In this respect, the authors highlight an increase in the average of Gini coefficient and underline that lockdowns contributed to the rise in inequalities both between and within countries.

Paulus and Tasseva (2017) emphasize that changes in population characteristics and market income increase poverty and inequality in EU countries. In addition, their results indicate that one of the largest inequality-reducing policy effects is in Romania. There are however still limited studies in the inequality of income distribution and its decomposition in Romania. One of the most popular inequality indexes, from the class of generalized entropy indexes, developed by the economist Henri Theil (1967) measures inequality of population between groups. Andrei et al. (2017) conducted a study on the inequality of income distribution by the source of income at the country level in Romania. Their results showed significant differences in the income distribution of the three sources of income: wages, capital income, and other sources. Also, using income tax data, Oancea et al. (2018) highlight that the

capital income is Pareto distributed in the upper tail. Moreover, decomposition analysis studies for Romania identify education, labour market status (Molnar, 2010) and remittances (Zamfir et al., 2010) as divers for income inequalities. In their study, Militaru and Stanila (2015) investigate the income inequality determinants in Romania. Education and status in employment are amongst the main determinants of income variability between households. However, the increased prevalence of single-mother families is only one of the major changes in the structure of the family over the years. It is well known that in Romania, mothers tend to have lower earnings compare to childless women (Glăvan, 2018). Such changes in the dynamic of the population add burden to the overall income inequalities between countries.

The aim of this research is to explore the role of education and its influence in the decomposition of income inequality in Romania. The present study adds knowledge in the literature by analysing the decomposition of individual net monthly income from main paid job using data for 2015 in Romania. In particular, a multivariate regression-based decomposition method is used to estimate the inequality contribution of variables included in the model. Socio-demographic factors, such as area of residency, age and education level are analysed for their impact on the income inequality. Furthermore, Theil indices are introduced through their belonging in the general entropy family and their decomposition into two components, one referring to the between groups and the other to the within groups part.

2. METHODOLOGY AND DATA

In this modelling study, microdata from UK Data Service collection (Eurofound, 2017), is used to examine the differences in income levels and inequality between individuals in Romania. A selection from the original dataset is carried out and several variables are computed mapping the employed work force, between 18 to 63 years of age. This selection includes the following variables depicted in Table 1.

Although there are a multitude studies following techniques that measure inequality, such as Lorenz curves, Blinder-Oaxaca decomposition (Glăvan, 2016), Generalized Entropy indices (Andrei et al.,2017), Atkinson indices (Atkinson, 2005; Gradín, 2020) and other scalar transformations. The Generalized Entropy decompositions class, and more specifically through Theil L. and Theil T. indexes, are of interest in the present study due to their straight interpretation and their additively decomposable measures (Shorrocks 1980) for which inequality is the sum of inequality between groups and the weighted sum of inequality within groups.

The widely used Gini coefficient proposed by Corrado Gini in 1912 ranks income distribution on a scale theoretically from 0 to 1, where 0 denote total equality of income and 1 total inequality. In respect to the General Entropy Indexes - $GEI(\alpha)$, two popular members of its class are: $GEI(0)$ -Theil L. or mean logarithmic deviation and $GEI(1)$ or Theil T. index. In addition to these measures, the variance of log income is another useful tool to measure income inequality.

Defined variables

Table 1

Variable name	Definition
wage	Wage in national currency [RON]
weights	Population weights
age_cat	Age (categorical), with levels: <25, 25-34, 35-44, 45-54, >=55
education_cat	Number of years in school, factor with levels: Low education, Medium education, High education
area_cat	Area, factor with levels: urban, rural
sex	Factor with levels: woman, man

The Theil index is an inequality measure, and its principle is related to the generalised entropy indices to calculate inequality or difference (Bellù and Liberati, 2006). Compared to the widely used Gini coefficient, the Theil index is decomposable, and therefore can be linked on the source of the inequalities. When the parameter α, driving the weights given to distances between cases in different parts of a distribution for the generalized entropy, is 1, it is obtained Theil's first measure or commonly known "Theil T.". When α takes the value of 0, Theil's second measure of inequality is obtained, also known as "Theil L." or the mean log deviation. Furthermore, with α taking the value of 2, the measure is called coefficient of variation.

Partitioning a population into $K \geq 1$ disjoint groups with *k* referring to *k*-th group, each group of population size n^k and mean income (wage) μ^k has a distribution of $y^k = \left(y_1^k, \ldots, y_{n^k}^k\right)$ and *i* identifies the *i*-th individual of vector *y*. According to Shorrocks (1980) the Generalised Entropy family $I_\alpha(y)$ is written as:

$$GEI(\alpha) = I_\alpha(y) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu}\right)^\alpha - 1\right] \quad (1)$$

with $w_{I_\alpha}^k$ as index specific weights being a function of group population shares n^k/n and relative wages μ^k/μ.

$$w_{I_\alpha}^k = \frac{n^k}{n}\left(\frac{\mu^k}{\mu}\right)^\alpha \quad (2)$$

With $\alpha = 0$, Mean log deviation or Theil L. index (Theil, 1967) is defined as:

$$I_0 = T_L = \frac{1}{n}\sum_{i=1}^{n} ln\frac{\mu}{y_i} \qquad (3)$$

with its index specific weights $w_{I_0}^k = \frac{n^k}{n}$.

For $\alpha = 1$, it should be noted that the Theil T. index or the first measure (Theil, 1967) is defined as:

$$I_1 = T_T = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{\mu} ln\frac{y_i}{\mu} \qquad (4)$$

with its index specific weights $w_{I_1}^k = \frac{n^k}{n}\frac{\mu^k}{\mu}$, where *n* is the population size, y_i is the income or wage of the individual *i*, μ is the overall mean wage of the population.

Both Theil L. and Theil T. indices can be decomposed into 2 parts, between-group $(T_b \; or \; L_b)$, and within-group $(T_w \; or \; L_w)$, components. The first term, $(T_b \; or \; L_b)$, captures the inequality of distribution of population incomes due to the differences of income distribution existing between the disjoint groups, while the second term, $(T_w \; or \; L_w)$, measures the inequality of income distribution as a result of the differences in the distribution of incomes within the disjoint groups of population.

Equations (5) and (6), further reflect both Theil T. and Theil L. decomposition (Akita, 2003; Liao, 2019):

$$T_T = T_b + T_w = \sum_{k=1}^{K} w_k ln\frac{\mu_k}{\mu} + \sum_{k=1}^{K} w_k \sum_{n=1}^{n_k} w_{ik} ln\frac{y_{ik}}{\mu_k} \qquad (5)$$

where μ_k or μ^k is the mean income of disjoint group *k*, w_k is the *k*-th group's income share represented as a proportion of sample or population total income, w_{ik} is the income share of the *i*-th individual in the *k*-th group, and y_{ik} is the *i*-th individual's income in group *k*.

$$T_L = L_T = L_b + L_w = \sum_{k=1}^{K} n_k ln\frac{\mu}{\mu_k} + \sum_{k=1}^{K} n_k \sum_{n=1}^{n_k} n_{ik} ln\frac{\mu_k}{x_{ik}} \qquad (6)$$

where n_k or n^k represents *k*-th group's group size proportion of the overall sample and respectively, n_{ik} is the proportion of the *i*-th case out of the *k*-th group.

Generalised Entropy family indices are more sensitive to differences in income shares among the poor or among the rich depending on the α parameter defining the $GEI(\alpha)$. The lower the α parameter value the more sensitive the index is to the differences at the bottom of the distribution (Jenkins and Van Kerm, 2008). Among the two indices Theil L. is known to be more sensitive to income differences in the lower end of the distribution with better decomposability properties, while Theil T. is more sensitive to the differences in the top end of the distribution. Therefore, a smaller Theil index denotes a smaller degree of inequality, and vice versa (Gradín, 2020).

In the present research the income inequality is computed using four different popular indexes: Gini, Theil L., Theil T. and variance of log income. To further inspect income inequality a regression-based decomposition method is applied following the methodology covered by Fields (2003) and Brewer and Wren-Lewis (2016).

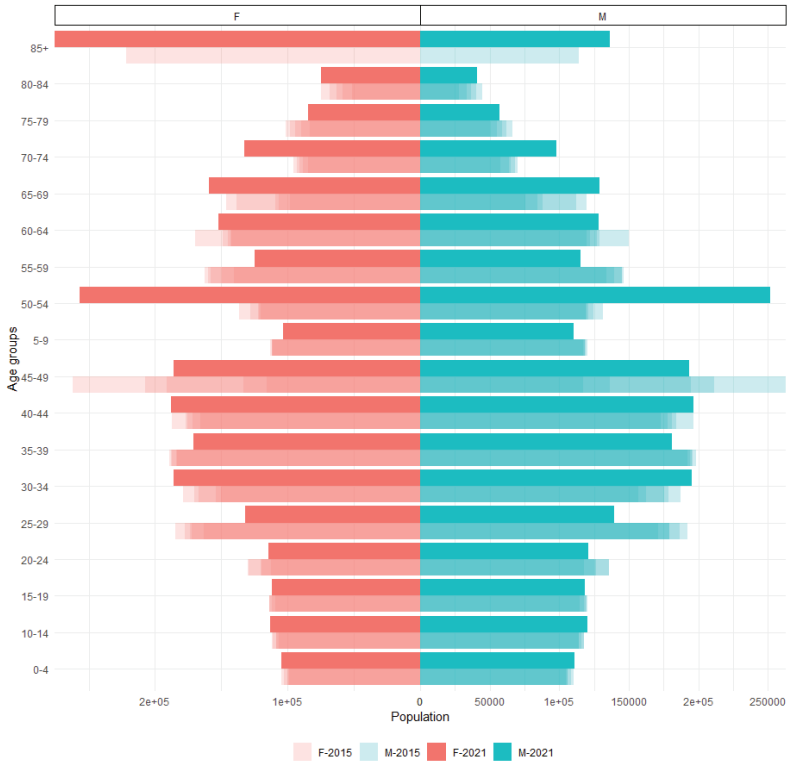
The analysis is performed using R software packages "dineq" (Schulenberg, 2018) and "IC2" (Plat, 2012). The regression-based decomposition method incorporates multiple variables to compute each variable's contribution to the wage inequality. Socio-economic along with demographic and territorial characteristics (such as. education, age, area and sex) of individuals are added to the function. A key feature of this method is its usage of value logarithmic transformation in the dependent variable, thus, only non-zero values are kept. The decomposition of both Theil L. and Theil T. indices follow the mathematical framework covered by Cowell (2000) and Elbers et al. (2005).

3. RESULTS

Using data provided by Romania NIS, Figure 1 shows that the share of population aged 65 and over was rising in 2021 compared to 2015, while the population under 25, respectively the working-age population is decreasing.

Population pyramid by age group and sex (in thousands), Romania in 2015 and 2021

Figure 1



Source: designed by the author based on data provided by the Romania NIS online database

Furthermore, the old-age dependency ratio is rising from 25.2 percent in 2015 to estimated 29.7 percent in 2021 (Eurostat, 2022a). Under the base line scenario, the old-aged dependency ratio is projected to rise at 57.8 percent by 2100 (Eurostat, 2022b). These changes in the structure of population foretell abrupt changes in the distribution of wages, thus increasing the burden of existing inequalities.

Table 1 illustrates, based on individual monthly income, the results for the inequality measured by four indices.

Results of the of 4 inequality measures

Table 1

Inequality measures	Gini	Theil L. (Mean log dev.)	Theil T.	Variance of log
Values	0.2640	0.1138	0.1296	0.2030

Source: designed by the author based on EWCS 2015 selection data

Considering the case when individuals are grouped according to several variables such as age, sex, area, educational level. Table 2 presents the decomposition of the inequality into the selected variable with education highlighting its main contribution.

The (relative) decomposition of the inequality into the different variables

Table 2

Variables	Age	Education	Area	Sex	Residual
Values	0.0098	0.1977	0.0596	0.0847	0.6482

Source: designed by the author based on EWCS 2015 selection data

To conclude this section, Table 3 helps in understanding the Ordinary Least Squares (OLS) estimates from the fitted regression equation used to decompose wage inequality.

OLS regression estimation used in the decomposition of inequality

Table 3

Variable	Estimate	Standard Error
intercept	7.04023***	0.06780
Age 25-34	0.14851*	0.05983
Age 35-44	0.18532**	0.05768
Age 45-54	0.17457**	0.05831
Age >=55	0.14310*	0.06596
Low education	-0.64735***	0.06704
Medium education	-0.40929***	0.03515
Urban area	0.20748***	0.03038
Masculine	0.27938***	0.02833

*Source: designed by the author based on EWCS 2015 selection data; Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Next, with equations (5) and (6) the overall degree of the inequality of wage distribution is explored through its two components, the within-group inequality and the difference between the groups of population. Furthermore,

computing both Theil T. and Theil L. decomposition, the results of the wage distribution for the three disjoint groups categorized by the educational level are presented in Table 4.

Theil indices and decomposition by education for Romania

Table 4

Education	Theil T.	Theil L.
Within-Educational Groups		
Low education	0.06	0.06
(% Contrib.)	(1.70)	(2.97)
Medium education	0.08	0.07
(% Contrib.)	(35.92)	(44.39)
High education	0.14	0.12
(% Contrib.)	(40.48)	(28.72)
Decomposition		
Within-Educational Groups	0.10	0.09
(% Contrib.)	(78.10)	(76.08)
Between-Educational Groups	0.03	0.03
(% Contrib.)	(21.90)	(23.92)
Total	0.13	0.11

Source: designed by the author based on EWCS 2015 selection data; % Contrib. is the share contribution of each group to the overall inequality.

Theil decomposition analysis by the attained level of education shows that the main contributing groups to the inequality are people with medium education and high education. The group of people with high education contributed with 40% to the inequality while for Theil L. decomposition the same group contributed with 28.72%. This contribution occurs through within-group component which contributes with 78.10% to the total inequality for Theil T. and 76.08% for Theil L. Educational groups with low education, on the contrary, contributed marginally to inequality with 1.7%.

4. CONCLUSIONS

In this research income inequality in Romania is computed based on a subset of microdata from UK Data Service collection. New perspectives linking the importance of education is revealed through analysing the decomposition of income inequality in the country.

First, to measure overall wage inequality four well known indices, Gini, Theil L., Theil T and Variance of log are computed. Second, the data are analysed using a multivariate regression-based decomposition method. As a result, education has the main contribution to the income inequality in Romania.

Next, using entropy class techniques, an individual-based decomposition of inequality applying Theil T and Theil L measures is performed. The two parts of inequality given by the between-group and within-group components are revealed. The first component, T_b with abs. value 0.03 and contribution 21.90% or L_b with same abs. value and 23.90% contribution, known also as the between-groups inequality term, captures the inequality of distribution of population incomes due to the differences of income distribution existing between the disjoint groups. The second component, T_w with abs. value 0.10 and contribution 78.10% or L_b with abs. value 0.09 and 76.08% contribution, noted as within-group term, measures the inequality of income distribution as a result of the differences in the distribution of incomes within the disjoint groups of population. Although the results for Theil T. and Theil L. are not far apart, it should be noted that in particular, the Theil's T. is more sensitive for change in the upper tail, while Theil's L. is more sensitive to changes that affect the lower end of the distribution.

In a next study more focus will be on investigating new data sources and larger series. An interesting further approach is to decompose the change of the mean log deviation between two data sets by population subgroups and examine its inequality components, changes between and within groups and the changes in the relative size of the groups. Moreover, the data will be re-examined using other entropy techniques and RIF decomposition as generalization of the Blinder-Oaxaca methodology to estimate the marginal effect of covariates on income distribution.

5. References

1. **Akita, T.**, 2003, *Decomposing Regional Income Inequality in China and Indonesia Using Two-stage Nested Theil Decomposition Method*. Annuals of Regional Science 37:55-77.
2. **Andrei, T., Oancea, B., Richmond, P., Dhesi, G., Herteliu, C.**, 2017, *Decomposition of the Inequality of Income Distribution by Income Types-Application for Romania*, *Entropy*, 19(9), 430.
3. **Atkinson, A.B.**, 2005, *Comparing the distribution of top incomes across countries*. J. Eur. Econ. Assoc. 3(2–3), 393–401.
4. **Bellù, LG, Liberati, P**, 2006, *Describing income inequality: Theil index and entropy class indexes*. Food and Agriculture Organization of the United Nations.
5. **Oancea, B., Pirjol, D., Andrei, T.**, 2018, *A Pareto upper tail for capital income distribution*, Physica A: Statistical Mechanics and its Applications, Volume 492, Pages 403-417, ISSN 0378-4371.
6. **Botev, N.**, 2012, *Population ageing in Central and Eastern Europe and its demographic and social context*. European Journal of Ageing, 9(1): 69–79.
7. **Brady, D. and Burroway, R.**, 2012, *Targeting, universalism and single mother poverty: A multilevel analysis across 18 affluent democracies*. Demography, 49, 719–746.
8. **Brewer M. and L. Wren-Lewis**, 2016, *Accounting for Changes in Income Inequality: Decomposition Analyses for the UK*, 1978–2008. Oxford Bulletin of economics and statistics, 78 (3), p. 289-322.

9. **Castellano, R., Musella, G. and Punzo, G**., 2017, *Structure of the labour market and wage inequality: evidence from European countries*. Qual Quant 51, 2191–2218.
10. **Cho, Y. and Newhouse, D.**, 2013, *How Did the Great Recession Affect Different Types of Workers? Evidence from 17 Middle-Income Countries*. World Development 41(0): 31–50.
11. **Cholezas, I. and Panos, T.**, 2007, *Earnings inequality in Europe: structure and patterns of inter-temporal changes*, IZA Discussion Paper, n. 2636.
12. **Cowell F.A.**, 2000, *"Measurement of Inequality"*. In Atkinson A.B., Bourguignon F. (Eds.) Handbook of Income Distribution. Elsevier, Vol. 1, pp. 87-166.
13. **Elbers C., Lanjouw P., Mistiaen J.A., Ozler B**., 2005, *Re-Interpreting Sub-Group Inequality Decompositions*. World Bank, World Bank Policy Research Working Paper 3687, 42 p.
14. **Eurofound**, 2017, *European Foundation for the Improvement of Living and Working Conditions*, European Working Conditions Survey, 2015. [data collection]. 4th Edition. UK Data Service. SN: 8098, http://doi.org/10.5255/UKDA-SN-8098-4.
15. **Eurostat**, 2022a, *Old-age-dependency ratio*, https://ec.europa.eu/eurostat/databrowser/view/tps00198/default/table?lang=en, accessed 28 March 2022.
16. **Eurostat**, 2022b, *Old-age dependency ratio increasing in the EU*, https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200713-1, accessed 28 March 2022.
17. **Fields, G. S.**, 2003, *Accounting for income inequality and its change: a new method, with application to the distribution of earnings in the United States*, Research in Labor Economics, 22, p. 1–38.
18. **Gerőházi, É., Hegedüs, J., Szemzö, H., Tosics, I., Tomay, K., Gere, L**., 2011, *The impact of European demographic trends on regional and urban development. Synthesis Report*, Budapest.
19. **Glăvan, IR. et al.**, 2016, *Application of Oaxaca decomposition in R. Case study: Roma Minority in Romania*, 28th IBIMA Confference.
20. **Glăvan, IR.**, 2018, *Parenthood Impact on Wage Gap in Romania*, Romanian Statistical Review nr. 4 / 2018, pages 110-120.
21. **Gradín, C.**, 2020, *Quantifying the Contribution of a Subpopulation to Inequality: An Application to Mozambique*. Journal of Economic Inequality, 18: 391–419.
22. **Guerin, B.**, 2013, *Demography & Inequality: How Europe's changing population will impact on income inequality*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR183.html.
23. **Huber, P.**, 2007, *Regional labour market developments in transition: a survey of the empirical literature*. The European journal of comparative economics, 4: 263–298.
24. **Jenkins, S.P. and Van Kerm, P.**, 2008, *"The Measurement of Economic Inequality"* in Salverda, W., Nolan, B. and Smeeding, T.M. (eds), The Oxford Handbook of Economic Inequality, Oxford University Press, Oxford.
25. **Kollmeyer, C.**, 2013, *Family Structure, Female Employment, and National Income Inequality: A Cross-National Study of 16 Western Countries*, European Sociological Review, Volume 29, Issue 4, Pages 816–827.
26. **Liao, T.F.**, 2019, *Individual Components of Three Inequality Measures for Analyzing Shapes of Inequality*. Sociological Methods & Research, first published on September 26, as https://doi.org/10.1177/0049124119875961.
27. **Maldonado, L.C. and Nieuwenhuis, R.**, 2015, *Family policies and single parent poverty in 18 OECD countries, 1978–2008*, Community, Work & Family, 18:4, 395-415.
28. **Matysiak, A., Sobotka, T. and Vignoli, D**., 2021, *The Great Recession and Fertility in Europe: A Sub-national Analysis*. Eur J Population 37, 29–64.

29. **Militaru, E., Stanila, L.**, 2015, *Income Variability in Romania: Decomposing Income Inequality by Household Characteristics, Procedia Economics and Finance*, Volume 26, Pages 227-233, ISSN 2212-5671.
30. **Molnar, M.**, 2010, **Income distribution in Romania Molnar**. MPRA Paper No. 30062.
31. **O'Higgins, N.**, 2012, *This time it's different? Youth labour markets during "the great recession."* Comparative Economic Studies, 54, 395-412.
32. **OECD**, 2013, *Revenue Statistics 2013*. OECD Publishing. https://www.oecd-ilibrary.org/taxation/revenue-statistics-2013_rev_stats-2013-en-fr, accessed 5 April 2022.
33. **Palomino, J.C., Rodríguez, J.G., Sebastian, R.**, 2020, *Wage inequality and poverty effects of lockdown and social distancing in Europe*, European Economic Review, Volume 129, 103564, ISSN 0014-2921.
34. **Paulus, A. and Tasseva, I.**, 2017, *Decomposition of changes in the EU income distribution in 2007-2011*, EUROMOD Working Paper, No. EM9/17, University of Essex, Institute for Social and Economic Research (ISER), Colchester.
35. **Plat, D.**, 2012, *IC2: Inequality and Concentration Indices and Curves. R package version 1.0-1*. https://CRAN.R-project.org/package=IC2.
36. **SCHULENBERG, R.**, 2018, *Package dineq. Vienna: Comprehensive R Archive Network*. https://cran.r-project.org/web/packages/dineq/index.html, accessed 4 August 2022.
37. **Shorrocks, A.F.**, 1980, *The class of additively decomposable inequality measures*. Econometrica. 48(3), 613–625.
38. **Theil, H.**, 1967, *"Economics and Information Theory"*. Chicago: Rand McNally and Company.
39. **Zamfir, A.M., Mocanu, C., Militaru, E. and Pirciog, S**., 2010, *Impact of Remittances on Income Inequalities in Romania*, In Schuerkens, U. (Eds.) "Globalization and Transformations of Social Inequality", Routledge Taylor& Francis Group, New York, London, pp. 58-75.