Ministry of Education and Science of Ukraine
Sumy State University
Education and Research Institute Business, Economics and Management
Department of Economic Cybernetics

BACHELOR'S QUALIFICATION WORK

on the topic «ECONOMIC AND MATHEMATICAL MODELING THE IMPACT OF DIGITALIZATION ON INCOME INEQUALITY»

Student IV course, group АБ-71a
speciality 051 «Economics»
(Economic Cybernetics)
Basanets Sofiia Romanivna
Supervisor PhD, as. prof. Bozhenko V.V.

Sumy – 2021

ABSTRACT

bachelor's qualification work on the topic

«ECONOMIC AND MATHEMATICAL MODELING THE IMPACT OF DIGITALIZATION ON INCOME INEQUALITY»

Student: Sofiia Basanets

The relevance of the topic chosen for research is determined by the fact that one of the key factors that determines the trajectory of the national economy is digitalization. New digital technologies are often viewed as a tipping point in terms of organizing production within value chains. This is because they give intangible assets a more prominent role in generating income. Digitalization also leads to the emergence of new types of jobs and employment, to a change in the nature and conditions of work, a change in professional requirements for the level of qualifications and is reflected in the functioning of labor markets, as well as in the international division of labor.

The purpose of the qualification work is to develop an economic and mathematical model to estimate the impact of digitalization on income distribution among the population.

The object of research is the processes of modeling in the system of income management.

The subject of research is mathematical methods and techniques for modeling the impact of digitalization on income inequality.

The objectives of the study are: consideration of the determinants of the spread of digital technologies and their transformational impact on the income distribution system; systematization of existing approaches to modeling the relationship between the development of digital technologies and human capital and its value; statement of the problem of modeling the impact of digitalization on income inequality; practical approbation of the econometric model; economic interpretation of the

obtained results; formation of proposals for the results of the proposed scientific and methodological approach.

To achieve the goal and objectives of the study, the following research methods were used: analysis, synthesis, tree clustering, k-means clustering, regression models of panel data.

The main scientific result of the qualifying bachelor's thesis is to improve the scientific and methodological approach to assessing the impact of digitalization on the state of income diversification based on the construction of regression models of panel data, which allowed to establish the relationship between selected processes. The obtained results can be used by public authorities, public organizations.

The results of approbation of the main provisions of the qualification work were considered at the International scientific-practical conference "Strategic guidelines for economic development, accounting, finance and law" (June 9, 2021, Poltava). The qualification work was performed within the research topic "Reforming the lifelong learning system in Ukraine to prevent labor emigration: a cooperative model of institutional partnership" (state registration number 0120U102001), funded by the State Budget of Ukraine.

Key words: digitization, income inequality, panel data, clustering.

The content of the qualification work is presented on 55 pages. The list of used sources from 40 names, placed on 4 pages. The work contains 9 tables, 7 figures, and 3 appendices.

Year of qualification work - 2021.

Year of work defense - 2021.

Ministry of Education and Science of Ukraine
Sumy State University
Education and Research Institute Business, Economics and Management
Department of Economic Cybernetics

APPROVED
Head of department
Dc.Sc., professor
____________ O.V.Kuzmenko
“__” ____________ 2021

TASKS FOR BACHELOR’S QUALIFICATION WORK
051 «Economics (Economic Cybernetics)»
student IV course, group АБ-71а
Basanets Sofia Romanivna

1. Topic «Economic and mathematical modeling the impact of digitalization on income inequality», approved by order of the University of 15th of March 2021 № 0382-III.
2. Deadline for submission of completed work by the student – 18th of June 2021.
3. The aim of the qualification work – to develop an economic and mathematical model to estimate the impact of digitalization on income distribution among the population.
4. The object of research is the processes of modeling in the system of income management.
5. The subject of research is mathematical methods and techniques for modeling the impact of digitalization on income inequality.
6. Qualification work is performed on the materials of the State Statistics Service of Ukraine, the World Bank, Eurostat, analytical reviews and scientific publications of domestic and foreign authors devoted to the study of income inequality.
7. Indicative plan of qualification work, terms of submission of sections to the head and the maintenance of tasks for performance of the set purpose
Section 1. Theoretical and methodological principles of modeling the influence of digitalization on income inequality – 7th of May 2021
In section 1 it is necessary to consider the essence of income inequality, to identify the impact of digitalization on income inequality, to analyze existing approaches and methods for modeling the impact of digital technologies on income distribution among the population, to develop their own economic and mathematical model.

Section 2. Practical implementation of the model, checking its adequacy and suggestions for its use – 21th of May 2021

In section 2 it is expedient to carry out practical approbation of the developed econometric model, to check its adequacy and to offer recommendations on results of calculations.

8. Work consultations

Section	Surname, initials and position consultant	Signiture, data	
		Task issued	task accepted
1			
2			

9. Assignment date of the task: 1st of March 2021

Supervisor of qualification work _____________ V.V.Bozhenko

Student ______________ S.R.Basanets

Plan

INTRODUCTION

One of the key factors that determines the trajectory of the national economy is digitalization. New digital technologies are often viewed as a tipping point in terms of organizing production within value chains. This is because they give intangible assets a more prominent role in generating income. Digitalization also leads to the emergence of new types of jobs and employment, to a change in the nature and conditions of work, a change in professional requirements for the level of qualifications and is reflected in the functioning of labor markets, as well as in the international division of labor. Thus, the paper analyzes the impact of digitalization on income inequality.

The purpose of the qualification work is to develop an economic and mathematical model to estimate the impact of digitalization on income distribution among the population.

The object of research is the processes of modeling in the system of income management.

The subject of research is mathematical methods and techniques for modeling the impact of digitalization on income inequality.

The objectives of the study are:

– consider the determinants of the spread of digital technologies and their transformational impact on income distribution;

– to systematize the existing approaches to modeling the connection between the development of digital technologies and the development of human capital;

– to set the problem of modeling the impact of digitalization on income inequality;

– to carry out practical approbation of the econometric model;

– provide an economic interpretation of the results;

– to form proposals for further use of the proposed scientific and methodological approach.

To achieve the goal and objectives of the study, the following research methods were used: analysis, synthesis, tree clustering, k-means clustering, regression models of panel data.

The main scientific result of the qualifying bachelor's thesis is to improve the scientific and methodological approach to assessing the impact of digitalization on the state of income diversification based on the construction of regression models of panel data, which allowed to establish the relationship between selected processes. The obtained results can be used by public authorities, public organizations.

The information base of the study is laws and regulations, official statistics of Eurostat, the World Bank, analytical reports and scientific publications on digitalization and welfare.

SECTION 1. THEORETICAL AND METHODOLOGICAL PRINCIPLES OF MODELING THE INFLUENCE OF DIGITALIZATION ON INCOME INEQUALITY

1.1. Analysis of income inequality in the digital context

Today, solving income inequality among different segments of the population is a problem not only in developing countries but also in developed countries. The main factors of differences in well-being are globalization, labor market reforms and technological progress (improving the mechanism of resource allocation and efficiency, rapid introduction of technologies and innovative technologies in various spheres of society, strengthening the financialization of the economy, intellectualization of production, and improving the quality of knowledge and skills).

Income inequality means a significant disproportion in income distribution between different members: individuals, groups, populations, or countries [18]. Income inequality should be considered as a complex dynamic system, which is influenced by different factors and spheres of influence. The level of unequal income distribution in a particular country depends on various factors, the main of which, in our opinion, include the intensification of globalization processes, rapid implementation of innovative technologies in various spheres of society, growing financialization of the economy, development of financial inclusion (gaining full access to key financial services in the country), as well as significant imbalances in the levels of social protection and security.

In practice, income inequality among individuals is measured by five indicators [20]:

– the Gini coefficient is based on the comparison of cumulative proportions of the population against cumulative proportions of income they receive, and it ranges between 0 in the case of perfect equality and 100 in the case of perfect inequality;

– S80/S20 is the ratio of the average income of the 20% richest to the 20% poorest;

– P90/P10 is the ratio of the upper bound value of the ninth decile (i.e. the 10% of people with the highest income) to that of the first decile;

– P50/P10 of median income to the upper bound value of the first decile;

– the Palma ratio is the share of all income received by the 10% of people with the highest disposable income divided by the share of all income received by the 40% of people with the lowest disposable income.

To analyze the level of income inequality globally, we will look at the Gini index during 1980-2019 (figure 1.1). Today, the level of income inequality globally is estimated at 42.3% and has decreased over the past 40 years by an average of only four percentage points.

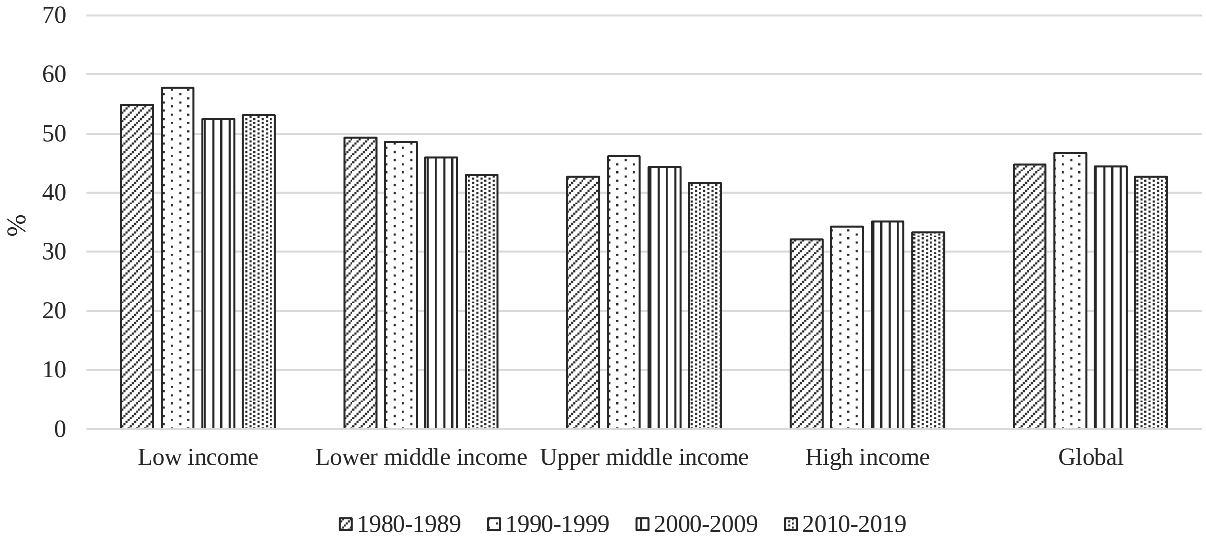

Figure 1.1 – The dynamics of the Gini index in countries depending on their income level [36]

Figure 1.1 shows that the most income is distributed unequally among the population of low-income countries. The most economically unfavorable for the distribution of welfare among different segments of the population for most countries (except for high-income countries) was the period from 1990-1999.

Despite adopting a set of measures by government agencies and international organizations aimed at reducing the income inequality in the world, this problem remains relevant. According to IMF experts [7], the main global drivers of income inequality are technological change and trade globalization. In this paper, we decided to analyze in more detail the impact of innovative technology (digitalization) on income inequality.

Digitalization is expected to affect production, employment and trade structures and require adaptation of existing strategic frameworks in different areas. Businesses and economies have been grappling for years to capture the full opportunities that digital technologies can offer. Even as they do so, the next wave of transformational technologies has arrived: accelerating progress in robotics, analytics, artificial intelligence (AI), and machine learning amount to a step-change in technical capabilities that will have profound consequences for business, for the economy, and society. So, potential economic gains from digital technologies are enormous, but with new opportunities come new challenges.

According to IMF, digitalization a wide range of new innovative information technology in business models and products, transforms the economy and social interactions [25]. The core technologies that matter for all markets, from production and commerce to finance, are: artificial intelligence, augmented reality, automation, blockchain, Internet of things, robotics, virtual reality, 3D printing.

Nowadays, automation and artificial intelligence have the most significant impact on the labor market development because [12]:

– machine-learning algorithms have progressed in recent years, especially through the development of deep learning and reinforcement-learning techniques based on neural networks;

– exponentially increasing computing capacity has become available to train larger and more complex models much faster. This compute capacity has been aggregated in hyper-scalable data centers and is being made much more accessible to users through the cloud;

– massive amounts of data that can be used to train machine learning models are being generated, for example through daily creation of billions of images, online click streams, voice and video, mobile locations, and sensors embedded in the Internet of Things.

Technology has been the dominant force in reshaping the demand for labor. Digital technologies and automation have shifted demand toward higher-level skills. Demand has shifted, in particular, away from routine, middle-level skills that are more vulnerable to automation, as in jobs like clerical work and repetitive production. Job markets have seen an increasing polarization, with the employment share of middle-skill jobs falling and that of higher-skill jobs, such as technical professionals and managers, rising. The employment share of low-skill jobs has also increased but mainly in nonroutine manual jobs in services such as personal care that are hard to automate.

As the demand for skills has shifted, supply has been slow to adapt. Shortages of higher-level skills demanded by the new technologies have prevented a broader diffusion of the innovations across firms. Workers with skills complementary with the new technologies have been clustered increasingly in leading firms at the technological frontier. Imbalances between skills demand and supply have fueled income inequality, by increasing the wage premia on higher-level skills.

The implementation of digital technologies also creates problems, costs and risks. Differences in access to low-cost digital technologies and limited opportunities for their effective use can lead to an unfair distribution of benefits. The widespread use of new technologies, automation and Internet platforms will lead to job losses, increased income inequality and increased market concentration and wealth. For example, ILO [19] estimates that labour income has dropped by 10.7 percent (or US$ 3.5 trillion) globally, only in the first three quarters of 2020 in comparison with the same period in 2019, and these losses are highest in the middle-income countries.

Highlighting wage inequality, C.Frey and M. Osborne [14] observed in their study of the US labour market that automation is expected to take away 47 percent of all jobs. Most of these being in the low and mid-skill levels

New information technology has led to improvements in productivity and well-being by leaps and bounds, but has also played a central role in driving up the skill premium, resulting in increased labor income inequality. This is because technological changes can disproportionately raise the demand for capital and skilled labor over low-skilled and unskilled labor by eliminating many jobs through automation or upgrading the skill level required to attain or keep those jobs [6,1]. To sum it up, digitalization affects the composition and nature of jobs available as well as relative wages (figure 1.2).

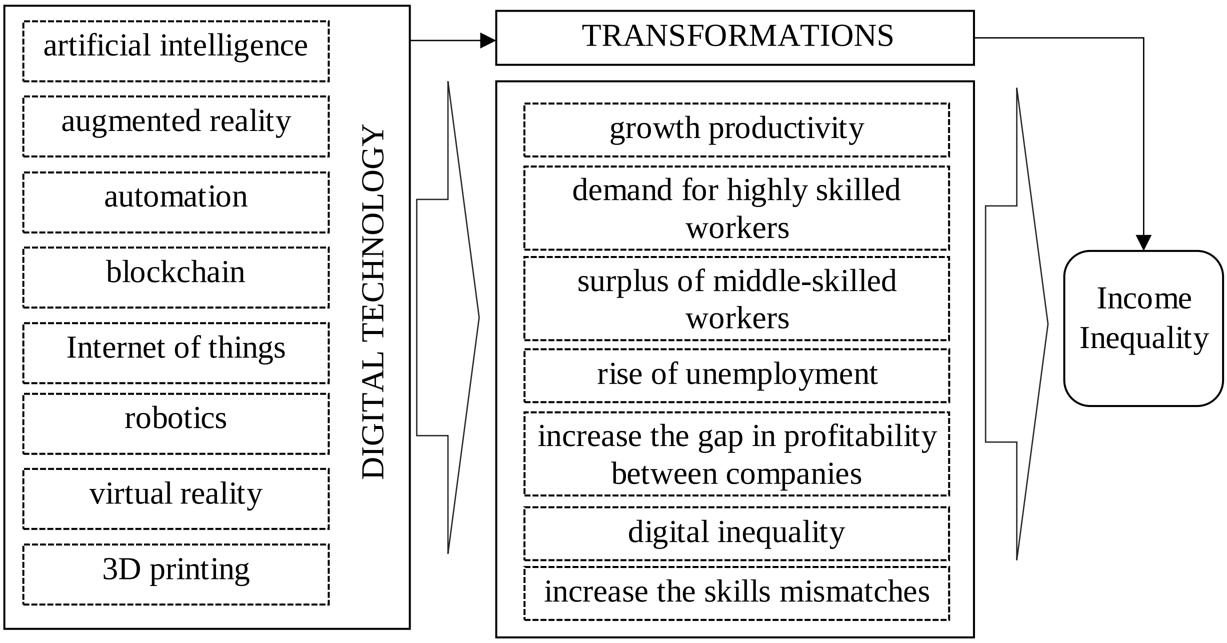

Figure 1.2 – The impact of digitalization on income inequality

So, the role technology plays in income and wealth inequality is complex and contested. Technology is a crucial driver of aggregate economic growth through productivity improvements, but its contribution to economic growth varies significantly across countries. Technology can also be a driver of income and wealth inequality because of its skills-bias nature and innovators capturing high rent.

1.2. Systematization of existing approaches to modeling the impact of digitalization on income inequality

Digital technologies are altering business models and how firms compete and grow. They are reshaping market structures. The change affects all markets, from production and commerce to finance. How the new technologies deployed across industries and firms has important implications for their economic impact and the distribution of rewards.

Technological innovation diffuses within economies and interacts with market conditions greatly for productivity growth and income distribution [9, 2]. The benefits of the new technologies have not been diffusing widely across firms. They have been captured for the most part by a relatively small number of larger firms. Productivity growth has been relatively strong in leading firms at the technological frontier.

The most noticeable impact of technology is the increased premium it places on skills. Modern technologies substitute tasks traditionally performed by unskilled workers while acting as a complement to skilled workers. Frey and Osborne [15] identify a wide range of occupations likely to be affected by automating routine and non-routine tasks. The abovementioned study concludes that low-skilled occupations such as factory, sales, and service jobs are more susceptible to technology substitution than high-skilled jobs.

The literature on the effects of the digital economy on employment provides various insights into the issue. On the one hand, a set of empirical highlights the debate on whether the digital revolution has created more jobs than they have displaced. Nottebohm et al. [28] suggests that the internet is a net job creator in both developed and developing economies, while more recent studies indicate that the threat of automation is more accurate, as advancements in cognitive computing and artificial intelligence are replacing work related to decision-making [10, 3535].

Interestingly, computerisation and changing technology has had significant impact on wage premium through the evolving composition of occupational employment in the last few decades [26]. The profound technological change has generated greater need for greater skills and education.

In paper [88] it was estimated the impact of ICT indicators on economic growth and reduction of inequality using panel data regression models. Authors found a negative relationship between the development of digital society and income inequality: an increase of 1% of ICT sector share in GDP will lead to a decrease of 0.27% of income inequality distribution. While another group of researchers who are part of the research group of IMF [30,7] empirically prove that technological progress had a greater impact on inequality. Jaumotte et al [30] analyzed technological development (the share of ICT capital in the total capital stock) and income inequality (Gini coefficient) in developing and high income countries using regression model. Technological progress thus increases the relative demand for higher skills, thereby exacerbating inequality in income.

The study [7] examines three groups of reasons for behind the dynamics of inequality: technological progress, international integration, government policy in relation to product and labor markets. The results of the study indicate that most of the decline in the share of labor in developed countries in the period 1993–2014. related to technological progress. For emerging economies, trade integration is the largest contributor to the income distribution, although the impact of technology is also significant.

Rey [29], Tselios [33], and Rodríguez-Pose and Tselios [31], Lin [24] have adopted spatial econometric models to study income inequality and found that income inequality levels were positively spatially correlated among neighbouring countries or regions. Therefore, the spatial effect on income is crucial. If the spatial effect is ignored, then the endogeneity problems of explanatory variables can induce biased model estimations. For instance, in paper [24**Error! Reference source not found.**], the effect between digital divide and the income inequality was estimated by a spatial quantile regression model. The estimation result showed that the

Internet usage exacerbate income disparity in low-income countries but improve income inequality in high-income countries.

Fuchs C. [16] used stepwise regression based on data on 126 countries for determining the influence of 11 factors (addressing socioeconomic, political, cultural, social, and technological issues) on digital divide

In paper [23] was estimated the impact of Fourth Industrial Revolution on the dynamics of income distribution in developed European countries using correlation and fourier analyses.

Thus, modern studies of systemic risk in most cases are based on the use of regression analysis, which allows to determine the degree and nature of the impact of individual indicators (including technological indicators) on the change of dependent variable. However, the analyzed models do not allow to clearly establish the nature of digital transformations (destructive or stimulative) on the income inequality.

1.3. Building mathematical model

Income as an indicator of the socio-economic development of the country is an important indicator of social welfare. However, a characteristic feature of all socio-economic systems is inequality in income, their differentiation, which is associated with significant differences in the position of members of society. Insignificant income differentiation stimulates economic development, and its high level negatively affects its pace, as it undermines incentives to work and reduces its efficiency. In addition, excessive income differentiation worsens the situation of the most vulnerable segments of the population, reduces life expectancy, worsens physical health, increases budget expenditures, and poses a threat of social conflict.

To find tools to reduce disparities in the distribution of income among different segments of the population, it is advisable to develop a methodological approach that involves assessing the impact of technological change on income inequality.

The object of the study is selected EU countries. In this investigation, we use such indicators to analyze income inequality as the Gini coefficient and Palma ratio. To characterize digitalization in the country, it is proposed to choose the following indicators: households - level of internet access, employed ICT specialists, use of computers and the internet by employees, value of e-commerce sales, and integration of internal processes.

The information base of the constructed model will be the annual data, the description of which is given in Table 1.1, indicating the possible range of fluctuations of the selected indicators.

Table 1.1 – Description of input data indicators

Indicator	Name of indicator	Measurement scale	Valid values	Source
Y1	Gini coefficient	units	(0;1)	OECD
Y2	Palma ratio	units	(0;1)	OECD
X1	Households - level of internet access	% of households	(0;100)	Eurostat
X2	Employed ICT specialists	% of total employment	(0;100)	Eurostat
X3	Use of computers and the internet by employees	% of total employment	(0;100)	Eurostat
X4	Value of e-commerce sales,	% of turnover	(0;100)	Eurostat
X5	Integration of internal processes	% of enterprises	(0;100)	Eurostat

The time range of our study is the period from 2011 to 2018. To solve the set tasks, we accumulated a balanced data panel in the context of 25 European Union countries.

The implementation of the proposed methodological approach to assessing the impact of digitalization on the income inequality will be carried out in stages, as shown in figure 1.3.

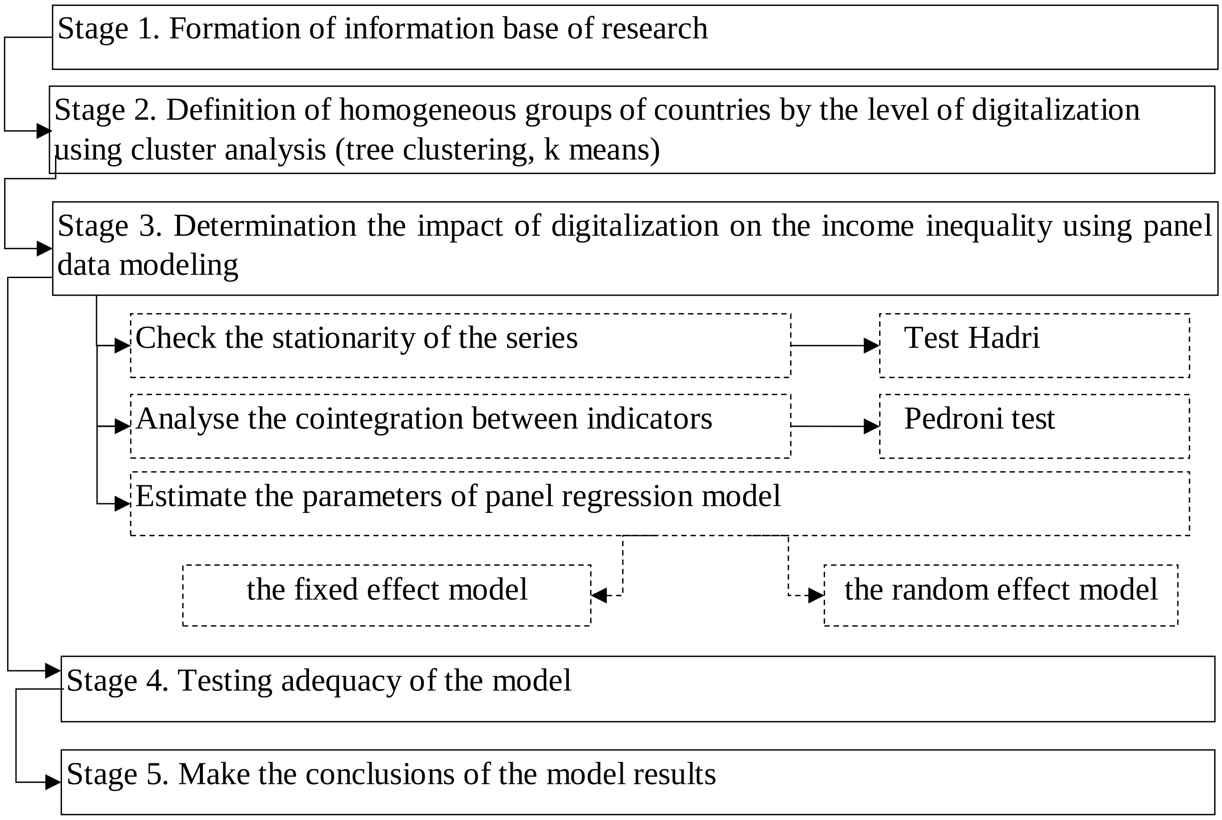

Figure 1.3 – The stages of estimation the impact of digitalization on the income inequality

We performed mathematical calculations and model construction using the software package STATISTICA and Eviews. These programs are easy to use the built-in functions, research tools will facilitate the process of design calculations.

Statement of the problem in terms of the subject area is given in table 1.2.

Table 1.2 – Basics of modeling the impact of digitalization on the income inequality

Elements	Descriprion
Problem	impact assessment of modern information and computer technologies on the population welfare
Affects on	social stability, investment attractiveness of the country, ensuring stable and sustainable economic growth of the country
The results of which are	construction of a methodological approach to assess the impact of digitalization on the state of income distribution among members of society
Advantages of the model	identify the most significant digital factors on the income distribution among the population, as well as analyze the individual effects by EU countries

The construction of economic and mathematical methods is possible in high-quality information resources that fully characterize digitalization. Considering the peculiarities of the business environment, opportunities, and threats from the introduction of information technology in financial relations, it is necessary to create a complete and high quality, sufficiently structured information and analytical base for making scientifically sound management decisions.

Thus, the reliability, objectivity, and completeness of the results obtained depend on the choice of the information base for the study and the pre-processing of statistical information. Based on the above indicators, the model's input parameters will be formed, determining the variables for assessing the patterns of income distribution between different segments of the population from the impact of digitalization.

SECTION 2. PRACTICAL IMPLEMENTATION OF THE MODEL, VERIFICATION OF ITS ADEQUACY, AND SUGGESTIONS FOR ITS USE

2.1. Clustering of EU countries by the level of digitalization of the economy

The development of digital technologies is uneven, and therefore its sphere of influence is also different for the economies of different countries. Under these conditions, there is a need to identify homogeneous groups of countries depending on the level of the digital economy. This condition will allow grouping relatively related objects of study, which will have specific patterns in their development. To solve this problem, we used the cluster analysis methods.

The task of cluster analysis is to sequentially divide the objects of study that are part of the set into m clusters (homogeneous groups), and the object should belong to only one cluster. Unlike most mathematical and statistical analysis methods, cluster analysis does not impose any restrictions on the format of input information. These methods allow the researcher to divide multidimensional sets of input data into homogeneous groups so that the objects within the group are similar to each other according to some criteria, and the objects from different groups differ from each other. In this case, the classification of objects is carried out simultaneously on several grounds on the basis of the introduction of a certain degree of total proximity on all grounds of classification.

The object of the study was 25 countries of the European Union. In our opinion, the following variables characterized the level of digitalization of the economy: households - level of internet access (x1), employed ICT specialists (x2), use of computers and the internet by employee (x3), value of e-commerce sales (x4), integration of internal processes (x5). Only 2018 was chosen for the study, as the development of digital technologies is dynamic, so it is advisable to consider only the latest available period with statistics.

The preparatory stage of cluster analysis is to bring the original statistical base to a single comparable form. With the help of the Data / Standartize tool of the

STATISTICA 6 package, the indicators characterizing the level of digitalization of the economy were normalized.

There are two groups of cluster analysis methods in the economic literature: hierarchical (agglomerative and divisional) and non-hierarchical. If agglomerative methods consistently combine elements into a single cluster, then divisional - on the contrary - divides the population into certain clusters.

Thus, in this study we used hierarchical-divisional methods, namely: tree clustering (tree clustering) and the method of k-means (k-means clustering).

The use of tree clustering allows you to group countries based on the calculation of distance or similarity between indicators. The following methods are used to determine the measure of the distance between objects depending on the scale of measurements: weighted Euclidean distance, city-block distance (Manhattan), Minkowski distance, Mahalanobis distance, percentage discrepancy, 1-Pearson correlation coefficient. However, we will use the Euclidean distance function:

$$p(x_i x_j) = \sqrt{\sum_{i=1}^{k}(x_{il} - x_{jl})^2} \qquad (2.1)$$

де x_{il} – standardized value of the i-th object on the l-th indicator;

x_{jl} – standardized value of the j-th object on the l-th indicator;

k – number of objects.

The final results of the calculation of Euclidean distances for EU countries as of 2018 are presented in the annex B.

Thus, based on the calculated values of Euclidean distances, a hierarchical cluster tree (dendrogram) is constructed, which demonstrates the stages of grouping countries (fig. 2.1 and Appendix B). The mechanism of construction of the dendogram is as follows: on the ordinate axis the distance between the selected objects is plotted with a gradual weakening of the unification criterion, while on the abscissa axis - the country. The result is the formation of clusters that have similar research objects in their structure.

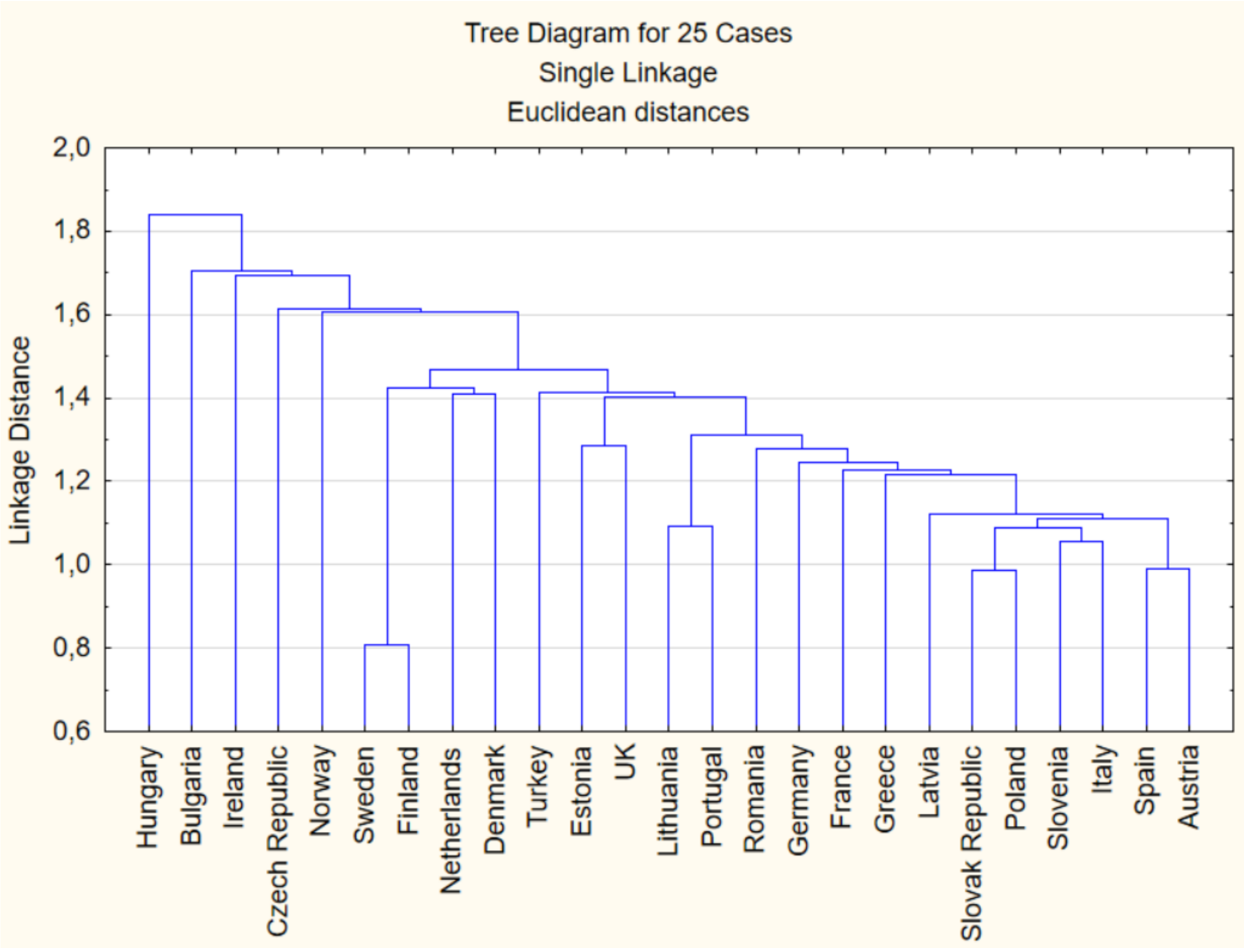

Figure 2.1 – Vertical dendrogram of clusters of EU countries by the level of digitalization of the economy as of 2018

The dendrogram analysis shows that it is expedient to single out only 2 clusters of countries, provided that the method of determining the distance between clusters is single linkage (nearest neighbor), and the degree of similarity between countries is Euclidean distance.

To further study the characteristics of EU countries in the development of digital technologies and their implementation in economic relations, the k-means method was used. The essence of the method is to construct the number of clusters k given by the researcher, such that the average values of all variables by objects of one of the clusters differ as much as possible from the average values of these variables by objects of other clusters. An essential problem in applying this cluster analysis method is to determine the number of clusters into which the initial set of observations will be divided. Based on the hierarchical clustering procedure, we note that the constructed dendrogram demonstrates the absence of fundamental differences in the development of the digital economy among EU countries. Therefore, it was assumed that the most likely number of clusters into which EU countries can be divided are 2 clusters.

The grouping of countries by the level of digitization by the method of k-means assumes the availability of data on the preliminary division of objects into groups, and also allows you to check the statistical significance between the selected clusters. The results of clustering of EU countries by the k-means method are presented in Table 2.1. Additional intermediate results describing clusters of EU countries by the k-means method are presented in the appendix.

Table 2.1 – The results of the clustering of EU countries by level of development of the digital economy as of 2018

Cluster	Countries	Euclidean distance	Countries	Euclidean distance
1 cluster - moderate development of the digital economy	Greece	0,720	Turkey	0,848
	Hungary	1,042	Latvia	0,473
	Italy	0,520	Lithuania	0,838
	Poland	0,296	Bulgaria	0,890
	Portugal	0,630	Romania	0,486
	Slovak Republic	0,536		
2 cluster - developed digital economy	Austria	0,547	Netherlands	0,806
	Czech Republic	0,871	Norway	0,527
	Denmark	0,767	Spain	0,727
	Finland	0,770	Sweden	0,835
	France	0,606	UK	0,695
	Germany	0,673	Estonia	0,794
	Ireland	1,005	Slovenia	0,556

In addition, table 2.1 shows the distances of the countries to the centers of the clusters in which they fell. The distance to the center of the cluster characterizes the degree of approach to the most typical representative in the cluster (to its geometric center); respectively, the smaller this distance, the more typical the cluster representative is the country.

The first cluster includes countries with moderate development of the digital economy, namely Greece, Hungary, Italy, Poland, Portugal, Slovak Republic, Turkey, Latvia, Lithuania, Bulgaria, Romania. Thus, the national governments of these countries are taking measures to develop digital technologies and their implementation in various spheres of public life, but the pace of these transformations is lower than in other EU countries.

The second cluster includes EU countries with developed digital economies: Austria, Czech Republic, Denmark, Finland, France, Germany, Ireland, Netherlands, Norway, Spain, Sweden, UK, Estonia, Slovenia.

Figure 2.2 shows the average values of variables for each cluster of EU countries according to the level of digitalization of the economy.

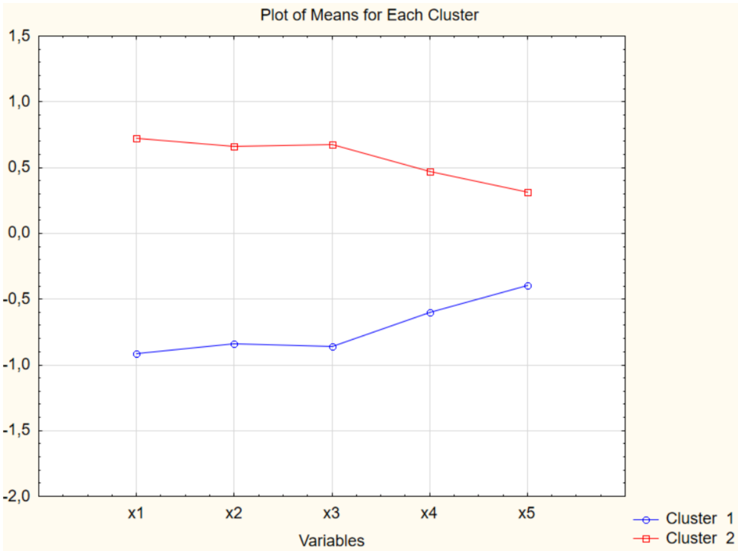

Figure 2.2 – The results of calculating the average values of indicators characterizing the digitalization of the economy, in terms of clusters of EU countries

The graph of average values of factor variables for clusters shows a significant difference between the first and second clusters in terms of households - level of internet access (x1), while the most similar clusters of countries in terms of ¬ integration of internal processes (x5).

Thus, using the mathematical and statistical method to group EU countries by the level of the digital economy allows us to move from a one-dimensional assessment of one of the indicators to a multidimensional space.

2.2. Assessing the impact of digitalization on the income inequality

To study the impact of digital technologies on the structure of income distribution between different population segments, we used appropriate econometric modeling methods. A panel data structure was used to build econometric models. When we have data for the considered objects for several periods, allows us to build deeper and more meaningful models and get answers to questions that are not available within, for example, models based only on country observation data in a fixed moment.

As digitalization and its impact on various spheres of public life have become relevant only in recent years, the formation of a good information base for the study is one of the main challenges of the study. One way to solve this problem is to model based on panel data.

The panel regression model involves data that displays information about the same set of objects over a series of consecutive periods. Thus, panel data is a combination of cross-sectional data and time-series data [39, 34, 4].

The main advantages of using the panel data structure are:

- models built using panel data allow to detect and analyze changes at the individual level, which cannot be done in the implementation of models with temporal or spatial data;

- a significant increase in the number of observations in the model is achieved, which reduces the threat of multicollinearity and significantly increases the degree of freedom;

- as a result of a combination of temporal and spatial measurements in the panel data, the efficiency of the estimates calculated in the models increases in comparison with the estimates obtained on the basis of individual models of time or spatial series.

In general, the model with panel data can be represented as [39]:

$$Y_{it} = \alpha + X_{it}\beta_{it} + \varepsilon_{it} \tag{2.2}$$

where Y_{it} – the value of the studied indicator for the i-th object in the i-th period of time;

X_{it} – order vector of explanatory variables (factors);

ε_{it} – perturbation for the i-th object in the i-th period of time;

α – scalar;

β_{it} – model parameters that measure the partial effects of the change in Hit in the period t for a certain and i-th period of time.

Model (2.2) has a general form, so it is advisable to introduce additional restrictions on the parameters of the model. The standard assumption valid for many empirical situations is the assumption of constancy of the parameters β_{it} for all values of t and and i. Under such conditions, model (2.2) takes the form

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \varepsilon_{it} \tag{2.3}$$

Model (2.3) is a general model of panel data (pooled model). Construction, estimation of parameters and research of such model occurs as at classical multifactor regression models. Note that this specification does not take into account the individual characteristics of the objects under study (within this study - countries).

Econometric models based on panel data, depending on the behavior of the perturbation components are divided into models with fixed effects and models with random effects. The fixed effects panel data model has the following general form:

$$Y_{it} = \mu_i + \beta \quad X_{it} + u_{it} \tag{2.4}$$

In addition, it is assumed that all X_{it} are independent of all u_{it}, and the perturbations u_{it} are independent equally distributed random variables with a mathematical expectation of zero and a constant variance.

The model with fixed effects should be applied to the data if there is a need to take into account unobservable factors that differ for different points in time.

If μ_i is represented as implementations of independent of random variables X_{it} with average distribution α_i and variance, then model (2.4) refers to random effects models:

$$Y_{it} = \alpha + \beta \ \ X_{it} + \mu_i + u_{it} \tag{2.5}$$

Estimation of parameters on the constructed models with panel data can be carried out by means of a 1MNK method. To establish the most adequate model, the Wald, Broysch-Pagan, and Hausman tests are used.

The proposed models should comply with the following principles [6]:

- adequacy - the ability to reflect the desired properties of the object;

- objectivity - compliance of the obtained conclusions based on the results of calculations with real conditions;

- simplicity - the model should contain only significant parameters;

- universality - the proposed method can be tested for another object of study.

Since modeling aims to use its results to make scientifically sound decisions, it is crucial to follow the stages of implementing the econometric model and economic interpretation of the obtained patterns.

The input information base for the construction of panel regression was grouped taking into account the results of cluster analysis, ie the panel data for cluster 1 (11 countries) and cluster 2 (14 countries) were formed. It should be noted that all variables that are included in the econometric model were prologarithmic to increase the normality of the distribution of residues and minimize the standard errors of the model. All mathematical calculations are performed in the program Eviews.

One of the main requirements for building an econometric model is the use of stationary time series. To check for the presence of single roots in the panel data, various tests are used: Levin-Lina-Chu, Hadri, Pesaran and Chin, panel analogues of the Dickie-Fuller tests.

Checking the presence of a single root in the panel data involves testing the null hypothesis, which assumes that the series is stationary at $p < 0.05$. The test results for stationarity for variables within 1 and 2 clusters are presented in tables 2.2 and 2.3, respectively, and the intermediate results in figures C.1-C.14, Annex C.

Table 2.2 – Test results of panel unit root tests within cluster 1

Indicator		Levin, Lin & Chu Test		IM, Pesaran and Shin Test		ADF - Fisher Chi-square		Conclusion
		statistics	p-value	statistics	p-value	statistics	p-value	
Y1	actual	-0,636	0,2624	0,766	0,778	18,496	0,676	the first level of integration
	1st differences	-4,296	0,0000*	-0,652	0,257	39,623	0,028*	
Y2	actual	-0,420	0,337	1,169	0,878	19,271	0,627	the first level of integration
	1st differences	-3,220	0,001*	-0,360	0,036*	44,653	0,014*	
X1	actual	-4,438	0,000*	0,917	0,820	15,433	0,843	the first level of integration
	1st differences	-6,894	0,000*	1,391	0,082	38,498	0,016*	
X2	actual	0,774	0,781	1,830	0,966	13,461	0,919	the first level of integration
	1st differences	-7,722	0,000*	-1,141	0,127	32,6528	0,047*	
X3	actual	-19,034	0,000*	-4,092	0,000*	34,727	0,041	the first level of integration
	1st differences	-12,129	0,000*	-4,121	0,000*	53,938	0,000*	
X4	actual	-3,904	0,000*	-0,572	0,284	26,772	0,220	the first level of integration
	1st differences	-6,614	0,000*	-2,264	0,012*	47,055	0,001*	
X5	actual	-32,009	0,000*	-8,461	0,000*	66,843	0,000*	no integration
	1st differences	-	-	-	-	-	-	

* – time series is stationary

Table 2.2 – Test results of panel unit root tests within cluster 2

Indicator		Levin, Lin & Chu Test		IM, Pesaran and Shin Test		ADF - Fisher Chi-square		Conclusion
		statistics	p-value	statistics	p-value	statistics	p-value	
Y1	actual	-2,438	0,007	-0,256	0,399	31,148	0,310	the first level of integration
	1st differences	-0,128	0,049*	-9,109	0,046*	29,505	0,387	
Y2	actual	-4,603	0,000*	-1,026	0,153	41,50	0,048*	no integration
	1st differences	-46,112	0,000*	-15,565	0,000*	69,224	0,000*	
X1	actual	-6,117	0,000*	0,513	0,696	25,254	0,614	the first level of integration
	1st differences	-50,490	0,000*	-10,208	0,000*	61,231	0,000*	
X2	actual	-6,406	0,000*	1,670	0,953	18,840	0,903	the first level of integration
	1st differences	-5,113	0,000*	-1,280	0,100	41,425	0,0491*	
X3	actual	-1,254	0,105	2,625	0,990	9,8098	0,999	the first level of integration
	1st differences	-8,025	0,000*	-1,724	0,042	47,609	0,012*	
X4	actual	-5,805	0,000*	0,21490	0,585	31,644	0,289	the first level of integration
	1st differences	-4,893	0,000*	-6,317	0,044*	42,697	0,037*	
X5	actual	-15,445	0,000*	-2,194	0,014	49,701	0,007	the first level of integration
	1st differences	-2,010	0,022*	0,475	0,009*	24,975	0,629	

* time series is stationary

According to the results of three tests to check the stationarity within the 1st and 2nd clusters, it was found that most variables need to take the first differences so that they turn from non-stationary to stationary.

Because the variables in the 1st and 2nd clusters of the study have different integration orders, it might not be cointegration relationships between the variables. However, we decided to test for cointegration using the Kao test. According to the Kao test, the null hypothesis is the lack of cointegration. The results of checking the cointegration between the analyzed variables within cluster 1 and 2 are shown in the figure 2.3.

1 cluster

Kao Residual Cointegration Test
Series: Y1_LN X1_LN X2_LN X3_LN X4_LN X5_LN
Sample: 2011 2018
Included observations: 88
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
User-specified lag length: 1
Newey-West automatic bandwidth selection and Bartlett kernel

	t-Statistic	Prob.
ADF	-2.050724	0.1201
Residual variance	0.001056	
HAC variance	0.000744	

Kao Residual Cointegration Test
Series: Y2_LN X1_LN X2_LN X3_LN X4_LN X5_LN
Sample: 2011 2018
Included observations: 88
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
User-specified lag length: 1
Newey-West automatic bandwidth selection and Bartlett kernel

	t-Statistic	Prob.
ADF	-1.721554	0.1426
Residual variance	0.003326	
HAC variance	0.002427	

2 cluster

Kao Residual Cointegration Test
Series: Y1_LN X1_LN X2_LN X3_LN X4_LN X5_LN
Sample: 2011 2018
Included observations: 112
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
User-specified lag length: 1
Newey-West automatic bandwidth selection and Bartlett kernel

	t-Statistic	Prob.
ADF	-0.650804	0.2576
Residual variance	0.000567	
HAC variance	0.000519	

Kao Residual Cointegration Test
Series: Y2_LN X1_LN X2_LN X3_LN X4_LN X5_LN
Sample: 2011 2018
Included observations: 112
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
User-specified lag length: 1
Newey-West automatic bandwidth selection and Bartlett kernel

	t-Statistic	Prob.
ADF	-0.809375	0.2091
Residual variance	0.001110	
HAC variance	0.000903	

Figure 2.3 – Results of the Kao test to check the cointegration between variables within clusters 1 and 2

The figure 2.3 confirms our judgments that there is no cointegration between the variables in the context of clusters 1 and 2 because the actual p-value, according to the Kao test results is more significant than 0.05.

The paper proposes to use two dependent variables that reflect the state of income distribution between different segments of the population: the Gini index (Y1) and the Palma index (Y2). While the independent variables that should be included in the econometric model are: households - level of internet access (x1), employed ICT specialists (x2), use of computers and the internet by employees (x3), the value of e-commerce sales (x4), integration of internal processes (x5).

Table 2.4 and 2.5 show the results of estimating panel regressions of two types: the model of panel data with fixed effects (fixed effects models), the model of panel data with random effects (random effects models).

Table 2.4 – The results of the impact assessment of digital factors on the income distribution within the 1st cluster of EU countries

	Dependent variable – Gini index (y_1)				Dependent variable – Palma index (y_2)			
	Fixed effects models		Random effects models		Fixed effects models		Random effects models	
	statistics	p-value	statistics	p-value	statistics	p-value	statistics	p-value
const	-0,004	0,649	-0,011	0,167	-0,002	0,902	-0,015	0,292
ln x_1 (1)	-0,017	0,899	0,076	0,483	-0,070	0,769	0,093	0,629
ln x_2 (1)	-0,006	0,920	0,011	0,853	-0,052	0,631	-0,014	0,891
ln x_3 (1)	0,079	0,410	0,122	0,177	0,085	0,612	0,177	0,266
ln x_4 (1)	0,003	0,876	0,007	0,686	0,004	0,908	0,012	0,695
ln x_5 (1)	0,041	0,078	0,043	0,049*	0,085	0,041*	0,086	0,025*
Indicators of adequacy								
R-squared	0,183		0,090		0,202		0,091	
Adjusted R-squared	-0,018		0,026		0,006		0,027	
F-statistic	0,909		11,402		1,030		11,416	
Prob (F-statistic)	0,559		0,034		0,438		0,029	

Table 2.5 – The results of the impact assessment of digital factors on the income distribution within the 2nd cluster of EU countries

	Dependent variable – Gini index (y_1)				Dependent variable – Palma index (y_2)			
	Fixed effects models		Random effects models		Fixed effects models		Random effects models	
	statistics	p-value	statistics	p-value	statistics	p-value	statistics	p-value
const	-1,261	0,000	-1,249	0,000	-0,025	0,954	-0,028	0,949
ln x_1 (1)	-0,107	0,213	-0,091	0,284	-0,150	0,201	-0,129	0,266
ln x_2 (1)	-0,160	0,002*	-0,145	0,003*	-0,328	0,000*	-0,305	0,000*
ln x_3 (1)	0,190	0,003*	0,166	0,006*	0,301	0,001*	0,274	0,001*
ln x_4 (1)	-0,001	0,951	0,002	0,929	0,020	0,494	0,023	0,414
ln x_5 (1)	-0,009	0,574	-0,015	0,340	-0,012	0,576	-0,021	0,324
Indicators of adequacy								
R-squared	0,955		0,194		0,962		0,301	
Adjusted R-squared	0,946		0,156		0,954		0,268	
F-statistic	109,327		5,098		129,837		9,127	
Prob (F-statistic)	0,000		0,000		0,000		0,000	

According to the results of the projects, it was established that the five independent winters were statistically significant at the level of 0.95 at the boundaries of 1 cluster є integration of internal processes (x5), those at the edges of the 2 cluster - employed ICT specialists (x2) and use of computers and the internet by employee (x3). To determine an adequate model at the boundaries of the skin cluster of the regions, the Hausman test should be scored. Hull hypothesis behind the Hausman test є the adolescence of victorious models with vivid panel data. The results of pairwise matching of models with fixed and specific effects are shown in the table 2.6.

Table 2.6 – Result of the Hausman test

1 cluster		2 cluster	
Dependent variable – Gini index (y_1)	Dependent variable – Palma index (y_2)	Dependent variable – Gini index (y_1)	Dependent variable – Palma index (y_2)
Chi-Sq. Statistic – 5,9878 Prob – 0,3074	Chi-Sq. Statistic – 7,0257 Prob – 0,2187	Chi-Sq. Statistic – 7,9432 Prob – 0,1594	Chi-Sq. Statistic – 8,4245 Prob – 0,1343

The table 2.6 shows that it is more expedient to use models with random effects (p-value for all unfolded variants is greater than 0.05). In the model with random effects, it is assumed that individual differences are random. This model can be considered as a compromise between the general regression, which imposes a strong limitation of homogeneity on all coefficients of the regression equation, and regression with fixed effects, which allows for each sample object to enter its own constant and thus take into account existing but unobservable heterogeneity.

One of the critical stages in building an econometric model is to check its adequacy. One of the key indicators for checking the adequacy of the constructed models is the coefficient of determination, Fisher's and Student's criterion. Therefore, the coefficient of determination for the constructed models with fumigation effects ranges up to 0.3. This indicates that the selected factors of digitalization do not significantly affect the equal distribution of income of the EU population. These results suggest that income differentiation within EU countries may be influenced by political, economic, geographical, demographic, and other factors that in one way or another affect the formation of household income. At the same time, according to Fisher's F-test, the obtained econometric models are statistically significant because prob <0.05.

To assess the normality of the distribution of the residues of the panel model, we can analyze the histogram of the distribution and the results of the Jarque-Bera test for models of panel data with random effects (fig. 2.4)

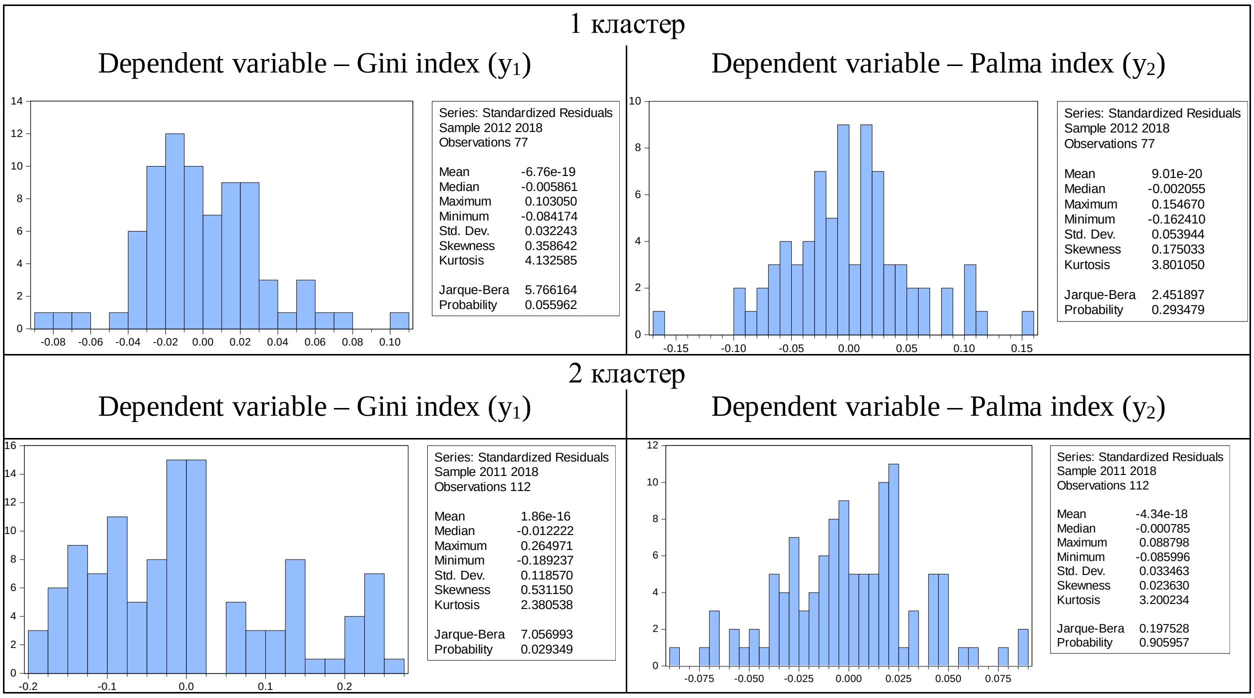

Figure 2.4 – Histogram of the distribution of residues

The data in Figure 2.5 suggest a normal distribution of residues in most constructed models (the level of significance of the Zharko-Ber criterion is greater than 0.05). The hypothesis regarding the normal distribution is rejected for the econometric model with the dependent variable y1 within the 2nd cluster.

The next criterion is to check the presence of correlation in the residues based on the Breusch-Pagap and Pesaran tests (fig. 2.5).

1 кластер

Dependent variable – Gini index (y_1)

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQ01_LN
Periods included: 7
Cross-sections included: 11
Total panel observations: 77
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	58.65311	55	0.3430
Pesaran scaled LM	-0.700498		0.4836
Pesaran CD	-1.679292		0.0931

Dependent variable – Palma index (y_2)

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQ01_Y2
Periods included: 7
Cross-sections included: 11
Total panel observations: 77
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	58.87088	55	0.3358
Pesaran scaled LM	-0.679735		0.4967
Pesaran CD	-1.542386		0.1230

2 кластер

Dependent variable – Gini index (y_1)

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQ01_Y1
Periods included: 7
Cross-sections included: 14
Total panel observations: 98
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	86.28702	91	0.6202
Pesaran scaled LM	-1.387098		0.1654
Pesaran CD	0.459516		0.6459

Dependent variable – Palma index (y_2)

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQ01_Y2
Periods included: 7
Cross-sections included: 14
Total panel observations: 98
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	124.4680	91	0.0114
Pesaran scaled LM	1.443064		0.1490
Pesaran CD	0.308988		0.7573

Figure 2.5 – The result of the correlation test in the residuals

The results of the test calculation allow us to accept the null hypothesis of the absence of cross-sectional correlation in the residues of the obtained model (values of the Breusch-Pagap and Pesaran tests are greater than 0.05).

Therefore, to assess the impact of digitalization on the uniformity of income distribution, it is advisable to use a panel data model with random effects, which is adequate. Based on the results of the constructed models, it is advisable to draw the following conclusions:

- for cluster 1 countries (moderate development of the digital economy), the integration of internal processes (x5) is a statistically significant factor of digitization (according to Student's criterion). Thus, an increase in the integration of internal processes by 1% causes an increase in the Gini index and Palma index by 0.043% and 0.086%, respectively, ie, there is an increase in uneven income distribution;

- for the countries of the 2nd cluster (developed digital economy), statistically, significant factors of digitalization were employed ICT specialists (x2) and the use of computers and the Internet by employees (x3). These factors have different effects on income differentiation among the population of the EU. In particular, in the EU countries, the increase in the level of employment in information and communication technologies by 1% led to a decrease in income inequality (Gini index by 0.145; Palma index by 0.305%). At the same time, an increase in the share of employees who use computers and the Internet by 1% causes an increase in the Gini index and Palma index by 0.166% and 0.274%, respectively.

2.3. Ways to reduce the income inuquality

The origins of the income inequality problem are the imperfection of state regulation of economic and social processes. In the monopolization of many spheres of economic activity, the spread of shadow employment, corruption, mergers of business, and government leads to social justice in wages and separation of wages.

The consequences of the unequal income distribution are []

– low living standards in the country;
– poverty;
– outflow of skilled labor abroad;
– social tensions;
– difficult crime situation;
– the prosperity of the shadow economy;
– high levels of corruption;

– political instability, which together lead to social restraint, economic development of the country in general and human capital in particular.

Overcoming economic inequality is possible with the help of government agencies, socialization, overcoming corruption, economic growth, and removing it from the shadow sector.

The experience of many Latin American countries shows that large-scale subsidy programs, progressive tax scales, strict restrictions on labor laws, which are designed to reduce the income gap, can have the opposite effect, so it is necessary to eliminate sources of product distortions (over-regulation and licensing), adjust credit and financial policy, to form an effective labor market [12, p. 114. Without state intervention in the distribution of income, economic inequality of the population cannot be overcome, so it is necessary to take such measures as:

– establishing a minimum wage at the level of the real subsistence level;

– providing social benefits and subsidies to the poor;

– stimulating the introduction of innovations in production, which will lead to an increase in wages and increase the share of wages in the product;

– reduction of burdensome taxation, which will help reduce the shadow sector of the economy and stabilize tax legislation;

– curbing inflation;

– improving the conditions for small business, which will create additional jobs;

– fight against corruption;

– orientation of the state on the formation and support of the middle class, etc.

SUMMARY

High levels of inequality impede the expansion of skills, limit economic and social mobility and human development, and as a result, inhibit economic growth. In addition, it fosters a sense of uncertainty, vulnerability, and insecurity, undermines trust in institutions and government, increases social discord and tensions, and generates violence and conflict.

One of the key factors that determines the trajectory of the national economy is digitalization. New digital technologies are often viewed as a tipping point in terms of organizing production within value chains. This is because they give intangible assets a more prominent role in generating income. Digitalization also leads to the emergence of new types of jobs and employment, to a change in the nature and conditions of work, a change in professional requirements for the level of qualifications and is reflected in the functioning of labor markets, as well as in the international division of labor. Thus, the paper analyzes the impact of digitalization on income inequality.

Based on tree clustering and k-means, the EU countries were clustered according to the level of digitalization, and the expediency of separating 2 clusters of countries was established. In our opinion, the following variables characterized the level of digitalization of the economy: households - level of internet access, employed ICT specialists, use of computers and the internet by employee, value of e-commerce sales, integration of internal processes. We have used such indicators to analyze income inequality as the Gini coefficient and Palma ratio.

The paper constructs a panel regression with random effects to formalize the relationship between the level of income distribution among the population in the country and the indicators of digitalization.

For cluster 1 countries (moderate development of the digital economy), the integration of internal processes is a statistically significant factor of digitization. Thus, an increase in the integration of internal processes by 1% causes an increase

in the Gini index and Palma index by 0.043% and 0.086%, respectively, ie, there is an increase in uneven income distribution.

For the countries of the 2nd cluster (developed digital economy), statistically, significant factors of digitalization were employed ICT specialists and the use of computers and the Internet by employees. These factors have different effects on income differentiation among the population of the EU. In particular, in the EU countries, the increase in the level of employment in information and communication technologies by 1% led to a decrease in income inequality (Gini index by 0.145; Palma index by 0.305%). At the same time, an increase in the share of employees who use computers and the Internet by 1% causes an increase in the Gini index and Palma index by 0.166% and 0.274%, respectively.

So, countries should consider adapting their education and training systems and policies for successful solving the problems arising on the path of transition to the digital economy. Governments should consider how they can adapt social protection systems in order to reduce the risks of increasing polarization and rising income inequality driven by digitalization.

LITERATURE

1. Acemoglu D. Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. Quarterly Journal of Economics. 1998. 113 (4): 1055–89.

2. Aghion P., Akcigit U., Bergeaud A., Blundell R., Hémous D. Innovation and Top Income Inequality. Review of Economic Studies. 2019. Vol. 86(1). P. 1-45.

3. Aliandali A., Tatahi M. Economic and Financial Modeling with Eviews: a guide for students and professionals. Springer International Publishing AG, 2018. 284 p.

4. Asteriou D., Stephen G. Applied Econometrics: a modern approach. Palgrave Macmillan, 2006. 397 p. – URL: http://www.doaei.com/Books/Applied_Econometrics__A_Modern_Approach_Using_Eviews_and_Microfit_Revised_Edition.pdf

5. Brooks C. Introductory Econometrics for Finance, Second Edition. Cambridge University Press, 2008. 648 p. – URL: http://www.afriheritage.org/TTT/3%20Brooks_Introductory%20Econometrics%20for%20Finance%20(2nd%20edition).pdf

6. Card D., DiNardo J.E. Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. NBER Working Paper. 2002. 8769, National Bureau of Economic Research, Cambridge, Massachusetts.

7. Causes and Consequences of Income Inequality: A Global Perspective Era Dabla-Norris, Kalpana Kochhar, Nujin Suphaphiphat, Frantisek Ricka, Evridiki Tsounta / International Monetary Fund. 2015. URL: https://www.imf.org/external/pubs/ft/sdn/2015/sdn1513.pdf

8. Cioacă S. I., Cristache S. E., Vuţă M., Marin E., & Vuţă M. Assessing the impact of ICT sector on sustainable development in the European Union: An empirical analysis using panel data. Sustainability. 2020. № 12(2). URL: https://doi.org/10.3390/su12020592

9. Comin D., Mestieri M. If Technology has Arrived Everywhere, Why has Income Diverged? American Economic Journal: Macroeconomics. 2018. 10(3): 137– 175.

10. Davenport T.H., Kirby J. Beyond automation. Harvard Business Review. 2015. № 93(6). P. 59-65

11. Demir A., Pesqué-Cela V., Altunbas Y., Murinde V. Fintech, financial inclusion and income inequality: a quantile regression approach. European Journal of Finance. 2020. URL: https://doi.org/10.1080/1351847X.2020.1772335

12. Digitalization, AI and the future of work: Imperatives for Europe. McKinsey Global Institute. 2017. URL: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/europe/ten%20imperatives%20for%20europe%20in%20the%20age%20of%20ai%20and%20automation/digitization-ai-and-the-future-of-work.ashx

13. Eurostat. URL: https://ec.europa.eu/eurostat/data/database

14. Frey C., Osborne M. The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change. 2017. Volume 114. P. 254-280. URL: https://doi.org/10.1016/j.techfore.2016.08.019

15. Frey C.B., Osborne M.A. The future of employment: how susceptible are jobs to computerisation?. Technological Forecasting and Social Change. 2017. № 114. P. 254- 280

16. Fuchs C. The role of income inequality in a multivariate cross-national analysis of the digital divide. Social Science Computer Review. 2009. Vol. 27(1). P. 41–58. URL: https://doi.org/10.1177/0894439308321628

17. Hanck Ch.,Arnold M., Gerber A., Schmelzer M. Introduction to Econometrics with R, 2018. 400 p.

18. Howard M.W., Carter V. J. Income inequality. Encyclopedia Britannica, 20 Aug. 2020. URL: https://www.britannica.com/topic/income-inequality.

19. ILO Monitor: COVID-19 and the world of work. Sixth edition Updated estimates and analysis. URL: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/briefingnote/wcms_755910.pdf

20. Income inequality Database. OECD. URL: https://data.oecd.org/inequality/income-inequality.htm

21. Kapoor A., Tucker C.E. Income Inequality and Digital Advertising Strategies: Evidence from a Field Experiment. SSRN Electronic Journal. 2017. URL: https://doi.org/10.2139/ssrn.3048553

22. Kozhan R. Financial Econometrics with Eviews. Roman Kozhan & Ventus Publishing ApS, 2010. 119 p.

23. Kuzmenko O., Roienko V. Nowcasting income inequality in the context of the Fourth Industrial Revolution. SocioEconomic Challenges, 2017. 1(1). P. 5-12. URL: http://doi.org/10.21272/sec.2017.1-01.

24. Lin C.-H. A., Lin H.-S., Hsu C.-P. Digital Divide and Income Inequality: A Spatial Analysis. Review of Economics & Finance. 2017. P. 31–43.

25. Measuring the Digital Economy. International Monetary Fund. 2018. URL: https://www.imf.org/en/Publications/Policy-Papers/Issues/2018/04/03/022818-measuring-the-digital-economy

26. Mishel L., Schmitt J., Shierholz H. Assessing the job polarization explanation of growing wage inequality. Economic Policy Institute. 2013. Working Paper.

27. Mohd Daud S.N., Ahmad A.H., Ngah W.A.S.W. Financialization, digital technology and income inequality. Applied Economics Letters. 2020. URL: https://doi.org/10.1080/13504851.2020.1808165

28. Nottebohm O., Manyika J., Bughin J., Chui M., Syed A. Online and upcoming: the Internet's impact on aspiring countries. 2012. McKinsey&Company Raport.

29. Rey S. J. Spatial analysis of regional income inequality. Spatially Integrated Social Science: Examples in Best Practice. 2004. P. 280-299.

30. Rising Income Inequality: Technology, or Trade and Financial Globalization? Prepared by Florence Jaumotte, Subir Lall, and Chris Papageorgiou. IMF Working Paper. 2008. WP/08/185 URL: https://www.imf.org/external/pubs/ft/wp/2008/wp08185.pdf

31. Rodríguez-Pose A., Tselios V. Education and income inequality in the regions of the European Union. Journal of Regional Science. 2009. Vol. 49 (3). P. 411-437.

32. Stock J. H., Watson M. W. Introduction to Econometrics, Third Edition. Addison-Wesley, 2016. 785 p. – URL: https://econometricsweb.files.wordpress.com/2016/11/stock-watson-econometrics-3rd-edition-ilovepdf-compressed.pdf

33. Tselios V. Income and educational inequalities in the regions of the European Union: Geographical spillovers under welfare state restrictions. Papers in Regional Science. 2008.Vol. 87 (3). P. 403-430.

34. Wooldridge J. M. Econometric Analysis of Cross Section and Panel Data (MIT Press), 2002. 392 p.

35. World Bank. World Development Report 2016: Digital Dividends, Washington DC.

36. World Income Inequality Database. WIID. URL: https://www.wider.unu.edu/database/wiid

37. Voloshina S.V., Skubilina A.V., Chebotarenko A.M. Income inequality of the population of Ukraine as an obstacle to the development of human capital. Economy and society. 2017. № 9. pp. 882–889.

38. Yeliseyeva L.V., Grisyuk O.V. Problems of income distribution. Scientific Bulletin of Kherson State University. 2016. № 16. S. 15–17.

39. Koval N.V. Socio-economic inequality in Ukraine and the world: assessment problems and solutions. Economy and state. 2016. № 2. S. 46–50.

40. Lukyanenko I.G., Gorodnichenko Y.O. Modern econometric methods in finance. Tutorial. - K.: Litera LTD, 2002.

ANNEXES

ANNEX A

SUMMARY

Basanets S.R. Economic and mathematical modeling the impact of digitalization on income inequality. – Qualification bachelor’s work. Academic and Research Institute of Business, Economics and Management Sumy State University, Sumy, 2021.

The paper analyzes the current state of the income distribution, examines the impact of digital technologies on household incomes, analyzes existing approaches and methods for modeling the relationship between these economic processes. The paper develops a scientific and methodological approach to clustering countries by the level of the digital economy and assessing the impact of digitalization on income distribution in terms of existing clusters based on the construction of panel regression models.

Key words: digitization, income inequality, panel data, clustering.

АНОТАЦІЯ

Басанець С.Р. Економіко-математичне моделювання впливу діджиталізації на нерівність доходів. – Кваліфікаційна бакалаврська робота. Навчально-науковий інститут бізнесу, економіки та менеджменту Сумського державного університету, Суми, 2021 р.

У роботі проаналізовано сучасний стан рівномірності розподілу доходів, проаналізовано вплив цифрових технологій на доходи населення, проаналізовано існуючі підходи та методи до моделювання зв’язку між даними економічними процесами. У роботі розроблено науково-методичний підхід до кластеризації країн за рівнем ровзитку цифрової економіки та оцінювання впливу цифровізації на рівномірність розподілу доходів у розрізі сформованих кластерів на основі побудови регресійних моделей панельних даних.

Ключові слова: цифровізація, нерівність доходів, панельні дані, кластеризація.

ANNEX B

The results of the clusterization of the regions of the EU for the level of digitalization of the economy by the method of k-means

Variable	Analysis of Variance (Spreadsheet2) Between SS	df	Within SS	df	F	signif. p
x1	16,47931	1	7,52069	23	50,39751	0,000000
x2	13,85820	1	10,14180	23	31,42819	0,000011
x3	14,62007	1	9,37993	23	35,84903	0,000004
x4	7,12025	1	16,87975	23	9,70191	0,004874
x5	3,09375	1	20,90625	23	3,40359	0,077971

Figure B.1– Analysis of Variance

Variable	Cluster Means (Spreadsheet2) Cluster No. 1	**Cluster No. 2**
x1	-0,915940	0,719667
x2	-0,839945	0,659957
x3	-0,862725	0,677855
x4	-0,602068	0,473053
x5	-0,396863	0,311821

Figure B.2 – Cluster Means

Cluster Number	Euclidean Distances between Clusters (Spreadsheet2) Distances below diagonal Squared distances above diagonal **No. 1**	No. 2
No. 1	0,000000	1,791285
No. 2	1,338389	0,000000

Figure B.3 – Euclidean Distances between Clusters

Continuation of Annex B

Variable	Descriptive Statistics for Cluster 1 (Spreadsheet2) Cluster contains 11 cases		
	Mean	Standard Deviation	**Variance**
x1	-0,915940	0,569546	0,324383
x2	-0,839945	0,526734	0,277448
x3	-0,862725	0,483134	0,233418
x4	-0,602068	0,877812	0,770553
x5	-0,396863	1,026879	1,054481

Figure B.4 – Descriptive Statistics for Cluster 1

Variable	Descriptive Statistics for Cluster 2 (Spreadsheet2) Cluster contains 14 cases		
	Mean	Standard Deviation	**Variance**
x1	0,719667	0,573576	0,328989
x2	0,659957	0,752806	0,566717
x3	0,677855	0,736194	0,541981
x4	0,473053	0,840065	0,705709
x5	0,311821	0,892767	0,797033

Figure B.5 – Descriptive Statistics for Cluster 2

ANNEX C

Results of stationary testing

Panel unit root test: Summary
Series: X1_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-4.43798	0.0000	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	0.91650	0.8203	11	66
ADF - Fisher Chi-square	15.4331	0.8430	11	66
PP - Fisher Chi-square	56.1758	0.0001	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X1_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-6.89394	0.0000	11	55
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-1.39192	0.0820	11	55
ADF - Fisher Chi-square	38.4976	0.0161	11	55
PP - Fisher Chi-square	89.6170	0.0000	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.1 – Results of stationary tests for households - level of internet access within 1 cluster (X1)

Panel unit root test: Summary
Series: X2_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	0.77378	0.7805	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	1.83033	0.9664	11	66

Continuation of Annex C

ADF - Fisher Chi-square	13.4614	0.9195	11	66
PP - Fisher Chi-square	38.3722	0.0166	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X2_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-7.72244	0.0000	11	55
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-1.14113	0.1269	11	55
ADF - Fisher Chi-square	32.6528	0.0469	11	55
PP - Fisher Chi-square	62.5124	0.0000	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.2 – Results of stationary tests for employed ICT specialists within 1 cluster (X2)

Panel unit root test: Summary
Series: X3_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-19.0341	0.0000	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-4.09201	0.0000	11	66
ADF - Fisher Chi-square	34.7265	0.0414	11	66
PP - Fisher Chi-square	26.1587	0.2449	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X3_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-12.1287	0.0000	11	55
Null: Unit root (assumes individual unit root process)				

Im, Pesaran and Shin W-stat	-4.12065	0.0000	11	55

Continuation of Annex C

ADF - Fisher Chi-square	53.9383	0.0002	11	55
PP - Fisher Chi-square	69.9591	0.0000	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.3 – Results of stationary tests for use of computers and the internet by employees within 1 cluster (X3)

Panel unit root test: Summary
Series: X4_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-3.90414	0.0000	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-0.57188	0.2837	11	66
ADF - Fisher Chi-square	26.7715	0.2200	11	66
PP - Fisher Chi-square	25.4767	0.2748	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X4_LN)
Date: 06/06/21 Time: 00:12
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-6.61377	0.0000	11	55
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-2.26425	0.0118	11	55
ADF - Fisher Chi-square	47.0550	0.0014	11	55
PP - Fisher Chi-square	99.0478	0.0000	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.4 – Results of stationary tests for value of e-commerce sales within 1 cluster (X4)

Continuation of Annex C

Panel unit root test: Summary
Series: X5_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-32.0091	0.0000	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-8.46063	0.0000	11	66
ADF - Fisher Chi-square	66.8430	0.0000	11	66
PP - Fisher Chi-square	89.9607	0.0000	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X5_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-17.4877	0.0000	11	55
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-6.43656	0.0000	11	55
ADF - Fisher Chi-square	73.8428	0.0000	11	55
PP - Fisher Chi-square	46.4445	0.0017	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.5 – Results of stationary tests for integration of internal processes within 1 cluster (X5)

Panel unit root test: Summary
Series: Y1_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-0.63589	0.2624	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	0.76558	0.7780	11	66
ADF - Fisher Chi-square	18.4960	0.6762	11	66
PP - Fisher Chi-square	33.5404	0.0547	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Continuation of Annex C

Panel unit root test: Summary
Series: D(Y1_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-4.29563	0.0000	11	55
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-0.65183	0.2573	11	55
ADF - Fisher Chi-square	39.6232	0.0279	11	55
PP - Fisher Chi-square	64.9555	0.0000	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.6 – Results of stationary tests for the Gini index within 1 cluster (Y1)

Panel unit root test: Summary
Series: Y2_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-0.41976	0.3373	11	66
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	1.16973	0.8789	11	66
ADF - Fisher Chi-square	19.2708	0.6286	11	66
PP - Fisher Chi-square	25.2322	0.2861	11	77

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(Y2_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-3.21985	0.0006	11	55
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-0.36021	0.0364	11	55
ADF - Fisher Chi-square	44.6526	0.0140	11	55
PP - Fisher Chi-square	64.6747	0.0000	11	66

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.7 – Results of stationary tests for the Palma index within 1 cluster (Y2)

Continuation of Annex C

Panel unit root test: Summary
Series: X1_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-6.11674	0.0000	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	0.51376	0.6963	14	84
ADF - Fisher Chi-square	25.2542	0.6140	14	84
PP - Fisher Chi-square	64.0731	0.0001	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X1_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-50.4901	0.0000	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-10.2077	0.0000	14	70
ADF - Fisher Chi-square	61.2312	0.0003	14	70
PP - Fisher Chi-square	82.8579	0.0000	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.8 – Results of stationary tests for households - level of internet access within 2 cluster (X1)

Panel unit root test: Summary
Series: X2_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-6.40555	0.0000	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	1.67068	0.9526	14	84
ADF - Fisher Chi-square	18.8404	0.9030	14	84
PP - Fisher Chi-square	37.8204	0.1018	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Continuation of Annex C

Panel unit root test: Summary
Series: D(X2_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-5.11255	0.0000	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-1.28006	0.1003	14	70
ADF - Fisher Chi-square	41.4251	0.0491	14	70
PP - Fisher Chi-square	118.201	0.0000	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.9 – Results of stationary tests for employed ICT specialists within 2 cluster (X2)

Panel unit root test: Summary
Series: X3_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-1.25409	0.1049	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	2.62582	0.9957	14	84
ADF - Fisher Chi-square	9.80975	0.9994	14	84
PP - Fisher Chi-square	18.1504	0.9223	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X3_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-8.02472	0.0000	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-1.72353	0.0424	14	70
ADF - Fisher Chi-square	47.6090	0.0118	14	70
PP - Fisher Chi-square	81.4180	0.0000	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.10 – Results of stationary tests for use of computers and the internet by employees within 2 cluster (X3)

Continuation of Annex C

Panel unit root test: Summary
Series: X4_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-5.80476	0.0000	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	0.21490	0.5851	14	84
ADF - Fisher Chi-square	31.6436	0.2893	14	84
PP - Fisher Chi-square	40.7448	0.0567	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(X4_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-4.89259	0.0000	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-6.31733	0.0439	14	70
ADF - Fisher Chi-square	42.6968	0.0372	14	70
PP - Fisher Chi-square	111.239	0.0000	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.

Figure C.11 – Results of stationary tests for value of e-commerce sales within 2 cluster (X4)

Panel unit root test: Summary
Series: X5_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-15.4448	0.0000	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-2.19359	0.0141	14	84
ADF - Fisher Chi-square	49.7006	0.0070	14	84

PP - Fisher Chi-square	56.2022	0.0012	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Continuation of Annex C

Panel unit root test: Summary
Series: D(X5_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-2.01044	0.0222	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	0.47543	0.0085	14	70
ADF - Fisher Chi-square	24.9753	0.6292	14	70
PP - Fisher Chi-square	66.8053	0.0001	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.12 – Results of stationary tests for integration of internal processes within 2 cluster (X5)

Panel unit root test: Summary
Series: Y1_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-2.43848	0.0074	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-0.25550	0.3992	14	84
ADF - Fisher Chi-square	31.1497	0.3105	14	84
PP - Fisher Chi-square	50.6281	0.0055	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(Y1_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-0.12814	0.4490	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-9.10928	0.0465	14	70
ADF - Fisher Chi-square	29.5050	0.3873	14	70

PP - Fisher Chi-square	94.5903	0.0000	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.13 – Results of stationary tests for the Gini index within 2 cluster (Y1)

Continuation of Annex C

Panel unit root test: Summary
Series: Y2_LN
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-4.60273	0.0000	14	84
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-1.02560	0.1525	14	84
ADF - Fisher Chi-square	41.5065	0.0482	14	84
PP - Fisher Chi-square	50.2596	0.0060	14	98

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Panel unit root test: Summary
Series: D(Y2_LN)
Sample: 2011 2018
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-46.1123	0.0000	14	70
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-15.5651	0.0000	14	70
ADF - Fisher Chi-square	69.2244	0.0000	14	70
PP - Fisher Chi-square	84.8951	0.0000	14	84

** Probabilities for Fisher tests are computed using an asymptotic Chi
-square distribution. All other tests assume asymptotic normality.

Figure C.14 – Results of stationary tests for the Palma index within 2 cluster (Y2)