sciendo

Statistical and Econometric Analysis of Selected Effects of COVID-19 Pandemic

doi:10.2478/mape-2021-0036
Date of submission to the Editor: 04/2021
Date of acceptance by the Editor: 06/2021

MAPE 2021, volume 4, issue 1, pp. 395-407

Wojciech Kempa
ORCID ID:0000-0001-9476-2070
Silesian University of Technology, **Poland**

Joanna Rydarowska-Kurzbauer
ORCID ID:0000-0001-7024-9088
Silesian University of Technology, **Poland**

Marzena Halama
Silesian University of Technology, **Poland**

Elżbieta Smuda
Silesian University of Technology, **Poland**

Maciej Biel
Silesian University of Technology, **Poland**

INTRODUCTION

One of the issues widely discussed in contemporary economic literature is that of the causes and nature of changes in economic activity. It can certainly be said that variability is an immanent feature of economic development. On the one hand, changes may create new opportunities for development and growth for economies, on the other hand, they may pose a serious threat to them. Although the history of research on the variability of economic processes is already quite long, there is still no unambiguous explanation of their causes. The literature mentions many factors triggering these changes, both of economic, political and social nature. However, these factors also change over time, and are often unpredictable, hence there is a constant need to study these phenomena. Currently, the COVID-19 pandemic is such a trigger. This pandemic situation is an unexpected and unpredictable event, with very strong consequences on a global scale. According to many analysts, it can be classified as a 'black swan event' (Taleb, 2008). According to this author, events of this type are distinguished by attributes such as: unpredictability and unusual nature, causing a huge impact on the economies. To explain pandemic as unpredictable and unusual situation, the results of a report prepared by the experts of World Economic Forum in Davos can cited. This report lists the top ten events that pose the most significant threat to world economic development that could occur in 2020. There is no pandemic has been identified, unlike other factors such as climate change or trade wars (see World Economic Forum Global Risks Survey 2019-2020). As regards the impact of a pandemic, it can be concluded that this phenomenon brings losses on many levels: it threatens health and human life,

Contact address: joanna.rydarowska-kurzbauer@polsl.pl

changes a family and professional life, and finally affects economic prosperity. Research on the scale and nature of these impacts is now the leading problem of many scientific papers. Looking at the statistics, the COVID-19 lockdown is one of the biggest shocks to GDP for many economies around the world (Buera et al., 2021). For this reason, it seems interesting to present the effects of the temporary lockdowns of a large part of the economies in response to the outbreak of the COVID-19 pandemic in many countries around the world. It is also reasonable to ask whether the pre-pandemic similarities or differences between economies have been distorted by the increasing incidence of COVID-19. Therefore, the aim of this paper is to examine the impact of the pandemic, in its initial phase, on the economic activity of selected European countries.

METHODOLOGY OF RESEARCH

The research are based on monthly or quarterly indicators of gross domestic product (GDP), unemployment rates and key indicators of the tourism sector, such as the numbers of accommodations in hotels, boarding houses, apartments. To present how COVID-19 has affected these macroeconomic variables, statistic data from the three periods are compared. Namely, data are collected from the pre-pandemic period, i.e. the fourth quarter of 2019 as the reference period, the second period covers the first quarter of 2020 and means the beginning of the pandemic, and the third one covers second quarter of 2020, during which the pandemic has spread to all over the world. Statistical data were obtained from 25 European countries such as: Austria, Belgium, Bulgaria, Cyprus, Czechia, Denmark, Estonia, Finland, France, Greece, Spain, Netherlands, Ireland, Lithuania, Luxembourg, Germany, Poland, Hungary, United Kingdom, Italy, Portugal, Slovakia, Slovenia, Switzerland, Sweden. Central Statistical Offices of these countries are the main data sources and for all calculations Statistica version 13.3. is used. There are many statistical methods that describe and measure correlation between variables. In this paper, the following statistical techniques have been selected: regression analysis, the hierarchical grouping of agglomerations, k-means method, and selected non-parametric tests (Kruskal-Wallis test for a selected group of countries and Kolmogorov-Smirnov test for a selected pair of countries).

One of the most important statistical tools used in modeling the direction and nature of the relationship between two or more statistical variables is classical regression analysis, executed with using the least squares method. Since in practice it very often happens that a multidimensional random sample has a normal (multivariate) distribution, as a rule, the search for a functional relationship between the examined variables begins with a linear function (Peck et al., 2008, Montgomery & Runger, 2003, Dekking et al., 2005). Of course, the determination of the structural parameters of the regression function, which is an analytical approximation of the unknown relationship between the variables, should be preceded by examining the strength of the correlation between the features. The squared classic Pearson correlation coefficient, calculated on the

basis of the results of a random sample, is used as a determination coefficient for regression fitting estimation. Among the classical techniques used in data mining agglomeration cluster analysis is frequently used. It is a grouping tool that allows you divide a set of objects into disjoint clusters that bring together objects similar to each other, while maintaining the greatest possible separation of individual clusters from each other (Dunham, 2003, Murphy, 2012). The process of agglomeration grouping begins with joining two objects that are the most "similar" to each other (in this way the first cluster is created) and ends with the building of one large cluster containing all analyzed objects. Of course, depending on the interpretative needs, the grouping process should be stopped at the appropriate moment (i.e. stopped at a specific number of clusters). Various methods are used here: the criterion of the maximum difference between distances at which successive joins are executed, Grabiński's measure, Mojena's rule etc.

An alternative clustering technique is the k-means method (Dunham, 2003, Murphy, 2012). Its essence is to establish a priori the target number of clusters into which the entire group of objects is to be divided. The k-means grouping process begins with the selection of the so-called initial cluster centers (different criteria can be used here) and is iterative. In the first step of the algorithm, the distances of individual objects from the designated centers are calculated. Objects are classified into the clusters to which they are "closest". In the next step of the iteration, the coordinates of new cluster centers are determined on the basis of the objects that were included in them. The distances of objects from the new cluster centers are recalculated. If there is such a need, the object is then transferred to the cluster to which it is "closer". This is how the remaining steps of the algorithm work. At some moment, the objects stop being moved anymore – they are in those clusters to the centers of which they have the shortest distances. Moreover, The Kolmogorov-Smirnov and Kruskal-Wallis statistical tests (Peck et al., 2008, Montgomery and Runger, 2003, Dekking et al., 2005) are used to verify the null hypotheses about the identical distribution of the unemployment rate in selected countries during the Covid-19 pandemic.

RESULTS AND DISCUSSION

This section presents the main results of the research carried out using the following statistical techniques: regression analysis, hierarchical agglomeration-type clustering, k-means method, and selected non-parametric tests.

Linear regression fitting

The regression analysis is performed using a one-dimensional linear model. The following variables are taken into account in the regression analysis:

- monthly changes in unemployment rate (as a dependent variable);
- monthly changes in number of accommodations in hotels, boarding houses, apartments (as a dependent variable);

- the daily changes in number of COVID-19 infections, aggregated to monthly values (as an independent variable).

Table 1 presents the values of the determination coefficients that determine the fitting of the obtained regression line to the empirical data. In addition, regression fittings is also given after removing the most outlier case, using the so-called the removed rest's criterion. The determination coefficient (a squared Pearson correlation coefficient) is calculated to examine the fitting of the linear regression line between the number of COVID-19 infections and the level of unemployment rate and between the number of COVID-19 infections and selected indicators of the tourism sector. The results are presented in Table 1. As it can be observed, values vary from country to country. The highest results are for Austria, Germany and Poland. This implies a strong positive correlation between the selected variables and the number of COVID-19 cases. On the other side there are conuntries with the lowest values of linear regression fitting. There are such countries as: Greece and Switzerland. Unfortunately there were not enough available data for Spain, Ireland, United Kingdom, Portugal, Slovenia, and therefore results for these countries are not included in the Table 1.

Table 1 Linear regression fitting summary

Country	**Number of accommodations in hotels, boarding houses, apartments**		**Unemployment rate**	
	Regression fitting	Regression fitting after removing the most atypical case	Regression fitting	Regression fitting after removing the most atypical case
Austria	93.5%	92.9%	77.2%	93.6%
Belgium	41.9%	88.7%	-	-
Bulgaria	73.3%	77.6%	87.8%	98.3%
Czechia	62.6%	91.4%	-	-
Cyprus	86.6%	99.0%	95.5%	97.1%
Denmark	-	-	98.6%	99%
Estonia	56.9%	93.5%	-	-
Finland	77.4%	77.7%	77.7%	96.7%
France	-	-	27.8%	33.2%
Germany	89.1%	89.9%	49.2%	69.1%
Greece	15%	69.6%	27%	83.8%
Hungary	75%	91.5%	20.4%	96.6%
Italy	51.9%	92.4%	65.3%	98.7%
Lithuania	64%	68.3%	98.3%	99.2%
Luxemburg	35.9%	77.9%	-	-
Poland	90.5%	99.4%	49.2%	69.9%
Netherlands	38.7%	50.7%	-	-
Slovakia	79.8%	92.7%	94.3%	95.3%
Switzerland	16%	58%	24%	70.5%

Source: own research based on Central Statistical Offices of selected European countries

The k-means algorithm

Using the k-means algorithm, 25 selected European countries are divided into separate groups (clusters), assuming k = 3, 4 and 5 as the target number of groups. The following variables are taken into account:

- quarterly changes in unemployment rate;
- quarterly changes in GDP;
- quarterly changes in number of accommodations in hotels, boarding houses, apartments;
- the daily changes in number of COVID-19 infections, aggregated to quarterly values.

Tables 2, 3 and 4 present groups of countries with similar responses profile to the COVID-19 pandemic, depending on the number of clusters applied.

Table 2 Summary of k-means clustering results for fourth quarter of 2019

Grouping result for k = 3		
Group 1	Group 2	Group 3
Austria Belgium Bulgaria Cyprus Finland France Ireland Lithuania Luxembourg Poland Portugal Slovakia Slovenia Sweden	Czechia Denmark Estonia Netherlands Germany Hungary United Kingdom Switzerland	Greece Spain Italy

Grouping result for k = 4			
Group 1	Group 2	Group 3	Group 4
Austria Belgium Bulgaria Cyprus Finland France Ireland Lithuania Poland Portugal Slovakia Slovenia Sweden	Luxembourg	Greece Spain Italy	Czechia Denmark Estonia Netherlands Germany Hungary United Kingdom Switzerland

Grouping result for k = 5				
Group 1	Group 2	Group 3	Group 4	Group 5
Greece Spain	Luxembourg	Austria Belgium Bulgaria Cyprus Finland France Ireland Lithuania Poland Portugal Slovakia Slovenia Sweden	Czechia Denmark Estonia Netherlands Germany Hungary United Kingdom Switzerland	Italy

Source: own research based on Central Statistical Offices of selected European countries

Table 3 Summary of k-means clustering results for the first quarter of 2020

Grouping result for k = 3		
Group 1	Group 2	Group 3
Spain Italy	France Netherlands Germany United Kingdom	Austria Belgium Bulgaria Cyprus Czechia Denmark Estonia Finland Greece Ireland Lithuania Luxembourg Poland Hungary Portugal Slovakia Slovenia Switzerland Sweden

Grouping result for k = 4			
Group 1	Group 2	Group 3	Group 4
Spain Italy	France Netherlands Germany	Belgium United Kingdom Switzerland	Austria Bulgaria Cyprus Czechia Denmark Estonia Finland Greece Ireland Lithuania Luxembourg Poland Hungary Portugal Slovakia Slovenia Sweden

Grouping result for k = 5				
Group 1	Group 2	Group 3	Group 4	Group 5
Spain Italy	France Netherlands Germany	Belgium United Kingdom Switzerland	Austria Portugal Sweden	Bulgaria Cyprus Czechia Denmark Estonia Finland Greece Ireland Lithuania Luxembourg Poland Hungary Slovakia Slovenia

Source: own research based on Central Statistical Offices of selected European countries

Table 2 shows the results before COVID-19 pandemic. It could be noticed that, regardless of the number of clusters applied, there is a clearly distinguishable group of similar economies in terms of the variables examined.
This group consists of Greece, Spain and Italy. There are also the recurrences of two groups of countries which have also been grouped independently on the number of clusters. The results are different, when comparing those from pre-pandemic period with the first and second quarter of 2020, (presented in Table 2 and 3 respectively). Table 3, referring to the first quarter of 2020, shows the maintenance of a group consisting of Italy and Spain.

Table 4 Summary of k-means clustering results for the second quarter of 2020

Grouping result for k = 3				
Group 1	Group 2		Group 3	
Germany United Kingdom Italy	Belgium France Portugal Sweden		Bulgaria Cyprus Czechia Denmark Estonia Finland Greece Netherlands Ireland Lithuania Luxembourg Poland Hungary Slovakia Slovenia Switzerland	
Grouping result for k = 4				
Group 1	Group 2	Group 3	Group 4	
Spain, Germany, United Kingdom	France Italy Poland	Austria Belgium Netherlands Ireland Portugal Switzerland Sweden	Bulgaria Cyprus Czechia Denmark Estonia Finland Greece Lithuania Luxembourg Hungary Slovakia Slovenia	
Grouping result for k = 5				
Group 1	Group 2	Group 3	Group 4	Group 5
Spain Germany United Kingdom	France Italy Poland	Belgium Portugal Sweden	Austria Denmark Netherlands Ireland Switzerland	Bulgaria Cyprus Czechia Estonia Finland Greece Lithuania Luxembourg Hungary Slovakia Slovenia

Source: own research based on Central Statistical Offices of selected European countries

However, among the remaining countries there is a new division due to similarities in response at the start of the pandemic. The first group of countries: France, the Netherlands and Germany (group 2 for k = 4 and 5), while the second included Belgium, the United Kingdom (UK) and Switzerland (group 3 for k = 4 and 5). It should also be added that group 1 in Table 3, is the group in which the pandemic caused the greatest change. In retrospect, there is known, that Italy was the first and the most affected European country in terms of a drastic increase in the spread of the disease. On the other hand, the lowest number of cases with a high GDP and low unemployment was combined in group 2 (Table 2 and 3). A completely different picture is presented in Table 4, which refers to the second quarter of 2020. In this period, completely new country groups were created, the first with Spain, Germany and the UK (group 1 for k = 4 and 5) and the second with France, Italy and Poland (group 2 for k = 4 and 5). At that time, countries of the first group had the highest number of of Covid-19 infections. There has been also a significant decline in GDP as well. As the number of sick people increased, the economic crisis deepened.
It can be stated that the k-means method used, allows to conclude that the development of the COVID-19 pandemic has a significant impact on the economic activity of selected European countries. A new grouping of countries is created due to the similarity in changes of the analysed macroeconomic variables. The analysis of the data also leads to the conclusion that the progressive increase in the incidence of the disease deepened the decline in GDP, output of the tourism sector and the unemployment rate.

The hierarchical clustering

Figures 1, 2 and 3 present the process of grouping 25 countries in terms of the similarity between the GDP value and the unemployment rate. The following variables are taken into account:

- quarterly changes in unemployment rate;
- quarterly changes in GDP;
- quarterly changes in number of accommodations in hotels, boarding houses, apartments;
- the daily changes in number of Covid-19 infections, aggregated to quarterly values.

Figure 1 shows the similarities in the level of economic activity of selected European countries, during the period before the COVID-19 pandemic. On the dendrogram of hierarchical clustering a large group of countries with small bond distances can be seen. This group, named as group 1 includes the following countries: Austria, France, Belgium, Luxembourg, Ireland, Slovenia, Bulgaria, Cyprus, Portugal, Sweden, Finland, Lithuania, Poland, Czechia, Switzerland, Denmark, Hungary, United Kingdom, Estonia, Germany, Netherlands. Countries such as Italy, Greece and Spain are clearly in contrast (reflecting larger distant bonding) to the other economies.

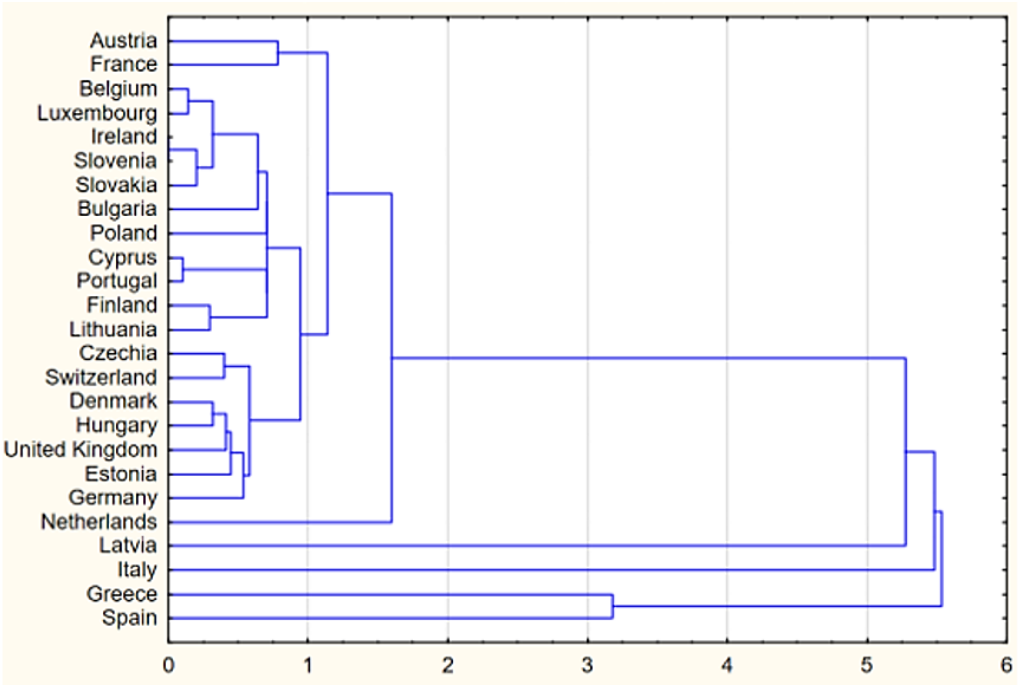

Fig. 1 The pairs of the most similar countries in fourth quarter of 2019

This result is consistent with those obtained by the k-means method (Table 2). During the first quarter of 2020 (Figure 2), the similarities in response to the pandemic are particularly visible in the following group 1 of countries: Slovakia, Lithuania, Hungary, Cyprus, Estonia, Slovenia, Finland, Greece, Luxembourg, Poland, Czechia, and Denmark. There are also ovbserved the other similarities among Belgium and Switherland. However, some countries stand out from the group 1.

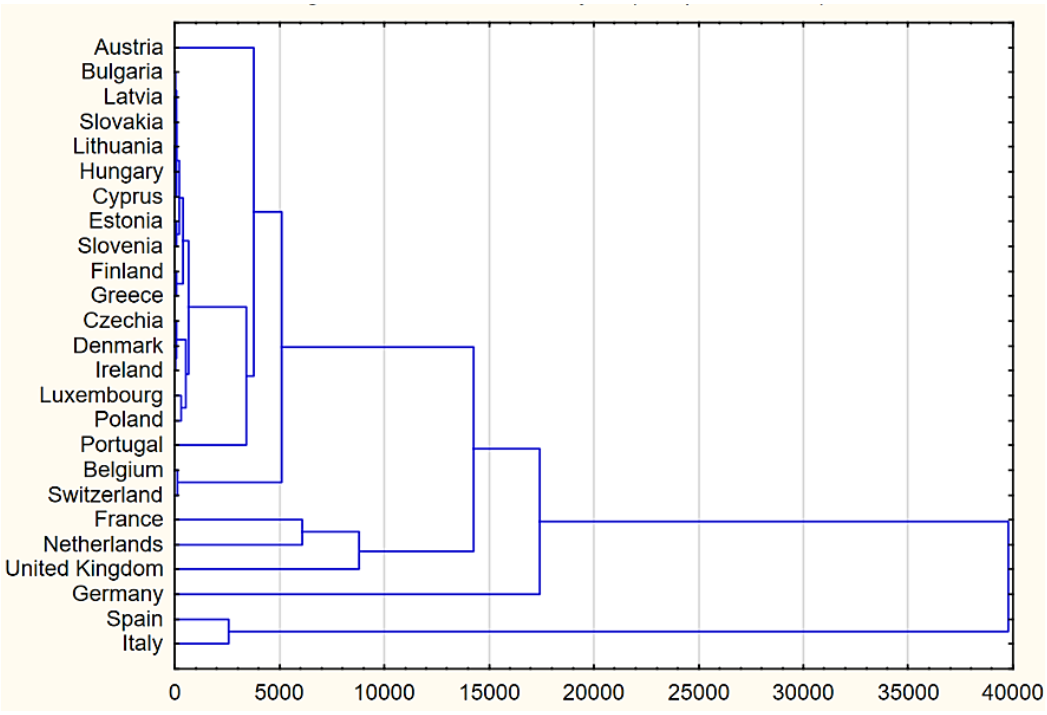

Fig. 2 The pairs of the most similar countries in first quarter of 2020

These are France, the United Kingdom, Germany, Spain, and Italy. So, the situation has changed when compared to the period before the COVID-19 pandemic. The clear similarity of the situation for Italy and Spain is apparent, and at the same time they are characterised by a large distance from the other economies. This is consistent with the results obtained using the k-means method. Hence, it can be stated that the most important variable of changing economic activity in this period, seems to be the number of COVID-19 infections. Figure 3 shows the situation characterising the second quarter of 2020.

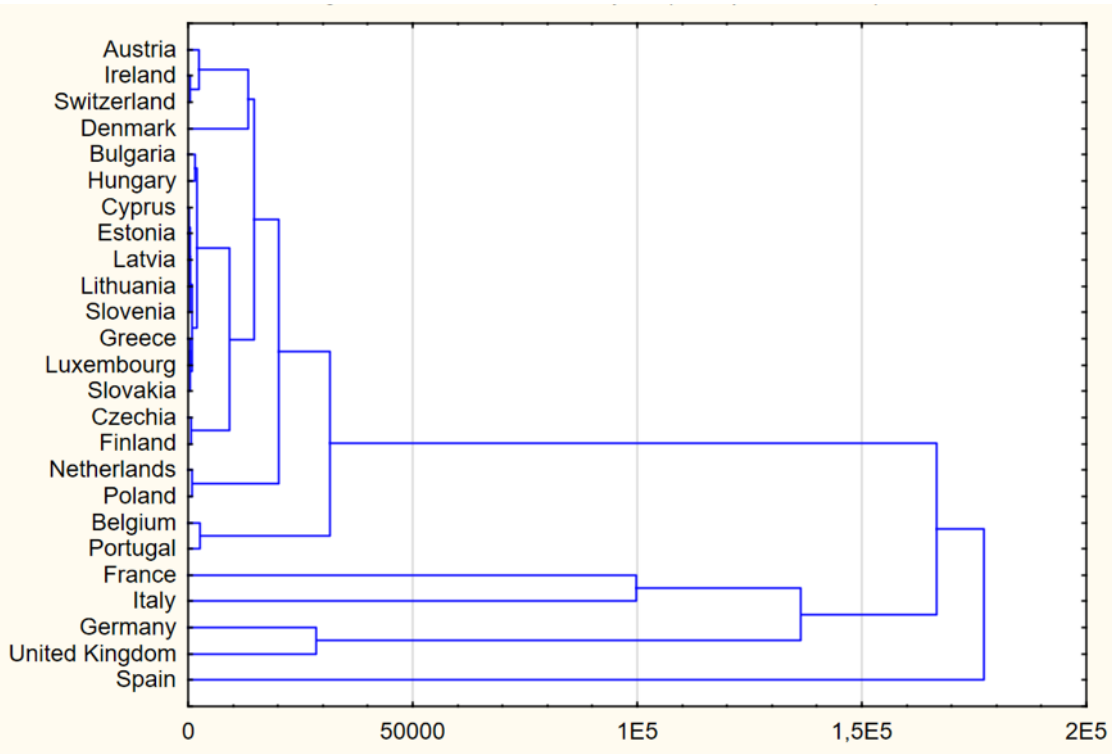

Fig. 3 The pairs of the most similar countries the second quarter 2020

The graph shows a slightly different clustering of countries compared to the previous period. Group 1 is basically unchanged. Countries such as Italy, France, Germany, UK and Spain again stand out from the other countries, which is also partly the result when using the k-means method. The similarity between Belgium and Portugal is also confirmed. It should also be noted that the situation of Poland has changed and it has been paired with the Netherlands, which is not consistent with the result obtained using the k-means method. But, again, like in k-means method, it can be stated that during the first two quartera of 2020, the increase in infection cases declines the macreconomic performance and changes the he distribution of similarities between economies.

Statistical tests

The Kolmogorov-Smirnov test examines the difference between cumulative distribution functions of two populations. This test is carried out for the comparison of the unemployment rate for the following pairs of countries: Poland

with Italy, Spain with United Kingdom, Germany with Spain, Belgium with Switzerland, Greece with Spain, and Spain with Italy. The results are always unambiguous. Table 6 presents the results. It can be noticed, that the p-value is very low, lower than 0.001. Therefore, taking the singificance level 0.05, the null hypothesis that the distributions of the unemployment rate in these countries in the analyzed period are the same has been rejected. Of course, rejecting the null hypothesis does not mean that the values of the unemployment rate cannot be similar in practical observations.

Table 6 The Kolmogorov-Smirnov test for comparison distributions of unemployment rate (countries are chosen basing on k-means clustering results for the second quarter of 2020)

Group 2 (k = 5), results for Poland and Italy									
variable	The Kolmogorov-Smirnov test, relative to the variable: state The marked results are significant with $p < ,05000$								
	max neg. sub.	max pos. sub.	p	Average Poland	Average Italy	stan. dev. Italy	stan. dev. Poland	N imp	N imp
Unemployment rate	-1.0000	0.00	p< 001	5.361538	9.40000	0.287340	0.832666	13	13
Group 1 (k = 3), results for Unitated Kingdom and Spain									
	The Kolmogorov-Smirnov test, relative to the variable: state The marked results are significant with $p < ,05000$								
	max neg. sub.	max pos. sub.	p	Average UK	Average Spain	stan. dev. UK	stan. dev. Spain	N Imp UK	N Imp S
Unemployment rate	-1.0000	0.00	p< 001	3.881538	14.06154	0.065044	0.256380	13	13
Group 1 (k = 4), results for Germany and Spain									
	The Kolmogorov-Smirnov test, relative to the variable: state The marked results are significant with $p < ,05000$								
	max neg. sub.	max pos. sub.	p	Average Germany	Average Spain	stan. dev. Germany	stan. dev. Spain	N Imp G	N Imp S
Unemployment rate	0.00	1.0000	p< 001	3.369231	14.06154	0.256380	0.406990	13	13

Source: own research based on Central Statistical Offices of selected European countries

Using the Kruskal-Wallis test, the null hypothesis about the identity of the unemployment rate distributions in the analyzed period for two groups of countries is verified. As a result of these tests, it is obtained the p-value equal to zero, which means that there are significant differences between the distributions of the unemployment rate in the selected countries. The results are presented in Tables 7 and 8.

Table 7 Kruskal-Wallis test for unemployment rate of the most similar countries (second quarter of 2020) (GROUP 4, k = 5, Q2)

Country	Kruskal-Wallis test: Unemployment rate Independent variable: State Kruskal-Wallis test: (4, N = 65) = 58.38904 p = 0.000			
	Code	N important	Sum of the rang	Average Rang
Austria	106	13	767.0000	59.0000000
Denmark	107	13	444.0000	34.1538462
Netherlands	108	13	259.0000	19.9230769
Ireland	109	13	572.0000	44.00000
Switzerland	110	13	103.0000	7.92308

Source: own research based on Central Statistical Offices of selected European countries

Table 8 Kruskal-Wallis test for unemployment rate of the most similar countries (second quarter of 2020) (GROUP 5, k = 5, Q2)

Country	Kruskal-Wallis test: Unemployment rate Independent variable: State Kruskal-Wallis test: H (10, N = 143) = 126.5632 p = 0.000			
	Code	N important	Sum of the rang	Average Rang
Bulgaria	101	13	942.500	72.5000
Cyprus	102	13	1160.000	89.2308
Czechia	103	13	105.000	8.0769
Estonia	104	13	539.000	41.4615
Finland	105	13	1147.500	88.2692
Greece	106	13	1781.000	137.0000
Lithuania	107	13	1540.000	118.4615
Luxembourg	108	13	835.000	64.2308
Hungary	109	13	249.000	19.1538
Slovakia	110	13	605.000	46.5385
Slovenia	111	13	1392.000	107.0769

Source: own research based on Central Statistical Offices of selected European countries

CONCLUSION

The paper examines the impact of the COVID-19 pandemic on macroeconomic activity in the selected European countries. In each of the tests there are obtained similar results. The results show the significant impact of the pandemic on the level of gross domestic product, unemployment rate and turism sector. In most cases, a correlation between number of COVID-19 infections and unemployment rate and GDP is observed. The statistical techniques used also allow to demonstrate the similarities and differences in the response of the economies to the COVID-19 pandemic.

REFERENCES

F.J. Buera, R.N. Fattal-Jaef, H. Hopenhayn, P.A. Neumeyer, Y. Shin (2021), The economic ripple effects of COVID, NBER Working Paper No 28704.

F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester (2005). A Modern Introduction to Probability and Statistics. London: Springer-Verlag.
M.H. Dunham (2003). Data Mining. Introductory and Advanced Topics. London: Prentice Hall/Pearson Education.
D.C. Montgomery, G.C. Runger (2003). Applied Statistics and Probability for Engineers. Third Edition. New York: John Wiley & Sons, Inc.
K.P. Murphy (2012). Machine Learning. A Probabilistic Perspective. Cambridge: The MIT Press.
R. Peck, Ch. Olsen, J. Devore (2008). Introduction to Statistics & Data Analysis. Third Edition. Belmont: Thomson Higher Education
N. Taleb (2008), The Black Swan. The impact of Highly Improbable, Penguine Books, New York (revised edition).
World Economic Forum Global Risks Survey 2019–2020 https:// reports.weforum.org/global-risks-report-2020/chapter-one-risks-landscape/ [Accessed 10 Apr. 2020]
F.J. Buera, R. N. Fattal-Jaef, H. Hopenhayn, P. A. Neumeyer, Y. Shin (2021), The economic ripple effects of COVID, NBER Working Paper No 28704.

Abstract: The paper examines the impact of the COVID-19 pandemic on macroeconomic activity in the selected European countries. The studies are based on monthly and quarterly indicators of GDP, unemployment rates and key indicators of the tourism sector. To present how COVID-19 has affected these macroeconomic variables, statistic data from the three periods are compared. Namely, data are collected from the pre-pandemic period, i.e. the fourth quarter of 2019 as the reference period, the second period covers the first quarter of 2020 and means the beginning of the pandemic, and the third one covers second quarter of 2020, during which the pandemic has spread to all the analyzed countries. The following statistical techniques are used in the research: regression analysis, the hierarchical grouping of agglomerations, k-means method, and selected non-parametric tests (Kruskal-Wallis test for a selected group of countries and Kolmogorov-Smirnov test for a selected pair of countries). The results show the significant impact of the pandemic on the level of gross domestic product, unemployment rate and turism sector. In most cases, a correlation between incidence of COVID-19 infections, unemployment rate and GDP is observed. The statistical techniques also allow to demonstrate the similarities and differences in the response of the economies to the COVID-19 pandemic. Central Statistical Offices of the selected countries are the main data source and for all calculations Statistica version 13.3. is used.

Keywords: cluster analysis, COVID-19 pandemic, GDP, regression, unemployment