Preliminary communication
UDC: 338:616.9-036.21COVID-19(4)
https://doi.org/10.18045/zbefri.2022.1.177

Impact of Covid-19 pandemic on economic performance of European Countries*

Marina Milanović[1], *Milan Stamenković*[2]

Abstract

Starting from the fact that the rapid spread of the SARS-CoV-2 virus and the implementation of social distancing strategies have dramatically affected all aspects of human lives at global, national, and micro levels, this paper focuses on examining the impact of the COVID-19 pandemic on the economic performance of selected European countries. To perceive and understand this impact, the authors applied a complex research methodology based on the combined application of suitable univariate and multivariate methods of statistical analysis. The classification of 40 European countries into different groups, in terms of the selected set of COVID-19 indicators in 2020, was performed through hierarchical agglomerative cluster analysis, while statistical evaluation of the quality of the obtained solution of a non-hierarchical procedure, based on the k-means method, was implemented. The classification consists of four clusters of countries identified as the "optimal" clustering solution. The authors conducted the analysis and comparison of profiles of the formed clusters of countries in terms of their average GDP growth rates in 2020 using the statistical methods of descriptive analysis and hypothesis testing. This study reveals that a cluster of countries with a relatively "lower" severity of the COVID-19 health consequences recorded a higher average GDP growth rate compared to groups of countries that suffered more serious consequences and vice versa. The obtained results, which indicate the connection between the

* Received: 27-01-2022; accepted: 27-06-2022

[1] *Assistant Professor, University of Kragujevac – Faculty of Economics, Department of Informatics and Quantitative methods, Liceja Knezevine Srbije 3, 34000 Kragujevac, Serbia. Scientific affiliation: application of quantitative methods in economics. Phone: +38134303536. E-mail: milanovicm@kg.ac.rs. ORCID: 0000-0002-6245-5313. Web page: https://www.ekfak.kg.ac.rs/sr/nastavnici/nastavnici-az?id=156&idd=453.*

[2] *Assistant Professor, University of Kragujevac – Faculty of Economics, Department of Informatics and Quantitative methods, Liceja Knezevine Srbije 3, 34000 Kragujevac, Serbia. Scientific affiliation: application of quantitative methods in economics. Phone: +38134303525. E-mail: m.stamenkovic@kg.ac.rs. ORCID: 0000-0003-0689-0369. Web page: https://www.ekfak.kg.ac.rs/sr/nastavnici/nastavnici-az?id=156&idd=382.*

magnitude of the negative health and economic consequences of the COVID-19 pandemic, can serve as additional support to policymakers in making decisions aimed at mitigating pandemic impacts and crisis management.

Key words: *Cluster analysis, COVID-19 indicators, GDP growth rate, European countries*

JEL classification: *C12, C38, I10, O47, O57*

1. Introduction

Every crisis, whether financial, economic, health, environmental, or any other, represents a phenomenon with negative (the short- and/or long-term) consequences in many aspects and levels. At the end of the second decade of the 21st century, the world faced an unexpected and an unprecedented crisis (Clemente-Suarez et al., 2021; Kempa et al., 2021). The corona-virus disease (widely known as COVID-19), caused by the SARS-CoV-2 virus, initially identified in December 2019 in Wuhan city of China (Hubei province), has spread worldwide in a short period of time. With the increase in the infected and laboratory-confirmed cases, this infectious disease was declared a pandemic by the World Health Organization (WHO) in March 2020.

However, the consequences of the pandemic outweigh the impact on public health and health systems. COVID-19 is not only the cause of health but also an economic and social crisis on a global scale, with significant implications for national and international security, which has, consequently, initiated appropriate responses at the international level by relevant organizations, as well as governments and policymakers around the world.

As the degree of impact of the pandemic varies from country to country, different countries have adopted and implemented different measures and strategies of social distancing and "lockdown" to address emerging problems. The undertaken measures (initially primarily focused on the health segment of the crisis and the suppression of the pandemic) have become one of the main causes of the reduction and slowdown of economic activity. In this regard, in many countries, the key negative consequences of these measures relate to a significant reduction in income, rising unemployment, and disruptions, mainly in transport, service, and manufacturing industries. The impact of the COVID-19 pandemic on economic activity is extremely broad: from dramatically diminished consumer discretionary spending to a "freeze" on business activities including capital budgets, hiring, and a reduction in everything but essential operational expenses (International Telecommunication Union, 2020). In other words, the alarming economic consequences are seen on the basis of data related to leading economic indicators on the reduction of production, employment, GDP, and others, at global, country, and micro levels.

Along with the serious consequences for public health and economic performance, perhaps one of the least expected (at first glance side) effects of the COVID-19 pandemic relates to the affirmation of the role of statistics and data. Actually in the current pandemic situation, statistical terms have become part of everyday vocabulary, and data are a key element in the selection and successful implementation of strategies and measures aimed at suppressing the spread of the virus and mitigating/overcoming its consequences. On this occasion, the director of the WHO, Tedros Adhanom Ghebreyesus, stated in one of his television appearances that "fighting a pandemic without data is as unintelligent as fighting blindfolded against fire" (Ljones, 2020). In this regard, in times of crisis, more than usual, reliable and trustworthy data and statistics are needed to make important decisions (Committee for the Coordination of Statistical Activities, 2020) and, in general, for efficient and effective crisis management.

According to the presented considerations, the research in this paper focuses on the negative impact of COVID-19 disease on the economic activity of selected European countries. In this context, the following research objectives were formulated: (1) a clear and thorough demonstration of statistically valid combined application of cluster analysis and hypothesis testing methods in the domain of the defined research subject; (2) classification of selected European countries into internally homogeneous / externally heterogeneous clusters based on the COVID-19 pandemic indicators, and (3) analysis and comparison of the profiles of the formed groups of countries from the perspective of the selected indicator of economic activity in 2020. Therefore, the research presented in the paper will test the following hypothesis: the decline in economic activity (measured by the average GDP growth rate) recorded in the cluster of European countries characterized by a relatively lower severity of health consequences of the COVID-19 pandemic is significantly smaller than the corresponding decrease recorded in the clusters of countries that suffered more serious health COVID-19 consequences. This paper makes several contributions to the literature. First, it provides a popularization of wide-ranging application possibilities of complementary usage of selected multivariate and univariate statistical methods, namely, cluster analysis and statistical hypothesis testing, in research of the economic consequences of the COVID-19 pandemic. Second, the results of this study can serve as a suitable basis for gaining a more complete insight into the COVID-19 pandemic health consequences and a better understanding of their implications on countries' economic activity. Finally, this study adds to the already extensive literature by filling the specific research gap, elaborated within Section 2 – Literature review.

Accordingly, the paper is structured as follows. After the introduction, the authors discuss several works related to the research subject in Section 2. Section 3 includes the used research methodology framework, while the description of the variables, sources, and definition of the spatial-temporal coverage of the data, as well as the obtained results of applied methodology, are presented in Section 4. Section 5 provides the results of comparison of the COVID-19 clusters of countries' profiles from the perspective of economic performance and discussion. Finally, concluding remarks are presented within the last Section 6.

2. Literature review

Various aspects and impacts of the COVID-19 pandemic have been the subject of numerous studies by the academic community. In this section of the paper, a brief overview of several bibliographic units relevant to the observed research area is presented, with a focus on (a) (macro) economic effects of the pandemic and (b) the role of statistical methods in the analysis of pandemic data, with a special reference to the application of cluster analysis.

Ehnts and Paetz (2021) consider and examine the economic consequences of the pandemic in the Euro Zone from the perspective of key macroeconomic indicators (i.e., real GDP growth rate and unemployment) and point out that without the intervention of governments and central banks around the world, the global economy would probably collapse. However, despite their quick response, compared to the global financial crisis 13 years ago, the short-term negative economic impacts of the pandemic are already stronger than those caused by the financial crisis. Praščević (2020) points out that the pandemic will have short-term and long-term macroeconomic effects and states that the macroeconomic costs and losses caused by the pandemic in the affected countries are related to GDP loss, rising employment rates, and deteriorating other macroeconomic indicators. Considering the alarming consequences that an increase in the unemployment rate has on the increment in poverty, inequality, and crime rates, Su et al. (2021) examine and quantify the impact of the pandemic on unemployment in selected, highly developed European economies, which are, surprisingly, among the top ten most vulnerable countries in the world in terms of the number of COVID-19 infected cases and confirmed deaths. The results of this study in the observed countries confirm the presence of causal relationships between COVID-19 variables and the unemployment rate.

Bhardwaj et al. (2020) consider, through the implementation of descriptive statistical analysis, the impact of the COVID-19 pandemic on GDP and unemployment rate in the world's largest economies (i.e., the United States of America, the People's Republic of China, the Republic of India), which had different strategies in the early stages of the pandemic. Pointing out the negative consequences of this impact, the same authors conclude that strategies of social distancing and a "lockdown" should be applied to fight against the spread of COVID-19 disease since a less negative impact was observed in the countries that have implemented this strategy. Kempa et al. (2021) examine the impact of the pandemic on macroeconomic activity in selected European countries and, for these purposes, demonstrate the application of different statistical methods. The obtained results confirm the significant impact of the pandemic on GDP, unemployment, and key indicators of the tourism sector. Similarly, in a study conducted by de la Fuente-Mella et al. (2021), statistical models have been developed to assess the effects of the crisis generated by the COVID-19 pandemic on the economic performance of certain countries, with the following variables included for the purpose of model specification: GDP growth rate, as a dependent variable and, COVID-19 cases per million inhabitants, Global Health Security Index (GHSI), risk factors for virus spread, country's membership in the OECD (yes / no) and GDP per capita, as independent variables.

Küçükefe (2020), based on data on the number of deaths per million inhabitants and the GDP growth rate for China and the OECD group of countries, develops a linear regression model and empirically shows the countries with the highest mortality rates experienced the largest economic decline. Also, the same author conducts cluster analysis using the k-means algorithm and finds three clusters of countries according to current account balances, GDP growth rate, and the number of COVID-19 deaths per million inhabitants. The analysis pinpoints that the countries with current account surplus managed to limit economic decline and mortality rates due to COVID-19 disease. Zarikas et al. (2020) present an approach that results in the grouping of countries on the basis of health aspects. It takes into account the following epidemic data: the number of active cases of COVID-19, the number of active cases per population, and the number of active cases per population and per area. These authors emphasize the importance of clustering results in terms of usefulness for experts and decision-makers in various fields, such as physicians and managers of the health sector, economy/finance experts, politicians, and even sociologists. Rizvi et al. (2021) present the application of an unsupervised k-means algorithm for grouping 79 countries based on 18 social, economic, health, and environmental variables (as factors related to the spread of COVID-19 disease) in order to implement the policies to control the widespread of disease. At the same time, in the same paper, using correlation analysis, the strength of the correlation between the selected non-COVID-19 and COVID-19 variables (i.e., the number of confirmed and dead cases) is evaluated.

Undoubtedly, separated research studies indicate that COVID-19 represents the most unusual recession in the modern world. Also, it was confirmed, in the methodological context, that clustering of countries to analyze different variables associated with a pandemic is a special topic of interest for researchers. In addition to the presented bibliographic units, mentioned statements and observations are confirmed by interesting research approaches presented, for example, in: Centre for Economic Policy Research (2020), Nicola et al. (2020), Nayak et al. (2021), Liu et al. (2021), Abdullah et al. (2021).

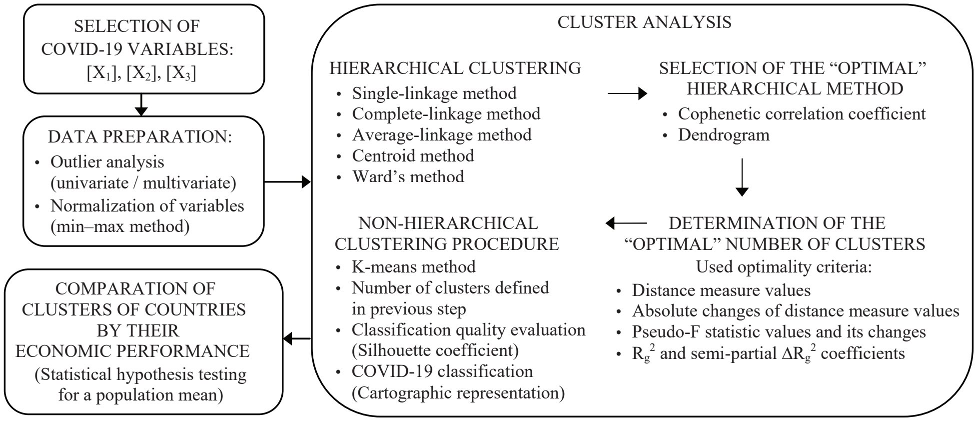
Regardless of the similarity of their research objectives, the common denominator of the previously described studies is the pronounced variability, present in terms of the following analytical issues: spatial-temporal scope of analysis, selection of groups of variables used (both, COVID-19 and economic, social or indicators of other development dimensions), applied statistical methods, and derived conclusions. In accordance with the aforementioned and defined objectives, the current research represents a specific combination of previously listed analytical aspects, which, according to the authors' knowledge, have not yet been exploited in the existing literature. The originality of this paper is reflected in the application of a specific methodological approach, based on the combined usage of cluster analysis and statistical hypothesis testing methods, in the domain of examining the relationship between COVID-19 health consequences and indicators of economic activity in selected European countries in 2020. By filling this research gap, the results of the conducted research, along with the conclusions drawn, will contribute to the enrichment of the literature, either through expansion or additional confirmation of the existing knowledge regarding the seriousness of economic consequences caused by the COVID-19 pandemic. A comparison of combinations of analytical issues emphasized in it. Previously described studies unequivocally confirm the statements regarding the originality of this study and the research niche it addresses.

3. Methodology

For the effective realization of formulated objectives, a complex research methodology is applied presented in Figure 1. The presented research framework is based on combined and complimentary usage of cluster analysis (CA), as the most frequently used method of multivariate statistical analysis, and statistical hypothesis testing about the population mean as a well-known method of statistical inference. More precisely, in the first step, a CA is used to investigate the interdependencies between carefully selected indicators of effects of the COVID-19 pandemic in 2020, and the discovery of a "natural", but hidden, classification structure within the analysed set of multivariate observations.

After statistical quality evaluation of the proposed classification of selected European countries into certain, previously unknown, a number groups, according to the used COVID-19 indicators, a comparison of the formed clusters of countries, in terms of their economic performance in 2020, is conducted using the procedure of statistical hypothesis testing. In accordance with the first research objective, as can be seen in the presented schematic representation, in conducting the described empirical research, special attention is dedicated to the examination of the fulfilment of assumptions on which the valid application of mentioned statistical methods is based.

Figure 1: Schematic representation of the used research methodology framework



Source: Authors' representation

Data analysis and the necessary statistical calculations were done using the statistical software package *IBM SPSS Statistics*, version 20, and *Microsoft Office Excel*. The final interpretation of the research results is complemented by adequate tabular, graphical, and cartographic representations.

4. Empirical data and analysis

In this section aspects of the conducted research in terms of selected COVID-19 variables, sources used as well as the temporal-spatial coverage of pandemic data are presented together with their descriptive statistics. Additionally, following the described two-stage methodological framework, after adequate preparation of input data and based on the use of listed objective statistical criteria of optimality and evaluation, the classification of 40 selected European countries into appropriate COVID-19 clusters, according to the severity of recorded health consequences of the pandemic, has been carried out.

4.1. Variables, sources of data, and temporal-spatial scope of the research

Following the defined objectives, for multivariate classification of selected European countries, daily data on the number of recorded positive cases of COVID-19 and the number of deaths caused by COVID-19 infections in the period from January to December 2020 were collected. The list of formed COVID-19 variables, used in cluster analysis, together with the procedure for determining their values, is given in Table 1. Data related to the health consequences of the COVID-19 pandemic and the size of the population of the analyzed countries were obtained from the electronic database of the European Centre for Disease Prevention and Control. To neutralize or mitigate the impact of the total demographic mass of individual countries on the outcome of CA and the resulting classification, instead of using absolute values, the values of COVID-19 variables are expressed per 100,000 inhabitants or in the form of an appropriate relative (%) participation. Taking into account the availability of data for the described COVID-19 variables, the following 40 European countries were included in the conducted empirical research, namely: EU-27 countries ([AUT], [BEL], [BGR], [CRO], [CYP], [CZE], [DNA], [EST], [FIN], [FRA], [GER], [GRE], [HUN], [IRL], [ITA], [LVA], [LTU], [LUX], [MLT], [NLD], [POL], [POR], [ROU], [SVK], [SLO], [ESP], [SWE]); candidate and potential candidate countries for EU membership ([SRB], [MKD], [MNE], [ALB], [BIH]), and countries that are not part of the EU by their decision ([ISL], [NOR], [GBR], [CHE], [MDA], [BLR], [UKR], [RUS]).

Table 1: List of the used COVID-19 indicators

Symbols	Variables	Units of measurement
X_1	COVID-19 cases per 100,000 inhabitants, in 2020	Number of cases
X_2	COVID-19 deaths per 100,000 inhabitants, in 2020	Number of deaths
X_3	COVID-19 mortality rate, in 2020	Percentage (%)

Notes related to the method and way of determining the values of individual COVID-19 variables:

- Values of variable X_1 are calculated as ratio of total number of confirmed COVID-19 cases in 2020 and total number of inhabitants for particular country, multiplied by 100,000;
- Values of variable X_2 are calculated as ratio of total number of confirmed COVID-19 deaths in 2020 and total number of inhabitants for particular country, multiplied by 100,000; and
- Values of variable X_3 are calculated as percentage participation of total number of confirmed COVID-19 deaths in total number of confirmed COVID-19 cases in 2020 for particular country;

Source: Authors' tabular representation

After classifying the observed countries into previously unknown clusters according to the severity of medical consequences of the COVID-19 pandemic in 2020, for subsequent analysis of the magnitude of economic consequences of the pandemic in their territories, by selected clusters, data on the recorded GDP growth rate in 2020, for each of the listed 40 countries, were also collected from the *EUROSTAT* electronic database.

4.2. Classification of European countries by COVID-19 variables

According to the described two-stage methodological research framework (Figure 1), using the presented COVID-19 variables, the authors conducted a cluster analysis of 40 European countries for their classification into internally homogeneous / externally heterogeneous COVID19–specific groups. In this sense, determining values of the arithmetic mean ($\bar{x}$), median (m_e), coefficient of variation (v), maximum (*max*) and minimum (*min*) of the used COVID-19 indicators' original values in 2020, are given in Table 2.

Table 2: Descriptive statistical measures of the used COVID-19 indicators

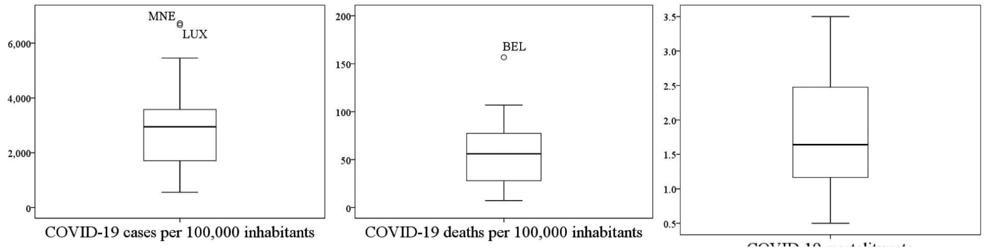
COVID-19 variables		$\bar{x}$	m_e	*min*	*max*	*v*
COVID-19 cases per 100,000 inhabitants	X_1	2,948.28	2,947.73	558.36 [FIN]	6,722.98 [LUX]	49.5%
COVID-19 deaths per 100,000 inhabitants	X_2	54.85	56.07	7.26 [NOR]	156.70 [BEL]	63.9%
COVID-19 mortality rate	X_3	1.83	1.64	0.50 [ISL]	3.50 [ITA]	46.7%

Source: Authors' calculations

Table 2, besides providing the basic description of the sample of countries and visual inspection of the statistical measures, also includes the following observations: (1) Approximate, almost identical values of positional and calculated measures of central tendency indicate, preliminary, the absence of one-dimensional outliers at the level of analyzed variables; and (2) High values of coefficients of variation, suggest the presence of a relatively high degree of heterogeneity of the analyzed countries in terms of values of individual COVID-19 variables, which indirectly meets the primary prerequisite for the justified application of CA. Also, since selected indicators are expressed in different measurement units, their normalization was conducted using the *min-max* method, thus converting original values of COVID-19 variables into normalized values ranging from 1 to 10.

Within the data pre-processing phase, the normalized values of COVID-19 indicators were examined for the possible presence of one-dimensional and/or multivariate non-standard observations, using graphical representations in the form of individual box plots and corresponding values of the Mahalanobis distance measure, respectively. In this sense, the resulting box plots for individual COVID-19 variables (Figure 2) suggest the absence of true one-dimensional outliers, thus confirming the validity of previously presented preliminary observations, made based on descriptive statistical measures.

Figure 2: Box-plots for individual COVID-19 variables



Source: Authors' representations

On the other hand, multivariate observations corresponding to the countries [BEL], [LUX], [IRL], [MNE], and [GRE] have been identified as outliers, as they are characterized by the values of Mahalanobis distance measure (i.e., $MD_{BEL} = 17.352$, $MD_{LUX} = 13.849$, $MD_{IRL} = 12.088$, $MD_{MNE} = 10.235$, $MD_{GRE} = 9.926$, respectively) which are above the defined critical values of the chi-square distribution ($\chi^2_{(3;\ 0.975)} = 9.348$). Precisely some of these countries have been singled out, on box plots, as one-dimensional potential outliers. The exclusion of identified multivariate outliers from the further analysis was not performed due to the importance and significance of taking into account their COVID-19 characteristics, but to see the magnitude of their atypical impact on clustering results, a CA was conducted again without their presence.

For the purpose of objective and statistically justified selection of the "most suitable" approach for classification of analysed countries, on pre-processed multivariate observations, using the squared Euclidean distance as an appropriate distance measure, the following hierarchical agglomerative clustering methods were applied: single-linkage, complete-linkage, average-linkage, centroid-linkage and Ward's method. By comparing the values of cophenetic correlation coefficient, calculated for the results of different clustering methods (Table 3), but also by visual analysis of the extracted dendrograms, the classification of countries obtained by the centroid method was singled out as the most interpretable and of the highest quality.

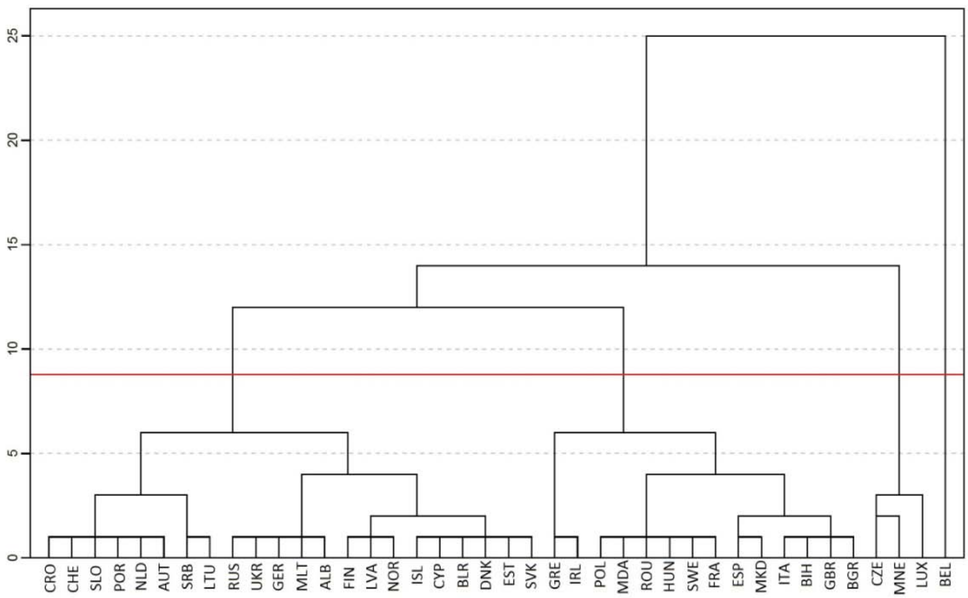
Table 3: Cophenetic correlation coefficient values for different clustering methods

Hierarchical agglomeration methods	single-linkage	complete-linkage	average-linkage	centroid-linkage	Ward's method
Cophenetic coefficient	0.489	0.437	0.609	*0.648*	0.572

Source: Authors' calculations

The complete structure of possible clustering solutions of the conducted hierarchical agglomerative classification of forty European countries, according to selected indicators of the severity of the COVID-19 pandemic consequences in 2020, is presented in Figure 3, in the form of a corresponding tree-diagram, known as a dendrogram.

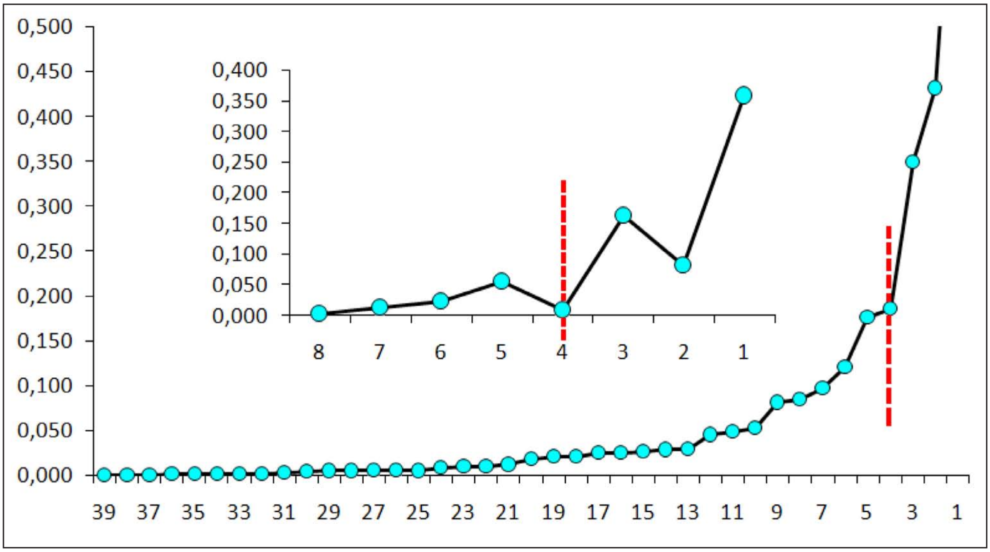
Figure 3: Graphical representation of the hierarchical structure of successively formed clusters



Source: Authors' representation

Since the development of a complete hierarchical tree is not the aim of conducted cluster analysis, already the identification and extraction of "natural" internally homogeneous and externally heterogeneous clusters, the selection of the "optimal" number of clusters within the presented dendrogram was made on the basis of appropriate optimality criteria, highlighted within the methodology framework. For this purpose, graphical representations of distance measure values, pseudo-F statistics and R_g^2 coefficient values, as well as their increments in successive steps of the agglomeration process, are given in Figures 4 and 5, respectively.

Figure 4: Graphical representations of distance measure values and its absolute changes



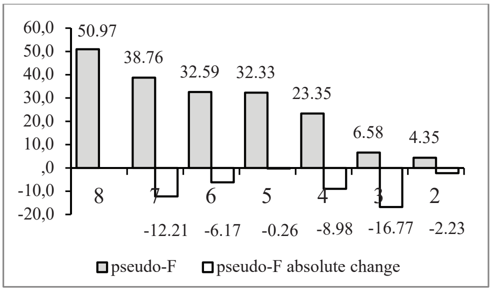
Source: Authors' representation

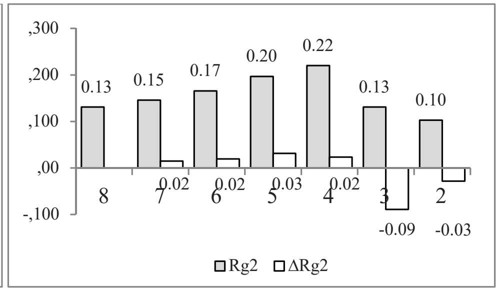
By visual analysis of the values of the used optimality criteria and corresponding graphical representations, the following conclusions can be drawn:

- The gradual but continuous, slight increase of the distance measure values between individual countries and / or groups of countries and, consequently, its absolute changes recorded in successive steps of the agglomeration process was interrupted at the time of forming a hierarchical solution consisting of three clusters (Figure 4). In fact, the value of distance measure at which merging groups of countries was performed in step 37 is 0.349 and is approximately twice the value recorded in the previous, 36th step of the agglomeration process (more precisely, 0.186). The magnitude of the observed increase in the distance value is even more evident if the fact that the increment of the distance value in step 37 (i.e. 0.163) is approximately 16 times higher than the comparable value recorded in step 36 (i.e. 0.010) is taken into account.
- By comparing the presented values of pseudo-F statistics (Figure 5, left), determined for solutions covering 8 to 2 clusters, the 37th step of the agglomeration process, in which a solution with three clusters is formed, was identified as a moment in which a drastic, clearly noticeable, negative change of pseudo-F statistic occurred (i.e. decrease from 23.35 to 6.58), followed, logically, by the largest (negative) increment of –16.77.

- Analogous to the previous conclusion, the first noticeable decrease of R_g^2 coefficient values, i.e., a significant increase of ΔR_g^2 semi-partial coefficient values, in relation to the formed solutions with a larger number of clusters, was recorded in step 37 (Figure 5, right).

Figure 5: Graphical representations of pseudo-F statistics (left) and R_g^2 / ΔR_g^2 values (right)





Source: Authors' representation

Since the described drastic change in the values of the used optimality criteria occurred as a result of merging highly heterogeneous groups, the number of clusters formed in a preceding step of agglomeration process, i.e. the classification consisting of four clusters of countries, is identified as the "optimal" clustering solution. The moment of formation of the "optimal" clustering solution and stopping the further hierarchical agglomeration is marked by a red horizontal line on the dendrogram. The distribution of countries by extracted clusters is presented in Table 4.

Table 4: Distribution of European countries by established clusters

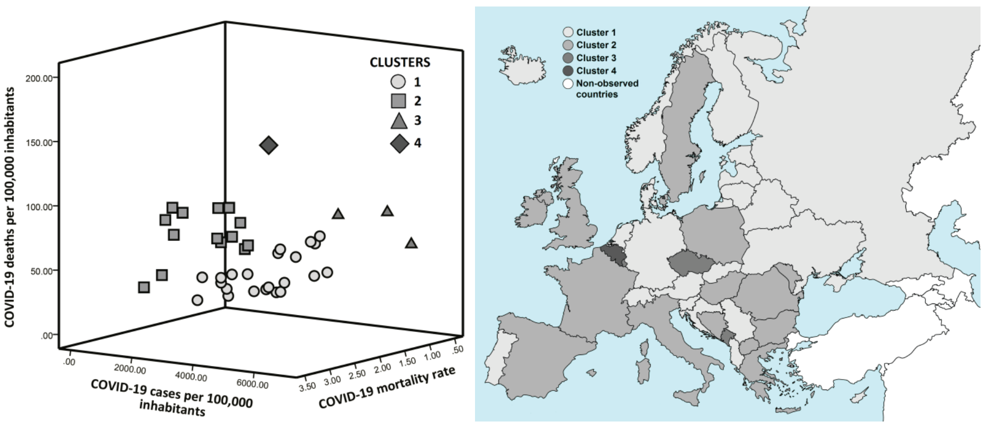
Cluster code	Severity of COVID-19 pandemic consequences	No. of countries	Countries within individual clusters
1	"Lower" level	22	[GER], [DNK], [FIN], [NLD], [POR], [AUT], [MLT], [SVK], [CYP], [LVA], [LTU], [EST], [SLO], [CRO], [NOR], [ISL], [CHE], [RUS], [BLR], [UKR], [SRB], [ALB]
2	Medium level	14	[FRA], [ESP], [ITA], [SWE], [GRE], [IRL], [ROU], [HUN], [POL], [BGR], [MKD], [BIH], [GBR], [MDA]
3	High level	3	[LUX], [CZE], [MNE]
4	Very high level	1	[BEL]

Source: Authors' tabular representation

To assess the magnitude of the impact of, by preliminary analysis identified, real multivariate outliers (i.e., [BEL], [LUX], [IRL], [MNE], and [GRE]) and [CZE], marked as suspected outlier, the hierarchical clustering procedure was re-conducted, without the mentioned observations. By comparing the obtained clustering solutions, no significant differences in the structure of the formed clusters were noticed, since these six multivariate observations were isolated, by the initial solution (Table 4) as members of the so-called outlier clusters (i.e., groups 3 and 4).

Visualization of the COVID-19 classification of 40 European countries, in the form of a 3D scatter diagram and cartographic representation, is given in Figure 6.

Figure 6: Visual representations of COVID-19 classification of selected European countries



Source: Authors' representations

For the purpose of statistical evaluation of quality and validity of hierarchical classification of European countries within four clusters, a non-hierarchical CA was carried out on normalized values of COVID-19 variables, using k-means method, since it, in contrast to hierarchical methods, allows reallocation of observation during the clustering procedure. More precisely, the comparison of hierarchical and non-hierarchical classification results with four clusters was performed based on the corresponding values of silhouette coefficient, calculated for individual clusters and the overall solution (Table 5).

Table 5: Silhouette coefficient values for different clustering solutions

Hierarchical clustering solution			Non-hierarchical clustering solution		
Cluster code	Number of countries	Silhouette coefficient	Cluster code	Number of countries	Silhouette coefficient
Cluster 1	22	0.495	Cluster 1	14	0.672
Cluster 2	14	0.556	Cluster 2	10	0.484
Cluster 3	3	0.839	Cluster 3	9	0.522
Cluster 4	1	1.000	Cluster 4	7	0.552
Overall	40	0.722	Overall	40	0.557

Source: Authors' calculations

The overall value of silhouette coefficient (0.722), as a statistical measure intended for a comprehensive evaluation of the quality of clustering results, calculated for hierarchical classification, is clearly higher than the comparable value of overall non-hierarchical alternative (i.e., 0.557), as a result of which it can be unequivocally concluded that the created hierarchical solution is characterized by better quality. In general, the quality level of the proposed hierarchical classification can be described as high, since the overall silhouette coefficient value (0.722) ranges from 0.70 to 1.00. The previous statement is also confirmed by the results of the comparison of silhouette values determined for individual clusters, since in the hierarchical alternative, in 3 out of 4 clusters, higher coefficient values are present.

5. Results and discussion

This section focuses on the interpretation of the formed COVID-19 clusters of countries, comparison of their profiles from the perspective of corresponding (average) GDP growth rates, and, consequently, discussion of the results of the evaluation process of the initially formulated research hypothesis.

5.1. Interpretation of the COVID-19 classification of European countries

For better insight and understanding of the magnitude of negative medical consequences caused by the COVID-19 pandemic in analyzed countries, in this section, the interpretation of the proposed classification (Table 4 and Figure 6) is performed, in terms of the average values of COVID-19 indicators in 2020 (Table 6). In addition, taking into account the relatively small size of clusters 3 and 4, as well as the fact that countries in their composition were identified as real / suspected multivariate outliers, for interpretation purposes, their integration into one common cluster (with code [3+4]) is done.

Table 6: Comparative review of average values of COVID-19 variables per clusters

Cluster code	Number of countries	COVID-19 cases per 100,000 inhabitants	COVID-19 deaths per 100,000 inhabitants	COVID-19 mortality rate
Cluster 1	22	2,446.24	31.70	1.26 %
Cluster 2	14	2,854.93	77.98	2.75 %
Cluster [3+4]	4	6,036.22	101.24	1.74 %
Overall	40	2,948.28	54.85	1.83 %

Source: Authors' calculations

Presented average values of COVID-19 indicators at the level of individual clusters fully justify the gradation of the severity of COVID-19 pandemic consequences and are used in assigning descriptive names to the formed clusters (Table 4). Specifically, the average number of COVID-19 confirmed cases per 100,000 inhabitants determined for countries in cluster 1 (i.e., "lower" severity of consequences) is smaller than the corresponding average values in cluster 2 (i.e., medium severity of consequences) and cluster [3+4] (i.e., high and very high level of severity) by approximately 14 % and 60 %, respectively. Similarly, the average value of this indicator in cluster 2 is 53 % lower than the comparable average in the cluster [3+4]. Generally, an identical ratio of average values of the considered groups of countries is present in terms of the variable COVID-19 deaths per 100,000 inhabitants, namely: cluster 1 < cluster 2, nearly 59%; cluster 1 < cluster [3+4] by ≈ 69%; and cluster 2 < cluster [3+4] by ≈ 23%. Also, it is important to note that only the average number of COVID-19 deaths per 100,000 inhabitants in cluster 1 is below the corresponding average of variable X2 for 40 European countries, compared to the other two clusters. The scale of the catastrophic consequences of the COVID-19 pandemic, recorded on the territory of countries within the cluster [3+4], is unequivocally evidenced by almost twice the average values of the first two COVID-19 indicators (X1 and X2) compared to the corresponding average values of 40 countries. Finally, viewed from the angle of the COVID-19 mortality rate (X3), the ranks assigned to the clusters based on correspondent average values are slightly different compared to the clusters ranking results in the case of the previous two COVID-19 indicators. More precisely, the average COVID-19 mortality rate of countries in cluster 2 is higher compared to the average values recorded in cluster 1 and [3+4], by approximately 118 % and 58 %, respectively. The mentioned value significantly exceeds the average COVID-19 mortality rate determined for 40 countries, in contrast to the comparable values at the level of the other two clusters.

5.2. COVID-19 pandemic and economic performance of European countries

In order to reveal / confirm the impact of previously described negative medical consequences of COVID-19 pandemic on economic performance of 40 European countries, a comparison of profiles of the formed clusters of countries from the perspective of average values of variable *GDP growth rates in 2020* (Table 7), as a key indicator of national economies' activity, is performed.

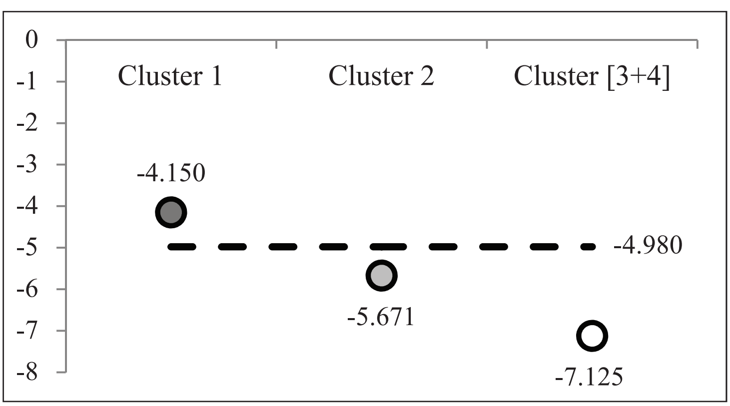
Table 7: Average and min/max values of the used economic indicator per formed clusters

Clusters		Severity of COVID-19 pandemic consequences	GDP growth rates in 2020 (in %)		
code	size		average	min	max
1	22	"Lower" level	–4.150	–9.0	–0.8
2	14	Medium level	–5.671	–11.0	+2.5
[3+4]	4	High and very high level	–7.125	–15.2	–1.3
Overall	40	-	–4.980	–15.2	+2.5

Source: Authors' calculations

A visual representation of the relation between average GDP growth rates per formed clusters and overall average GDP growth rate of 40 European countries is illustrated in Figure 7.

Figure 7: Average GDP growth rates per clusters (circles) and overall average (dashed line)



Source: Authors' representation

By comparing the presented (average) values of selected economic indicator, the following observations can be singled out:

- The average GDP growth rate of 40 observed European countries is a negative value.
- Individual average values of GDP growth rate in all three clusters are also negative.
- The average decline in economic activity in 2020, recorded in the countries within cluster 1, is smaller compared to the recorded size of the recession of the national economies classified within cluster 2, especially cluster [3+4].
- Compared to the average GDP growth rate determined for all 40 countries (–4.98 %), only cluster 1, in terms of corresponding value (–4.15 %) is positioned above (Figure 7), as opposed to cluster 2 and cluster [3+4], which are below the overall average.
- Although below the average growth rate calculated for 40 countries, the average percentage decline of GDP determined at the level of 14 countries in cluster 2 is still about 20% lower than the corresponding average of countries allocated within the cluster [3+4].
- The ranking of the formed clusters according to the achieved average GDP growth rates is completely identical to the results of their ranking conducted in terms of the severity of recorded health consequences of the COVID-19 pandemic, of course, taking into account the nature of the relationship between these two (i.e., medical and economic) phenomena. More precisely, it is evident that countries distributed within cluster 1 (i.e., relatively “lower” degree of severity of the COVID-19 consequences) also recorded the “highest” average GDP growth rate (i.e., the lowest negative value). Also, the highest negative average GDP growth rate is characteristic of countries, which, according to CA results, are allocated within the cluster that suffered the most serious consequences of the COVID-19 pandemic.

In order to examine the statistical significance of previously determined (practical) differences between the average GDP growth rates, at the level of individual clusters of countries, the parametric one-sample t test is applied. The justification of its application was confirmed by the results of testing hypotheses on the normality of distribution of variable GDP growth rate at the level of individual clusters (Table 8). In fact, in the case of all three groups of countries, the results suggest that the null hypothesis, which claims that distribution of GDP growth rates at the population level of specific (COVID-19) categories of countries is normal, cannot be rejected, as the resulting p-values are less than test significance level, $\alpha = 0.05$.

Table 8: Results of one-dimensional distribution normality testing

Economic variable	Clusters		Anderson-Darling normality test			
	code	size	statistic	p-value	α	decision
GDP growth rate in 2020	1	22	0.332	0.485	0.05	H_o
	2	14	0.290	0.558	0.05	H_o
	[3+4]	4	0.307	0.347	0.05	H_o

Source: Authors' calculations

The results of one-sample t test in the context of a defined research question, together with the corresponding statistical hypotheses and the logic of their formulation, are given in Table 9.

Table 9: Statistical hypotheses and results of one-sample t test

Research problem definition			One-sample t test			
Clusters compared	Alternative hypothesis	Hypothetical value (μ_0)	statistic	p-value	α	decision
cluster 1 vs. cluster 2	H_1: $\mu_1 > \mu_0$	–5.671	4.357	0.00014	0.05	H_1
cluster 1 vs. cluster [3+4]	H_1: $\mu_1 > \mu_0$	–7.125	8.523	0.00001	0.05	H_1
cluster 2 vs. cluster [3+4]	H_1: $\mu_2 > \mu_0$	–7.125	1.827	0.04537	0.05	H_1

Note: Parameter hypothetical values were determined on the basis of corresponding values of sample statistics.

Source: Authors' calculations

Since the obtained p-values are less than the selected type I error rate ($\alpha = 0.05$), the presented results of conducted procedure of testing hypotheses about the population mean unequivocally suggest the acceptance of the following (alternative) statistical assumptions:

- The average GDP growth rate of countries in which a relatively "lower" degree of severity of consequences of the COVID-19 pandemic in 2020 was recorded is statistically significantly higher than $\mu_0 = –5.671$, and $\mu_0 = –7.125$, and thus, indirectly, than the average GDP growth rate determined for countries with medium, i.e., high and very high severity of COVID-19 health consequences.
- The average GDP growth rate of countries in which a medium level of severity of consequences of the COVID-19 pandemic in 2020 was recorded is statistically significantly higher than $\mu_0 = –7.125$, and thus, indirectly, than the average GDP growth rate determined for countries with recorded high and very high level of severity of COVID-19 consequences.

Although there are numerous studies dealing with the analysis of economic consequences of the COVID-19 pandemic, regardless of the similarity of the objectives, their results are not comparable to those presented in this study, primarily due to pronounced methodological differences, used combinations of COVID-19 and economic indicators, and, finally, the temporal-spatial coverage of data. In other words, due to the listed differences, a concrete comparison of empirical results would not be valid. Nevertheless, the obtained results confirm the well-known fact about the negative impacts of the COVID-19 pandemic on the economic activity of national economies (e.g. Praščević, 2020; Su et al., 2021; Bhardwaj et al., 2020; Wren-Lewis, 2020; Nicola et al., 2020), but also complement it by confirming the presence of statistically significant direct relationship, in terms of the magnitude of those impacts, that is a higher degree of severity of COVID-19 pandemic (health) consequences → higher negative economic consequences, measured by GDP growth rate. Küçükefe (2020) came to similar findings, but examined the relationship only between the number of confirmed deaths per million inhabitants and GDP growth rate in the example of China and the OECD group of countries, using a non-hierarchical clustering procedure and regression analysis. In addition to the above, the specific combination of statistical methods applied to carefully selected indicators further emphasizes the originality of the conducted research.

6. Conclusions

The current COVID-19 pandemic, in parallel with the implications of unprecedented proportions on public health, has also caused serious economic consequences. Following the defined subject and formulated objectives of the research, a two-stage statistical analysis was conducted (based on the combined application of hierarchical agglomerative clustering procedure and statistical hypotheses testing for population mean), aimed at examination of the relationship and magnitude of health and economic consequences of the COVID-19 pandemic on the example of 40 selected European countries. In this sense, the results of a hierarchical classification of observed countries, obtained using carefully selected pandemic indicators that quantify different health aspects of the crisis, reveal/confirm the presence of pronounced differences between the formed clusters of countries in terms of COVID-19 consequences, since the solution includes four clusters singled out as optimal. Descriptive names assigned to each of them reflect the observed rank of the degree of severity of the consequences caused by the pandemic.

Finally, the results of the second stage of conducted empirical research unequivocally confirm the presence of a direct relationship between the magnitude of the health and economic consequences of the COVID-19 pandemic at the level of analysed European countries. More precisely, countries with a relatively "lower" severity of the COVID-19 pandemic consequences recorded the "highest" (i.e., lowest negative) average GDP growth rate in relation to the clusters of countries that suffered more serious consequences of the COVID-19 pandemic, i.e., countries characterized by medium and high / very high degree of severity of COVID-19 pandemic (health) consequences. The statistical significance of the previous finding was confirmed by the results of statistical hypotheses testing.

Generally, the obtained results can serve as an additional support to policymakers in making decisions aimed at mitigating pandemic impacts and crisis management. In other words, the empirically confirmed direct association between the severity of health and economic consequences of the COVID-19 pandemic may suggest the official state institutions conduct a subsequent analysis of the capacity of the healthcare system (including available medical equipment and trained medical personnel), as well as the speed and manner of its response in pandemic conditions. Corrective measures aimed at reducing the mortality rate and slowing the spread of the infection, according to the findings of this research, will indirectly contribute to mitigating the negative economic consequences caused by this and some new, future pandemics. With the presented overview of related research studies, this paper s main contribution is applying an innovative methodological framework designed to examine the relationship between economic and COVID-19 indicators. In addition, the advantage of the applied methodological framework, from the perspective of ensuring objectivity and scientific verification of results, is reflected in the use of statistically based criteria in choosing the optimal hierarchical agglomerative method and best quality clustering solution, in contrast to the approach based on the researcher's subjective choice. The practical significance of the conducted analysis and applied methodology contributes to changing the spatial and temporal scope of the research in the context of future research. The obtained results can serve as a suitable basis for gaining a complete insight into the COVID-19 pandemic and its economic implications. For future research, not only updating and expanding the existing database are necessary by including new variables and new statistical methods. Additionally, one of the key limitations of this research is the omission of the economic effects of implemented anti-pandemic measures and measures supporting economic activity in the conducted analysis. Since the mentioned measures, in the context of the COVID-19 pandemic, can significantly affect the economic performance of countries, considering their effects can represent one of the possible directions of future research efforts, too.

References

Abdullah, D. et al. (2021) "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data" *Quality & Quantity*, Vol. 56, pp. 1283–1291. https://doi.org/10.1007/s11135-021-01176-w.

Bhardwaj, G. et al. (2020) "Statistical Analysis for Socio-Economic Impacts of Covid-19 Pandemic" *Easy Chair Preprint*, [Internet], No. 3661. Available at: <*https://easychair.org/publications/preprint/JzM8*> [Accessed: September 18, 2021]

Centre for Economic Policy Research (CEPR) (2020) "*Economics in the Time of COVID-19*", edited by Baldwin, R. & Weder di Mauro, B., [Internet], London: CEPR Press. Available at: <*http://cepr.org/sites/default/files/news/COVID-19.pdf*> [Accessed: September 9, 2021]

Clemente-Suarez, V.J. et al. (2021) "The Impact of the COVID-19 Pandemic on Social, Health, and Economy" *Sustainability*, Vol. 13, No. 11, https://doi.org/10.3390/su13116314.

Committee for the Coordination of Statistical Activities (CCSA) (2020) "*How COVID-19 is changing the world: a statistical perspective*", [Internet], Vol. 2, New York: CCSA. Available at: <*https://unstats.un.org/unsd/ccsa/pubs/*> [Accessed: October 3, 2021]

Ehnts, D. & Paetz, M. (2021) "COVID-19 and its economic consequences for the Euro Area" *Eurasian Economic Review*, Vol. 11, No. 2, pp. 227–249, https://doi.org/10.1007/s40822-020-00159-w.

de la Fuente-Mella, H. et al. (2021) "Modeling COVID-19 Cases Statistically and Evaluating Their Effect on the Economy of Countries" *Mathematics*, Vol. 9, No. 13, https://doi.org/10.3390/math9131558.

International Telecommunication Union (ITU) (2020) "*Economic Impact of COVID-19 on digital infrastructure*" Report of an Economic Experts Roundtable organized by ITU, [Internet], Geneva: ITU. Available at: <*http://handle.itu.int/11.1002/pub/816783a2-en*> [Accessed: September 20, 2021]

Kempa, W. et al. (2021) "Statistical and Econometric Analysis of Selected Effects of COVID-19 Pandemic" *Multidisciplinary Aspects of Production Engineering*, Vol. 4, No. 1, pp. 395–407, https://doi.org/10.2478/mape-2021-0036.

Küçükefe, B. (2020) "Clustering Macroeconomic Impact of COVID-19 in OECD Countries and China" *Journal of Research in Economics, Politics & Finance*, Vol. 5, Special Issue, pp. 280–291, https://doi.org/10.30784/epfad.811289.

Liu, N., Xu, Z. & Skare, M. (2021) "The research on COVID-19 and economy from 2019 to 2020: analysis from the perspective of bibliometrics" *Oeconomia Copernicana*, Vol. 12, No. 2, pp. 217–268, https://doi.org/10.24136/oc.2021.009.

Ljones, O. (2020) "Pandemic and official statistics; some comments on recent COVID-19 experiences" *Statistical Journal of the IAOS*, Vol. 36, No. 2, pp. 285–290, https://doi.org/10.3233/SJI-200672.

Nayak, S.R. et al. (2021) "A statistical analysis of COVID-19 using Gaussian and probabilistic model" *Journal of Interdisciplinary Mathematics*, Vol. 24, No. 1, pp. 19–32, https://doi.org/10.1080/09720502.2020.1833442.

Nicola, M. et al. (2020) “The socio-economic implications of the corona virus pandemic (COVID-19): A review” *International Journal of Surgery*, Vol. 78, pp. 185–193, https://doi.org/10.1016/j.ijsu.2020.04.018.

Praščević, A. (2020) “Economic Shock of the COVID-19 Pandemic – Turning Point in Global Economic Developments” *Ekonomske Ideje i Praksa*, [Internet], Vol. 2020, No. 37, pp. 7–22. Available at: <*http://www.ekof.bg.ac.rs/wp-content/uploads/2014/10/011.pdf*> [Accessed: August 23, 2021]

Rizvi, S.A., Umair, M. & Cheema, M.A. (2021) “Clustering of countries for COVID-19 cases based on disease prevalence, health systems and environmental indicators” *Chaos, Solitons and Fractals*, Vol. 151, https://doi.org/10.1016/j.chaos.2021.111240.

Su, C.-W. et al. (2021) “COVID-19 pandemic and unemployment dynamics in European economies” *Economic Research-Ekonomska istraživanja*, ahead-of-print, https://doi.org/10.1080/1331677X.2021.1912627.

Wren-Lewis, S. (2020) “The economic effects of a pandemic” In Baldwin, R. & Weder di Mauro, B. (Eds), *Economics in the Time of COVID-19*, London: CEPR Press.

Zarikas, V. et al. (2020) “Clustering analysis of countries using the COVID-19 cases dataset” *Data in Brief,* Vol. 31, https://doi.org/10.1016/j.dib.2020.105787

Sources of data:

The European Centre for Disease Prevention and Control electronic database. Available at: <*https://www.ecdc.europa.eu/en*> [Accessed: July 9, 2021]

The EUROSTAT electronic database. Available at: <*https://ec.europa.eu/eurostat*> [Accessed: July 5, 2021]

Utjecaj pandemije Covid-19 na gospodarski učinak Europskih zemalja

Marina Milanović[1], *Milan Stamenković*[2]

Sažetak

Polazeći od činjenice da su brzo širenje virusa SARS-CoV-2 i provedba strategija socijalnog distanciranja dramatično utjecali na sve aspekte života ljudi na globalnoj, nacionalnoj i mikro razini, ovaj rad se fokusira na ispitivanje utjecaja pandemije COVID-19 na gospodarski učinak odabranih europskih zemalja. Kako bi se uočio i razumio ovaj utjecaj, korištena je složena metodologija istraživanja koja se temelji na kombiniranoj primjeni prikladnih univarijatnih i multivarijatnih metoda statističke analize. Klasifikacija 40 europskih zemalja u različite grupe u pogledu odabranog seta COVID-19 pokazatelja u 2020. godini izvršena je primjenom hijerarhijske aglomerativne klaster analize, dok je za statističku ocjenu kvalitete dobivenog rješenja korišten ne-hijerarhijski postupak baziran na k-means metodi. Klasifikacija koja se sastoji od četiri klastera zemalja identificirana je kao "optimalno" rješenje. Analiza i usporedba profila formiranih klastera zemalja s obzirom na njihove prosječne stope rasta BDP-a u 2020. godini provodi se statističkim metodama deskriptivne analize i testiranja hipoteza. Provedeno istraživanje otkriva da je klaster zemalja s relativno "manjom" ozbiljnošću COVID-19 zdravstvenih posljedica zabilježio višu prosječnu stopu rasta BDP-a u usporedbi s klasterima zemalja koje su pretrpjele teže posljedice, i obratno. Dobiveni rezultati koji upućuju na povezanost veličine negativnih zdravstvenih i gospodarskih posljedica COVID-19 pandemije mogu poslužiti kao dodatna potpora kreatorima politike u donošenju odluka usmjerenih na ublažavanje posljedica pandemije i suzbijanje krize.

Ključne riječi: *klaster analiza, COVID-19 pokazatelji, stopa rasta BDP-a, europske države*

JEL klasifikacija*: C12, C38, I10, O47, O57*

[1] *Docentica, Sveučilište u Kragujevcu – Ekonomski fakultet, Odsjek za informatiku i kvantitativne metode, Liceja Knezevine Srbije 3, 34000 Kragujevac, Srbija. Znanstveni interes: primjena kvantitativnih metoda u ekonomiji. Tel.: +38134303536. E-mail: milanovicm@kg.ac.rs. 0000-0002-6245-5313. Web-stranica: https://www.ekfak.kg.ac.rs/sr/nastavnici/nastavnici-az?id=156&idd=453.*

[2] *Docent, Sveučilište u Kragujevcu – Ekonomski fakultet, Odsjek za informatiku i kvantitativne metode, Liceja Knezevine Srbije 3, 34000 Kragujevac, Srbija. Znanstveni interes: primjena kvantitativnih metoda u ekonomiji. Tel.: +38134303525. E-mail: m.stamenkovic@kg.ac.rs. 0000-0003-0689-0369. Web-stranica: https://www.ekfak.kg.ac.rs/sr/nastavnici/nastavnici-az?id=156&idd=382.*