APPLICATION OF THE HIERARCHICAL CLUSTER METHOD IN THE FIELD OF TAX POLICY OF EU

Mária Stachová – René Duda – Pavol Kráľ

Abstract

Our contribution is focused on application of agglomerative hierarchical clustering in analysis of tax policies of European Union countries. By applying Ward linkage function we identify countries similar in terms of their tax policies in the form of compact clusters. The resulting clusters are interpreted based on similarities and differences in tax policies between countries forming these clusters. Principal component analysis and scatterplots of the principal components are used to further enhanced the interpretation. Finally, for selected counties we discuss possible improvements to reach the level of the most successful, in terms of tax revenue, countries.

We use data set from Eurostat and OECD database and the analysis is based on ESA (European System of National and Regional Accounts) methodology. Years 2016 and 2020 were utilized. The year 2016 represents tax policies typical for expansion stage of the economic cycle and the year 2020 should reflect potential changes in tax policies instigated by the COVID-19 pandemic.

Key words: Hierarchical cluster analysis, Tax policy, EU countries.

JEL Code: C38, G33

Introduction

Tax policies of EU countries are often discussed with European commission. According to Pippin and Tosun (2016), the primary goal of such discussions is an attempt to harmonize tax policies within EU as much as possible. Currently, a relatively high level of harmonization is present in case of indirect taxation, especially in case of consumption taxes and values added taxes, where individual rates of these taxes are very similar among various countries. There are also significant attempts to increase the level of harmonization aiming for a common judicial framework and common consolidated tax base (Verbeken, 2021), (Kubátová, 2003), (Feranecová et al., 2017). On the other hand, personal income taxes are still fully controlled by individual EU member states and can be used, for example, to attract foreign investments by

offering minimal tax burden on labor. Such practices have potentially detrimental effects on tax harmonization process. To better facilitate further harmonization in tax policies, it is vital to understand similarities and differences in tax policies and incomes among EU members (Stiglitz & Rosengard, 2015), (Paternoster, 2019).

The aim of our paper is to provide a partial insight in the matter by identify groups of EU countries similar with respect to their tax policies and tax incomes

1 Data and Methodology

Our data set consists of tax income data of 27 EU members collected in years 2016 and 2020. The year 2016 was selected to represent the period of economic growth preceding COVID-19 pandemic. The year 2020 data is supposed to be also including partial information of COVID-19 impacts on economic performance of the EU member states measured by GDP and thus on tax incomes. Applying ESA (2010) methodology, we work with the following tax categories: Value Added Tax (VAT), Production and Import Taxes, Personal Income Tax, Corporate Income Tax, Capital Tax, Other Taxes, Social Contributions which form majority of tax incomes in the selected countries. Such selection of tax categories should provide us with insight in structure of tax burden and, indirectly, in tax policies. For example, we can observe if a state focuses on indirect taxes such as VAT, production and import taxes, or direct taxes where it applies more taxes on wealth both individuals and corporations. We express each tax category income for a particular country as a percentage of total tax incomes of the country.

We applied Ward method of agglomerative clustering (Murtagh & Legendre, 2014) as this method was capable of handling outliers present in the input data and provides reasonable granular and interpretable solution. Statistical software R 4.1.2. (R Core Team, 2021) and package “cluster” (Maechler et al., 2021) were used to generate the result. The Ward agglomerative method is utilizing distance between clusters minimizing their within-cluster heterogeneity. This distance is defined as an increase in the sum of squares of deviations from the mean vector. Two clusters with the minimal distance are merged in each step. This method produces very similar clusters and eliminates small clusters. This method is described in more details in (Kaufman & Rousseeuw, 2005), (Trebuna & Halcinova 2012) and (Aggarwal & Reddy, 2014).

2 The result of hierarchical clustering of data collected in 2016

Descriptive statistics of our data set can be found in Table 1. We see there rather high range of values in individual tax categories. Observed differences between means and medians together higher variability of variables point towards presence of outliers for personal income tax, capital taxes and other taxes also supported by variation coefficient exceeding 0.5. Variability of other taxes is rather small.

Tab. 1: Descriptive statistics of data, year 2016

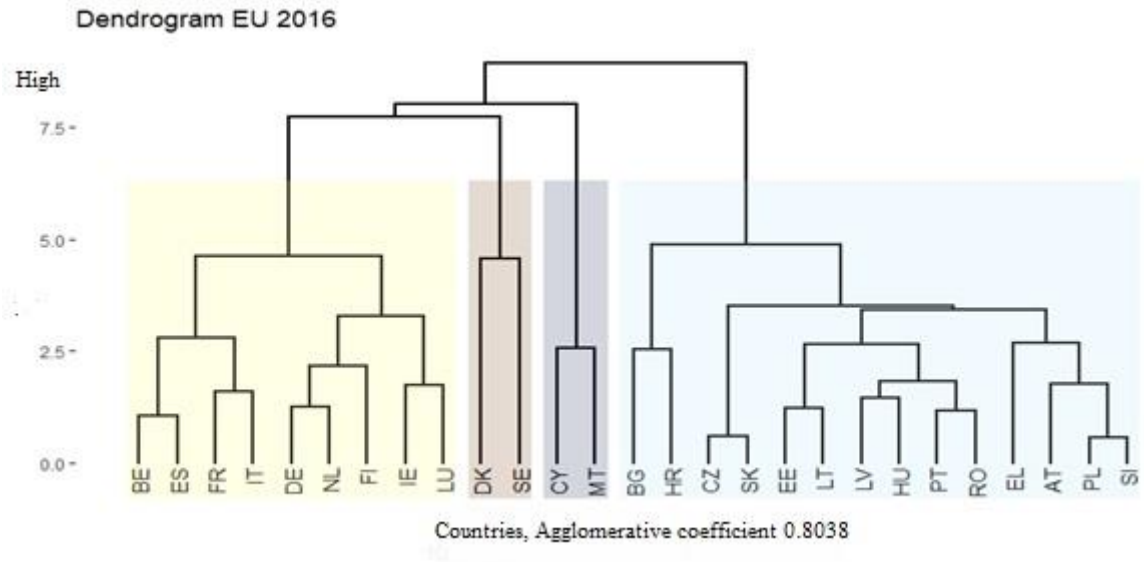
	VAT	Production and Import Taxes	Personal Income Tax	Corporate Income Tax	Capital Tax	Other Taxes	Social Contributions
n	27	27	27	27	27	27	27
Missing	0	0	0	0	3	0	0
Minimum	0.142	0.097	0.074	0.042	0	0.002	0.020
Maximum	0.340	0.298	0.542	0.174	0.017	0.064	0.433
Range	0.198	0.201	0.468	0.132	0.017	0.062	0.413
Sum	5.759	4.705	5.287	2.049	0.105	0.596	8.498
Median	0.206	0.172	0.181	0.061	0.001	0.021	0.339
Mean	0.213	0.174	0.196	0.076	0.004	0.022	0.315
Variance	0.002	0.002	0.010	0.001	0	0	0.010
Standard deviation	0.049	0.040	0.098	0.035	0.005	0.014	0.1
Variation coefficient of variation	0.229	0.229	0.500	0.455	1.273	0.617	0.319
Skewness	0.663	0.894	1.613	1.580	1.302	0.984	-1.327
Kurtosis	0.047	1.504	3.373	1.770	0.802	1.132	1.410

Source: Own computation based on Eurostat data

Descriptive statistics are coherent with dendrogram depicting tax income based clusters of EU countries (see Figure 1). Agglomerative coefficient of quality of the clustering structure for the year 2016 data equals 0.808, i.e., our structure of clusters is appropriate. In Dendrogram we see four clusters with varying number of countries in each of them. The first cluster contains economic leaders of Western Europe with highest proportion of personal income taxes on total tax incomes and social contributions above average. On the other hand, contribution of VAT and Production and Import Taxes in these countries is rather small comparing to countries in remaining clusters. The second cluster contains Denmark and Sweden. These two countries are

outliers with respect to Personal Income Tax which is 54% for Denmark and almost 37% for Sweden respectively. On the other hand, these countries have a very small Social Contributions accounting only for 2% for Denmark and 7.4% for Sweden, respectively. The third cluster is formed by Malta and Cyprus. These countries also have a smaller proportion of tax incomes from Social Contributions but exhibit a high proportion of Corporate Taxes on the total tax income.

Fig. 1: Dendrogram of EU countries, year 2016

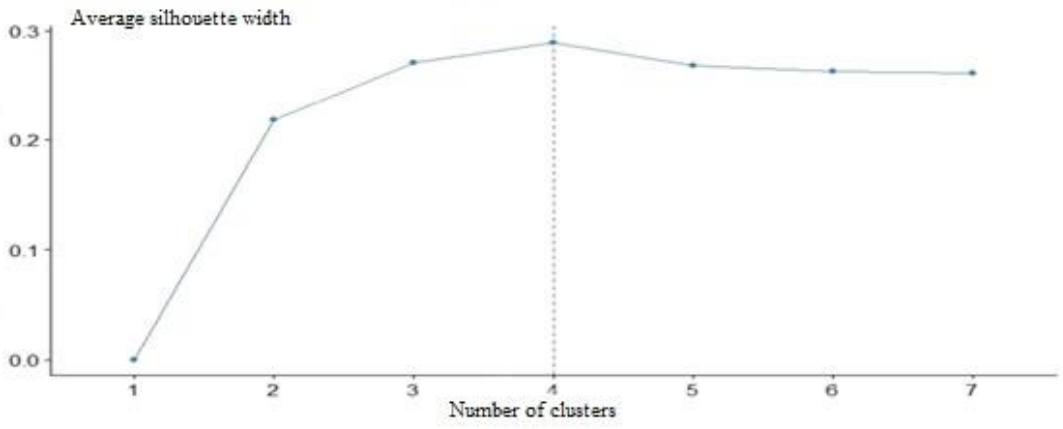


Source: author's work

The last (forth) cluster is formed by the Balkan countries, Baltic states, V4 countries, Greece, Austria, and Portugal. These countries are characterized by average and above average social contributions, VAT and below average Personal and Corporate Income Tax, thus focusing on indirect taxation. Let us mention also a significant similarity between Czech Republic and Slovakia forming a separate cluster at initial stages of clustering.

The appropriate number of clusters, 4 in our case, we identified based on cluster outline widths depicted in Figure 2.

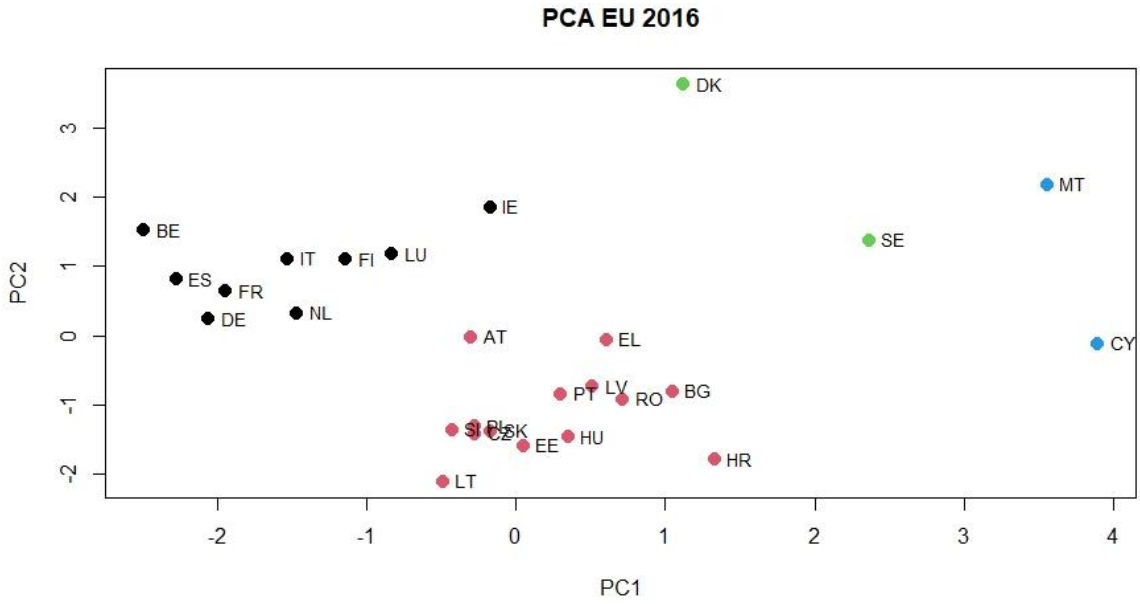
Fig. 2: Average silhouette width, year 2016



Source: author's work

In order to visualize the identified clusters, we applied Pricipal Component Analysis to create a representation of clustered countries in two dimesions. The resulting scatter plot based on two main pricipal components can be found in Figure 3

Fig. 3: Two-dimensional plot of clusters, year 2016



Source: author 's work

3 The result of hierarchical clustering of data collected in 2020

Table 2 list descriptive statistics for tax data from the year 2020 and shows as a similar picture to Table 1 for tax data from the year 2016.

Dendrogram (Figure 4) presents results of clustering for the year 2020 with coefficient of quality of the clustering structure equals 0.792. We see there again four clusters of countries. The biggest cluster is formed by V4, Balkan and Baltic countries. Another big cluster is formed

by countries mainly from West and South Europe. A smaller, four-member cluster consists of Ireland, Luxembourg, Malta, and Cyprus. Denmark and Sweden again form a standalone cluster with high similarities in tax incomes coming from Social Contributions.

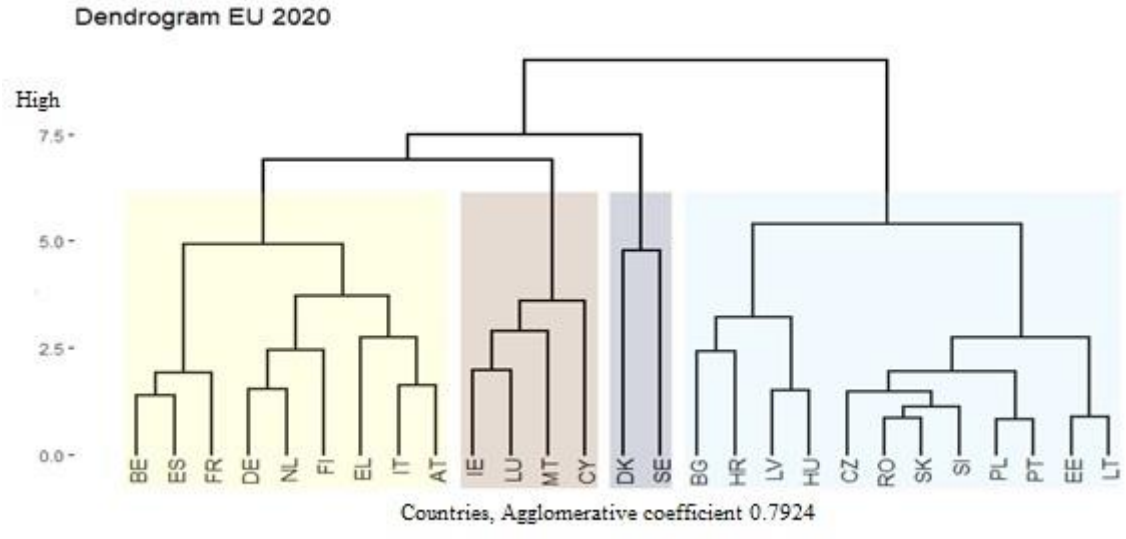
The cluster outline widths (Figure 5) again indicate four cluster as the optimal value for the number of cluster although the mean value of the cluster outline widths coefficient corresponding to that number of clusters dropped slightly to the value 0.25.

Tab. 2: Descriptive statistics of data, year 2020

	VAT	Production and Import Taxes	Personal Income Tax	Corporate Income Tax	Capital Tax	Other Taxes	Social Contributions
n	27	27	27	27	27	27	27
Missing	0	0	0	0	4	0	0
Minimum	0.140	0.094	0.087	0.023	0	0.003	0.017
Maximum	0.338	0.287	0.565	0.169	0.016	0.054	0.453
Range	0.197	0.193	0.478	0.147	0.016	0.052	0.435
Sum	5.670	4.332	5.624	1.878	0.099	0.545	8.851
Median	0.210	0.153	0.203	0.057	0.001	0.016	0.342
Mean	0.210	0.160	0.208	0.070	0.004	0.020	0.328
Variance	0.003	0.002	0.010	0.001	0	0	0.011
Standard deviation	0.050	0.041	0.100	0.036	0.005	0.012	0.103
Variation coefficient of variation	0.239	0.256	0.480	0.519	1.243	0.606	0.316
Skewness	0.549	1.153	1.530	1.328	1.067	0.913	-1.376
Kurtosis	-0.245	1.356	3.543	1.038	0.022	0.365	1.751

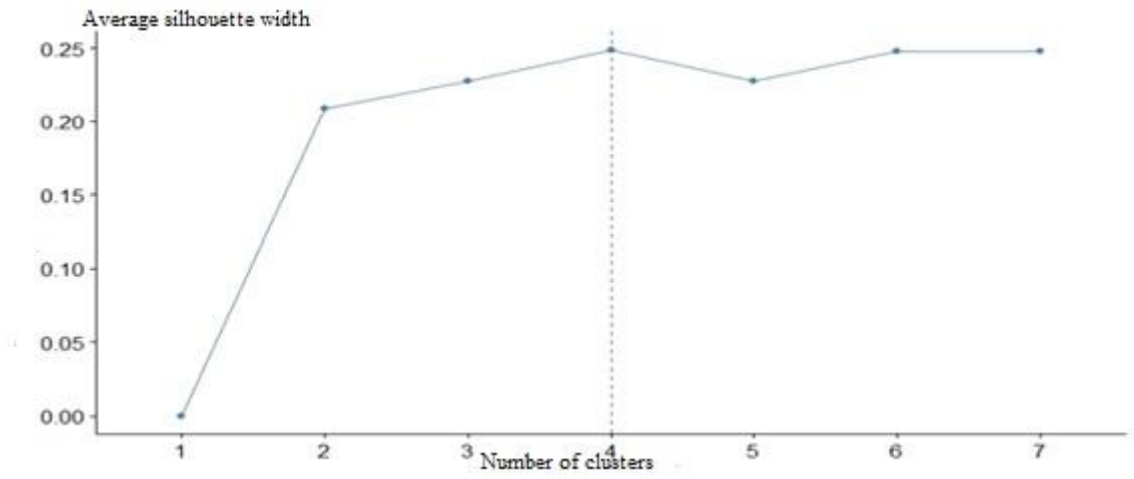
Source: Own computation based on Eurostat data

Fig. 4: Dendrogram of EU countries, year 2020



Source: author's work

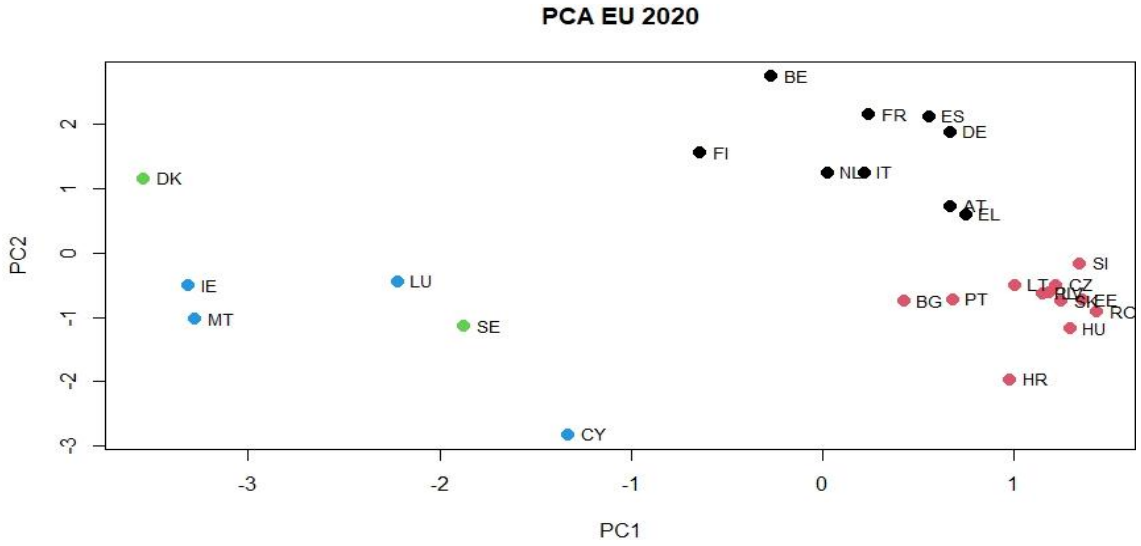
Fig. 5: Average silhouette width, year 2020



Source: author's work

The results and interpretations outlined above are also supported by visualization of the countries in two dimensions via scatter plot (Figure 6) where we again see Sweden close to the cluster of four countries although forming a cluster on its own. The scatter plot reveals that the other clusters are highly heterogeneous as demonstrated via the depicted distances. The least heterogeneous cluster can be identified in the right bottom corner of the graph.

Fig. 6: Two-dimensional plot of clusters, year 2020



Source: author's work

Conclusion

In our contribution, we clustered EU countries into compact groups determined by similarities in tax incomes and thus indirectly in tax policies. When comparing the years 2016 and 2020, we see just a small adjustment in clusters which reflects minimal changes in tax policies of individual countries. It was apparent that the highest level of harmonization was demonstrated in indirect taxation as the corresponding group of taxes were not crucial to identify the clusters. The clustering solution was mainly impacted Personal and Corporate Income Taxes as well as Social Contributions. The identified clusters did not reflect just regional similarities but also similarities in economies of the countries. Most similarities were observed among V4, Baltic and Balkan EU members. The second cluster was formed by biggest EU economies. Denmark and Sweden formed their own cluster, but we can, via visualization of the clusters utilizing PCA, view them as outliers. The last cluster was formed by Cyprus and Malta, and, eventually, Luxembourg and Ireland which have in common quite low tax burden and this cluster could be interpreted as a cluster of tax heavens.

We see transfer of majority of tax burden towards indirect taxation as the main potential driver of improvement in tax policies of EU states. This long-term process already started but seem to be progressing slowly so far. At the same integration of EU has progressed mainly in trading of goods and services. Therefore, it is quite natural to expect that production and import taxed as well as corporate income taxes would be similar and at least partially harmonized. Recommendation to move towards indirect taxation is also motivated by the fact that it reflects

taxation of wealth through more significant taxation of consumption of goods and services. In the periods of economic growth and prosperity, it may be recommended to increase taxation of consumption and thus correct excessive purchasing power of the population and at the same time move towards the surplus state budget. It includes VAT and specific selective taxes imposed on selected goods. Due to the current level of harmonization of EU, possibility to purchase less expensive goods and services fin neighboring countries may be partially mitigated and thus not affecting significantly domestic producers and retailers.

The future work will focus on the period 2021-2022 where effects of the COVID-19 pandemic as well as economic crisis driven by the war in Ukraine should be fully demonstrated..

Acknowledgment

The work has been supported by project VEGA No. 1/0150/21 (Profit testing of pension insurance products).

References

Aggarwal, Ch., & Reddy, Ch. (2014). *Data Clustering Algorithms and Applications*. New York: CRC Press Tylor & Francis Group. 622 p.

European Commission, (2010). *European system of accounts – ESA2010*. Luxembourg: Publications Office of the European Union. 652 p.

Feranecová, A., et al. (2017). Possibilities of harmonization of direct taxes in the EU. Sumy: Business Perspectives. *Investment Management and Financial Innovations*, 14, p. 191-199.

Kaufman, L., & Rousseeeuw, P. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis.* New Jersey: John Wiley & Sons, Inc. 342 p.

Kubatová, K. (2003). *Daňová teorie a politika.* Praha: ASPI Publishing.

Maechler, M., Rousseeuw, O., Struyf, A., Hubert, M., Hornik, K. (2021). R package "cluster ": Cluster Analysis Basic and Extensions. R package version 2.1.2. online: https://CRAN.R-project.org /package=cluster

Murtagh, F., & Legendre, P., (2014), Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?, *Journal of Classification*, 31(3), p. 274-295.

Paternoster, D. (2019). *Všeobecná daňová politika*. [cit. 2021-11-20]. online: https://www.europarl.europa.eu/factsheets/sk/sheet/92/vseobecna-danova-politika

Pippin, E. S., & Tosun, S. M. (2016). Tax harmonization in the European Union and the Eurozone: a multilateral analysis of tax systems. *Accounting and Information Systems*. p. 438-461. [cit. 2021-11-21]. online: https://scholarworks.unr.edu/handle/11714/911

R Core Team. (2021). *A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. [cit. 2022-03-07]. online: https://www.R-project.org

Stiglitz, J. E., & Rosengard, J. K. (2015). *Economics of the Public Sector* 4. Edition. New York – London: W.W. Norton Company, Inc., 961p.

Trebuna, P., & Halcinova, J. (2012). Experimental modelling of the cluster analysis process, *Modelling of Mechanical and Mechatronics Systems*, 48, p. 673-678.

Verbeken, D. (2021). *Hospodárska a menová únia, daňová politika a politika hospodárskej súťaže: Všeobecná daňová politika*. [cit. 2021-11-12]. online: https://www.europarl.europa.eu/factsheets/sk/sheet/92/vseobecna-danova-politika

Contact

Mária Stachová

Faculty of Economics, Matej Bel University

Tajovského 10

975 90 Banska Bystrica, Slovakia

maria.stachova@umb.sk

René Duda

Faculty of Economics, Matej Bel University

Tajovského 10

975 90 Banska Bystrica, Slovakia

renedudaa@gmail.com

Pavol Kráľ

Faculty of Economics, Matej Bel University

Tajovského 10

975 90 Banska Bystrica, Slovakia

pavol.kral@umb.sk