Cluster analysis of the EU banking sector based on EBA Risk Indicators

Patrik Mihalech[1] and Martina Košíková[2]

[1] University of Economics in Bratislava, Faculty of Economic Informatics/Department of Statistics, Dolnozemská cesta 1/b, Bratislava, 852 35 Slovak Republic
[2] University of Economics in Bratislava, Faculty of Economic Informatics/Department of Statistics, Dolnozemská cesta 1/b, Bratislava, 852 35 Slovak Republic

patrik.mihalech@euba.sk
martina.kosikova@euba.sk

https://doi.org/10.53465/EDAMBA.2021.9788022549301.306-316

Abstract. Banking sector plays a key role in financial system of every developed country. To know possible weaknesses proper risk management is necessary. European Banking Authority (EBA) is the arterial institution in attempt to consolidate risk management among different countries of European Union. EBA discloses on quarterly basis various Key Risk Indicators (KRIs) for all EU member countries. The goal of this paper is to analyze chosen KRIs of all EU countries and based on distances and similarities among them, insert them into homogenous groups. The purpose of the analysis is to seek insights into different countries bank's sector and finding similarities among them, which might not be visible at the first glance. For the research, both hierarchical and non- hierarchical cluster analysis were performed. Results show that we could observe four groups of states which could be, with a little generalization, labeled as eastern countries, southern countries, northern countries and middle and core countries of EU, based on analyzed KRIs.

Keywords: cluster analysis, EU banking sector, risk indicators.

JEL classification: *C 38, G 21.*

1 Introduction

Banks are important institutions of financial sector, and their economic health is essential for stable growth of national economy. To know strengths and weaknesses of banks it necessary to monitor their risk exposure. For banks in European Union, one of the most important institutions that monitors various risk data from all member states is the European Banking Authority (EBA). EBA is specialized agency of the European Union set up to achieve a more integrated approach to banking supervision across the

EU [6]. One of the core tasks of EBA is to establish a single set of rules applicable to all banking institutions in the EU in the same manner. This is also precondition for an EU single market in the banking sector.

One of the main responsibilities of EBA is to monitor banking risk. In February 2011, EBA started collecting statistical information of 55 banks across 20 European Economic Area (EEA) countries [5]. From these data EBA constructed Key Risk Indicators (KRIs). These KRIs are ratios, which are expected to provide early warning signs of trends, potential risks and vulnerabilities in the EU banking sector. Information regarding KRIs started to be published by EBA quarterly as EBA Risk Dashboards [7] in order to provide general information regarding risk factors of banking sector of EEA countries.

The aim of this paper is to find similarities among countries of the European Union based on EBA KRIs by usage of the cluster analysis. Analysis is performed on EBA Risk Dashboard data set as of December 2020 on chosen KRIs. Based on the results of the analysis, the goal is to compare bank system among EU countries and find insights on similarities and differences through countries.

2 Definition of chosen KRIs

From available indicators in EBA Risk Dashboard we have decided to choose five, which are expected to cover different areas of banking sector's risk appetite and financial profitability on country basis. These indicators, as defined by EBA [7], are:

1. Capital Adequacy Ratio (CAR),
2. Leverage Ratio (LR),
3. Return on Equity (RoE),
4. Loans-to-deposit ratio for households and non-financial corporations (LtD),
5. Liquidity Coverage Ratio (LCR).

Capital Adequacy Ratio is a measurement of bank's available capital which is expressed as a percentage of bank's risk weighted assets. Its purpose is to protect depositors and promote stability and efficiency of financial system. Capital used to calculate the capital adequacy ratio is according to BCBS standards [3] divided into two tiers. First tier consists of Common Equity Tier 1 and other Tier 1 capital and consists of equity capital, ordinary share capital, intangible assets and audited revenue reserves. Tier 2 capital consists of unaudited reserves and general loss reserves [13]. Risk weighted assets are bank's balance and off-balance sheet exposures weighted according to risk. Capital adequacy ratio is calculated as following:

$$CAR = \frac{Tier\ 1\ capital + Tier\ 2\ capital}{Risk\ Weighted\ Assets} \tag{1}$$

Leverage ratio is defined as the capital measure divided by the exposure measure and is expressed as a percentage:

$$LR = \frac{capital\ measure}{exposure\ measure} \tag{2}$$

The capital measure is total Tier 1 capital. The exposure measure is the sum of balance sheet exposures, derivatives exposures, securities financing transaction exposures and off-balance sheet items [9]. According to BCBS [1] LR is intended to restrict the build-up leverage in the banking sector to avoid destabilizing deleveraging processes that can damage the broader financial system and the economy and is supposed to reinforce the risk-based capital requirements with a simple, non-risk-based "backstop" measure. Leverage ratio must exceed 3% and high percentage means that bank have sufficient amount of capital to cover its risk exposure.

Return on equity is a financial measure of how effectively a bank generates profit from the money that investors have put into the business. ROE is calculated by dividing net income by total shareholders' equity:

$$ROE = \frac{net\ income}{average\ shareholder's\ equity} \tag{3}$$

In banking sector ROE was found out to be a better metric at assessing the market value and growth than earning per share growth widely used in other sectors. Investors are interested in having ROE as high as possible.

Loans-to-deposit ratio for households and non-financial corporations (LtD) helps assess bank's liquidity position. According to EBA methodology [5] LtD is calculated as total loans and advances divided by total liabilities and gives an indication for which share of loans is funded by depositors:

$$LtD = \frac{total\ loans\ and\ advances}{total\ deposits} \tag{4}$$

LtD is expressed as percentage and value highly above 100% indicates that bank uses extensively other sources of funding than deposits.

Liquidity coverage ratio is designed to ensure that a bank maintains an adequate level of unencumbered, high-quality assets that can be converted into cash to meet its liquidity needs for a 30-day time horizon under an acute liquidity stress scenario specified by supervisors [4]. LCR is defined as followed:

$$LCR = \frac{stock\ of\ High\ Quality\ Liquid\ Assets\ (HQLA)}{total\ net\ cash\ outflow\ over\ next\ 30\ calendar\ days} \tag{5}$$

Asset to be eligible as an HQLA must be liquid during the stress period and easily convertible into cash without significant loss of the value. General guideline for an asset to be considered as an HQLA is to be eligible as collateral for the central bank's liquidity facilities [8]. Total net cash outflows as defined according to BCBS [2] is total expected cash outflows minus total expected cash inflows in the specified stress scenario for the subsequent 30 calendar days. LCR must be maintained above 100% by regulatory limit (for banks operating in EU legislative since 1.1.2018, even though former plan of BCBS was to fully implement LCR at amount of above 100% since year 2019).

3 Methodology and data description

Cluster analysis is a multivariate statistic method which purpose is grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. These groups are called clusters and cluster analysis is an important tool with respect to multivariate exploratory data analysis. According to [10] cluster analysis differs from other methods of classification such as discriminant analysis where classification pertains to known number of groups and the operational objective is to assign new observations to one of these groups.

In Cluster analysis grouping is done based on similarities or distances. Dissimilarity measures can be divided into four groups [12]:

- measures of distance,
- coefficient of association,
- correlation coefficient,
- probability measures of similarities.

Most of statistical packages (including programming language R used in this article) supports measures of distances (dissimilarities). Given two objects X and Y in a *p* dimensional space, a dissimilarity measure follows these conditions:

1. d(X,Y) ≥ 0 for all objects X and Y,
2. if d(X,Y) = 0, than X = Y,
3. d(X,Y) = d(Y,X).

Most commonly used metrics to compute distances are [10]:

- **Euclidean distance** – geometric distance in the multidimensional space:

$$d_{ij} = \sqrt{\sum_{k=1}^{K} \left(x_{ik} - x_{jk}\right)^2} \tag{6}$$

- **Manhattan distance** – average difference across dimensions:

$$d_{ij} = \sum_{k=1}^{K} \left|x_{ik} - x_{jk}\right| \tag{7}$$

- **Mahalanobis distance** – eliminates influence of difference in variability of variables and influence of correlated variables:

$$d_{ij} = \left(x_i - x_j\right)^T * \boldsymbol{S}^{-1} * \left(x_i - x_j\right) \tag{8}$$

Where $\boldsymbol{S}^{-1}$ stands for unified sample covariance matrix.

Methods of clustering are divided into two categories: Hierarchical and Non-Hierarchical. **Hierarchical cluster analysis** proceeds either by a series of mergers or successive divisions. Agglomerative hierarchical method starts with individual objects and most similar objects are grouped and these initial group are merged according to their similarities. Divisive hierarchical method works opposite direction, when a single initial object is divided into subgroups such that the object in one subgroup is further from the object in the other subgroup [11].
Objects can be clustered together based on linkage methods. Final results of the cluster analysis are very dependent from chosen method. Among most used linkage methods are following [12]:

- **Single linkage** (nearest-neighbor) – historical method. Each step combines two clusters that contain the closest pair of elements not yet belonging to the same cluster as each other.
- **Complete linkage** (furthest-neighbor) – each step combines two clusters based on distance between two elements that are farthest away from each other. Clusters with the shortest of these distances are merged together.
- **Average linkage** – distance between each pair of observations in each cluster is added up and divided by the number of pairs to get an average inter-cluster distance.
- **Centroid distance** – clusters with lowest distance between their centroids are merged together.
- **Ward's method** – different and the most used method which tend to produce homogenous clusters of relative same size and shape and tends to avoid small clusters. Ward's criterion minimizes the total within-cluster variance. To implement this method, it is necessary to find at each step pair of clusters that leads to minimum increase in total within-cluster variance after merging.

3.1 Data description

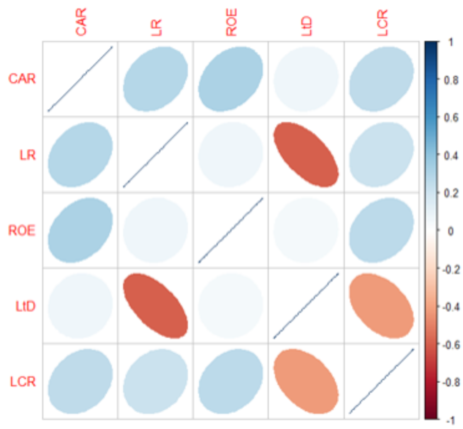
Data used for the cluster analysis are taken from EBA Risk Dashboard Q4 2020 [7] regarding 27 countries of the EU and related to 5 chosen KRIs. This dataset was chosen because it provides us the latest available data in the time of the research and already captures possible impacts of the Covid-19 pandemic. All KRIs are disclosed as percentage. However, given fact that average values among KRIs varies significantly and some KRIs are not allowed to achieve values greater than 100 % by definition and for some it is possible (or it is directly expected of them, like LCR), dataset was standardized according to (9), where x is the original feature vector, $\bar{x}$ is the mean of that feature vector, and σ is it's standard deviation, with purpose to avoid stronger impact of some KRIs over others.

$$x' = \frac{x - \bar{x}}{\sigma} \tag{9}$$

EBA dataset provide average KRIs of biggest banks by country. Information is provided quarterly usually with 4 months of delay.

4 Results

At first, KRIs after regularization were tested for presence of correlation. High level correlation among chosen KRIs might corrupt results of cluster analysis due to correlation among distances. Results are shown on Fig. 1. On significance level 0,01 no correlation is present among KRIs, but correlation among Leverage ratio and Loan-to-deposit ratio (p = 0,0009). Given fact, that only one correlation relationship is statistically significant, we decided to carry on analysis on underlying data. Performing of principal components analysis or factor analysis prior clustering with purpose to obtain linear independent variables would lead to bigger information losses (in terms of interpretability) than gains this approach would provide.



Pearson's correlation coefficient

	CAR	LR	ROE	LtD	LCR
CAR	1.00	0.28	0.32	0.07	0.25
LR	0.28	1.00	0.07	-0.60	0.21
ROE	0.32	0.07	1.00	0.04	0.26
LtD	0.07	-0.60	0.04	1.00	-0.42
LCR	0.25	0.21	0.26	-0.42	1.00

P-values

	CAR	LR	ROE	LtD	LCR
CAR		0.1504	0.1073	0.7387	0.2026
LR	0.1504		0.7368	0.0009	0.2896
ROE	0.1073	0.7368		0.8404	0.1887
LtD	0.7387	**0.0009**	0.8404		0.0281
LCR	0.2026	0.2896	0.1887	0.0281	

Fig. 13. Correlation coefficients of chosen KRIs and p-values of correlation tests in R.

Different linkage methods were used during the analysis with different number of clusters. Best results were achieved by usage of Ward's method and division into four clusters. These clusters were roughly the same size and were stable. Dendrogram showing division of countries into four clusters based on Euclidean distances by usage of Ward's method in hierarchical clustering is shown on Fig. 2. Division of countries based on clusters and their KRIs are shown on Fig. 3 and more insights provide summary table (Fig. 4) of average values of KRIs by clusters and their comparison to total average among all countries.

Results shows interesting cluster differences in Return on equity indicator. Average ROE among all EU countries is 0,43%. However, countries in second cluster (including Bulgaria, Estonia, Croatia, Lithuania, Latvia and Romania) have average ROE of 8 % which means, that banks in these countries achieved much higher profits in Covid-19 influenced year 2020 than banks in the rest of the EU. As a contrast, there is cluster 3 (consisting of Cyprus, Spain, Greece, Ireland, Malta, Portugal, and Poland) with negative or null ROE at best. This cluster consists mainly of south EU countries which were hit by pandemic the worst and results shows that their banking system was severely hit as well. Countries in cluster 3 have also smallest average CAR ratio (19 %) which might indicate they might be lacking capital in case of ongoing crisis.

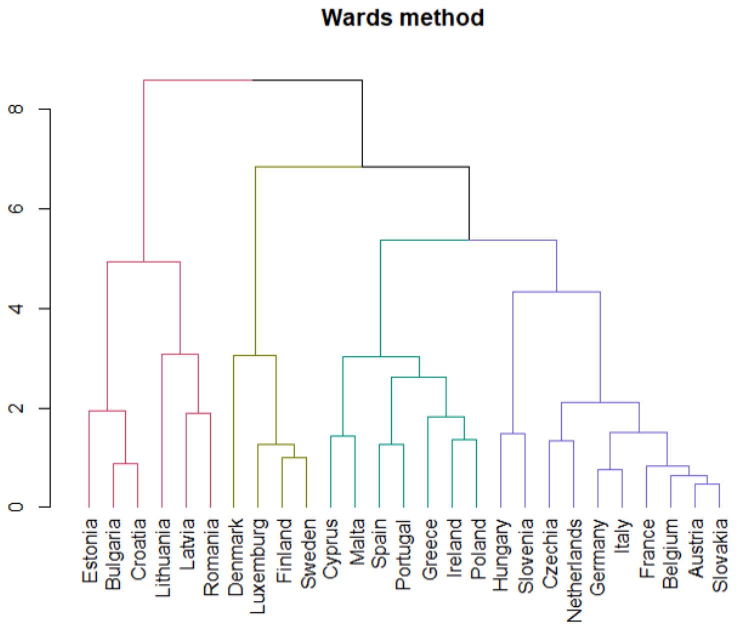


Fig. 2. Dendrogram of EU countries KRIs based on Ward's method of Hierarchical clustering (standardized distances on y-axis).

Moving back to cluster 2, another insight into these eastern European counties is that they have highest average of all KRIs, but Loan-to-deposit. Especially high is LCR indicator (average of 335% being 2.5x times higher than EU average) indicating high amount of liquid assets on bank's balances. We might expect that high ROE in hand with LCR might indicate that banks in these countries possess high amount of government bonds (probably with intention to help financing state debt due crisis) and bond spreads in these countries are significantly higher than, let us say, in Germany. This might help us explain, why these countries performed so well. Also given fact, that cluster 2 countries have highest CAR, meaning they have sufficient capital, we conclude banks in these countries are in good shape to absorb impact of the crisis. Fist cluster consist mainly of middle EU countries (like Czechia, Slovakia, Hungry and Austria) and the most important economics such as Germany, France, Italy, Belgium and Netherlands. Countries in cluster 1 are close to overall EU average in analyzed indicators. They are generally in good shape and their average values shows no extreme values (even though individual country KRIs on Fig. 3 might find some exceptions among particular countries). The last is cluster 4. This cluster we could label as northern countries because it consists of Denmark, Finland and Sweden. The last country in this smaller cluster is Luxembourg. All these countries belong among the best developed countries in the world and their bank's balance sheet structure is, as the analysis shown, way different from the rest of EU countries. They have exceptionally high Loan-to-deposit ratio related to the rest of the countries. Average LtD in cluster 4 is 203 %, while total EU average as of December 2020 was 107 %.

Cluster	Country	CAR	LR	ROE	LtD	LCR
1	Austria	20%	7%	5%	97%	184%
1	Belgium	21%	7%	5%	97%	183%
1	Czechia	24%	7%	7%	77%	166%
1	Germany	20%	5%	0%	122%	158%
1	France	19%	6%	4%	106%	167%
1	Hungary	18%	9%	10%	75%	218%
1	Italy	20%	7%	0%	97%	187%
1	Netherlands	23%	5%	3%	113%	165%
1	Slovenia	17%	8%	17%	61%	294%
1	Slovakia	19%	7%	6%	105%	184%
2	Bulgaria	24%	12%	6%	67%	251%
2	Estonia	29%	10%	7%	99%	183%
2	Croatia	25%	13%	5%	68%	176%
2	Lithuania	24%	7%	11%	63%	763%
2	Latvia	29%	9%	7%	66%	420%
2	Romania	24%	10%	11%	57%	357%
3	Cyprus	20%	9%	-4%	56%	312%
3	Spain	17%	6%	-4%	105%	187%
3	Greece	17%	9%	-7%	78%	175%
3	Ireland	23%	10%	-3%	82%	172%
3	Malta	23%	7%	-3%	52%	383%
3	Poland	19%	10%	0%	84%	220%
3	Portugal	18%	7%	0%	79%	244%
4	Denmark	23%	5%	4%	294%	184%
4	Finland	21%	6%	6%	185%	172%
4	Luxembourg	24%	7%	6%	151%	152%
4	Sweden	23%	5%	9%	182%	169%

Fig. 3. Country's KRIs and cluster segmentation.

Average by cluster	CAR	LR	ROE	LtD	LCR
1	20%	7%	6%	95%	191%
2	26%	10%	8%	70%	335%
3	19%	8%	-3%	77%	242%
4	23%	6%	6%	203%	169%
Total average	*21%*	*8%*	*0%*	*107%*	*221%*

Cluster average to *Total average*					
1	95%	89%	1332%	89%	86%
2	121%	134%	1844%	65%	152%
3	92%	108%	-706%	71%	110%
4	109%	74%	1430%	189%	77%

Fig. 4. Average KRIs by clusters and comparison to total KRIs averages.

This means, that banks in these regions provided much more loans, than they posses deposits from customers (they have basically twice more loans than deposits). This shows us two things. First is, that households in these countries tend to be highly indebted and the second is diversion from standard source of funding in banking sector, which are deposits. Banks have to gain deposits from customers in order to obtain funding, which they can use to lend loans. However, given long-time low interest rates caused by central banks with purpose to support economy leads to situation, when banks can issue covered bonds under extremely favorable conditions (for example in Denmark there were cases, when banks were able to issue covered bonds with negative interest, meaning they will pay less on maturity, than they borrowed). This all caused shift from standard sources of funding for banking industry, especially in these four countries. The rest of the KRIs of cluster 4 are quite comparable with total averages. They have slightly lowest LCR average, which might be caused by fact, that government bonds of these countries tend to yield very small interest (most likely negative) and therefore banks are not interested in possessing high amount of eligible assets in contrast to cluster 2 countries. For better visualization of clusters, choropleth map of EU is shown on Fig. 5.

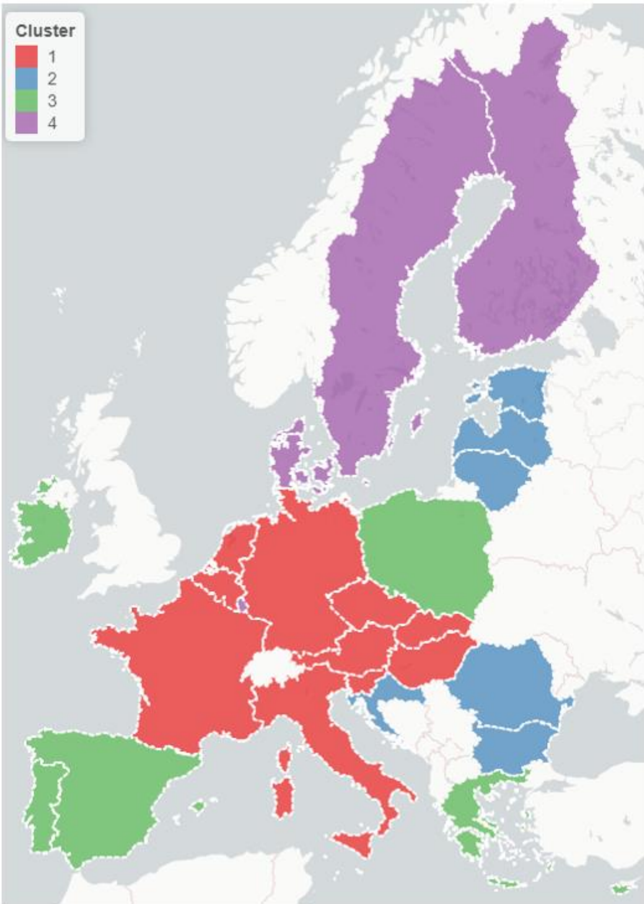


Fig. 5. Cluster segmentation based on hierarchical cluster analysis.

5 Discussion and conclusion

In this paper we focused on EU banking sector based on Risk Indicators which are on periodical basis disclosed by European Banking Authority with intent to compare different countries and find similarities and dissimilarities among them. For this purpose, cluster analysis was performed. Results from hierarchical cluster analysis are shown in this paper by usage of Ward's method and four clusters. During the analysis also non-hierarchical k-means algorithm with 4 centers was used in order to determine clusters and check feasibility of the hierarchical analysis. Results were the same as for hierarchical analysis and therefore only these are shown in the paper.

We divided countries into 4 clusters based on similarity among KRIs. First cluster consist mostly of middle and western European countries, which shows fairly average values of chosen KRIs. Second cluster we labeled as eastern EU countries, with surprisingly high return on equity and liquidity coverage ratio. Third cluster consists mostly of southern EU countries and is specific with very low profits (mostly losses) as shown by negative ROE and lowest available capital (CAR). This cluster is the most exposed to any upcoming crisis. Fourth and last cluster can be considered as north EU with specific of very high ratio of loans to deposits, depicting specifics of the country's banking sectors.

Analysis shows results as of December 2020. Some KRIs are quite significantly changing in time and for further research, comparison with 2021 results is suggested with purpose to identify changes during the year caused by development of coronavirus crisis and global economy. In the upcoming year, huge increase of banks balance sheet is expected, beside increased lending, also by increased drawing of TLTRO III (Targeted longer-term refinancing operations) from European Central Bank with purpose to borrow money under very favorable conditions. These operations (also significantly used by Slovak banks) will have huge impact on future development of several KRIs, such as Encumbrance ratio and Loan-to-Deposits ratio. Given their secured nature in terms of encumbered securities provided as a necessary collateral for central bank, also changes in LCR and LR are expected. Given fact, that balance sheet of banks in cluster 4 countries already bears high amount of encumbered collateral, we expect, that shift in other countries KRIs will be visible on year-to-year basis at the end of 2021 and movement in cluster average KRIs of first to third clusters will be closer to cluster 4 average KRIs.

Acknowledgement

This paper was elaborated within the framework of the project VEGA 1/0561/21 The impact of the COVID-19 crisis on business demography and employment in the Slovak Republic and the EU and project KEGA 007EU-4/2020 Interaktívna a interdisciplinárna výučba predmetov Služby a Inovácie v cestovnom ruchu s využitím informačných technológií.

References

1. Basel Committee on Banking Supervision: Laverage Ratio (2019).
2. Basel Committee on Banking Supervision: Liquidity Coverage Ratio (2019).
3. Basel Committee on Banking Supervision: Risk-based capital requirements (2019).
4. Castagna, A., Fede F.: Measuring and managing Liquidity risk. 1st edn. Whiley, London (2013).
5. European Banking Authority: The EBA Methodological Guide, https://www.eba.europa.eu/sites/default/documents/files/documents/10180/1380571/02c01596-5122-4d45-ac3e-23eb2f79f032/EBA%20Methodological%20Guide%20-%20Risk%20Indicators%20and%20DRAT.pdf?retry=1, last accessed 2021/07/05.
6. European Banking Authority: The European Banking Authority at Glance. https://www.eba.europa.eu/sites/default/documents/files/documents/10180/1401372/e8686db2-6390-4c52-ad06-bc8d24b7aeb5/EBA%20AT%20A%20GLANCE.pdf?retry=1, last accessed: 2021/07/05.
7. EBA Risk Dashboard, https://www.eba.europa.eu/risk-analysis-and-data/risk-dashboard, last accessed 2021/07/05.
8. P. Farahvash: Asset-Liability and Liquidity Management. 1st edn. Whiley, London (2020).
9. Hull, J. C.: Risk Management and Financial Institutions. 5th edn. Whiley, London (2018).
10. Král, P. et al.: Viacrozmerné štatistické metódy so zameraním na riešenie problémov ekonomickej praxe. Univerzita Mateja Bela (2009).
11. Nikam, V., Jhajhria, A., Pal, S: Quantitative Methods for Social Sciences. ICAR (2019).
12. Stankovičová I., Vojtková, M.: Viacrozmerné štatistické metódy. Wolters Kluwer (2007).
13. Teplý, P.: The Level of Capital and the Value of EU Banks under Basel III. Praque Economic Papers 23(2):143-161.