

Empirical Paper

Michał Bernardelli[✉], Zbigniew Korzeb^{*}, Paweł Niedziółka[✉]

What drives the profitability of banking sectors in the European Union? The machine learning approach

<https://doi.org/10.2478/ijme-2024-0022>

Received: March 17, 2024; accepted: June 28, 2024

Abstract:

The study aims to establish patterns of relations between the profitability of the European Union (EU) banking sectors between 2007 and 2021 and sets of variables appropriate for clusters of countries into which the 27 countries of the EU are divided. The random forest method is deployed to identify the factors influencing the value of the return on equity. Shapley additive explanations are exploited to add interpretability to the results. The results show that the sets of variables shaping the profitability of banking sectors in the EU grouped by use of sovereign rating criterion are different. However, there are variables common to all banking sectors. These include cost efficiency and default risk. The study's novelty lies in the reliance on a broad spectrum of explanatory variables assigned to three groups of factors, reference to all EU countries, and decomposition of the sample to identify similarities among the determinants of profitability.

Keywords: banking sector, machine learning, profitability, random forest, SHAP

JEL Classification: G01, G21, G30, Z10

1 Introduction

The profitability of bank operations is one of the hot topics of research by academics and market practitioners, including investors. First, the profitability of the banking sector of the host country is a determinant of foreign direct investment (FDI) in the financial industry [Focarelli and Pozzolo, 2000]. Second, profitability affects the dynamics of regulatory capital accumulation, which drives the pace of lending [Bustamante et al., 2019]. Finally, knowledge of the determinants of the banking sector's profitability makes it possible to answer whether, in what area and to what extent, the macroeconomic or regulatory environment changes alter bank performance [Rahman et al., 2015]. In the last 15 years, the world has experienced the subprime crisis, the sovereign debt crisis, the migration crisis, and the COVID-19 pandemic. This study primarily focuses on the profitability of banking sectors in the European Union (EU) countries over the period covering most of the above events, that is, 2007–2021. The choice of research period is therefore driven by the intention to test whether regulatory action in response to the subprime and pandemic crises has altered the importance of individual determinants of profitability of European commercial banks. Previous studies [e.g., Bongini et al., 2019; Bonaccorsi di Patti and Palazzo, 2020] have focused on the post-subprime crisis

***Corresponding author: Zbigniew Korzeb**, Department of Management, Economy and Finance, Białystok University of Technology, Kleosin, Poland. E-mail: z.korzeb@pb.edu.pl

Michał Bernardelli, Collegium of Economic Analysis, SGH Warsaw School of Economics, Warsaw, Poland.

Paweł Niedziółka, Collegium of Socio-Economics, SGH Warsaw School of Economics, Warsaw, Poland.

 Open Access. © 2024 Bernardelli et al., published by Sciendo.

 This work is licensed under the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

period. This study extends the period of analysis to include the pandemic period. The choice of the EU stems from the intention to conduct the study in a single regulatory and supervisory environment.

We narrow our study to the identification of profitability determinants, deliberately separating this sphere from performance measures, in those considering the market value of the bank (e.g., Tobin's Q). Like Yüksel et al. [2018], we have chosen return on equity (ROE) as a measure of profitability because this measure lacks the disadvantage that characterizes the commonly used formula for determining the return on assets (ROA), that is, net income/total assets. In the ROA equation, liabilities include debt and equity investors' contributions, whereas net income reflects only the return to equity investors. By contrast, adjusting the ROA calculation to make the numerator include the income of creditors and owners would require information on interest expenses and tax rates by year, and these are not present in the databases that we have used. The omission of the net interest margin (NIM), on the other hand, is due to the lack of data for a broad sample of banks and the relatively small number of studies that verify this measure of profitability.

In this study's framework, EU countries are combined into four groups according to their characteristics. According to the criterion of similarity of trajectories of change for variables, each variable is assigned to the three groups (macroeconomic, banking, and governance). The random forest model is used within each country group, with ROE as a dependent variable. Using the machine learning (ML) approach allows for high accuracy of ROE assessment. Identification of potential ROE determinants, however, is not straightforward due to the lack of direct interpretability of the results of ML models. A modern SHapley Additive exPlanations (SHAP) technique for examining interpretability proposed by Lundberg and Lee [2017] is used to overcome these explainability problems.

The novelty of this study consists in concentrating on the entire banking sectors of all EU countries rather than individual banks, the period between the subprime and pandemic crises. This is also the first study to highlight the importance of globalization, economic freedom, and the risk of banking sector insolvency as variables determining banking sector profitability.

The rest of the article is organized as follows: in Section 2, we review relevant literature and develop hypotheses. Section 3 describes the methodology and data. Section 4 reports our empirical results, the next one discusses them while Section 6 covers conclusions limitations and directions for further studies.

2 Literature review and hypotheses development

2.1 Macroeconomic and business environment variables

Bonaccorsi di Patti and Palazzo [2020] suggest that **GDP growth** is one of the main factors supporting bank profitability. In turn, Mondol and Wadud [2022] prove the negative impact of GDP growth on banks' ROA. Kok et al. [2015] add that the development of economic activity increases the demand for banking products and thus contributes to banks' net interest income and fee income. Although we have not found any studies in the literature on the direct impact of public debt on bank profitability, however since the increase in **public debt** weakens economic growth [Heimberger, 2021], we expect that the growth of public debt indirectly contributes to the deterioration of bank profitability.

Sekwati and Dagume [2023] state that **inflation** and unemployment have a negative impact on GDP growth. Since GDP growth is positively correlated with the bank profitability, a negative impact of inflation on bank results can be expected. Inflation causes a decrease in the part of the income that can be used for debt service, while capital and interest payments will grow up. This thesis is confirmed by Pradhan [2016]. On the other hand, however higher inflation is accompanied by greater interest rates, and these, due to the possibility of leveraging the result in products such as cash loans and increasing the differential between the reference rate and the actual cost of funding, contribute to improving the bank's net result [Caliskan and Kirer-Silva-Lecuna; 2020; Saif-Alyousfi, 2020]. Since employment has a positive impact on bank profitability [Genay and Podjasek, 2014] and labor cost moderation helps employment creation [Pierluigi and Roma, 2008], a negative influence of **labor costs** on bank performance can be expected.

Brahmaiah and Ranajee [2018] find that deposits/GDP positively affect ROA and ROE of banks. This suggests that the broader category, that is, **savings**, will also be positively correlated with the profitability of the banking sector.

Based on the inconclusive results of the studies cited above, the following hypothesis is formulated:

H1: The GDP growth, inflation, public debt, savings, and labor costs influence European banking sectors' profitability.

Teixeira et al. [2019] highlight that higher levels of the institutional environment or stricter banking regulation typically reduce banks' profitability. However, they note that this effect weakens considerably during crises. It implies that banking sectors subject to more rigorous regulation and monitoring by supervisory institutions are less exposed to the adverse effects of emergencies. Asteriou et al. [2021] find that corruption and transparency have a negative effect on bank profitability while economic freedom and regulation positively impact the performance of banks. Nasreen et al. [2020] state that financial globalization hinders the process of financial sector development, as expressed by the dimensions of the financial market. Athari [2021] pays attention to the global economic policy uncertainty as a factor determining profitability of banks. Based on the findings mentioned above, the following hypothesis is tested:

H2: Banking sectors' profitability in the EU depends on the business and regulatory environment as well as openness of the economy.

2.2 Bank-specific and structural determinants

Mondol and Wadud [2022] find that the level of profitability is determined mainly by the bank's **credit risk management** quality. It is also confirmed by the study of Elkedag et al. [2020]. According to them, the most important factor (besides real GDP growth) is the level of non-performing loans (NPL). Menicucci and Paolucci [2016] argue that higher loan loss provisions result in lower profitability levels. Bongini et al. [2019] conclude that the main reason for the dynamic changes in the profitability of European banks is the credit policy pursued. Therefore, the main factor influencing banks' profitability during crises is the application of conservative lending policies and the effective recovery of NPL. Karkowska and Niedziółka [2019] also demonstrate that the profitability of European banks is positively affected by their credit policies. Hasanov et al. [2018] present a positive impact of bank **liquidity risk** on its profitability. In contrast to the results of the studies mentioned above, Yuan et al. [2022] and Mondol and Wadud [2022] indicate a negative relation between liquidity and profitability. The **cost efficiency** of banks and the level of operating costs are pretty often cited as determinants of the profitability of banking sectors. For example, Mehzabin et al. [2023] suggest that the impact of operational efficiency on profitability is positive. Also, Neves et al. [2020] prove that cost management is one of the main drivers of bank performance. Khan [2022] proves a significant relationship between operating efficiency and ROE. The positive impact of operating efficiency on ROA and ROE is confirmed by Mondol and Wadud [2022].

As the quality of risk management and cost efficiency in individual banks translates into indicators describing the risk of the banking sectors formed by these banks, it can be expected that:

H3: The banking sector's risk affects its profitability.

H4: The ROE of European banking sectors is impacted by their cost-efficiency ratios.

The positive impact of bank **capital** on its profitability is confirmed *inter alia* by Menicucci and Paolucci [2016] and Saif-Alyousfi [2020]. Regarding capital adequacy, Khan [2022] proves its negative impact on bank ROA, whereas Mondol and Wadud [2022] come to the opposite conclusion. Rossi et al. [2018] attempt to identify the determinants of bank profitability in nine European countries to see evident changes in the context of the onset of the global crisis. In their view, "after a period of 'irrational exuberance' in which credit growth and high leverage are seen as proper and fast ways to boost profitability, a sound financial structure and a wiser and objective credit portfolio management have become the main drivers to ensure higher returns." Having in mind the aforementioned results, we expect that:

H5: The capitalization of the banking sector is an essential determinant of the return on capital in this industry.

Yuanita [2019] and Horobet et al. [2021] argue that a higher **concentration** ratio measured by the Herfindahl–Hirschman Index (HHI) and illustrating banking sector structure is associated with lower profitability, while results achieved by Silalahi et al. [2015] suggest the opposite direction of the relation. Also, Keil [2017] points to the negative impact of the resurgence of market concentration on the sector's profitability. Having these ambiguous results in mind, we presume that:

H6: The concentration of the banking sector determines its profitability.

With fresh memories of the subprime crisis, the sovereign debt crisis, the COVID-19 crisis and the current economic slowdown, energy market shock, rapidly rising inflation, and tense geopolitical situation, it is also worth citing the results of studies dedicated to the impact of these extreme events on banking sector performance. Examining the resilience of Polish banks to the pandemic crisis, Karkowska et al. [2023] reveal that the largest banks are the most resistant ones to the consequences of the pandemic. In turn, Bernardelli et al. [2021], analyzing the behavior of banks during the COVID-19 crisis, prove that large retail banks have been less affected than medium-sized ones with relatively meaningful corporate portfolios.

Furthermore, given the heterogeneity of the sets of explanatory variables of relevance to the profitability of the different groups of banks (distinguished based on geography, business model, financial condition, and others) as well as different structures of EU banking sectors, we expect that:

H7: Different sets of explanatory variables influence the profitability of clusters of European banking sectors.

3 Data and methodology

The study is based on variables from 2007 to 2021 for all EU-27 countries, which are assigned to three categories (Table 1).

The sovereign ratings as of the end of 2021 granted by S&P, Moody's and Fitch Ratings are considered and then normalized on a scale of 0–21 points. In the next step, the averages of the ratings mentioned above are determined and assigned to four ranges (<14; <14.00–16.49>; <16.50–18>; and <18.01–21>). Based on this approach, 27 EU countries are divided into four groups:

- Group 1: Bulgaria, Croatia, Cyprus, Greece, Hungary, Italy, Portugal, and Romania.
- Group 2: Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia, and Spain.
- Group 3: Czech Republic, Estonia, and Ireland.
- Group 4: Austria, Belgium, Denmark, Finland, France, Germany, Luxembourg, Netherlands, and Sweden.

The above division is based on the criterion of assigning a given banking sector to a cluster characterized by a rating that falls within a certain range with the same spread. The exception is the first band containing the lowest ratings. The ROE is chosen to approximate the profitability rates of the banking sectors (Figure 1).

The crises that occurred within the years 2008–2013 affected the banking systems of individual countries to varying degrees. Hence, it is decided to omit the observation of countries in the year when the absolute value of ROE is >25, treating them as outliers. Detailed descriptive statistics for ROE are provided in Table 2.

Machine learning methods (ML) are usually often characterized by greater accuracy than classic econometric approaches regarding prediction or fitting quality [Pérez-Pons et al., 2021]. These models are often presented as the so-called “black boxes” due to the problematic interpretability of the results and the influence strength of individual factors contributing to these results. A frequently used technique for examining interpretability is SHAP, which are proposed by Lundberg and Lee [2017]. Shapley values determine the importance of a given variable by comparing the model's predictions with and without that variable, considering any possible ordering of the remaining variables. It is a valuable substitute for the interpretability of the econometric models, providing insights into the determinants of the phenomena under consideration. As regards the advantages of this methodology, the prediction is fairly distributed among the feature values. Instead of comparing a prediction to the average prediction of the entire dataset, it can be confronted with a subset or even with a single data point. SHAP requires a lot of computing time,

Table 1. Taxonomy of variables

Symbol	Description	Source
Bank-specific variables		
Assets	Total assets of the banking sector	ECB
Dyn_Assets	Asset dynamics in the banking sector	ECB
Liquid	Liquid assets to deposits and short-term funding (%)	World Bank
CIR	Bank cost-to-income ratio (%)	World Bank
Bank_Conc	Assets of the three largest commercial banks as a share of total commercial banking assets	World Bank
Foreign	Share (total assets) of foreign credit institutions	ECB
NPL	Bank NPL to gross loans (%)	World Bank
Z score	Bank Z-score	World Bank
Macroeconomic variables		
GDP_Growth	GDP growth (annual %)	World Bank
Gov_Debt	Government debt (consolidated) (as % of GDP)	ECB, Eurostat
PPI	Producer Prices (2010 = 100)	Eurostat
Labor	Labor cost for LCI (compensation of employees plus taxes minus subsidies) in euro	Eurostat
Savings	Gross savings (% of GDP)	ECB, Eurostat
Business environment variables		
GLOB	KOF Globalization Index	KOF Swiss Economic Institute
Freedom	An index that measures economic freedom based on trade freedom, business freedom, investment freedom, and property rights	Heritage Foundation

Source: Own elaboration.

NPL, non-performing loans; PPI, Producer Price Index.

and it can produce unintuitive feature attributions, which seem to be the disadvantage of this approach [Molnar, 2023].

Modeling using the set of all variables is performed independently for each group of countries. The modeling consists of two stages. The first stage is the selection and application of an ML model. The second stage is to assess the degree of influence of explanatory variables on the explained variable ROE using the SHAP value. Among many available ML methods, the random forest model is chosen. The random forest algorithm has been applied in many areas to make better predictions, estimations, and decisions. A version of the random forest algorithm used in modern applications was developed by Breiman [2001]. By integrating the combination of the bagging concept [Breiman, 1994] and the random selection of features introduced by Ho [1995], constructing the set of decision trees with controlled variance is possible. This research uses a fast implementation of random forests from the Ranger package written in R language (R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form). This implementation is particularly suited for high-dimensional data.

4 Results

The analysis of the results of the random forest models is performed on many levels. First, the quality of the models measured by the root-mean-square error (RMSE) is compared for data in the considered groups of countries from all available years, and the results are given in Table 2. RMSEs are in the range of 1.7–2.9, which – concerning the range of ROE (from 6 to 132, see Table 2) – seems to be an excellent approximation. Fitting accuracy and the associated RMSE vary for different countries. Descriptive statistics per country are presented in Table 3. For most countries, the absolute value of the average RMSE is <1. Only Latvia and

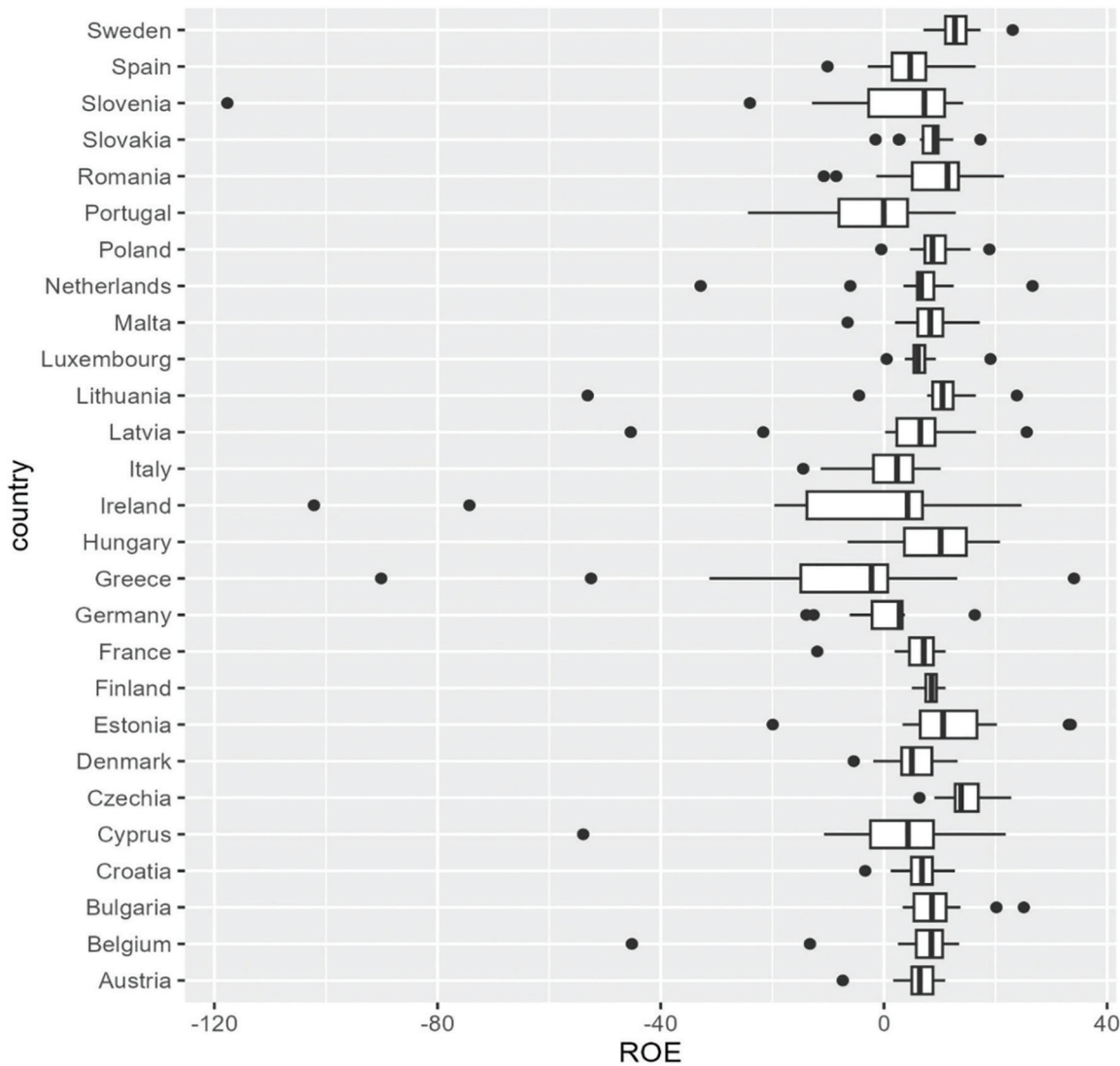


Figure 1. Distribution of ROE values for individual countries in 2007–2021.

Source: own calculations. ROE, return on equity.

Ireland are above that threshold. In the case of Ireland, the main reason for worse model accuracy is the lower number of observations due to missing data compared with other countries. The more significant variance of results influences the slightly larger average error in the case of Latvia (see the SD column in Table 4), which is also observed for Slovenia, Ireland, and Romania. The discrepancy of random forest model estimates from the actual ROE values exceeds 10 for Cyprus, Estonia, Hungary, Lithuania, and Portugal. A visible relationship exists between the range of ROE values and the range of estimation errors for individual countries (see columns with range values in Tables 2 and 4). Spearman's correlation coefficient for these values is 0.76 (p -value = $4.332e-06$).

Additionally, Table 3 shows a summary of R^2 , also known as explained variance or coefficient of determination. It is computed on out-of-bag data and indicates what percentage of the variance in the ROE values the random forest regression model explains. For the first and the fourth groups, R^2 exceeds 50%. The other two groups are worse, but still, the model explains more than one-third of the variance in the ROE values.

The second part of the analysis concerns determining the strength of the influence of the considered factors on the explained variable. The SHAP approach is used in which each feature assigns an importance

Table 2. ROE descriptive statistics per country

Group 1				Group 2			
Country	Mean	SD	Range	Country	Mean	SD	Range
Bulgaria	9.72	6.09	21.70	Latvia	2.78	16.59	70.96
Croatia	6.36	3.93	16.06	Lithuania	6.67	17.52	76.91
Cyprus	0.89	17.12	75.72	Malta	7.89	5.45	23.66
Greece	-11.06	29.19	124.16	Poland	9.35	4.52	19.37
Hungary	8.37	8.39	27.32	Slovakia	8.11	4.39	18.78
Italy	0.69	7.15	24.61	Slovenia	-4.62	32.98	131.88
Portugal	-2.43	10.18	37.25	Spain	4.33	6.53	26.53
Romania	8.31	9.09	32.28				
Group 3				Group 4			
Country	Mean	SD	Range	Country	Mean	SD	Range
Czech Republic	14.39	4.13	16.43	Austria	5.89	4.66	18.36
Estonia	11.49	12.68	53.39	Belgium	3.63	14.90	58.66
Ireland	-9.68	34.18	126.80	Denmark	5.32	4.92	18.61
				Finland	8.29	1.70	6.05
				France	5.69	5.57	22.99
				Germany	0.33	7.34	30.19
				Luxembourg	6.81	3.99	18.60
				Netherlands	5.22	12.41	59.49
				Sweden	13.30	3.70	15.96

Source: own calculations.

ROE, return on equity.

Table 3. Values of RMSE and R^2 for random forest models per group of countries

Measure	Group 1	Group 2	Group 3	Group 4
RMSE	2.33	2.59	2.90	1.70
R^2 [%]	57.28	37.50	33.51	52.44

Source: own calculations.

RMSE, root-mean-square error.

value to the particular feature. It determines the extent to which removing a variable reduces the accuracy of the prediction. The values alone do not carry information, but in the context of other values, information can be obtained about the relative strength of the influence of individual variables on the ROE variable. The higher the value, the more significant the impact. The bar plots in Figure 2 show SHAP feature importance, calculated as the average absolute SHAP value per feature for each group of countries independently. The top 10 features with the most significant importance are plotted.

As a robustness check, the method of division into groups has been changed by limiting it to the S&P ranking at the end of 2021. Due to the exact composition of countries, the results for Groups 1 and 4 have not changed. Moving Slovenia from Group 2 to Group 3 resulted in a slight improvement in the RMSE error in Group 2 (mean RMSE of 2.07), at the cost of worsening the mean RMSE in Group 3 to 2.99. The increase in the number of observations in Group 2 also resulted in a significant increase in R^2 from 33.51 to as much as 49.85. At the same time, the decreasing number of observations in Group 2 (which still includes many countries) only slightly reduces the R^2 coefficient to 37.26. The stability of the results concerning the original division into groups proves the approach used well.

Table 4. Random forest model errors per group of countries

Group 1				Group 2			
Country	Mean	SD	Range	Country	Mean	SD	Range
Bulgaria	0.19	1.50	6.22	Latvia	−1.02	3.81	16.74
Croatia	−0.02	0.79	3.15	Lithuania	0.57	2.85	13.22
Cyprus	−0.19	2.90	10.98	Malta	−0.12	1.37	5.75
Greece	0.36	2.16	7.19	Poland	0.29	1.25	5.23
Hungary	0.42	2.96	10.90	Slovakia	0.19	1.97	7.41
Italy	0.16	2.15	7.14	Slovenia	0.27	3.79	15.02
Portugal	−0.76	2.79	10.35	Spain	−0.01	2.38	9.20
Romania	0.23	2.93	11.10				
Group 3				Group 4			
Country	Mean	SD	Range	Country	Mean	SD	Range
Czech Republic	0.37	1.03	3.98	Austria	−0.03	1.98	7.95
Estonia	−0.65	3.47	13.44	Belgium	−0.55	2.31	9.54
Ireland	−1.81	4.19	11.73	Denmark	−0.55	1.41	5.38
				Finland	0.03	0.65	2.15
				France	−0.03	1.94	8.23
				Germany	−0.54	2.44	9.20
				Luxembourg	−0.12	0.92	3.38
				Netherlands	0.27	1.13	4.06
				Sweden	0.63	1.64	6.84

Source: own calculations.

5 Discussion

Econometric models have the advantage of interpretability but, at the same time, generally have weaker predictive ability and much more restrictive requirements for meeting applicability assumptions. ML methods supplemented with the SHAP method are an excellent alternative [Pérez-Pons et al., 2021; Molnar, 2023]. The combination of these two methods in the analysis of financial data has so far been rare in scientific research. Meanwhile, the results of the empirical analysis described in this article prove the effectiveness of this approach. The application of the SHAP framework for interpreting the output of the random forest model made it possible to isolate macroeconomic and banking variables, where, most likely, there is a causal relationship linking these variables to the ROE of the banking sectors in the analyzed period of 2007–2021. Thus, the results show the mechanism of action of the studied variables with an additional indication of the strength of the relationship between the analyzed variables and ROE.

The order of variables regarding the strength of influence affecting the ROE value differs depending on the group of countries, which allows confirming H7. The Government debt/GDP and Liquidity variables lead in Group 1, EU-27 in Group 2, and the Cost/Income (C/I) in Groups 3–4. The C/I and Z-Score variables appear in the top five in each group, with three groups having the most significant impact on the value of the explained variable. Leaders in Groups 1, 3, and 4 stand out significantly in the strength of their influence. Group 2 has no such drastic difference between the first and subsequent variables.

Demonstrating the significant impact of cost efficiency on the banking sector's profitability allows H4 to be confirmed. Given banks' relatively stable capital base (the denominator of ROE), the influence of operating costs relative to income on the banking sector's profitability is quite apparent. Therefore, the gradual reduction of the C/I ratio is becoming an integral part of the banks' strategies [Broby, 2021].

The importance of the C/I ratio is also emphasized by Mehzabin et al. [2023], Khan [2022], Neves et al. [2020], and among others.

The importance of cost-effectiveness is particularly significant in the third and fourth group of countries (economies with relatively high sovereign ratings), which can be linked to the limited possibilities of realizing additional income in an environment of high interest rates following inflation. It results from the position of the Producer Price Index (PPI) is significantly higher for groups of countries with weaker ratings.

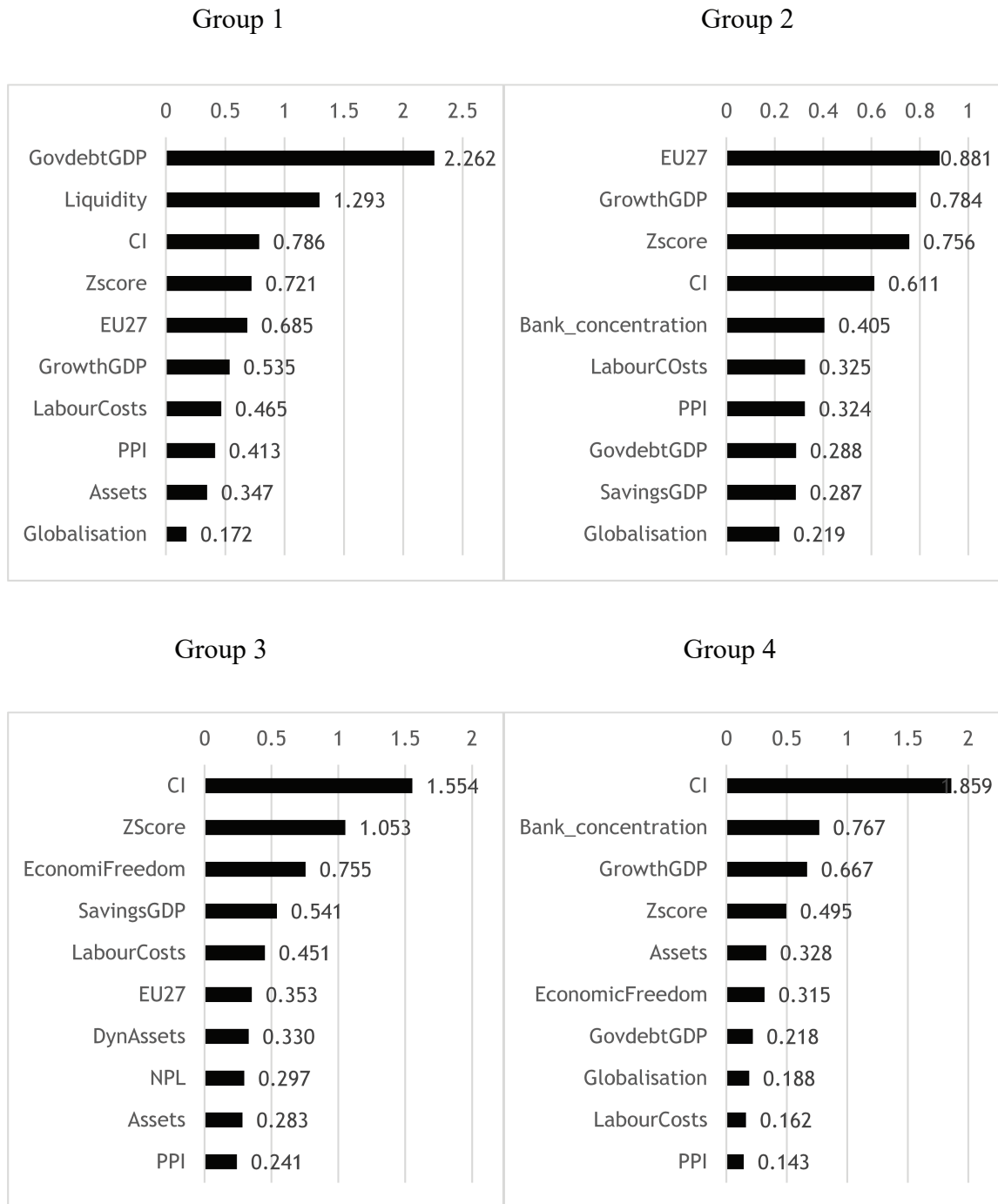


Figure 2. SHAP features importance per group of countries.

Source: own calculations. NPL, non-performing loans; PPI, Producer Price Index; SHAP, SHapley Additive exPlanations.

In turn, demonstrating the significance of the Z-Score, a measure of the banking sector's default risk, allows confirmation H3. A similar relationship is proven by Mushafiq et al. [2022]. A relatively higher cost of funding accompanies higher default risk, which ultimately determines the pricing of financial products and services in a given market. This relationship can also be explained from the perspective of the banking system's abundance of capital. The lower its value, the greater the default risk and, simultaneously, *ceteris paribus*, the higher the level of ROE in light of the Z-Score concept. The relevance of the Z-score in the context of the crucial role of own funds in its calculation allows positive verification of H5.

Our results also indicate the relatively high importance of the share of foreign ownership in banking sector assets as a determinant of profitability. This variable is among the five most significant in three groups of countries, and the finding of this relationship is in line with the results obtained by Othmani [2022], Gupta et al. [2020], and Rahman et al. [2020], and among others. The effect of sovereign credit rating on foreign ownership demonstrated by Korzeb et al. [2023], in turn, confirms the validity of combining macroeconomic and bank-specific variables as determinants of bank profitability. This factor is important primarily in the groups of countries with relatively low ratings, where the share of foreign ownership in banking sector assets is higher than it is in the case of the countries that make up the fourth group. In the cluster mentioned above, the foreign ownership is so low that it is difficult to indicate its impact on the banking sector's profitability.

The concentration of the banking sector as a variable with a significant impact on its profitability is included in the list of the five most essential factors in two clusters of countries, which allows partial positive verification of H6. Studies by Yuanita [2019], Horobet et al. [2021], and Silalahi et al. [2015] support our results. The higher concentration is accompanied by less competition, and the oligopolistic structure of the banking sector translates into relatively higher profitability for huge banks [Vera-Gilces et al., 2020].

Liquidity opposes profitability and is particularly important in relatively low-rated countries, at greater risk of speculative capital outflows. This low sovereign rating also determines the ability and cost of funding for banks in these countries. Our results, confirming H3, are in line with the findings of Yuan et al. [2022], Mondol and Wadud [2022], Caliskan and Kirer-Silva-Lecuna [2020], and Hasanov et al. [2018].

Among the macroeconomic variables, GDP growth and PPI are significantly affecting the profitability of banking sectors, which allows for positive verification of H1. Additionally, our study reveals the importance of such macroeconomic variables like government debt/GDP and labor costs. The government debt/GDP is the dominant determinant in the first group of countries, that is, economies with the lowest average sovereign ratings. They are characterized by relatively high public sector debt, and some of them have experienced problems with financial sector stability (e.g., Cyprus, Greece, Hungary), which has been sustained by state support or bail-outs from foreign investors.

The significance of the impact of GDP growth on bank profitability is demonstrated by Mondol and Wadud [2022], Saif-Alyousfi [2020], and Bonaccorsi di Patti and Palazzo [2020]. Labor costs and savings also have a relatively significant impact on the profitability of the banking sectors, but this applies only to the third group of countries. It is formed by countries with a relatively high share of foreign capital in the ownership structures of banks, so economic freedom can be considered an incentive for foreign investors, who can realize higher income in such a regulatory and tax environment than in their home countries.

Inflation as a factor affecting the profitability of the banking sector is, in turn, pointed out by Saif-Alyousfi [2020] and Caliskan and Kirer-Silva-Lecuna [2020]. The higher interest rates accompany higher inflation. Leveraging the result in products such as cash loans and increasing the differential between the reference rate and the actual cost of funding contributes to improving the bank's net result.

It is to be expected that the legal and regulatory environment and the quality and enforceability of a country's existing laws, as well as political stability and the extent of economic freedom (our study confirms its impact on banking sector profitability which allows to verify positively H2), will have a positive impact on the profitability. However, this issue requires in-depth research. Therefore, the conducted analyses will provide a basis for further detailed research focusing on: (i) building a model describing the determinants of the profitability of the banking sectors of the EU-27 countries, taking into account the strength, shape, and direction of this impact in individual subperiods and groups of banking sectors, (ii) explaining the reasons for the emergence of outliers and diagnosing the idiosyncratic determinants of such phenomena,

and (iii) identifying the reasons for the differentiation and belonging of individual countries to the formed clusters using additional factors besides macroeconomic and banking factors, such as ESG and cultural factors.

The performance of the banking sector is vital to investors (since it affects enterprise value), then the determinants of its volatility should also be in the orbit of their interest. Supervisory institutions should also be among the potential recipients of the results of this study since profitability affects capital adequacy and liquidity ratios. The profitability of a banking sector is also essential from the point of view of fiscal and economic policy since banking sector profits impact future lending. The banking sector's profitability also determines the possibility of using the banking sector to mitigate the adverse social effects of recession or inflation.

6 Conclusions

This article establishes a list of potential determinants of profitability of banking sectors. Based on the random forest method and the SHAP values, the strength of influence of each factor within the particular group of countries is established. To our knowledge, this is one of the first studies focusing on entire banking sectors rather than individual banks, covering all EU countries, the period between the subprime and pandemic crises, and the first study emphasizing the importance of the globalization, economic freedom, and default risk of the banking sector.

This study and its results entail certain limitations. The first is the high level of data aggregation (analysis of entire banking sectors), which may cause some distortion of the picture resulting from the fact that the results of the industry are dominated either by the financials of a single or a few large commercial banks or by the significant underperformance or overperformance of one of the banks, which does not necessarily belong to the largest institutions in the sector. The second limitation is that sectoral results have not been cleaned of the effects of exceptional or one-off events (e.g., the option crisis in Poland and the conversion of FX mortgages in Hungary). Additionally, only European banking systems are analyzed. These sectors, given certain criteria, are similar to each other that may narrow the spectrum of candidates for profitability determinants and cause exclusion of variables like cultural factors or Economic Policy Uncertainty Index (EPUE).

As regards the future directions of research, it would be interesting to compare the results of the study that uses ML and SHAP methodology with other (traditional) approaches like linear regression.

Acknowledgments

This work was supported by International Academic Partnership Programme No. BPI/PST/2021/1/00011/U/00001 with the Polish National Agency for Academic Exchange.

References

- Asteriou, D., Pilbeam, K., Tomuleasa, I. (2021), The impact of corruption, economic freedom, regulation and transparency on bank profitability and bank stability: evidence from the Eurozone area, *Journal of Economic Behavior & Organization*, Vol. 184, pp. 150–177, <https://doi.org/10.1016/j.jebo.2020.08.023>.
- Athari, S.A. (2021), Domestic political risk, global economic policy uncertainty, and banks' profitability: evidence from Ukrainian banks, *Post-Communist Economies*, Vol. 33, No. 4, pp. 458–483, <https://doi.org/10.1080/14631377.2020.1745563>.
- Bernardelli, M., Korzeb, Z., Niedziółka, P. (2021), The banking sector as the absorber of the COVID-19 crisis? Economic consequences: perception of WSE investors, *Oeconomia Copernicana*, Vol. 12, No. 2, pp. 335–374, <https://doi.org/10.24136/oc.2021.012>.
- Bonaccorsi di Patti, E., Palazzo, F. (2020), Bank profitability and macroeconomic conditions: Are business models different? *Economic Notes*, Vol. 49, No. 2, e12155, <https://doi.org/10.1111/ecno.12155>.

- Bongini, P., Cucinelli, D., Di Battista, M.L., Nieri, L. (2019), Profitability shocks and recovery in time of crisis evidence from European banks, *Finance Research Letters*, Vol. 30, pp. 233–239, <https://doi.org/10.1016/j.frl.2018.10.003>.
- Brahmaiah, B., Ranajee (2018), Factors influencing profitability of banks in India, *Theoretical Economics Letters*, Vol. 8, pp. 3046–3061, <https://doi.org/10.4236/tel.2018.814189>.
- Breiman, L. (1994), Bagging predictors, Technical Report, Department of Statistics, University of California Berkeley, 421.
- Breiman, L. (2001), Random forests, *Machine Learning*, Vol. 45, No. 1, pp. 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Broby, D. (2021), Financial technology and the future of banking, *Financial Innovation*, Vol. 7, No. 47, <https://doi.org/10.1186/s40854-021-00264-y>.
- Bustamante, J., Cuba, W., Nivin, R. (2019), Determinants of credit growth and the bank-lending channel in Peru: a loan level analysis, *BIS Working Papers*, No. 803, retrieved from <https://www.bis.org/publ/work803.htm> [2nd February 2024].
- Caliskan, M.M.T., Kirer-Silva-Lecuna, H. (2020), The determinants of banking sector profitability in Turkey, *Business and Economics Research Journal*, Vol. 11, pp. 161–167, <https://doi.org/10.20409/berj.2020.242>.
- Elekdag, S., Malik, S., & Mitra, S. (2020). Breaking the bank? A probabilistic assessment of Euro area bank profitability. *Journal of Banking & Finance*, 120(C), 105949, <https://doi.org/10.1016/j.jbankfin.2020.105949>
- Focarelli, D., Pozzolo, A. (2000), The determinants of cross-border bank shareholdings: an analysis with bank-level data from OECD countries, *Federal Reserve Bank of Chicago, Proceedings*, pp. 199–232.
- Genay, H., Podjasek, R. (2014), What is the impact of a low interest rate environment on bank profitability? *Chicago Fed Letter*, No. 324, retrieved from <https://www.chicagofed.org/publications/chicago-fed-letter/2014/july-324> [4th May 2024].
- Gupta, N., Mahakud, J., McMillan, D. (2020), Ownership, bank size, capitalisation and bank performance: evidence from India, *Cogent Economics & Finance*, Vol. 8, No. 1, <https://doi.org/10.1080/23322039.2020.1808282>.
- Hasanov, F.J., Bayramli, N., Al-Musehel, N. (2018), Bank-specific and macroeconomic determinants of bank profitability: Evidence from an oil-dependent economy, *International Journal Financial Studies*, Vol. 6, No. 78, pp. 2–21, <https://doi.org/10.3390/ijfs6030078>.
- Heimberger, P. (2021), Do higher public debt levels reduce economic growth? FMM Working Paper, No. 74, Hans-Böckler-Stiftung, Macroeconomic Policy Institute (IMK), Forum for Macroeconomics and Macroeconomic Policies (FMM), Düsseldorf.
- Ho, T.K. (1995), Random decision forests, *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, 14–16 August 1995, pp. 278–282.
- Horobet, A., Radulescu, M., Belascu, L., Dita, S.M. (2021), Determinants of bank profitability in CEE countries: evidence from GMM panel data estimates, *Journal of Risk and Financial Management*, Vol. 14, p. 307, <https://doi.org/10.3390/jrfm14070307>.
- Karkowska, R., Korzeb, Z., Matysek-Jędrych, A., Niedziółka, P. (2023), Banking, risk and crises in Europe: from the global financial crisis to COVID-19, 1st edn, London & New York, Routledge, <https://doi.org/10.4324/9781003351160>.
- Karkowska, R., Niedziółka, P. (2019), Rentowność banków komercyjnych a ich płynność w kontekście implementacji ilościowych norm płynności rekomendowanych przez Bazylejski Komitet ds. Nadzoru Bankowego, *Bank i Kredyt*, Vol. 50, No. 2, pp. 149–172.
- Keil, J. (2017), Explaining the concentration-profitability paradox, *Review of Political Economy*, Vol. 29, No. 2, pp. 209–231, <https://doi.org/10.1080/09538259.2017.1295945>.
- Khan, S. (2022), Determinants of banks profitability: an evidence from GCC countries, *Journal of Central Banking Theory and Practice*, Vol. 11, No. 3, pp. 99–116, <https://doi.org/10.2478/jcbtp-2022-0025>.
- Kok, C., Móré, C., Pancaro, C. (2015), Bank profitability challenges in euro area banks: the role of cyclical and structural factors. *Financial Stability Review*, retrieved from https://www.ecb.europa.eu/pub/pdf/fsr/art/ecb.fsrart201505_02.en.pdf [31st March 2024].
- Korzeb, Z., Niedziółka, P., Nistor, S. (2023), Sovereign creditworthiness and bank foreign ownership. An empirical investigation of the European banking sector, *Journal of International Financial Markets, Institutions and Money*, Vol. 89, p. 101857, <https://doi.org/10.1016/j.intfin.2023.101857>.
- Lundberg, S., Lee, S.-L. (2017), A unified approach to interpreting model predictions, *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4–9 December 2017, pp. 1–10, retrieved from https://www.researchgate.net/publication/317062430_A_Unified_Approach_to_Interpreting_Model_Predictions [10th January 2024].
- Mehzabin, S., Shahriar, A., Hoque, M.N., Wanke, P., Azad, M.A.K. (2023), The effect of capital structure, operating efficiency and non-interest income on bank profitability: new evidence from Asia, *Asian Journal of Economics and Banking*, Vol. 7, No. 1, pp. 25–44, <https://doi.org/10.1108/AJEB-03-2022-0036>.
- Menicucci, E., Paolucci, G. (2016), The determinants of bank profitability: empirical evidence from European banking sector, *Journal of Financial Reporting and Accounting*, Vol. 14, No. 1, pp. 86–115, <https://doi.org/10.1108/JFRA-05-2015-0060>.
- Molnar, Ch. (2023), Interpretable machine learning. A guide for making black box models explainable, retrieved from <https://christophm.github.io/interpretable-ml-book/index.html> [4th May 2023].

- Mondol, D.K., Wadud, Md.A. (2022), Determinants of profitability of commercial banking in Bangladesh: a panel analysis, *International Journal of Statistical Sciences*, Vol. 22, No. 1, pp. 115–143, retrieved from <https://csa.ru.ac.bd/stat-ijss/journals/v-22-iss-1-06/> [3rd November 2023].
- Mushafiq, M., Sami, S.A., Sohail, M.K., Sindhu, M.I. (2022), Merton-type default risk and financial performance: the dynamic panel moderation of firm size, *Journal of Economic and Administrative Sciences*, Vol. 40, No. 2, pp. 168–181, <https://doi.org/10.1108/JEAS-09-2021-0181>.
- Nasreen, S., Mahalik, M.K., Shahbaz, M., Abbas, Q. (2020), How do financial globalisation, institutions and economic growth impact financial sector development in European countries? *Research in International Business and Finance*, Vol. 54, p. 101247, <https://doi.org/10.1016/j.ribaf.2020.101247>.
- Neves, M.E., Proença, C., Dias, A. (2020), Bank profitability and efficiency in Portugal and Spain: a non-linearity approach, *Journal of Risk and Financial Management*, Vol. 13, No. 11, p. 284, <https://doi.org/10.3390/jrfm13110284>.
- Othmani, H. (2022), Foreign institutional ownership and bank performance: evidence from Tunisia, *African Development Review*, Vol. 34, pp. 42–53, <https://doi.org/10.1111/1467-8268.12625>.
- Pérez-Pons, M., Parra-Dominguez, J., Omatu, S., Herrera-Viedma, E., Corchado, J. (2021), Machine learning and traditional econometric models: a systematic mapping study, *Journal of Artificial Intelligence and Soft Computing Research*, Vol. 12, No. 2, pp. 79–100, <https://doi.org/10.2478/jaiscr-2022-0006>.
- Pierluigi, B., Roma, M. (2008), Labour cost and employment across euro area countries and sectors, ECB Working Paper, No. 912, retrieved from <https://ssrn.com/abstract=1147097> or <http://dx.doi.org/10.2139/ssrn.1147097> [5th May 2024].
- Pradhan, R.S. (2016), Bank specific and macroeconomic determinants of bank profitability: a case of Nepal, retrieved from <https://ssrn.com/abstract=2793484> or <http://dx.doi.org/10.2139/ssrn.2793484> [8th May 2024].
- Rahman, H., Yousaf, M.W., Tabassum, N. (2020), Bank-specific and macroeconomic determinants of profitability: a revisit of Pakistani banking sector under dynamic panel data approach, *International Journal of Financial Studies*, Vol. 8, No. 3, p. 42, <https://doi.org/10.3390/ijfs8030042>.
- Rahman, M., Hamid, Md., Khan, Md. (2015), Determinants of bank profitability: empirical evidence from Bangladesh, *International Journal of Business and Management*, Vol. 10, pp. 135–149, <https://doi.org/10.5539/ijbm.v10n8p135>.
- Rossi, S., Borroni, M., Lippi, A., Piva, M. (2018), Determinants of bank profitability in the euro area: what has changed during the recent financial crisis? *International Business Research*, Vol. 11, No. 5, pp. 18–27, <https://doi.org/10.5539/ibr.v11n5p18>.
- Saif-Alyousfi, A.Y.H. (2020), Determinants of bank profitability: evidence from 47 Asian countries, *Journal of Economic Studies*, Vol. 49, No. 1, pp. 44–60, <https://doi.org/10.1108/JES-05-2020-0215>.
- Sekwati, D., Dagume, M.A. (2023), Effect of unemployment and inflation on economic growth in South Africa, *International Journal of Economics and Financial Issues*, Vol. 13, No. 1, pp. 35–45, <https://doi.org/10.32479/ijefi.13447>.
- Silalahi, T., Manurung, A., Hidayat, Y. (2015), The Market structure of the bank, its performance and the macroprudential policy, *Buletin Ekonomi Moneter Dan Perbankan*, Vol. 18, No. 1, pp. 45–60, <https://doi.org/10.21098/bemp.v18i1.514>.
- Teixeira, J.C.A., Silva, F.J.F., Costa, F.A.T., Martins, D.M.C., da Graça Batista, M. (2019), Banks' profitability, institutions, and regulation in the context of the financial crisis, *International Journal of Finance & Economics*, Vol. 25, No. 2, pp. 297–320, <https://doi.org/10.1002/ijfe.1753>.
- Vera-Gilces, P., Camino-Mogro, S., Ordeñana-Rodríguez, X., Cornejo-Marcos, G. (2020), A look inside banking profitability: evidence from a dollarised emerging country, *The Quarterly Review of Economics and Finance*, Vol. 75, <https://doi.org/10.1016/j.qref.2019.05.002>.
- Yuan, D., Gazi, M.A.I., Harymawan, I., Dhar, B.K., Hossain, A.I. (2022), Profitability determining factors of banking sector: panel data analysis of commercial banks in South Asian countries, *Frontiers in Psychology*, Vol. 13, 1000412, <https://doi.org/10.3389/fpsyg.2022.1000412>.
- Yuanita, N. (2019), Competition and bank profitability, *Economic Structures*, Vol. 8, p. 31, <https://doi.org/10.1186/s40008-019-0164-0>.
- Yüksel, S., Mukhtarov, S., Mammadov, E., Özsarı, M. (2018), Determinants of profitability in the banking sector: an analysis of post-soviet countries, *Economies*, Vol. 6, No. 41, pp. 2–15, <https://doi.org/10.3390/economies6030041>.