Kristóf, T. & Virág, M. (2022). What drives financial competitiveness of industrial sectors in Visegrad Four countries? Evidence by use of machine learning techniques. *Journal of Competitiveness,* 14(4), 117–136. https://doi.org/10.7441/joc.2022.04.07

What drives financial competitiveness of industrial sectors in Visegrad Four countries? Evidence by use of machine learning techniques

▪ *Tamás Kristóf, Miklós Virág*

Abstract

This article presents machine learning (ML)-based empirical research with a specific focus on the financial competitiveness of different industrial sectors in Visegrad Four (V4) countries. Financial competitiveness is measured by the two most widely applied profitability ratios: return on assets (ROA) and return on equity (ROE). Several sectoral average financial ratios are considered as input variables from the 4 countries and 27 sectors, with data collected between 2016-2020 in a cross-sectional approach. Explorative data analysis reveals that the three strongest clustering features of V4 sector-level financial data are found in country classification, total assets per employee, and gross margin ratios. Hypothesis examination has justified a view that drivers of financial competitiveness are not necessarily identical to factors explaining variance between sectoral average financial ratios. Six methods have been applied to develop predictive models for ROA and ROE. Results demonstrate that the traditional generalized linear model (GENLIN) delivers insufficient predictive power despite fulfilment of each statistical assumption. The k-nearest neighbor (KNN) and random forest (RF) methods are demonstrated to be the best ML techniques to predict the sectoral financial competitiveness of V4 companies. Beyond country classification, the best predictors of ROA and ROE at the V4 sectoral level are found in income margin, turnover, and leverage ratios as compressed components by use of principal component analysis (PCA). The article also provides added value to literature on sectoral and financial competitiveness research, analysis of financial features of V4 companies, and the efficient application of ML methods.

Keywords: corporate competitiveness, sectoral performance, financial ratios, predictive modeling, Visegrad Four
JEL Classification: C45, C58, G30, G32, L25

Received: September, 2022
1st Revision: October, 2022
Accepted: November, 2022

1. INTRODUCTION

Financial competitiveness and financial performance are fundamental targets of corporate financial research projects. As measured by profitability, capital adequacy, turnover, liquidity, solvency, operational efficiency and leverage, financial competitiveness also represents the attainment of corporate financial performance. In line with findings from academic literature, financial competitiveness can also be adequately proxied by profitability measures acting as indicators of financial performance (Alarussi & Ghao, 2021; Khazaei, 2021). Accordingly, financial competitiveness can be effectively measured by return on assets (ROA) and return on equity (ROE), as demonstrated by various previous empirical studies (i.e., Ali et al., 2020; Sachin & Rajesh, 2022).

V4 denotes a group of four countries in the Central-Eastern Europe region sharing not only a similar history but also a similar economic environment whereby ideal targets for investigating financial competitiveness in international empirical research for different purposes can be accomplished. The aim of this article is, therefore, to empirically examine the relationship between V4 sector-level financial ratios and financial competitiveness. In spite of some publications having proposed the application of industrial mean financial ratios for different modeling purposes (Mioduchowska-Jaroszewicz, 2019), no empirical study can be located in the literature which might consider sectoral average financial ratios in V4 countries in order to meet similar research objectives. This aspect has, therefore, been identified as a research gap to be resolved within the framework of specific V4-level empirical research.

The article applies a diverse range of research methods in order to meet the research aim. Explorative data analysis is performed by use of three alternative multivariate clustering methods to identify the strongest clustering features of V4 sector-level financial data. Principal component analysis (PCA) is then applied to perform data reduction in the context of sectoral average financial ratios. By using the reduced dataset, six methods are applied to predict ROA and ROE, respectively. Results demonstrate that the ROA of V4 sectors can be predicted more reliably than ROE regardless of the applied method. It emerged that traditional parametric generalized linear models (GENLIN) could not provide high prediction performance levels, so state-of-the-art ML techniques are needed to generate better predictions in line with findings of a similar scope of empirical studies in the literature (i.e., Elamir, 2021; Green & Zhao, 2022). The k-nearest neighbor (KNN) and random forest (RF) methods are demonstrated to be the best ML techniques to predict the sector-level financial competitiveness of V4 companies.

This article can be positioned in academic literature on competitiveness as a multinational empirical study laying emphasis on studying significant factors regarding competitiveness. It focuses on researching competitiveness of sectors and companies by offering innovative approaches and research methods. Unique comparative empirical research in the field of corporate finance is presented by the findings of this article. Added value to the overall literature corpus is presented in the deepening of know-how on sectoral and financial competitiveness research, exploration of recent financial features of V4 companies, and the results of competing various techniques to perform reliable predictions of financial competitiveness.

The article is structured by the initial presentation of a theoretical background section providing a literature review on financial competitiveness with a focus on interrelationships between financial ratios and state-of-the-art challenges of researching financial competitiveness prediction. The following research objective, methodology and data section formulates research questions, introduces the applied research methods, and analyzes details of data collection. The results and discussion section evaluates the results of data exploration, formulates hypotheses, develops predictive models, and evaluates predictive power. The conclusion section summarizes empirical results, evaluates the added value and limitations of the research exercise, and sets future research objectives.

2. THEORETICAL BACKGROUND

Financial competitiveness research is strongly related to corporate performance and financial ratio analysis in the form of formulating prediction models (Kiseláková et al., 2018). Moreover, financial ratios have traditionally been used as indicators of corporate performance (Kliestik et al., 2020). Valaskova et al. (2021b) further suggest that mutual dependence exists between corporate financial health and earnings management. ROA and ROE are also widely applied as performance indicators as a means of researching financial competitiveness from all over the world (Lassala et al., 2017; Alshehhi et al., 2018; Ali et al., 2020; Keskin et al., 2020; Liu et al., 2021; Okafor et al., 2021; Zhang et al., 2021; Sachin & Rajesh, 2022).

A substantial number of publications have already examined factors influencing ROA and ROE with a broad focus on diverse financial ratios. In terms of financial competitiveness analysis, ROE can be broken down into three components: operating efficiency, assets turnover, and leverage (Jałowiecki, 2018). According to Larasati and Purwanto (2022), the debt-to-equity ratio is the factor with the most significant influence on profitability. However, leverage has a substantially negative effect on both ROA and ROE (Lenka, 2017; Daryanto et al., 2018; Nanda & Panda, 2018). Profitability can also be a strong determining factor in corporate capital structure (Rahayu et al., 2020). Gross and net income margins are important indicators of corporate financial competitiveness (Manogna & Mishra, 2022; Nariswari & Nugraha, 2020). Better liquidity levels might enhance financial competitiveness (Nanda & Panda, 2018).

Akgün and Karatas (2021) detected a negative relationship between working capital and profitability. While inventory turnover management can adversely affect profitability (Garba et al., 2020), some empirical studies have indicated that inventory turnover compared to total assets turnover has insignificantly influenced ROA (Larasati & Purwanto, 2022). Results of other studies have concluded that working capital financing displayed an inverted U-shape relationship with corporate profitability (Setianto et al., 2022). The latter variable demonstrates that financial competitiveness can have both non-linear and non-monotonic relationships with specific financial characteristics. Beyond financial ratios, market indicators and macroeconomic factors can be effectively applied to predict financial competitiveness (Vieira et al., 2019).

Valaskova et al. (2021a) found that the economic sector is one of the most important determinant factors of earnings management in V4 countries. Previous empirical results have demonstrated that the statistical analysis of corporate financial competitiveness can be improved by considering sectoral mean financial ratios (Mioduchowska-Jaroszewicz, 2019). In addition, multivariate clustering can yield flexible grouping of companies by using financial ratios within the framework of explorative data analysis (Ding et al., 2019; Herman et al., 2022).

The increasing availability of online financial databases has given impetus to accomplish financial competitiveness research at the international level. Nowadays, a wide range of historical financial data of companies can be retrieved globally. In parallel with the opportunity to consider big data, the application of ML methods in financial competitiveness analysis has increased in recent years (Shirota & Morita, 2020; Elamir, 2021; Popa et al., 2021; Green & Zhao, 2022). Publications have drawn attention to the superior predictive power of ML techniques over conventional parametric methods to explain and estimate financial performance (Sezer et al., 2020; Milana & Ashta, 2021). Despite the superior predictive performance of ML models, they often face a 'black box' problem (Zednik, 2021), for which a suitable good practice solution is to quantify the impact of variables via means of sensitivity analysis (Manogna & Mishra, 2022).

According to our experience, financial competitiveness research results can be widely read in academic literature with a focus on analyzing corporate data within one or few selected sectors, within one or few selected countries, or within one or few selected regions. However, the usage of sector-level financial data as observations is very limited in researching competitiveness. Since V4 countries can be regarded as well-developed to provide reliable corporate financial data to researchable online international databases, they can be positioned as an ideal research object for this article. To our knowledge, no empirical study has been published to this point to explicitly consider sectoral average financial ratios in order to perform financial competitiveness research for V4 countries and sectors within them. Having identified this research gap in this article, we perform novel empirical research by using contemporary ML methods as a means of enriching know-how of comparative empirical research in this field.

3. RESEARCH OBJECTIVE, METHODOLOGY AND DATA

A singular research objective has been set in line with previous findings and arguments to empirically examine the relationship between V4 sector-level financial ratios and financial competitiveness. The process of our empirical research approach is outlined in Figure 1.

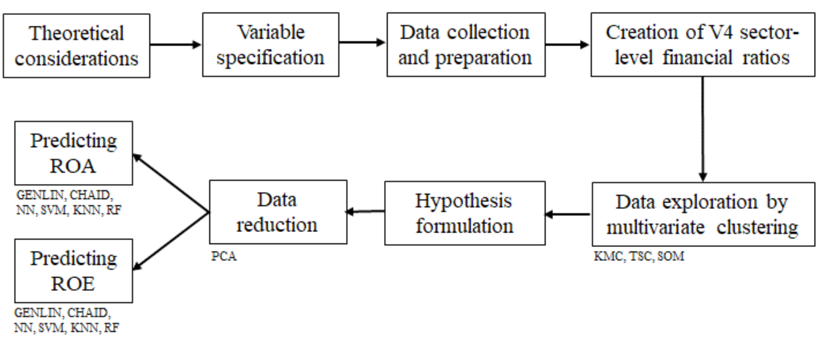


Fig. 1 – The process of empirical research. Source: own work

3.1 Data collection

In order to conduct empirical research, data was collected from Moody's Analytics Orbis Europe database, which is a reliable and widely used source of corporate data. Data retrieval from Orbis Europe was performed in June 2022 with active corporate entities considered as being located in Poland, Czechia, Slovakia, and Hungary, constituting the four V4 countries for this study. A minimum 1 M EUR turnover level for a V4 company was set as a minimum requirement to be considered in our database. To ensure the commensurability of financial data, it was decided to exclude 'Banking, Insurance and Financial Services' and 'Public Administration, Education, Health Social Services' sectors from further analysis. Table 1 summarizes the composition of observed financial data by country throughout the course of the analyzed period.

Tab. 1 – Distribution of companies by country. Source: Orbis Europe

Country	2016	2017	2018	2019	2020
CZ	26,709	27,316	27,649	27,740	21,500
HU	21,239	25,250	27,769	29,990	28,323
PL	46,228	53,359	56,894	60,085	55,824
SK	14,930	16,660	16,815	17,510	16,757
Total	109,106	122,585	129,127	135,325	122,404

The sectoral breakdown of analyzed companies is summarized in Table 2 as per Bureau Van Dijk (BVD) sectors, which is a widely applied sectoral classification source used in financial analysis. Most observations in each year of the surveyed period are classified in the Wholesale sector.

Tab. 2 – Breakdown of the database by sector. Source: Orbis Europe

Sector	2016	2017	2018	2019	2020
Agriculture, Horticulture & Livestock	4,540	5,130	5,140	5,175	4,919
Biotechnology and Life Sciences	459	501	520	545	549
Business Services	11,463	13,250	14,544	15,690	13,989
Chemicals, Petroleum, Rubber & Plastic	3,335	3,582	3,707	3,734	3,606
Communications	492	556	580	616	597
Computer Hardware	61	69	73	85	77
Computer Software	2,145	2,599	2,922	3,229	3,166
Construction	9,698	11,724	13,533	15,003	13,481
Food & Tobacco Manufacturing	3,649	3,882	3,924	4,016	3,787
Industrial, Electric & Electronic Machinery	4,499	4,865	5,097	5,245	4,804
Information Services	68	76	85	86	57
Leather, Stone, Clay & Glass products	1,170	1,288	1,345	1,373	1,262
Media & Broadcasting	527	595	652	722	617
Metals & Metal Products	5,651	6,205	6,489	6,562	5,885
Mining & Extraction	438	486	521	527	481
Miscellaneous Manufacturing	319	346	350	377	333
Printing & Publishing	888	941	965	997	850
Property Services	5,205	6,002	6,412	6,957	6,210

Retail	10,509	11,756	11,996	12,362	11,510
Textiles & Clothing Manufacturing	1,056	1,123	1,137	1,152	1,038
Transport Manufacturing	1,018	1,097	1,159	1,203	1,132
Transport, Freight & Storage	6,820	7,613	8,097	8,420	7,641
Travel, Personal & Leisure	4,289	5,130	5,400	5,828	3,543
Utilities	2,051	2,188	2,234	2,312	2,234
Waste Management & Treatment	1,021	1,144	1,190	1,292	1,243
Wholesale	25,307	27,730	28,270	28,983	26,764
Wood, Furniture & Paper Manufacturing	2,428	2,707	2,785	2,834	2,629
Total	109,106	122,585	129,127	135,325	122,404

In order to derive comparable financial ratios for observed V4 companies, each financial data item was expressed in Euros for the surveyed V4 countries, of which three do not use the Euro as a national currency. In addition, records with missing data were excluded from financial ratio mean calculations. The following financial ratios were selected to analyze financial competitiveness beyond country and sector classification.

Tab. 3 – Applied financial ratios. Source: own work

Profitability ratios	Operational ratios
• ROA using Net income • ROE using Net income • Profit margin • Gross margin • EBITDA margin • EBIT margin • Cash flow / Operating revenue	• Net assets turnover • Interest cover • Stock turnover • Collection period days • Credit period days
Structure ratios	**Per employee ratios**
• Current ratio • Liquidity ratio • Shareholders liquidity ratio • Solvency ratio Asset based • Solvency ratio Liability based • Gearing	• Profit per employee (EUR, 000) • Operating revenue per employee (EUR, 000) • Costs of employees / Operating revenue • Average cost per employee (EUR, 000) • Shareholders' funds per employee (EUR, 000) • Working capital per employee (EUR, 000) • Total assets per employee (EUR, 000)

In line with the main objective of the article, research was pursued by considering the sectoral average financial ratios of the 4 countries and 27 sectors for the period of 2016-2020 (5 years) in a cross-sectional approach. Overall, 540 observations for further analysis were derived from this process.

3.2 Methodology

Explorative data analysis was initially performed in the database in order to explore features of sector-level financial behavior within V4 companies by applying multivariate clustering techniques. The fundamental question of multivariate data exploration is to determine which combination of clustering of variables per sector can best discriminate observations from each other. Since no single target variable can be considered in clustering, the decisive factor in the developed models was drawn from the clustering power of sector classification without entering sector classification itself as an input variable in order to formulate clusters.

Three widely applied clustering techniques were attempted: k-means clustering (KMC), two-step clustering (TSC), and self-organizing maps (SOM). When selecting the best model, the sector variable was evaluated in the context of each other variable as a means of locating the most favorable grouping.

The KMC algorithm is probably the most known and used clustering method. It partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centroid) by minimizing the extent of within-cluster variance (Chao et al., 2022). The TSC initially compresses input data into a manageable set of sub-clusters. It then uses hierarchical clustering to gradually merge sub-clusters into larger clusters, whereby it is not a prerequisite to pre-select the number of clusters. It can also handle categorical variables (Popa et al., 2022) which can be used to test several cluster solutions and select the most suitable. SOM based methodology belongs to the family of unsupervised neural networks (Wehrens & Kruisselbrink, 2018). The basic units of SOM are formed of neurons organized into an input and a two-dimensional output layer. Input neurons are connected to output neurons through associated weights (Olszewski, 2021).

After evaluating the results of clustering, PCA was applied to perform dimensional reduction and handle multi-collinearity in the context of sectoral average financial ratios. PCA identifies a reduced set of variables thus representing the original data in a lower-dimensional subspace with limited information loss (Dugger et al., 2022).

Predictive models were then developed to explain financial competitiveness. Since the range of input variables incorporates the country as a categorical variable, traditional linear regression modeling was not applied. In addition, traditional classification methods are inadequate to develop models to comply with the idiosyncrasies of the database since ROA and ROE formed continuous target variables. The following methods were utilized:

- Generalized linear model (GENLIN)
- Chi-squared automatic interaction detector (CHAID)
- Neural networks (NN)
- Support vector machine (SVM)

- K nearest neighbor (KNN)
- Random forest (RF)

The GENLIN model comprises a flexible generalization of traditional linear regression in such a way that the dependent variable is linearly related to features through a specified link function. It enables the magnitude of the variance of each measurement so that it can be a function of its predicted value (Zuniga et al., 2021) and also allows for the dependent variable to possess a non-normal distribution.

CHAID is a classification and regression method for building decision trees by using chi-squared statistics to identify optimal splits. It principally examines relationships between input and target variables and tests their significance with the chi-squared independence test (Durica et al., 2019). Stopping rules control when to stop creating further splits in the decision tree.

The use of NNs is intended to model the way in which the human nervous system operates. It incorporates interconnected nodes which can recognize hidden relationships and thus make predictions through learning. The structure of a NN is organized into an input layer, one or more hidden layers, and an output layer. Examples are presented with predicted outcomes constantly compared to known outputs, and weights are adjusted to provide the best estimation (Wu et al., 2022). Overtraining is avoided by separating the database into a training and testing subset, and the structure of NN is optimized to obtain the best result in the testing subset.

The SVM tool forms a robust regression method that maximizes the predictive power of a model without overfitting the training data. It also maps data onto a high-dimensional space feature in such a way that data items can be categorized, even when they are not otherwise linearly separable. Data are further transformed by a hyperplane separator with the use of kernel functions (Cervantes et al., 2020).

The KNN algorithm forms a method for classifying or predicting instances based on their similarity to other instances. The distance is a measure of dissimilarity, whereby close cases are regarded as neighbors (Gallego et al., 2022). The number of nearest neighbors can be specified by the selection of K. When using KNN to predict a continuous target variable, the average or median target value of the nearest neighbors is used to obtain the predicted value.

The RF method is used to develop an ensemble model consisting of multiple decision trees. At the individual tree-level, it employs classification and regression tree (CART) methodology, which uses recursive partitioning to split records with similar output field values into segments. The termination criteria for decision tree building are also identical to that of CHAID. In ensemble learning, RF applies a bootstrap aggregation (bagging) tool to generate samples different from each other with replacement permitted (Athey et al., 2019). These samples are given to multiple learners, and results from each are combined by equal voting to get the final prediction.

The added value of variables in each model was evaluated by use of the predictor importance indicator. It can be determined by computing the reduction in the variance of the target variable attributable to each input variable via sensitivity analysis expressed by means of normalized sensitivity (Saltelli et al., 2019). This indicator can be used for all the previously discussed methods. The predictive power of the model in terms of predicted and actual values was evaluated by use of the well-known Pearson correlation and standard error.

4. RESULTS AND DISCUSSION

4.1 Results of data exploration

The use of the KMC method created five clusters. The sector variable attained an importance value of 0.194, thereby demonstrating weak clustering power. Application of SOM resulted in eight clusters, of which four displayed a fragmented feature unfavorable for modeling purposes. The sector variable possessed a very low (0.059) impact value in terms of SOM-based clustering.

In contrast, the use of TSC resulted in the creation of five clusters whereby the impact of sector variables became strong with an importance value of 0.620. As the aim of explorative data analysis is to detect variance between sectoral variables for further analysis, the results of the TSC were considered. It is further noted that TSC can adequately handle features of categorical variables, which is crucial when applying country and sector classification values in the current research project.

In overall terms, it can be argued that the formulated five clusters comply with the idiosyncrasies and size of the database. The smallest cluster contains 53 observations (9.8%), whereas the largest cluster contains 208 observations (38.5%). Hence, the largest cluster is 3.925 times larger than its smallest counterpart.

Results demonstrate that the three strongest clustering features are located in country classification, total assets per employee, and gross margin values. It is interesting to observe the dominance of size-efficiency and profit margin indicators within variables to formulate clusters. Surprisingly, key ratios in financial competitiveness analysis, such as ROE, ROA, liquidity ratios, current ratios, net assets turnover, interest cover, and collection periods represented low clustering power in terms of differentiating sectoral level observations. Figure 2 evaluates the rankings of variable importance in the formulated clusters.

From analyzing the characteristics of the five formulated clusters, the following conclusions can be drawn:

- The largest cluster (3) incorporates a relatively high share of Polish sectoral ratios. Profit margin ratios are generally low, and the cluster possesses the second-worst total assets volume per employee.
- The second cluster (2) is dominated by Hungarian sectoral ratios. Gross margin is the highest ratio in this cluster; however, total assets volume per employee is mid-level in comparison with cluster (3).
- The third cluster (5) mostly consists of Slovak sectoral observations. The gross profit margin is the second worst of all observations, and total assets per employee are lower than average.
- The fourth cluster (4) is highly diversified for each country; however, most observations emanate from Czechia. Total assets per employee are outstandingly high, and the gross profit margin value significantly exceeds the average.
- The fifth cluster, as the smallest (1), is substantially diversified for each country. However, the share of Slovak observations is higher than average. Total assets per employee are the lowest in this cluster, but the gross profit margin value is average.

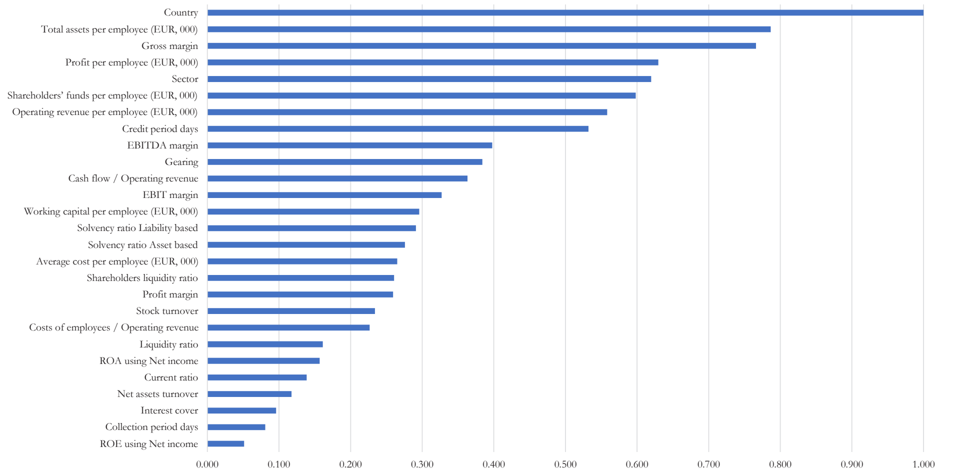


Fig. 2 – The importance of variables in creating clusters. Source: own research

The following hypotheses have been formulated from the preceding results of explorative data analysis in order to support the foundations of empirical research:

- Hypothesis 1: The country effect is significant in order to explain sectoral competitiveness of V4 companies.
- Hypothesis 2: Efficiency indicators per employee of V4 companies are significant in order to explain sectoral competitiveness.
- Hypothesis 3: Income margin indicators of V4 companies are significant in order to explain sectoral competitiveness.

ROA and ROE have been specified as target variables representing the financial competitiveness of V4 sectors. The mean of ROA in the database is 8.107 (standard deviation: 4.051), and that of ROE is 18.624 (standard deviation: 11.580), thus exhibiting substantially higher variance of ROE among V4 sectoral data.

4.2 Results of data reduction

By use of PCA to achieve dimensional reduction of input variables, 22 financial ratios were compressed into 7 components. Each financial ratio group was reduced into separate components as outlined in Table 4 with requirements for creating components as follows:

- Significant and strong correlation exists between variables.
- Variables have similar business meanings in terms of definition.
- Eigenvalues should exceed 1.000.
- The Kaiser-Meier-Olkin (KMO) sample adequacy ratio should be at least 50%.
- The component should be significant by use of the Bartlett chi-squared test.

Tab. 4 – The results of PCA. Source: own research

Component	Compressed variables	KMO	Bartlett chi-squared	Eigen-value	Explained variance %
PCA_1_INCOME_MARGIN	Profit margin, Gross margin, EBITDA margin, EBIT margin, Cash flow Operating revenue	0.579	2894.083 (p=0.000)	3.586	71.724
PCA_2_TURNOVER	Net assets turnover, Stock turnover	0.500	18.833 (p=0.000)	1.186	59.278
PCA_3_LIQUIDITY	Interest cover, Current ratio, Liquidity ratio	0.517	799.858 (p=0.000)	1.994	66.477
PCA_4_CRED_COLL_PERIOD	Credit period days, Collection period days	0.500	219.650 (p=0.000)	1.579	78.959
PCA_5_LEVERAGE	Solvency ratio Asset based, Solvency ratio Liability based, Gearing	0.667	446.751 (p=0.000)	2.054	68.480
PCA_6_PER_EMPLOYEE_INDICATORS	Profit per employee (EUR,000), Operating revenue per employee (EUR,000), Shareholders funds per employee (EUR,000), Working capital per employee (EUR,000), Total assets per employee (EUR,000)	0.792	2380.209 (p=0.000)	3.618	72.351
PCA_7_COST_PER_EMPLOYEE	Cost of employees Operating revenue, Average cost of employee (EUR,000)	0.500	60.829 (p=0.000)	1.327	66.356

It was impossible to incorporate the shareholders liquidity ratio into any single component for statistical and professional reasons. Accordingly, it has remained in its original format as a separate input variable for further analysis.

4.3 Results of predictive modeling

The country classification, the seven compressed components, and the shareholders liquidity ratio represented the range of input variables in each model. In the following narrative, details of modeling to set hyper-parameters and to optimize parameters when developing models for predicting ROA and ROE are summarized. Furthermore, the importance of predictors in each model is evaluated to test the validity of the formulated hypotheses.

The identity link function was selected in the use of the GENLIN models. The scale parameter method represented the maximum likelihood estimation, where the covariance-matrix formed the model-based estimator. The model type only considered main effects without two-way interactions.

Table 5 summarizes the parameters and test results of the GENLIN-based ROA model. Country classification is a significant model variable, together with the leverage component, the turnover component, the shareholders liquidity ratio, the income margin component, and the cost per employee component, all ranked by the strength of the Wald test. Other variables were insignificant. The impact of belonging to a Slovak industrial sector has become neutral and insignificant in the model. The Pearson chi-squared model goodness of fit value is 4885.507 (p=0.000), and the Akaike information criterion is 2741.782, while the Omnibus likelihood chi-squared ratio of the model is 320.586 (p=0.000).

Tab. 5 – Parameters and test results of the GENLIN ROA model. Source: own research

Variable	Beta	Standard error	95% Wald confidence interval		Hypothesis test	
			Lower	Upper	Wald test	p-value
(Intercept)	4.679	0.523	3.654	5.704	80.077	0.000
[Country=CZ]	0.729	0.483	-0.218	1.675	2.278	0.131
[Country=HU]	3.092	0.557	2.000	4.184	30.785	0.000
[Country=PL]	3.031	0.496	2.059	4.003	37.342	0.000
[Country=SK]	0.000	.	.	.	.	.
Shareholders_liquid-ity_ratio	0.037	0.005	0.027	0.047	54.748	0.000
PCA_1_INCOME_MARGIN	0.835	0.139	0.563	1.108	36.142	0.000
PCA_2_TURN-OVER	1.338	0.172	1.000	1.675	60.279	0.000
PCA_5_LEVER-AGE	1.629	0.208	1.223	2.036	61.664	0.000
PCA_7_COST_PER_EMPLOYEE	-0.542	0.152	-0.839	-0.244	12.766	0.000
(Scale)	9.047	0.551	8.030	10.193		

The Pearson correlation-based predictive power of the model is 0.669 (p=0.000), whereas the standard error between predicted and actual values is 3.013, as noted in comparison to other models in Table 8. It is important to note that normalized predictor importance statistics located only 3 important predictors in the model as compared to other models listed in Table 7.

Table 6 evaluates the features of the GENLIN-based ROE model. Country classification is again significant, together with the turnover component, the shareholders liquidity ratio, the income margin component, the cost per employee component, the leverage component, and the credit collections period component, respectively. As before, the impact of belonging to a Slovak industrial sector is not significant in the model. The Pearson chi-squared model goodness of fit is 49877.070 (p=0.000). The Akaike information criterion is 3998.357, and the Omnibus likelihood ratio chi-squared of the model is 200.273 (p=0.000).

Tab. 6 – Parameters and testing of the GENLIN ROE model. Source: own research

Variable	Beta	Standard error	95% Wald confidence interval		Hypothesis test	
			Lower	Upper	Wald test	p-value
(Intercept)	7.903	1.732	4.508	11.299	20.814	0.000
[Country=CZ]	5.200	1.695	1.877	8.523	9.407	0.002
[Country=HU]	7.543	2.151	3.327	11.758	12.299	0.000
[Country=PL]	10.385	1.586	7.277	13.493	42.896	0.000
[Country=SK]	0.000	.	.	.	.	.
Shareholders_liquidity_ratio	0.106	0.016	0.075	0.138	44.377	0.000
PCA_1_INCOME_MARGIN	1.987	0.447	1.111	2.863	19.766	0.000
PCA_2_TURNOVER	4.996	0.555	3.908	6.084	80.967	0.000
PCA_4_CRED_COLL_PERIOD	1.532	0.692	0.175	2.889	4.897	0.027
PCA_5_LEVERAGE	1.848	0.675	0.526	3.170	7.506	0.006
PCA_7_COST_PER_EMPLOYEE	-2.055	0.488	-3.011	-1.099	17.735	0.000
(Scale)	92.365	5.621	81.979	104.066		

The Pearson correlation-based predictive power of the model is 0.557 (p=0.000), whereas the standard error is 9.628, as noted in comparison to other models in Table 8. It is again interesting to observe that normalized predictor importance statistics located only 3 important predictors in the model as compared to other models listed in Table 7.

It can be concluded that in line with these findings, research has been pursued by the application of ML techniques to develop better predictive models, although GENLIN models meet all statistical assumptions due to their limited predictive performance levels.

Termination rules for CHAID were parameterized to possess at least 5% of records in 'parent branches,' and 4% in 'child branches' of decision trees. The significance level for splitting and merging trees was set at 0.05, and the maximum tree depth was set at 5. The ROA model possessed 3, and the ROE model possessed 4 important predictors, respectively, while the impact of other variables was negligible.

The multilayer perceptron training algorithm was applied to develop the NN models. Due to the limited number of input variables and instances, one hidden layer was adequate for modeling purposes with the sigmoid applied as an activation function. To prevent overtraining and to optimize the parameters, a 30% testing subset was separated from the database. The final network structure was saved where the least error was measured in the testing subset. The ROA model possessed 6, and the ROE model possessed 5 important predictors, respectively, while other features had weak impact levels.

The SVM models were developed by considering the radial basis kernel function (RBF). The regularization parameter (C) was set at 10, the regression precision value (epsilon) at 0.1, the RBF gamma value at 0.1, and the termination criterion specified as 1.0E-3. The ROA model possessed 4, and the ROE model possessed 2 important variables, respectively, with other predictors demonstrating weak added value.

KNN models were developed by testing K values between 3 and 7. The best K value for the ROA model was 5, whereas it was 6 for the ROE model based on the sum of squares error statistic. Distance computation was performed by use of the Euclidean metric with features weighted by importance when calculating distances. The prediction of continuous target variables was performed by averaging the nearest neighbor values, and then a ten-fold cross- validation exercise was performed to back-test results. Both the ROA and the ROE models possessed 8 important predictors.

The RF models considered 100 ensemble decision trees with the maximum depth of each tree limited to 5. The maximum number of bins for each splitting level was set at 10, with the minimum size of a 'child branch' node required to be at least 15. Tree building was terminated in cases where accuracy did not improve. The final prediction was calculated by the mean of predictions generated by the 100 trees. The ROA model possessed 6, and the ROE model possessed 4 important predictors, respectively.

Table 7 summarizes the 3 most important predictors of the 12 models by utilizing the results of normalized predictor importance statistics. It demonstrates that beyond country classification, income margin, turnover, and leverage components are the best predictors of ROA and ROE. Neither can the impact of the shareholders liquidity ratio be neglected. The per-employee indicators component as the object of Hypothesis 2 is only represented in the NN-ROE model.

Tab. 7 – Top 3 input variables in the models by predictor importance. Source: own work

Method	Rank of predictor importance	
	ROA models	ROE models
GENLIN	Leverage component, country classification, income margin component	Turnover component, country classification, income margin component
CHAID	Country classification, income margin component, turnover component	Turnover component, country classification, income margin component
NN	Turnover component, Shareholders liquidity ratio, leverage component	Shareholders liquidity ratio, cost per employee component, per employee indicators component
SVM	Country classification, leverage component, income margin component	Turnover component, country classification, leverage component
KNN	Shareholders liquidity ratio, income margin component, country classification	Leverage component, liquidity component, country classification
RF	Country classification, turnover component, leverage component	Turnover component, country classification, income margin component

The predictive power of the models was evaluated by using the Pearson correlation and standard error values to compare predicted and actual values. Results are summarized in Table 8.

Tab. 8 – Evaluation of predictive performance. Source: own research

Method	ROA models		ROE models	
	Pearson correlation	Standard error	Pearson correlation	Standard error
GENLIN	0.669 (p=0.000)	3.013	0.557 (p=0.000)	9.628
CHAID	0.664 (p=0.000)	3.032	0.605 (p=0.000)	9.229
NN	0.808 (p=0.000)	2.389	0.648 (p=0.000)	8.824
SVM	0.685 (p=0.000)	2.955	0.514 (p=0.000)	9.942
KNN	0.889 (p=0.000)	1.857	0.700 (p=0.000)	8.276
RF	0.816 (p=0.000)	2.346	0.773 (p=0.000)	7.357

It can be concluded that the KNN model substantially outperformed others in predicting ROA. RF and NN also demonstrated favorable predictive power; however, SVM, CHAID, and GENLIN models indicated weaker performance levels. Model performance to predict ROE generally lags behind the predictive power of ROA models for each method. The RF model became the best tool to estimate ROE, followed by the KNN and NN models. CHAID, GENLIN, and SVM models again produced substantially worse results.

The annex presents the distribution of ROA and ROE predictions made by the six models. The underlying reason for the relatively unusual picture of the CHAID model is that static decision trees can produce a limited combination of variable categories resulting in one prediction per combination.

5. CONCLUSION

Empirical results have demonstrated that the financial competitiveness of V4 sectors can be explained and predicted by a wide range of features and methods. Twelve predictive models have been developed by using six methods resulting in quite different model designs revealing diverse drivers of sectoral competitiveness. It can also be concluded that ROA of V4 sectors can be predicted more reliably than ROE regardless of the applied method. ROE has a greater variance than ROA, and the distribution of ROA and ROE predictions made by the six methods is remarkably diverse.

Since GENLIN, as a traditional parametric regression method, could not provide high prediction performance despite the significance of its parameters and the model, various ML techniques were applied to generate a better prediction. It can be convincingly argued that KNN and RF have been demonstrated to be the best methods of predicting the sectoral financial competitiveness of V4 companies.

Empirical results have also revealed that the most important predictors of sectorial competitiveness in V4 are not necessarily the same as those explaining higher variance between observations without considering ROA and ROE as target variables in this study. Normalized predictor importance statistics have demonstrated that country classification, income margin, turnover, and leverage components have emerged as the best predictors of ROA and ROE, followed by the shareholders liquidity ratio. Per-employee indicators did not become important predictors of financial competitiveness. Based on these findings, the results of the hypothesis examination can be evaluated as follows:

- Hypothesis 1 is accepted, as the country effect has proven to be significant and relevant in all models with high predictive power as a means of explaining the sectoral competitiveness of V4 companies.
- Hypothesis 2 is rejected, as per-employee indicators expressed by the relevant created component were not found to be important variables in terms of explaining the sectoral competitiveness of V4 companies.
- Hypothesis 3 is accepted, as income margin indicators expressed by the relevant created component are regarded as significant and/or relevant predictors in most of the models in order to explain the sectoral competitiveness of V4 companies.

The added value of this study to the developing corpus of academic literature in this field lies in the deepening of know-how on sectoral and financial competitiveness research, exploration of recent financial features of V4 companies, and the results of six competing ML methods to generate reliable predictions. A limitation of this study due to the application of sector-level ratios is that the database consists of 540 observations which cannot be considered as a large sample. For future research, it would be necessary to extend surveys to other countries (preferably to EU-27 countries) and to cover longer historical periods.

References

1. Akgün, A. I., & Karatas, M. A. (2021). Investigating the relationship between working capital management and business performance: Evidence from the 2008 financial crisis of EU-28. *International Journal of Managerial Finance,* 17(4), 545–567. https://doi.org/10.1108/IJMF-08-2019-0294
2. Alarussi, A. S., & Gao, X. (2021). Determinants of profitability in Chinese companies. *International Journal of Emerging Markets,* ahead of print. https://doi.org/10.1108/IJOEM-04-2021-0539
3. Ali, S. A., Yassin, M., & AbuRaya, R. (2020). The impact of firm characteristics on corporate financial performance in emerging markets: Evidence from Egypt. *International Journal of Customer Relationship Marketing and Management,* 11(4), 70–89. https://doi.org/10.4018/IJCRMM.2020100105
4. Alshehhi, A., Nobanee, H., & Khare, N. (2018). The impact of sustainability practices on corporate financial performance: Literature trends and future research potential. *Sustainability*, 10(2), 494. https://doi.org/10.3390/su10020494
5. Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *Annals of Statistics,* 47(2), 1148–1178. https://doi.org/10.1214/18-AOS1709
6. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408(September), 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

7. Chao, D., Xinyang, C., Shengije, Y., Ronga, C., & Yifan, Y. (2022). Credit ratings of Chinese households using factor scores and k-means clustering method. *International Review of Economics & Finance,* 78(March), 309–320. https://doi.org/10.1016/j.iref.2021.12.014

8. Daryanto, W., Samidi, S., & Siregar, D. (2018). The impact of financial liquidity and leverage on financial performance: Evidence from property and real estate enterprises in Indonesia. *Management Science Letters*, 8(12), 1345–1352. https://doi.org/10.5267/j.msl.2018.9.005

9. Ding, K., Peng, X., & Wang, Y. (2019). A machine learning-based peer selection method with financial ratios. *Accounting Horizons*, 33(3), 75–87. https://doi.org/10.2308/acch-52454

10. Dugger, Z., Halverson, G., McCrory, B., & Claudio, D. (2022). Principal component analysis in MCDM: An exercise in pilot selection. *Expert Systems with Applications*, 188(February), 115984. https://doi.org/10.1016/j.eswa.2021.115984

11. Durica, M., Frnda, J., & Svabova, L. (2019). Decision tree based model of business failure prediction for Polish companies. *Oeconomia Copernicana,* 10(3), 453–469. https://doi.org/10.24136/oc.2019.022

12. Elamir, E. A. (2021). Boosting algorithms to analyse firm's performance based on return on equity: An explanatory study. *International Journal of Computing and Digital Systems,* 10(1), 1031–1041. http://dx.doi.org/10.12785/ijcds/100193

13. Gallego, A. J., Rico-Juan, J. R., & Valero-Mas, J. J. (2022). Efficient k-nearest neighbor search based on clustering and adaptive k values. *Pattern Recognition,* 122(February), 108356. https://doi.org/10.1016/j.patcog.2021.108356

14. Garba, S., Mourad, B., & Chamo, M. A. (2020). The effect of inventory turnover period on the profitability of listed Nigerian conglomerate companies. *International Journal of Financial Research,* 11(2), 287–292. https://doi.org/10.5430/ijfr.v11n2p287

15. Green, J., & Zhao, W. (2022). Forecasting earnings and returns: A review of recent advancements. *The Journal of Finance and Data Science,* 8(November), 120–137. https://doi.org/10.1016/j.jfds.2022.04.004

16. Herman, E., Zsido, K. E., & Fenyves, V. (2022). Cluster analysis with k-mean versus k-medoid in financial performance evaluation. *Applied Sciences*, 12(16), 7985. https://doi.org/10.3390/app12167985

17. Jałowiecki P. (2018). Dupont analysis of Polish agri-food industry selected sectors: Logistic aspects. *Research in Logistics & Production,* 8(3), 217–227. https://doi.org/10.21008/j.2083-4950.2018.8.3.2

18. Keskin A. İ., Dincer, B., & Dincer, C. (2020). Exploring the impact of sustainability on corporate financial performance using discriminant analysis. *Sustainability*, 12(6), 2346. https://doi.org/10.3390/su12062346

19. Khazaei, M. (2021). Relationship of profitability of world's top companies with entrepreneurship, competitiveness, and business environment indicators. *Applied Economics,* 53(23), 2584–2597. https://doi.org/10.1080/00036846.2020.1859455

20. Kiseláková, D., Šofranková, B., Čabinová, V., & Šoltésová, J. (2018). Analysis of enterprise performance and competitiveness to streamline managerial decisions. *Polish Journal of Management Studies,* 17(2), 101–111. https://doi.org/10.17512/pjms.2018.17.2.09

21. Kliestik, T., Valaskova, K., Lazaroiu, G., Kovacova, M., & Vrbka, J. (2020). Remaining financially healthy and competitive: The role of financial predictors. *Journal of Competitiveness,* 12(1), 74–92. https://doi.org/10.7441/joc.2020.01.05

22. Larasati, C. I., & Purwanto, P. (2022). How financial ratios and firm size affect profitability: Evidence from food and beverages industry in Indonesia. *The Winners,* 23(1), 43–50. https://doi.org/10.21512/tw.v23i1.7099

23. Lassala C., Apetrei, A., & Sapena, J. (2017). Sustainability matter and financial performance of companies. *Sustainability*, 9(9), 1498. https://doi.org/10.3390/su9091498

24. Lenka, S. (2017). The relationship between company returns and leverage depending on the business sector: Empirical evidence from the Czech Republic. *Journal of Competitiveness*, 9(3), 98–110. https://doi.org/10.7441/joc.2017.03.07

25. Liu, Y., Saleem, S., Shabbir, R., Shabbir, M. S., Irshad, A., & Khan, S. (2021). The relationship between corporate social responsibility and financial performance: A moderate role of fintech technology. *Environmental Science and Pollution Research*, 28(16), 20174–20187. https://doi.org/10.1007/s11356-020-11822-9

26. Manogna, R. L., & Mishra, A. K. (2022). Measuring financial performance of Indian manufacturing firms: Application of decision tree algorithms. *Measuring Business Excellence,* 26(3), 288–307. https://doi.org/10.1108/MBE-05-2020-0073

27. Milana, C., & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: A survey of the literature. *Briefings in Entrepreneurial Finance,* 30(3), 189–209. https://doi.org/10.1002/jsc.2403

28. Mioduchowska-Jaroszewicz, E. (2019). Usefulness of sectoral means in financial analyses of enterprises. *Financial Internet Quarterly*, 15(4), 55–72. https://doi.org/10.2478/fiqf-2019-0028

29. Nanda, S., & Panda, A. K. (2018). The determinants of corporate profitability: an investigation of Indian manufacturing firms. *International Journal of Emerging Markets*, 13(1), 66–86. https://doi.org/10.1108/IJoEM-01-2017-0013

30. Nariswari, T. N., & Nugraha, N. M. (2020). Profit growth: Impact of net profit margin, gross profit margin and total assets turnover. *International Journal of Finance & Banking Studies,* 9(4), 87–96. https://doi.org/10.20525/ijfbs.v9i4.937

31. Okafor, A., Adeleye, B. N., & Adusei, M. (2021). Corporate social responsibility and financial performance: Evidence from U.S tech firms. *Journal of Cleaner Production*, 292(April), 126078. https://doi.org/10.1016/j.jclepro.2021.126078

32. Olszewski, D. (2021). A data-scattering-preserving adaptive self-organizing map. *Engineering Applications of Artificial Intelligence,* 105(October), 104420. https://doi.org/10.1016/j.engappai.2021.104420

33. Popa, D. C. S., Popa, D. N., Bogdan, V., & Simut, R. (2021). Composite financial performance index prediction: A neural networks approach. *Journal of Business Economics and Management,* 22(2), 277–296. https://doi.org/10.3846/jbem.2021.14000

34. Popa, D. N., Bogdan, V., Sabau Popa, C. D., Belenesi, M., & Badulescu, A. (2022). Performance mapping in two-step cluster analysis through ESEG disclosures and EPS. *Kybernetes*, 51(3), 98–118. https://doi.org/10.1108/K-08-2021-0672

35. Rahayu, M., Suhadak, S., & Saifi, M. (2020). The reciprocal relationship between profitability and capital structure and its impacts on the corporate values of manufacturing companies in Indonesia. *International Journal of Productivity and Performance Management,* 69(2), 236–251. https://doi.org/10.1108/IJPPM-05-2018-0196

36. Sachin, N., & Rajesh, R. (2022). An empirical study of supply chain sustainability with financial performances of Indian firms. *Environment, Development and Sustainability,* 24(5), 6577–6601. https://doi.org/10.1007/s10668-021-01717-1

37. Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software,* 114(April), 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

38. Setianto, R. H., Sipayung, R. S., & Azman-Saini, W. (2022). Working capital financing and corporate profitability in the ASEAN region: The role of financial development. *Entrepreneurial Business and Economics Review,* 10(1), 51–64. https://doi.org/10.15678/EBER.2022.100104

39. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review 2005–2019. *Applied Soft Computing,* 90(May), 106181. https://doi.org/10.1016/j.asoc.2020.106181

40. Shirota, Y., & Morita, M. (2020). Performance analysis of Japanese manufacturing industry before and after Lehman shock. *International Journal of Trade, Economics and Finance,* 11(5), 87–91. https://doi.org/10.18178/ijtef.2020.11.5.672

41. Valaskova, K., Adamko, P., Michalikova, K. F., & Macek, J. (2021a). Quo Vadis, earnings management? Analysis of manipulation determinants in Central European environment. *Oeconomia Copernicana,* 12(3), 631–669. https://doi.org/10.24136/oc.2021.021

42. Valaskova, K., Androniceanu, A-M., Zvarikova, K., & Olah, J. (2021b). Bonds between earnings management and corporate financial stability in the context of the competitive ability of enterprises. *Journal of Competitiveness*, 13(4), 167–184. https://doi.org/10.7441/joc.2021.04.10

43. Vieira, E. S., Neves, M. E., & Dias, A. G. (2019). Determinants of Portuguese firms' financial performance: Panel data evidence. *International Journal of Productivity and Performance Management,* 68(7), 1323–1342. https://doi.org/10.1108/IJPPM-06-2018-0210

44. Wehrens, R., & Kruisselbrink, J. (2018). Flexible self-organizing maps in Kohonen 3.0. *Journal of Statistical Software,* 87(7), 1–18. https://doi.org/10.18637/jss.v087.i07

45. Wu, D., Ma, X., & Olson, D. L. (2022). Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems,* 159(August), 113814. https://doi.org/10.1016/j.dss.2022.113814

46. Zhang, D., Xie, J., & Sikveland, M. (2021). Tourism seasonality and hotel firms' financial performance: Evidence from Norway. *Current Issues in Tourism,* 24(21), 3021–3039. https://doi.org/10.1080/13683500.2020.1857714

47. Zuniga, F., Kozubowski, T. J., & Panorska, A. K. (2021). A generalized linear model for multivariate events. *Journal of Computational and Applied Mathematics*, 398(December), 113655. https://doi.org/10.1016/j.cam.2021.113655

48. Zednik, C. (2021). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology,* 34(June), 265–288. https://doi.org/10.1007/s13347-019-00382-7

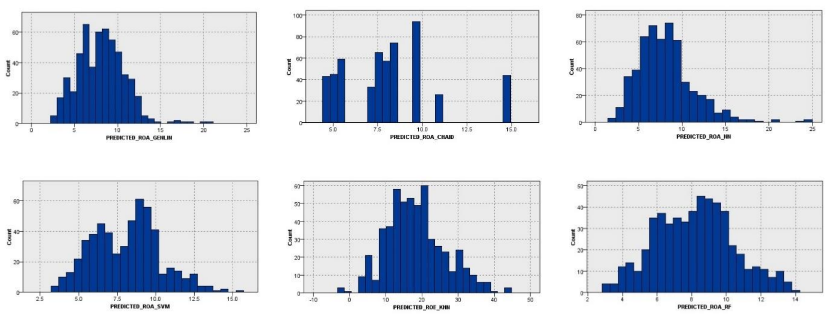
Contact information

Tamás Kristóf, Dr. Habil.
Corvinus University of Budapest
Institute of Entrepreneurship and Innovation
Hungary
E-mail: tamas.kristof@uni-corvinus.hu
ORCID: 0000-0003-2805-4900

prof. Miklós Virág, Dr. Habil.
Corvinus University of Budapest
Institute of Entrepreneurship and Innovation
Hungary
E-mail: miklos.virag@uni-corvinus.hu
ORCID: 0000-0002-6002-7729

ANNEX

The distribution of ROA predictions made by the six models:



The distribution of ROE predictions made by the six models: