

Using Kohonen networks in the analysis of transport companies in the Czech Republic

Tomáš Krulický^{1,*}

¹University of Žilina, Faculty of Operation and Economics of Transport and Communications, Univerzitná 1, 01026 Žilina, Slovak Republic

Abstract. The transport sector has a significant impact on the performance of the Czech economy. Transport companies, of course, have their own specificities, whether they deal with ecology or the financial and economic situation. It is precisely the economic position of a transport company that needs to be analysed in order to identify the need for change, to predict the further development of such company. For analysis, a variety of methods is used, of which artificial neural networks are a very interesting and effective tool. The aim of this paper is to make a cluster analysis of transport companies operating in the Czech Republic based on this tool. The data of the financial statements of transport companies in the Czech Republic in 2016 are taken into account. Only some items from the financial statements are selected for analysis. The file is then subjected to a cluster analysis, specifically using the Kohonen networks – Statistica software. In accordance with the methodology of the contribution, the data is divided into three sets - training, testing and validation. Companies were divided into clusters in the 10x10 Kohonen Map. Some clusters are significant in terms of number of companies. These clusters are further analysed. Specific conclusions are made: A larger company generates, on average, a higher operating profit, larger companies achieve higher ROE and, in the case of a larger company, the financial leverage acts more positively.

Key words: Kohonen networks, transport companies, cluster analysis, neural networks, financial situation

1 Introduction

The importance of company analysis lies in defining needs and solutions that bring value to stakeholders. Company analysis is a disciplined approach to implementing and managing changes in organizations, whether they are beneficial businesses, governments or non-profit organizations [1]. Company analysis is used to identify and formulate the need for changes in organization functioning and the simplification of these changes. By making effective use of company analyses, it is possible to ensure that a company improves the way it is doing its business [2].

Company analysis can be measured in different ways, but Kohonen networks are considered to be a very attractive model. Kohonen networks are types of neural networks

* Corresponding author: krulicky@mail.vstecb.cz

that can be used to group sets of data into different groups [3]. The data is grouped so that the records within the group tend to be the same and the records in the different groups are different. Many experimental results show that Kohonen's networks are very effective for assessing companies [4].

A Kohonen network consists of an input layer that is completely interconnected with the output layer and learns by self-organization, without a teacher. This network has a very wide range of uses, as it is an alternative network that is applicable to most neural network calculations. However, it is mainly used for audio editing, speech processing, photos, videos, security applications, and allows the projection of highly dimensional data to lower dimensional data [5].

Kohonen Networks will be used in this paper for the analysis of transport companies in the Czech Republic. The transport sector has a significant impact on the performance of the Czech economy and is an indispensable part of the national economy. It is a field that requires strategic vision and timely and courageous, though often financially demanding, decisions [6]. The transport industry includes a wide range of means of transport. According to the type of transport routes, we distinguish road, air, rail, water, combined transport, etc. The transport industry includes all companies transporting persons or goods, by land, sea or air, from one place to another [7].

However, the transport sector has been one of the main factors contributing to environmental pollution in recent years. Much research works on energy consumption analysis. Ou et al. [8] for example, compared fuel types, including alternative fuel systems, which lead to greater fuel technology development. This development can, according to Hao et al. [9] help to reduce oil consumption in transport.

Evaluating the financial situation of companies is a complex and multidimensional process. First of all, the following aspects of business activity should be considered: profitability, financial liquidity, operational efficiency, debt and market value. Certainly there are many financial ratios that are used to measure these five aspects. However, it should be noted that different ratios can provide conflicting signals that could be observed during the recent economic crisis in the transport sector. This situation may cause problems in assessing the financial situation of the transport companies examined [10]. According to Pánek [11], it is very useful to use multidimensional statistics in studies on the financial situation of transport companies that allow for the determination of the synthetic rate that will replace the many sets of characteristics of the surveyed entity with one aggregate variable. Synthetic measure allows for the measurement of a multidimensional phenomenon, which is the financial condition of the transport company.

Compared to other companies, a transport company does not store the transport services offered in its inventory and its investments require considerable costs and a long construction time, which can affect the efficiency of the organization in the long term [12]. A comprehensive expenditure accounting system for transport companies should describe in detail all of the costs incurred, all services provided, how all of the funds of participating agencies were expended and evenly divide the costs between those participating agencies by allocating costs according to services received [13]. To this end, an accounting scheme has to be developed that can monitor all types of expenditure related to transport services [14]. The accounting approach recommended and used by successful business operations and transport systems is called full cost accounting. The use of full cost accounting means that all the costs of providing transport services are taken into account and that all different types of collection are recorded [15].

As far as freight and passenger transport in the Czech Republic is concerned, the number of goods transported grew steadily in freight transport from CZK 435,450 thousand tons to 549,086 thousand tonnes, from 2012 to 2015. However, 2016 saw a decline of almost 10,000 thousand tonnes. In passenger transport, however, the situation was different.

In 2012 the number of persons transported was 2,448,956 thousand persons, higher than in 2013-2015, by almost 60,000 thousand persons. However, in 2016, the number of persons transported grew compared to 2012 by 50,000 thousand persons. Transport performance in freight transport has steadily increased from 2012 to 2015 by 8 0000 million tkm. In 2016, however, the transport performance dropped from 76,000 million tkm to 68,000 million tkm. The passenger transport performance in all monitored years grew steadily, from 42 723 million pkm to 46 702 million pkm [16].

The aim of this paper is to perform a cluster analysis of transport companies.

2 Data and methods

For the purposes of this paper, a data set of 3989 companies operating in the transport sector in 2016 will be created. The set of companies was selected randomly from the Albertina database. The complete financial statements of these companies will be available (all financial information is in thousands of CZK). For the analysis, only some items will be used:

1. Total assets: This is information about the size of the company, namely the volume of assets (fixed, current, other).
2. Fixed assets (also long-term assets): The item describes the extent of the involvement of fixed assets and gives an idea of the level of technology used and the projected volume of outputs. It can be assumed that the higher the volume of fixed assets, the more the transport machinery (cars, aircraft, etc.) is available to the company. Of course, it is necessary to mention that this will not affect the means of production that the company has in terms of leasing. Such long-term assets are indirectly reflected in other variables - in outputs, and above all in the form of the economic result of an economic unit.
3. Current assets: represent the amount of funds, receivables and inventories in the company. It is in most cases the so-called working capital, ie those asset components that change their substance in less than one year and which are directly involved in the core business.
4. Own resources: reflect the level of business risk of business owners. This is, therefore, the amount of funds that owners can lose in case of bankruptcy.
5. Foreign resources: It is a capital that does not have a share in the company's business management. In a way, it reflects the external view of the potential success of the company.
6. Performance: characterizes the main activities of the business, ie production.
7. Value Added: This is the value added by a company to its core business of basic input material.
8. Operating result: Provides information on how the business is successful in its core business.
9. ROE: return on equity means an assessment of the equity of a company, that is, the capital provided to the business owner.
10. Profit before tax: describes the success or failure of a company in its entire business.

These are not generators of value but rather characteristics that may have, albeit indirect, influence on the performance of a company.

He file will then be subjected to cluster analysis, specifically using Kohonen networks. For cluster analysis, Dell's Statistica software will be used in version 12. Data Mining will be used as a specific neural network tool. Here we select neural networks without a teacher – Kohonen's networks. We will specify the data for analysis. In all cases, these are continuous predictors. The file will be divided into three parts:

1. Training Data Set: Represents 70% of companies in the set. The Kohonen network will be created on this dataset.
2. Testing Data Set: This is 15% of the original set of companies. Using this data set, we will verify the parameters of the generated Kohonen network.
3. Validation dataset: 15% of companies in the file will also be involved. Using this dataset, we will also test the Kohonen network to see whether it is or is not usable.

We will determine the topological length and the topological width of the Kohonen map at 10. The number of iterations of the calculation will be set at 10,000. However, I will recall that the level of error is decisive. Unless there is an additional iteration to improve the Kohonen network parameters, the training will be terminated before the 10,000th iteration is completed. If the network parameters are improved even at 10,000 iterations, we need to repeat the process and set the higher value of the required iterations to make sure the result is the best possible.

3 Results

In accordance with the methodology of the contribution, the data was divided into three sets - training, testing and validation. The descriptive characteristics of all three sets and the original data set are given in Table 1.

Table 1. Descriptive characteristics of the dataset

Samples	Total assets	Fixed assets	Current assets	Own resources	Foreign resources	Perform.	Value added	Operating result	ROE	Profit before tax
Minimum (Training)	-533	-65	-704	-255943	-19155	-9	-7167334	-2465847	-	-1089
Maximum (Training)	72746546	66982500	14350000	55525850	45481000	7042804	3627035	6170000	319,000	4326000
Average (Training)	124438	86387	37161	67443	56422	22438	17526	7466	0,153	2116
Standard deviation (Training)	1850267	1674838	343483	1161169	970913	240585	187133	141499	7,923	84268
Minimum (Testing)	-167	-1156	-167	-22814	-310	-1184	-121017	-133373	-33,000	-385000
Maximum (Testing)	73664000	64656000	9008000	33946000	39718000	1263852	5168222	1371000	26,476	1152000
Average (Testing)	227116	178974	46690	105445	121110	13907	26831	8724	0,037	1391
Standard deviation (Testing)	3165045	2760381	411755	1459102	1708704	69514	221384	69524	2,267	49714
Minimum (Valid.)	-113	-637	-457	-586014	-31	0	-129229	-142348	-25,513	-382
Maximum (Valid.)	79078485	72006551	6935054	63040526	15241527	18071252	11372726	1569062	23,444	111167
Average (Validation)	182192	139408	41069	124394	55292	44604	41926	8277	0,150	234
Standard deviation (Validation)	264989	170133	106565	176070	95440	170274	89501	27764	1,919	5667
Minimum (Overall)	-533	-1156	-704	-586014	-19155	-1184	-7167334	-2465847	-	-385000
Maximum (Overall)	79078485	72006551	14350000	63040526	45481000	18071252	11372726	6170000	319,000	4326000
Average (Overall)	148488	108215	39175	81678	65950	24482	22579	7776	0,135	1725
Standard deviation (Overall)	2339005	2098590	352679	1503140	1076596	351253	257576	125890	6,736	73110

Source: Author.

Ideally, the characteristics of the partial sets of data would be approximately the same. Since companies have been randomly assigned, it is not entirely so. However, the random distribution of data may not be an error in the result.

Based on the assignment, the Kohonen network was calculated. It is also referred to as SOFM 10-100 in the text. The companies were divided into individual clusters within the 10 x 10 topological grid. The frequency of individual clusters is evident from the three-

dimensional graph in Figure 1. It is evident from the figure that the largest number of companies in the individual topological grid fields is in the position (4, 10) then in position (4, 9). The following ones are (5, 9) and (6, 10). Overall, we can say that the largest number of nearly 4000 companies monitored is in the sectors (4, 10), (6, 10) and (3, 7), (6, 7).

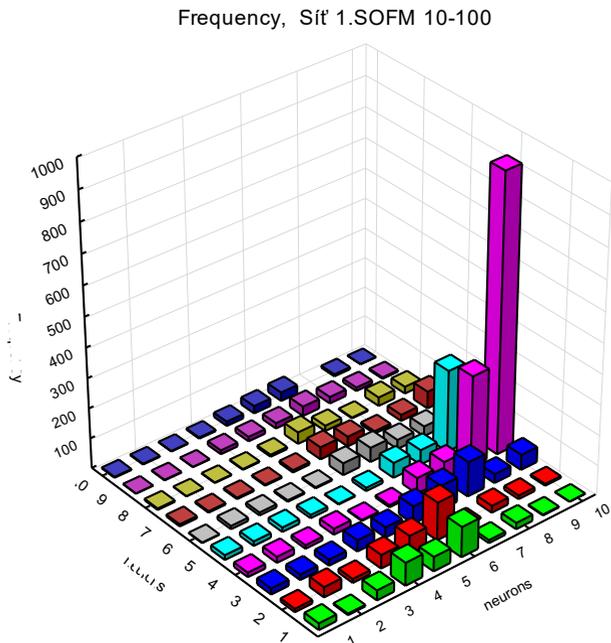


Fig. 1. Graph of company frequencies in the topological grid

Source: Author.

The results shown in Table 2 are more specific.

Table 2. SOFM 10-100 Network frequency Table

	Network: 1.SOFM 10-100 / Samples: Training, Testing, Validation									
	1	2	3	4	5	6	7	8	9	10
1	22	5	33	76	48	100	10	21	3	1
2	11	34	13	45	61	125	3	23	12	6
3	20	18	20	37	34	58	81	115	30	51
4	15	20	12	20	3	7	53	62	304	919
5	16	15	15	11	3	11	50	49	271	137
6	2	11	10	5	4	40	48	31	33	167
7	7	5	5	4	8	41	33	8	21	57
8	5	1	2	3	7	39	18	3	29	23
9	1	2	1	13	15	18	32	22	12	5
10	3	3	2	4	15	25	29	0	7	1

Source: Author.

Even from the table, it is advisable to focus on three clusters, namely (4, 10), (4, 9) and (5, 9). This is a very interesting set of data representing more than one third of all companies in the file. Dominance may be more evident from the Kohonen graph in Figure 2.

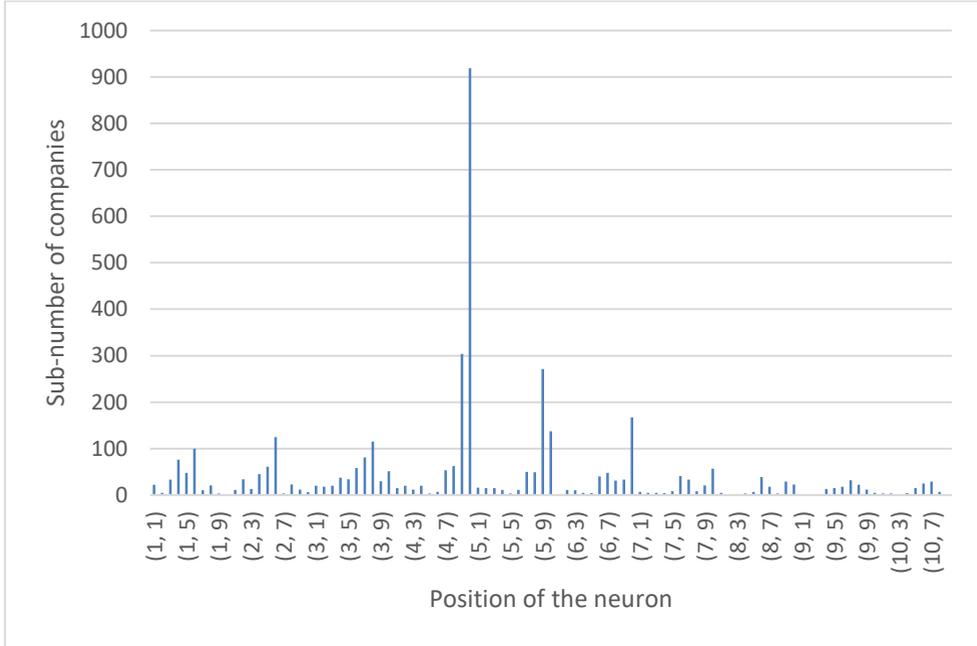


Fig. 2. Kohonen graph for selected cells of the SOFM 10-100 network

Source: Author.

While the data set is rather heterogeneous, cluster analysis suggests that there are characteristics in which the companies are similar.

Cluster (4, 10)

The cluster (4, 10) of the topological grid encompasses 919 companies. It is the largest set of data from the partial clusters. From the point of view of input parameters, we can derive the following average values:

1. Total assets: 1,686,400 thousand CZK
2. Total fixed assets: 810,882 thousand CZK,
3. Total current assets: CZK 823,749 thousand CZK,,
4. Total liabilities: 1,686,400 thousand CZK,
5. Total equity: 224,781 thousand CZK,
6. Foreign resources total: 1 471 796 thousand CZK,
7. Total performance: 38,515 thousand CZK,
8. Total value added: CZK 256,189 thousand CZK,
9. Operating result: 13 988 thousand CZK,
10. ROE: 10.5%.

From the average we can derive a typical business for this cluster of companies. It is a small company whose operating result is positive, yet close to zero. Despite this, its existence is not jeopardized. Business automation is relatively low. The company holds a relatively large volume of current assets (probably inventory and receivables).

When it comes to the homogeneous cluster of companies, we can verify the validity of basic logical links. The relationship between total assets and the level of operating income seems interesting (see Figure 3 below).

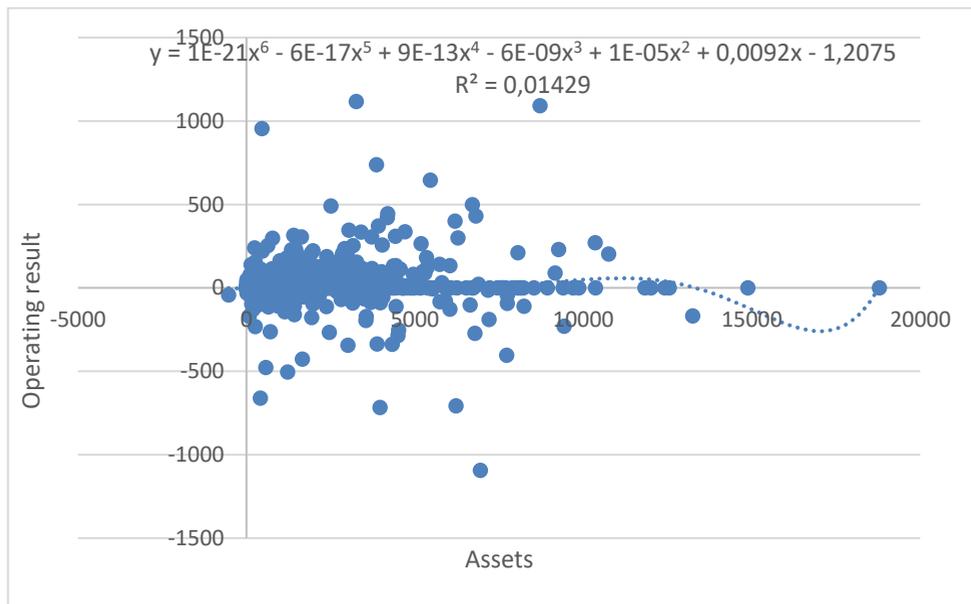


Fig. 3. Relation of total assets and operating profit of the cluster (4, 10)

Note: All values in graphs are in thousands of CZK.

Source: Author.

Total assets and operating result correlate to close to 0.12. Therefore, there can be no significant dependence of the two variables. Yet the data is interleaved by a regression curve, a 6th order polynomial. Based on this, it is possible to estimate the development of the operating result on the basis of the size of the company determined by the volume of total assets.

Figure 4 shows the relationship of foreign resources and operating income.

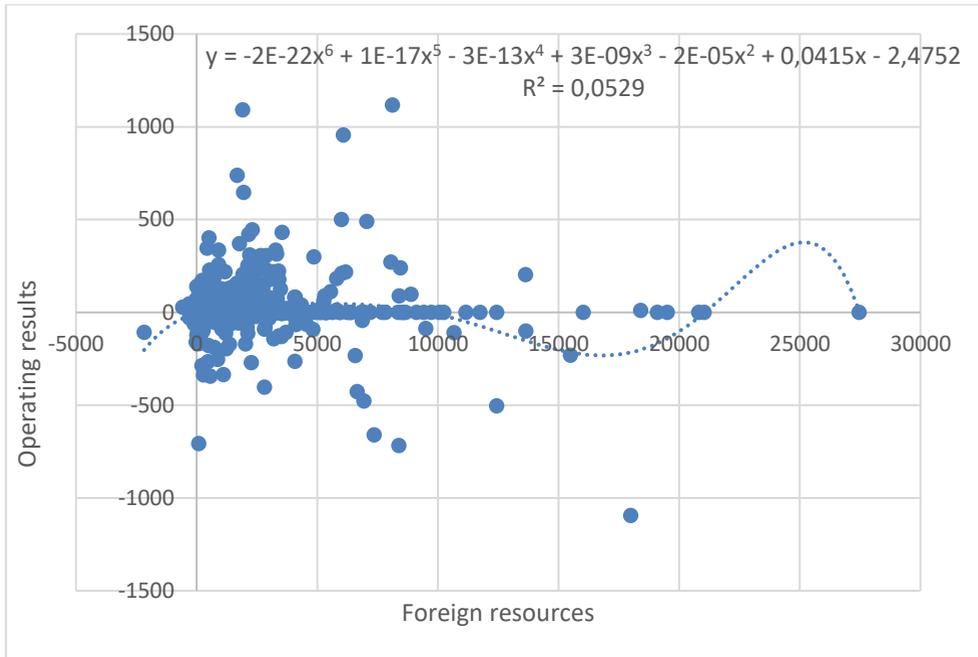


Fig. 4. Foreign resources and the operating result of the cluster (4, 10)

Source: Author.

The relationship can show us the effect of the financial leverage, assuming a relatively high correlation between the two variables. But this is only minimal. Yet, the graph points are interleaved by the 6th order polynomial. It can be seen that the ideal debt ratio of companies in the cluster (4, 10) is up to 2 million CZK. The average debt will then be up to 100,000 CZK.

Cluster (4, 9)

The cluster (4, 9) of the topological grid occupies the imaginary second position in frequency. It has 304 companies. From the point of view of input parameters we can deduce a typical representative of this cluster:

1. Total assets: 2 698 333 thousand CZK,
2. Total fixed assets: 877,032 thousand CZK,
3. Total current assets: 1,741,898 thousand CZK,
4. Total liabilities: 2 698 333 thousand CZK,
5. Total equity: 1,037,298 thousand CZK,
6. Foreign resources total: 1,638,506 thousand CZK,
7. Total performance: 15 037 thousand CZK,
8. Total value added: 312 950 thousand CZK,
9. Operating result: 41,158 thousand CZK,
10. ROE: 17.9%

A typical company of this cluster is five times larger than the company cluster (4, 10) (meaning from the point of view of the total assets). This is a company with a lower degree of automation and a higher share of the circulation of total assets than in the case of cluster (4, 10). A larger difference between the two clusters can be identified in the case of ROE, which is 17.9% in the cluster (4, 9) vs. 10.5% of the cluster (4, 10). Even in this case, it is certainly interesting to explore the relationship of some variables.

Figure 5 shows the relationship between total assets and operating result.

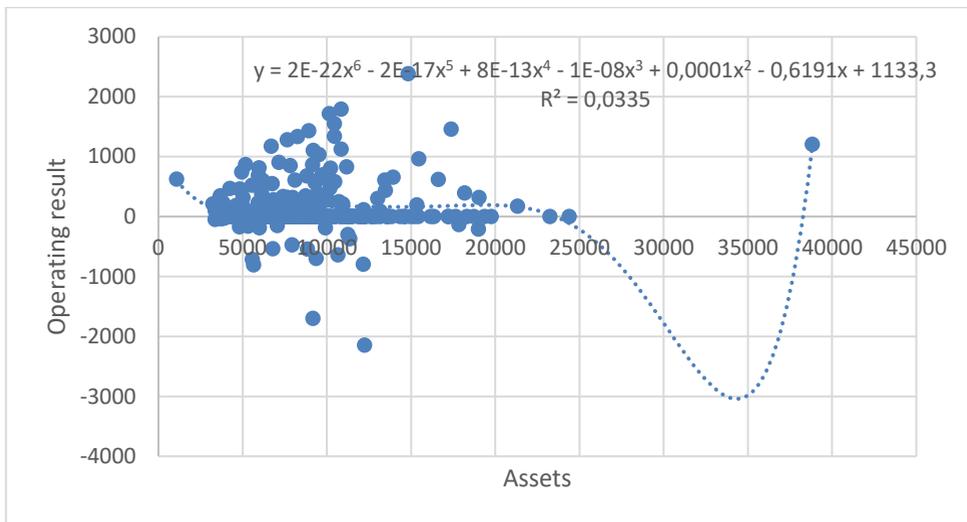


Fig. 5. Relationship of total assets and operating result of cluster (4, 9)

Source: Author.

The determination coefficient is 0.033. We can not talk about the dependencies of the two variables either. However, in this cluster we can see the fact that a higher volume of assets implies a higher operating result. The trendline has a 6th order polynomial again.

Figure 6 illustrates the relationship of foreign resources and operating result.

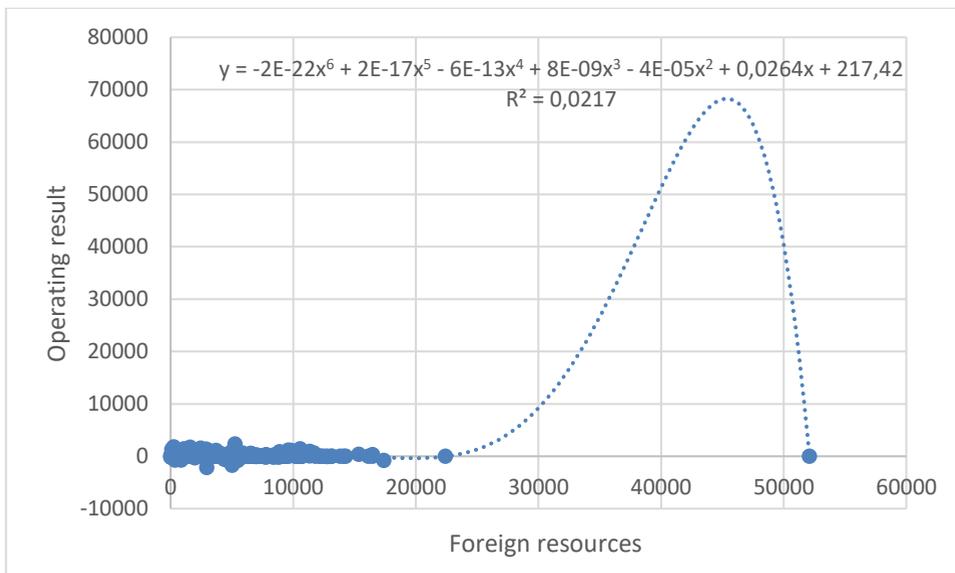


Fig. 6. Foreign resources and the operating result of the cluster (4, 9)

Source: Author.

Even in this case, the correlation coefficient is not high. However, even in this case, we can say that a higher debt ratio may mean a higher operating result.

Cluster (5, 9)

To illustrate the overall situation, we also present the characteristics of the third most significant cluster, (5, 9). The cluster contains 271 transport companies. If we take the total assets as a measure of the size of the company, we find that the companies of this cluster are approximately 1.5 times larger than those of the company cluster (4, 10), but 4 times smaller than the company cluster (4, 9):

1. Total assets: 567,190 thousand CZK,
2. Total fixed assets: 228,634 thousand CZK,
3. Total current assets: 328,622 thousand CZK
4. Total liabilities: 567,190 thousand CZK,
5. Total equity: 65,818 thousand CZK,
6. Foreign resources total: 496,975 thousand CZK,
7. Total performance: 48,996 thousand CZK,
8. Total value added: 218,318 thousand CZK,
9. Operating result: 27,713 thousand CZK,
10. ROE average: 14.5%

Even from the point of view of average profitability, clusters (5, 9) can be in terms of success categorized among companies of clusters (4, 10) and (4, 9).

4 Conclusion

The aim of the paper was to carry out a cluster analysis of transport companies.

The aim of the paper was fulfilled.

A cluster analysis was performed using neural networks without a teacher – Kohonen networks. Based on the cluster analysis, the companies were divided into clusters in the Kohonen map (10 x 10 clusters). Some clusters are significant in terms of number of companies. The most represented are clusters (4, 10), (4, 9) and (5, 9). These clusters concentrated more than a third of all companies operating in transport in 2016. These clusters have therefore been subjected to further analysis. In the case of the evaluated clusters, the typical representatives of these clusters were defined. They varied by their size, the volume of operating income generated and the return on equity. The companies of cluster (4, 10) are small businesses and generate lower operating profits compared to the remaining significant clusters. A typical company of cluster (4, 9) is five times larger than that of the cluster (4, 10). Such a company also achieved a significantly higher ROE, above 17%. The companies of cluster (5, 9) are larger than the companies of cluster (4, 10), but smaller than those of cluster (4, 9). Even in the case of ROE, they reach average values among the three analysed clusters. From the analysis we can make three very specific conclusions:

1. A larger company (or a company with a larger volume of assets) generates, on average, a higher operating result.
2. Larger companies achieve higher ROE rates.
3. In the case of larger company, the financial leverage is more positive, i.e. the more the growth of the volume of the foreign capital increases, the higher operating result and the profitability of the equity.

The scale of the work did not allow it, but it would be interesting to prepare a report on the state of the Czech industry in the range of all clusters of the Kohonen map. We could identify the most profitable companies for the entire industry, identify declining businesses, or others.

References

1. R. Andekina, R. Rakhmetova, Financial Analysis and Diagnostics of the Company. *Procedia Economics and Finance*, **5**, 50-57, (2013)
2. M. Slusarczyk Antosz, N. H. Morales Merchán, Análisis de las estrategias empresariales y de las TIC. *3C Empresa Investigación y pensamiento crítico*, **5**(1), 29-46, (2016)
3. M. Vochozka, J. Horák, Comparison of neural networks and regression time series in prediction of silver price development. *Proceedings of the 9th International Scientific Conference Company Diagnostics, Controlling, and Logistics*, 271-278, (2018)
4. X. Han, L. Wang, Stock Company Comprehensive Assessment Model Based on Kohonen Network. *2008 Second International Conference on Genetic and Evolutionary Computing*, 185-188, (2008)
5. M. Vochozka, Z. Rowland, V. Stehel, P. Šuleř, J. Vrbka, *Modelování nákladů podniku pomocí neuronových sítí*. České Budějovice: Institute of Technology and Business in České Budějovice, (2016)
6. J. Horák, M. Vochozka, V. Machová, Evaluation of transport and forwarding companies by means neural networks. *Proceedings of the International Scientific Conference for Ph.D. students and assistants – QUARE*, 377-385, (2018)
7. S. M. Wagner, Innovation management in the German transportation industry. *Journal of Business Logistics*, **29**(2), 215-231, (2008)
8. X. Ou, X. Yan, X. Zhang, Z. Liu, Life-cycle analysis on energy consumption and GHG emission intensities of alternative vehicle fuels in China. *Applied Energy*, **90**(1), 218-224, (2012)
9. H. Hao, H. Wang, L. Song, X. Li, M. Ouyang, Energy consumption and GHG emissions of GTL fuel by LCA: Results from eight demonstration transit buses in Beijing. *Applied Energy*, **87**(10), 3212-3217, (2010)
10. J. Jedrzejczak-Gas, P. Kuzdowicz, Evaluation of the financial condition of the companies of transport and storage section in times of economic crisis - taxonomic analysis. *Management*, **20**(1), (2016)
11. T. Pánek, J. Zwierzchowski, *Statystyczne metody wielowymiarowej analizy porównawczej: teoria i zastosowania*. Warszawa: Oficyna Wydawnicza, Szkoła Główna Handlowa, (2013)
12. E. Quinet, R. W. Vickerman, *Principles of transport economics*. Northampton, MA: Edward Elgar, (2005)
13. O. Ponisciaková, *Some Aspects of Costs Management in the Transport Companies*. Lithuania, 763-766, (2016)
14. Transportation Accounting: How Understanding Transportation and Freight Costs Leads to Strategic Thinking and More Profitable Business. *Cerasis* [online]. 2014 [cit. 2018-07-16]. Available at: <http://cerasis.com/2014/08/06/transportation-accounting/> (2014)
15. Transportation accounting fundamentals. *The national academies press: Sharing the Costs of Human Services Transportation* [online]. 2011 [cit. 2018-07-16]. Available at: <https://www.nap.edu/read/14490/chapter/7> (2011)
16. Přeprava věcí a osob, přepravní výkony. *Český statistický úřad: Veřejná databáze – Czech Statistical Office* [online]. 2018 [cit. 2018-07-16]. Available at: <https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=vystup-objekt&pvo=DOP05->

[D&z=T&f=TABULKA&skupId=1613&katalog=31028&pvo=DOP05-D&c=v3~8 RP2016](#)(2018)