UNIVERSITY OF NIS FACULTY OF MECHANICAL ENGINEERING
THE SEVENTH INTERNATIONAL CONFERENCE
TRANSPORT AND LOGISTICS
til 2021

APPLICATION OF PCA-DEA MODEL FOR EVALUATING THE EFFICIENCY OF A TRANSPORT COMPANY

Smiljka Miškić [1]
Eldina Huskanović [1]
Marko Vasiljević [1]
[1)] *University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Bosnia and Herzegovina*

Abstract

Due to increased competitiveness, transport companies need to work constantly on improving their services and promptly respond to all challenges. However, other than a job well done, it is very important that the company demonstrates positive results at the end of the financial year. For these reasons, it has been determined the efficiency of a transport company including a period of eight years since it provides a more realistic picture of business performance. The PCA-DEA model was applied in order to evaluate the efficiency of a transport company on the basis of six input and four output parameters. The number of vehicles, number of drivers, number of operating hours, vehicle maintenance costs, fuel costs per total kilometers traveled and transport staff costs are inputs for the PCA-DEA model, while the total number of deliveries, total quantity transported, total number of kilometers traveled, and profit are outputs. Based on all these indicators, the company operated most efficiently in the period 2014-2017, followed by less efficient results. In addition, the results obtained provide guidelines for further business operations.

Keywords: *PCA-DEA, efficiency, transport.*

1 INTRODUCTION

The goal of every transport company is efficient logistics solutions that primarily relate to timely delivery of goods, quality and reliability. However, transport companies in BiH have additional aggravating circumstances due to legal procedures that apply to countries that are not members of the European Union. Although these are limiting factors, it is constantly being worked on improving business performance so that customers are satisfied with the end service. In addition to the satisfaction of transport service users, equally important is the performance of the transport company itself. In order to determine business performance, various input and output parameters are analyzed. In this paper, the efficiency of a transport company from BiH in a period of eight years (2013-2020) is determined by considering a total of 10 parameters that represent a combination of inputs and outputs and relate to quantitative parameters expressed in numbers and cost parameters. In order to solve this problem, it is used a combination of DEA (Data envelopment analysis) model and PCA (Principal component analysis) model. It has been defined six input parameters (inputs) and four output parameters (outputs) used for evaluating the efficiency of the transport company. The defined inputs are: the total number of vehicles, number of drivers, number of operating hours, vehicle maintenance costs, fuel costs per total kilometers traveled and transport staff costs. The following outputs are also defined: the total number of deliveries, total quantity transported, total number of kilometers traveled and profit. In addition to the introductory considerations in Section 1, the paper is structured through five other sections. In Section 2, it is given a review of the literature with reference to previous application of the models used in this paper. Section 3 presents the research methodology consisting of DEA model and PCA model. The application procedure and basic characteristics of the models are explained in detail. DEA is a specially designed technique for measuring the efficiency of complex units where it is not clearly expressed which input parameters and to what extent participate in creating a particular output parameter. The Principal Component Analysis is a technique of creating new variables that are linear combinations of initial variables. PCA is used to increase the discrimination power of the DEA model. Section 4 presents an analysis of the transport system of the company observed. In Section 5, it is integrated a model for determining the efficiency of the transport company, which is based on the PCA and DEA model. In Section 6, the last section, concluding remarks are given with an emphasis on the contribution of the paper.

2 LITERATURE REVIEW

The measurement of business efficiency of a transport company was performed using an integrated PCA-DEA model. The models have been applied in many studies for a purpose of measuring the efficiency of different areas of transport. Some of the studies will be given below.
Andrejić [2] believes that for successful measurement of efficiency in logistics it is necessary to consider a large number of inputs and outputs that are different in nature (financial, technical, environmental, energy, social, etc.) and expressed in different units. In this regard, it is possible to measure energy, environmental, cost and other types of efficiency in logistics. Davoudabadi et al. [7] apply an integrated PCA-DEA model in order to measure efficiency and select the best supplier. The PCA approach is used to reduce dimensionality and correlations between criteria, while the DEA is used to rank suppliers. A similar study was conducted by Karami et al. [10]. The PCA-DEA model was used in combination with other

methods in order to evaluate and select suppliers in the garment industry. Decisions related to the purchase and supply of raw materials play a key role in business logistics. In this regard, it is very important to develop models that will enable the selection of the best supplier and thus long-term business cooperation with the desired suppliers without unforeseen changes in supply that may have a negative impact on the business. Andrejić and Kilibarda [4] measure global logistics efficiency using the PCA-DEA approach. It has been concluded that the proposed model can be used to assess logistics activities at a global level and to improve current approaches. Layeb et al. [11] conduct an analysis of the efficiency of a logistics service provider operating in Tunisia. The paper uses the PCA approach to select appropriate indicators, followed by the DEA method for measuring the efficiency of all warehousing and transport activities. The DEA model greatly contributes to the improvement of business operations since it enables the identification of efficient or inefficient companies. Accordingly, it is possible to react in time and provide guidelines for potential improvement in the efficiency of each company that has been identified as inefficient, i.e. it is possible to determine how much each business element should be reduced or increased in order for a given company to become relatively efficient. Deng et al. [8] apply an integrated PCA-DEA approach with the aim of assessing carbon emissions and negative impacts when implementing logistics activities. The PCA model was applied to reduce the dimensionality of the indicators observed, and then the DEA method was applied to measure and assess logistics performance with and without carbon emission constraints in 30 provinces in China. The authors emphasize the advantage of the DEA model, i.e. the possibility of analyzing mutually comparable units despite heterogeneous data, which are expressed in different units of measurement and affecting business efficiency.in different ways.

3 METHODOLOGY

The methodology of the paper is shown in Figure 1, where it is easiest to see what efficiency evaluation models have been applied and what steps have been taken to determine the business efficiency of a representative transport company.

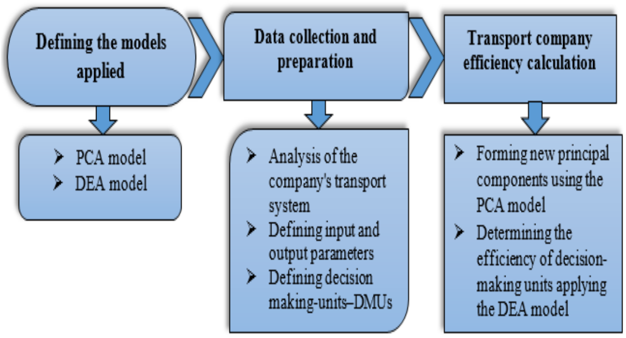

***Fig 1**. Applied methodology*

The paper analyzes the efficiency of a transport company from Bosnia and Herzegovina, which provides services of domestic and international transport of goods. After defining the models applied, it was performed an analysis of the transport system of a representative transport company in the last eight years. Six input parameters and four output parameters were defined. The defined inputs are: the total number of vehicles, number of drivers, number of operating hours, vehicle maintenance costs, fuel costs per total kilometers traveled and transport staff costs. The defined outputs are: the total number of deliveries, total quantity transported, total number of kilometers traveled and profit. Data on input and output values for all eight decision-making units were collected. The data collected represent the basis for the realization of the next phase, i.e. basis for determining the efficiency of the company using an integrated PCA-DEA model. DEA is used to evaluate the relative efficiency of a homogeneous set of DMUs characterized by multiple inputs and outputs. If the value of DEA=1, the model shows efficiency, and if the value of DEA<1, the model shows inefficiency. PCA is used to further strengthen the discrimination power of the DEA model by creating new principal variables that represent linear combinations of initial variables. The main steps in the analysis of principal components are: standardization of variables, computation of the matrix of correlations between all initial standardized variables, finding the eigenvalues of principal components, rejection of the components that are carriers of proportionally small share of variance [3].

3.1. DEA model

DEA is a specially designed non-parametric technique based on linear programming for measuring the efficiency of complex units where it is not clearly expressed which input parameters and to what extent participate in creating a particular output parameter. The DEA method measures the relative efficiency of units by constructing an empirical efficiency limit based on the data on the inputs used and the obtained outputs of all units. Best practice units, the ones that determine the efficiency limit, gain a grade of 1, and the degree of technical inefficiency of other units is calculated based on the distance of their input-output ratio to the efficiency limit [6]. CCR is the most basic DEA model. The model was presented by Charnes et al. [5]. This section presents two models of DEA CCR models that were used to obtain the values of alternatives, i.e. DMUs, according to an input-oriented model (min) and according to an output-oriented model (max). The DEA CCR input oriented model (min) is (1):

$$DEA_{input} = min \sum_{i=1}^{m} w_i x_{i-input} \tag{1}$$

with constraints (2):

$$\sum_{i=1}^{m} w_i x_{ij} - \sum_{i=m+1}^{m+s} w_i y_{ij} \geq 0, \quad j = 1, \dots, n$$

$$\sum_{i=m+1}^{m+s} w_i y_{i-output} = 1$$

$$w_i \geq 0, \; i = 1, \dots, m+s \tag{2}$$

The DMU consists of m input parameters for each alternative x_{ij}, while s represents the output parameters for each alternative y_{ij}, taking into account the weights of the parameters denoted by w_i. Also, n represents the total number of DMUs. The DEA CCR output oriented model (max) is (3):

$$DEA_{output} = max \sum_{i=m+1}^{m+s} w_i y_{i-output} \quad (3)$$

with constraints (4):

$$-\left(\sum_{i=1}^{m} w_i x_{ij}\right) + \sum_{i=m+1}^{m+s} w_i y_{ij} \leq 0, \quad j = 1, \ldots, n$$

$$\sum_{i=1}^{m} w_i y_{i-input} = 1$$

$$w_i \geq 0, \; i = 1, \ldots, m+s \quad (4)$$

3.2. PCA model

Using the PCA model, initial variables are transformed into new principal variables that represent linear combinations of initial variables. The first principal component covers most of the variance of an original data set, and the following components include the part of the variance that is not covered by previously singled out components [12].

In this way, dimensionality is reduced, and it is achieved greater visibility and simplification of a large number of data. The analysis of the principal components is completed through four main steps [3]:

1. standardization of variables;
2. computation of the matrix of correlations between all initial standardized variables;
3. finding the eigenvalues of the principal components;
4. rejection of the components that are carriers of a proportionally small share of variance (usually the first several components carry 80%–90% of the total variance).

The main goal of the PCA model application is to summarize and analyze the linear correlation of a large number of differently distributed, quantitative, mutually correlated variables, into a smaller number of new variables, with minimal loss of information. The results of the principal components can be used for further explanation of results.

4 ANALYSIS OF THE TRANSPORT SYSTEM OF THE COMPANY

In this section, it is performed an analysis of the transport system of a representative transport company. Due to data protection, the name of the company will not be stated nor any other data that may reveal the identity of the company. The company is engaged, as already mentioned, in domestic and international freight transport. It provides transport of containers, packaged goods – ADR dangerous goods, and transport of goods that require controlled temperature, i.e. transport by refrigerators. When analyzing the transport system of the company, attention is focused on monitoring changes in the following indicators: total number of vehicles, number of drivers, number of operating hours, vehicle maintenance costs, fuel costs per total kilometers traveled, transport staff costs, total number of deliveries, total quantity transported, total number of kilometers traveled, profit and turnover for the month of April in the period from 2013 to 2020.

When it comes to the number of vehicles, the company had the most vehicles available in 2018 (58 vehicles), and the least in 2020 (27 vehicles). The number of drivers employed by the company during the period observed ranges from 30 to 61. A graphical representation of the number of operating hours and the total quantity transported is given in Figure 2. The largest number of operating hours was achieved in 2018 as well as the total number of kilometers traveled, when the number of operating hours was 146,000, while the number of kilometers traveled was 5,856,000. In contrast, the lowest number of operating hours was observed in 2020, when 73,000 operating hours were achieved. When it comes to this parameter, the highest number of kilometers traveled was recorded in 2018, then in 2017, 2019, 2015, 2014 and 2013, where that number was 4,900,000 km, 4,704,000 km, 4,608,000 km, 4,224,000 km, 3,700,000 km and 3,000,000 km by years, respectively.

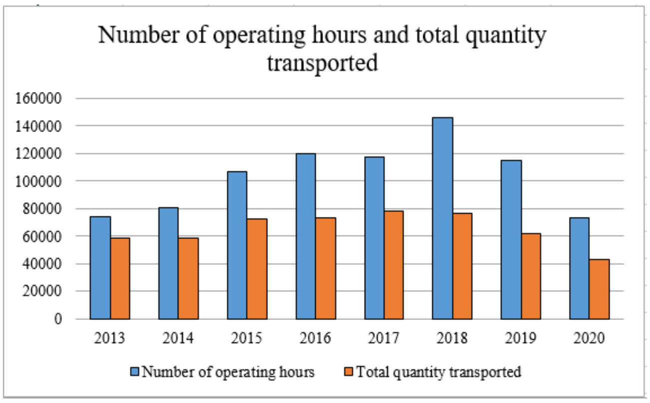

Fig. 2 *Graphical representation of the number of operating hours and the total quantity of goods transported*

Observing the total quantity transported, it can be noticed that the largest amount of goods was transported in 2017, more precisely 78,600 tons, and then in 2018, when 76,980 tons were transported. The smallest amount of goods was transported in 2014 (58,340 t) and in 2020 (43,360 t). When analyzing the total number of deliveries, we come to the conclusion that the largest number of deliveries was done in 2017 (3,930), followed by 2018 (3,849) and 2016 (3,672). The lowest number of deliveries was recorded in 2020 with 2,168 deliveries. Figure 3 shows a graphical representation of vehicle maintenance costs.

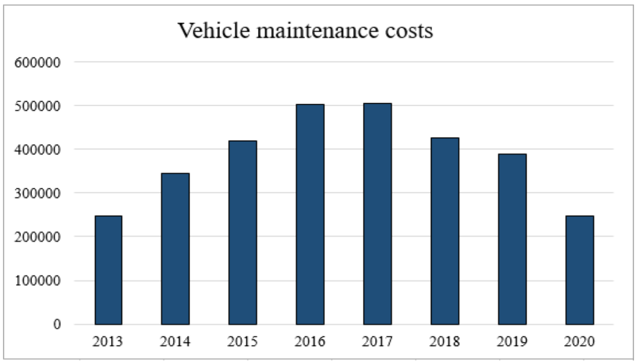

Fig. 3 *Graphical representation of vehicle maintenance costs*

Observing vehicle maintenance costs, it can be noticed that they had the highest value in 2017 (BAM 505,906), followed by 2016 (BAM 503,687) and 2018 (BAM 427,432). Slightly lower vehicle maintenance costs compared to the previously given were observed in 2015 (BAM 419,855), while the lowest maintenance costs were recorded in 2013 and 2020 (BAM 208,505). The highest fuel costs per total kilometers traveled are at the beginning of the period observed, i.e. in 2013 (0.67), followed by 2014 (0.60) and 2017 (0.53). In the period from 2018 to 2020, there is very little difference between fuel costs, where the value of costs is 0.46; 0.42 and 0.45. Transport staff costs, in the period observed, range from BAM 981,066 to BAM 219,0264. At the beginning of the period observed, the transport staff costs are BAM 1,463,427, while in 2016 they reach a maximum value of BAM 2,190,264. Since 2017, transport staff costs have been reduced, and in 2019 they have a minimum value of BAM 981,066 KM. Profit is the difference between income and expenditure. A graphical representation of profit is given in Figure 4.

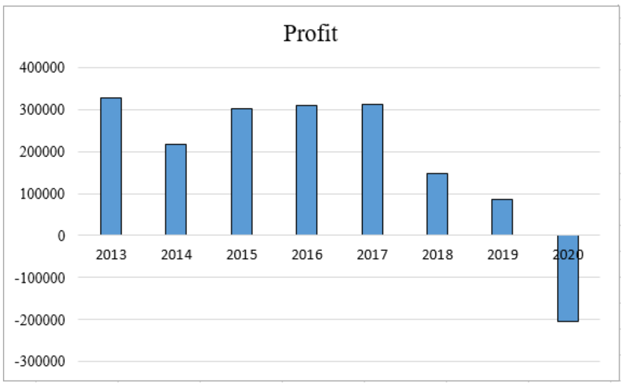

Fig. 4 *Graphical representation of profit*

The company made the maximum profit of BAM 328,675 in 2013. In the following year, 2014, the profit was BAM 217,144, while, in 2015, the profit was BAM 302,445. Compared to the previously observed year, in 2016 and 2017, it was noticed smaller differences in profit, where a profit of BAM 309,331 and BAM 313,002 was achieved. After that, in 2018 and 2019, it was recorded a significantly lower value of the profit of BAM 148,554 and BAM 86,743, respectively. As it has already been said, profit represents the difference between income and expenditure, i.e. costs, therefore it is desirable that revenues be greater than costs. However, if the revenues are less than the costs, the profit has a negative value, i.e. the company has a loss, as is the case in 2020 with the analyzed company TC1. So, the company operates with a loss of BAM 205,389. Many factors can have a negative impact on profit, and in this case, the cause of the negative value of profit comes from the company itself. The company achieved a new business venture that includes the purchase of a new enterprise (not engaged in transport), and for that reason it directed large funds to financing the new company, which eventually led to the loss recorded in 2020. In addition to the parameters given, the turnover for the month of April was also analyzed, and a graphical representation of the turnover for this month is given in Figure 5.

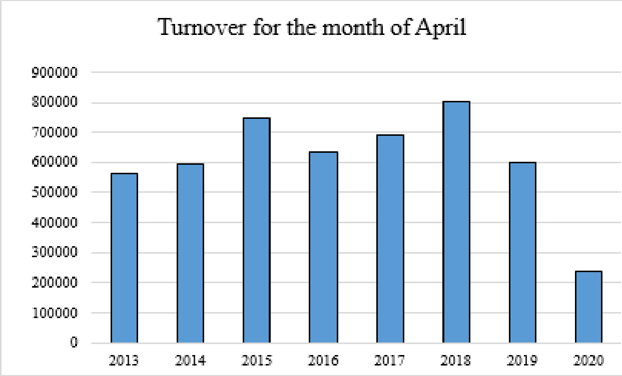

Fig. 5 *Graphical representation of turnover for April*

The company TC1 achieved the largest turnover of BAM 801,861 in April 2018, and then in April 2015 (BAM 745,398) and 2017 (BAM 692,151). In April 2016, the recorded turnover was BAM 633,099. In the same month, at the beginning of the period observed, the turnover was BAM 563,386, while in the following year in the same month, the turnover was BAM 594,002. The turnover for the given month in 2019 was BAM 599,948, and in 2020 it was noticed the lowest value of turnover of BAM 237,183.

During the analysis of the transport system of the company TC1, i.e. when monitoring the changes of given indicators, it can be noticed that most of the analyzed indicators have a minimum value at the end of the period observed, i.e. 2020. In addition to numerous factors, the negative influence of the COVID-19 pandemic had the hugest impact on the previously given results.

5 PCA-DEA MODEL FOR EVALUATING THE EFFICIENCY OF TRANSPORT COMPANIES

This section presents in detail the application of the model for determining the efficiency of transport companies, which is based on the PCA-DEA approach. The main goal of this research is to determine the efficiency of the observed transport company in a period of eight years, i.e. identify efficient or inefficient financial years, and thus provide a more detailed insight into the business performance and possible guidelines for improving the efficiency of transport companies. To analyze the efficiency, as mentioned, it has been used the DEA method, which evaluates the relative efficiency of a homogeneous set of DMUs characterized by multiple input and output parameters.

The Principal Component Analysis has also been applied, i.e. the PCA model that enables the creation of new principal variables that are linear combinations of initial variables, and thus contributes to the reduction of dimensionality and increase the discrimination power of the DEA model. Thus, the PCA-DEA model is applied in order to solve the problem of a relatively small number of DMUs in relation to a relatively large number of parameters, i.e. when the application of the DEA method for assessing the efficiency of companies does not provide adequate results. In this case, it is necessary to increase the discrimination power of the DEA method by using PCA as a method to reduce the number of parameters with minimal loss of information from an initial set of parameters. In this way, the analysis becomes simpler because a large number of parameters are reduced to a smaller one, retaining almost all the information contained in an initial data set [9]. At the very beginning, it has been defined input and output parameters, which represent the basis for calculating the efficiency of the transport company. Six input parameters and four output parameters have been defined as previously stated and explained. After defining the input-output parameters, it has been collected the data on the values of the defined parameters for all eight years, representing decision-making units in this case. Table 1 shows the values of the input and output parameters.

Table 1 Values of input-output parameters

Year - DMU	Input parameters						Output parameters			
	Number of vehicles	Number of drivers	Number of operating hours	Vehicle maintenance costs	Fuel costs per kilometers traveled	Transport staff costs	Total number of deliveries	Quantity transported	Kilometers traveled	Profit
2013	30	30	74000	248505	0.67	1463427	2931	58620	3000000	328675
2014	32	34	81000	344852	0.60	1769294	2917	58340	3700000	217144
2015	44	44	107000	419855	0.51	1904486	3611	72220	4224000	302445
2016	48	49	119600	503687	0.44	2190264	3672	73440	4704000	309331
2017	47	49	117600	505906	0.53	2016107	3930	78600	4900000	313002
2018	58	61	146000	427432	0.46	1785322	3849	76980	5856000	148554
2019	45	48	115000	390278	0.42	981066	3126	62250	4608000	86743
2020	27	30	73000	248505	0.45	925000	2168	43360	2880000	-205389

The following step involves applying PCA to the defined input and output parameters in order to create principal components. Using a statistical program – SPSS, eight decision-making units have been analyzed, and the average values and values of standard deviation by variables are shown in Table 2.

Table 2 Descriptive statistics

Variables		Average	Standard deviation
Input	Number of vehicles	41.38	10.66
	Number of drivers	43.13	10.96
	Number of operating hours	104150.00	25945.38
	Vehicle maintenance costs	386127.50	100435.85
	Fuel costs per kilometers traveled	0.51	0.09
	Transport staff costs	1629370.75	467169.04
Output	Total number of deliveries	3275.50	600.37
	Total quantity transported	65476.25	12017.49
	Kilometers traveled	4234000.00	1005452.56
	Profit	187563.12	181519.59

After that, it is necessary to check the suitability of the correlation matrix for factorization by KMO test, the value of which should be equal to or greater than 0.6. In our example it is 0.579 or 0.592, which is on the verge of acceptability. By applying the Bartlett's Test of Sphericity, it is checked whether the variables are correlated, i.e. it is tested the null hypothesis that all correlations between the variables are equal to zero. Accordingly, the value of the Bartlett's test should be less than 0.001, and in our example it is zero, which means that the application of factor analysis is possible. During extraction, i.e. separation, the communality of variables should be greater than 0.4, i.e. greater than 40%, so that the variables would be considered for further analysis. In our case, all variables have been included in further analysis since the obtained communality values range from 0.862 to 0.993, which is significantly higher than the previously defined limit of 40%.

By applying the PCA technique, out of the set of input parameters, it has been singled out two principal components which contain 90% of the information of the original set of parameters. Additionally, two principal components have been singled out from the set of output parameters. The creation of the first principal component of the input is most influenced by the first variable, i.e. "Number of vehicles", while the second principal component of the output is most affected by "Transport staff costs". "Total number of deliveries" and "Total quantity transported" have the greatest influence on the creation of the first principal component of the output, while "Profit" has the greatest influence on the second principal component of the output. The next step is to apply the DEA model in order to determine the efficiency of decision-making units. The values obtained when calculating the efficiency using the Excel Solver for the PCA-DEA model are given in Table 3.

After applying the DEA model, all decision-making units show efficiency, while after applying the integrated PCA-DEA model, which is based on three principal input components and two principal output components, the eighth decision-making unit shows inefficiency.

Table 3 Results of PCA-DEA model

	DEA	PCA - DEA		
	6 - 4	3 - 2	2 - 2	2 - 1
DMU_1	1.000	1.000	1.000	1.000
DMU_2	1.000	1.000	0.925	0.899
DMU_3	1.000	1.000	1.000	0.970
DMU_4	1.000	1.000	0.981	0.915
DMU_5	1.000	1.000	1.000	0.972
DMU_6	1.000	1.000	0.982	0.967
DMU_7	1.000	1.000	1.000	1.000
DMU_8	1.000	0.959	0.954	0.807

The model based on two input components and two output components shows the efficiency of four decision-making units (DMU_1, DMU_3, DMU_5 and DMU_7). It means that the company TC1 operated efficiently in 2013, 2015, 2017 and 2019. Finally, the model based on two principal input components and one output component shows efficiency only in 2013 and 2019.

6 CONCLUSION

The significance of measuring the business efficiency of transport companies has become increasingly important in recent years and greatly contributes to the persistence and increase of the company's competitive advantage in the market. The paper analyzes the business efficiency of a transport company by applying the integrated PCA-DEA in a period from 2013 to 2020. As previously mentioned, efficiency is measured as the amount of output produced per input, and, accordingly, it has been defined six inputs and four outputs which are the basis for calculating efficiency using the integrated model. The advantage of the DEA model is the measurement of efficiency considering many different input-output parameters, and the possibility of identifying efficient or inefficient decision-making units. In the paper, it is also used the PCA model by which initial variables have been transformed into new principal components that contain 90% of the total variance, and thus increase the strength, i.e. discrimination power of the DEA model. It is proved by the fact that after applying the DEA model, all eight decision-making units show efficiency, while after applying the integrated PCA-DEA model, which is based on three principal input components and two principal output components, the eighth decision-making unit shows inefficiency. The PCA-DEA model based on two input components and two output components shows the efficiency of four decision-making units and finally the model based on two principal input components and one output component shows efficiency only in 2013 and 2019. The main goal, as well as contribution, of this research is to determine the efficiency of the transport company in a longer period of time and to identify efficient or inefficient years of business performance. The above can greatly contribute to improving the efficiency of the overall business performance of the company, and ultimately increase the quality of transport services and competitiveness in the market.

REFERENCES

1. Adler, N., & Golany, B. (2007). Pca-Dea. *In Modeling data irregularities and structural complexities in data envelopment analysis* (pp. 139-153). Springer, Boston, MA.
2. Andrejić, M. M. (2013). *Merenje efikasnosti u logistici.* Vojnotehnički glasnik/Military Technical Courier, 61(2), 84-104.
3. Andrejić, M. M. (2015). *Modeli merenja i unapređenja efikasnosti logističkih procesa distribucije proizvoda*. Univerztet u Beogradu
4. Andrejić, M. M., & Kilibarda, M. J. (2016). *Measuring global logistics efficiency using PCA-DEA approach.* Tehnika, *71*(5), 733-740.
5. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). *Measuring the efficiency of decision making units.* European journal of operational research, 2(6), 429-444.
6. Cooper, W. W., Seiford, L. M., & Tone, K. (2006). *Introduction to data envelopment analysis and its uses: with DEA-solver software and references*. Springer Science & Business Media.
7. Davoudabadi, R., Mousavi, S. M., & Sharifi, E. (2020). *An integrated weighting and ranking model based on entropy, DEA and PCA considering two aggregation approaches for resilient supplier selection problem.* Journal of Computational Science, 40, 101074.
8. Deng, F., Xu, L., Fang, Y., Gong, Q., & Li, Z. (2020). PCA-DEA-*Tobit regression assessment with carbon emission constraints of China's logistics industry.* Journal of Cleaner Production, 271, 122548.
9. Ivanović, A. S., Macura, D. V., & Knežević, N. L. (2019). *A hybrid model for measuring the efficiency of transport infrastructure projects.* Tehnika, 74(6), 849-857.
10. Karami, S., Ghasemy Yaghin, R., & Mousazadegan, F. (2021). *Supplier selection and evaluation in the garment supply chain: an integrated DEA–PCA–VIKOR approach.* The Journal of the Textile Institute, 112(4), 578-595.
11. Layeb, S. B., Omrane, N. A., Siala, J. C., & Chaabani, D. (2020, December). *Toward a PCA-DEA based Decision Support System: A case study of a third-party logistics provider from Tunisia.* In 2020 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET) (pp. 294-299). IEEE.
12. Lozanov-Crvenković, Z. (2019). *Analiza glavnih komponenti sa primerima u statističkim softverima RI statistica.*

Contact address:
Smiljka Miškić
University of East Sarajevo, Faculty of transport and engeneering
Doboj, Bosnia and Herzegovina
Vojvode Mišića 52
E-mail: miskicsmiljka@gmail.com