

Model to Predict Survival of Transportation and Shipping Companies

Model za predviđanje opstanka prijevoznih i brodarskih kompanija

Marek Vochozka

Department of Economics and Management
Institute of Technology and Business
České Budějovice, Czech Republic
e-mail: vochozka@mail.vstecb.cz

Jarmila Straková

Department of Economics and Management
Institute of Technology and Business
České Budějovice, Czech Republic
e-mail: strakova@mail.vstecb.cz

Jan Váchal

Department of Economics and Management
Institute of Technology and Business
České Budějovice, Czech Republic
e-mail: vachal@mail.vstecb.cz

DOI 10.17818/NM/2015/SI4

UDK 656.61:338.2

Original scientific paper / Izvorni znanstveni rad
Paper accepted / Rukopis primljen: 17. 3. 2015.

Summary

The article focuses on the issue of evaluation of economic sustainability of transportation and shipping companies. The economic crisis, which is currently abating in many world's countries, has affected all industries, including transportation and shipping. This important sector of national economy with a multifunctional character has been influenced by the economic recession very significantly and its consequences are visible even now. For this reason there is a growing demand for methods allowing comprehensive evaluation of businesses. This article contributes to the topic by describing a procedure for development of a model for comprehensive evaluation of transportation and shipping companies. The model makes it possible to monitor, to evaluate and to signal financial health and performance of such companies. Finally, the model is validated and its predictive power is assessed. The proposed model is an output from research conducted by the Institute of Technology and Business in České Budějovice, Czech Republic.

KEY WORDS

transportation company
shipping company
economic sustainability
bankruptcy prediction model

Sažetak

Ovaj rad bavi se problemom procjene ekonomske održivosti prijevoznih i brodarskih tvrtki. Ekonomska kriza, koja trenutno vlada u mnogim zemljama diljem svijeta, utjecala je na sve industrijske grane, uključujući prijevoz i brodarstvo. Ovaj važni sektor nacionalne ekonomije koji ima multifunkcionalni karakter pod značajnim je utjecajem ekonomske recesije a njezine posljedice se i danas osjećaju. Zbog toga dolazi do sve veće potrebe za metodama koje bi omogućile sveobuhvatnu procjenu poslovanja. Ovaj rad doprinosi ovoj temi dajući opis postupka za razvoj modela sveobuhvatne procjene prijevoznih i brodarskih tvrtki. Model omogućuje praćenje, procjenu i signalizaciju financijskog zdravlja i izvedbe ovih tvrtki. U završnom dijelu rada, model se valorizira i utvrđuje se njegova moć predviđanja. Predloženi model nastao je kao rezultat istraživanja provedenog u Institutu tehnologije i poslovanja u Českim Budejovicama, Republika Češka.

KLJUČNE RIJEČI

prijevozna tvrtka
brodarska tvrtka
ekonomska održivost
model predviđanja bankrota

INTRODUCTION

In 2008, the economic crisis spread worldwide from the United States. Companies were getting into financial distress and started going bankrupt. Then a domino effect resulted in a situation in which companies and managers in practically all sectors had to deal with the problems. Transportation and shipping was one of the industries in which the economic crisis struck most severely. This was caused, among other factors, by the multifunctional character of this important sector of national economy and also as a result of provable negative synergic impacts of the other industries on transportation and

shipping activities. Therefore there was a significant increase of demand from those companies for models capable of predicting the ability of those companies to survive financial distress.

Existing methods, such as Altman Z-score, Neumaier's indexes, Tamari model, Taffler index, Grünwald index, Kralicek quick test, have used companies from various industries in an identical manner. They did not discern between production companies, building contractors or transportation and shipping companies. However, the implementation in the

transportation sector has shown the necessity to accentuate specific features of sectors in which the companies operate.

Therefore a hypothesis can be formulated that a model developed for transportation and shipping companies will be more accurate and its predictive power will be stronger.

ANALYSIS OF THE CURRENT STATE OF THE RESEARCHED PROBLEM

Traditional statistical methods were for many years widely used for development of one-dimensional discriminant analyses [1], [2], [3] and they made it possible to divide companies into the so-called failing and prospering. A much more widespread statistical method is a multi-dimensional discriminant analysis followed by a logit analysis [4]. Numerous classical statistical methods were developed for bankruptcy prediction and they were described particularly by the following authors: Mohd-Sulaiman [1], Garcia-Gallego & Mures-Quintana [5], Teng & Bhatia & Anwar [6], Sun & Huang & He [7], Waszkowski [8], Yazdanfar [9], De Andrés & Landajo & Lorca [10], Lin & Yeh & Lee [11] and Wu & Hsu [4]. The methods can be briefly characterized as follows:

Recent bankruptcy prediction models were developed using logit analysis, probit analysis and linear probability models (LPM). The methods were used to develop evaluation models of conditional probability [1], [12], consisting of a combination of variables with the best capability to discern between the groups of failing and prospering companies. A non-linear estimate of the maximum probability in a logistic analysis was used to estimate parameters of the following logit-model [13], [14] (Eq. 1):

$$P_i(X_i) = \frac{1}{1 + \exp^{-(B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_n X_{in})}} = \frac{1}{1 + \exp^{-(D_i)}} \quad (1)$$

where

$P_1(X_i)$ = probability of failure with regard to the vectors of attributes X_i ;

B_j = attribute coefficient j , for $j = 1, \dots, n$ and B_0 = limited section,

X_{ij} = attribute values j (for $j = 1, \dots, n$) for company i ,

D_i = "logit" for company i .

OBJECTIVE, WORK PROCEDURE AND METHODOLOGY

The objective is to propose a procedure for development of a model to predict survival of transportation and shipping companies. The testing set consisted of transportation companies and was made up of absolute and relative indicators of all companies dealing with transport in the Czech Republic (section H, classification CZ-NACE) in 2003–2012. The set of data was generated from the Albertina database and it contained 12,930 entry lines. Each line contained 150 characteristics of each of the companies (figures from the financial statement, indexes of profitability, activity, liquidity, indebtedness etc.).

The following assumptions were used to devise the new model [15-18]:

- The bankruptcy prediction model will be developed - the

dependent variables will be only 0 or 1),

- Absolute indicators will be used for the model,
- The analyzed data do not have to follow the normal distribution,
- The model development will use an iterative process repeated in cycles to ensure its further improvement,
- The analyzed group is a representative sample of the investigated population,
- The objective is to create a model that is as simple as possible and that sufficiently well explains behavior of the dependent variable. The dependent variable is binary.
- The model will feature generalizing properties,
- Quantitative variables of discrete nature (e.g. numbers of employees) will be viewed as continual variables (the employed software does not permit any better approach),
- Every model is, in its own way and to a smaller or greater extent, insufficient, inaccurate and distorting reality but some models are (or can be) more useful than others.

MATHEMATICAL MODEL OF THE PROBLEM AND ITS SOLUTION

With regard to the characteristic and specific features of transportation and shipping companies, the attention was focused on the so-called binary logistic regression.

MATHEMATICAL MODEL OF THE ASSIGNMENT PROBLEM

In this case, the value of the dependent variable can equal only two values. The dependent binary variable in our context will be defined as follows:

$$Z = \begin{cases} 1 & \text{if the company succeeds} \\ 0 & \text{if the company fails} \end{cases}$$

with the probabilities $\Pr(Z=1)=\pi$ and $\Pr(Z=0)=1-\pi$. If we have n of such mutually non-correlating quantities Z_1, Z_2, \dots, Z_n , where $\Pr(Z_j=1)=\pi$ for $j=1, 2, \dots, n$, then a random quantity Y can be defined, represented by the sum of all n random quantities Z_j , $j=1, 2, \dots, n$. The quantity Y therefore represents the total number of "successes" from n completed tests. In this case the probability density of Y , i.e. probability density function, can be expressed as follows (Eq. 2):

$$\Pr(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad y = 0, 1, 2, \dots, n \quad (2)$$

We can anticipate even more general situation in which we obtain, from experiments or observations, n of such independent values y_1, y_2, \dots, y_n . The values will respectively represent numbers of "successes" in n different groups. If we want to subsequently study relative frequencies of the "successes" in the individual groups, depending on various explanatory variables, we can use generalized linear models and determine probability π_j through the model (Eq. 3):

$$g(\pi_j) = \hat{\mathbf{a}}' \mathbf{x}_j \quad (3)$$

The symbol \mathbf{x}_j represents a column vector of explanatory variables for the j^{th} observation and $\hat{\mathbf{a}}$ is a column vector of parameters we look for. The function $g(\cdot)$ then represents the

so-called link function [14]. The simplest method to predict relative frequency of success consists in the assumption that the so-called canonical link function $g(\cdot)$ is an identical function. This will result in a classical linear model in the form $\pi_i = \beta'x_i$. However, it is not very suitable and one can use a distribution function where [17], [18] (Eq. 4):

$$\pi = \int_{-\infty}^t f(s) ds \quad (4)$$

The probability density function $f(s)$ is called a tolerance distribution. When choosing a suitable tolerance distribution we can formulate e.g. logit or probit models [15]. For probability modeling it is convenient to use the logit link function and the resulting model can be then recorded as follows (Eq. 5):

$$\text{logit } \pi(x_i) = h \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \hat{\mathbf{a}}'x_i \quad (5)$$

The goal of logistic regression is to model the conditional mean value of the dependent variable y at certain values of \mathbf{x} , i.e. by means of a logistic function. This can be formally expressed as follows (Eq. 6):

$$E(y | \mathbf{x}) = \pi(\mathbf{x}) = \frac{\exp(\hat{\mathbf{a}}'x)}{1 + \exp(\hat{\mathbf{a}}'x)} \quad (6)$$

In order to estimate unknown parameters of the logistic model, identified with the symbol of β , it is possible to use the maximum likelihood method [16]. The principle of the method can be described as follows. If we consider one pair of measurements obtained by observation or experiment, i.e. (x_i, y_i) , then the contribution of information contained in this pair to the likelihood function can be expressed as follows (Eq. 7):

$$\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i} \quad (7)$$

because $y_i \sim \text{Bi}(1, \pi(x_i))$. If we further assume that the individual measurements are mutually independent then the likelihood function can be expressed as a product of all those contributions because the function is essentially a joint probability function, i.e. we will get the following likelihood function (Eq. 8):

$$l(\hat{\mathbf{a}}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i} \quad (8)$$

MODEL PROJECTION

- a) Preparation of the data set
- b) Primary screening of variables
- c) Model development

Logistic regression was used and its results were analyzed by means of significance tests of regression coefficients. Subsequently, the model was evaluated for suitability and predictive power. The tests of likelihood ratio type 1 are shown in Table No. 1; the variables were then selected again and more tests of likelihood ratio followed (see Tables 2 and 3). The resulting model to predict survival of transportation and shipping companies was then formulated as follows (Eq. 9):

$$\text{logit } \pi(x_i) = \log \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \hat{\mathbf{a}}'x_i = \quad (9)$$

$0.095064582 - 0.061965429 \times \text{equity capital in thousands CZK}_i - 0.44632997 \times \text{registered capital in thousands CZK}_i - 1.062014871 \times \text{current liabilities}_i - 0.002490906 \times \text{profit/loss from ordinary activities}_i - 0.536248625 \times \text{share of receivable on current assets in } \%_i - 0.25903599 \times \text{payment time of payables from trade in days}_i - 0.095270924 \times \text{quick test}_i + 0.194857318 \times \text{index of working capital in } \%_i + 0.205345429 \times \text{capital coefficient of added value in } \%_i - 0.149379695 \times \text{long-term liabilities in percents of liabilities in } \%_i - 0.213213898 \times \text{current liabilities in percentage of liabilities } \%_i - 0.191000612 \times \text{long-term credits and loans in percentage of liabilities in } \%_i$.

The threshold value 0.523748 was determined based on a sensitivity analyses. Transportation companies with the model value 0.523748 and higher will probably survive potential financial distress. In the opposite case the companies will experience problems and will go bankrupt.

VALIDATION OF THE MODEL ON A SPECIFIC CASE SOLUTION IN PRACTICE

The quality of the model was evaluated with the Hosmer-Lemeshow test, confusion matrix and Receiver Operating Characteristic curve (ROC).

Table 1 Quality of model fitting - Hosmer-Lemeshow test (Part 1)

Response	Bankruptcy – Quality of fitting : Hosmer-Lemeshow Test (model) Distribution : BINOMIC, Link function: LOGIT HosmerLemeshow= 3.3596, p value = 0.9098										
	Group 1a	Group 2a	Group 3a	Group 4a	Group 5a	Group 6a	Group 7a	Group 8a	Group 9a	Group 10	Row Tot.
0: Observed	19.0	9.0	4.0	3.0	2.0	3.0	1.0	0.0	0.0	0.0	41
Expected	20.0	6.2	4.2	3.4	2.7	2.1	1.5	0.7	0.1	0.0	
1: Observed	23.0	33.0	38.0	39.0	40.0	39.0	41.0	42.0	42.0	46.0	383
Expected	22.0	35.8	37.8	38.6	39.3	39.9	40.5	41.3	41.9	46.0	
All groups	42.0	42	42	42.0	42.0	42.0	42.0	42.0	42.0	46.0	424

Source: authors

Table 2 Quality of model fitting - Hosmer-Lemeshow test (Part 2)

Stat.	Bankruptcy-Statistics of fitting quality (model) Distribution : BINOMIC, Link function: LOGIT Modeled probability that bankruptcy = 0		
	SV	Stat.	Stat/sv
Deviation	405	187.9935	0,464182
Deviance in scale	405	187.9935	0.464182
Pearson´s Chi2	405	306.8376	0.757624
Scaled P. Chi2	405	306.8376	0.757624
AIC		225.9935	
BIC		302.9385	
Cox-Snell R2		0.1748	
Nagelkerke R2		0.3717	
Log-likelihood		-93.9968	

Source: authors

The following items had the best results of likelihood ratio test type 1 (probability less than 0.05): 1. Registered capital - thous.CZK, 2. Payables from trade, 3. Current liabilities total, 4. Financial profit/loss - thous. CZK, 5. Profit/loss from ordinary activities - thous. CZK, 6. Share of receivables on current assets - %, 7. Repayment time of payables from trade – days, 8. Payroll expenses per employee - thous. CZK/month, 9. Quick test, 10. Working capital index - %, 11. Capital coefficient of added value - %, 12. Other assets in % of assets - %, 13. Current liabilities in % of liabilities - %, 14. Bank credits and loans in % of liabilities - %.

Likelihood ratio test type 3 is following. Similarity of both types of likelihood ratio tests are significant. Only a few items are different: 1. Equity capital - thous. CZK, 2. Registered capital - thous. CZK, 3. Payable from trade, 4. Current liabilities total, 5. Profit/loss from ordinary activities - thous. CZK, 6. Payroll expenses per employee - thous. CZK/month, 7. Quick test, 8. Capital coefficient of added value - %, 9. Other assets in % of assets - %, 10. Long-term liabilities in % of liabilities - %, 11. Current liabilities in % of liabilities - %, 12 Bank credits and loans in % of liabilities - %.

The results of the Hosmer-Lemeshow test and the other characteristics indicate suitability and information capability of the proposed model for the transportation and shipping sector. The confusion matrix for training data is provided in Table No. 3.

Table 3 Confusion matrix for training data

Observation	Predicted YES	Predicted NO	Percentage of correct predictions
YES	379	4	98.9 %
NO	31	10	24.3 %

Source: authors

Internal elements of the confusion matrix – contingency tables have their own names: a)TP – True positive (Observed YES; Predicted YES) number of correctly classified cases of bankruptcy. b) FP – False positive (Observed YES; Predicted NO) number of incorrectly classified cases of bankruptcy.c.)TN – True negative (Observed NO; Predicted NO) number of correctly classified cases, in which the company is OK. d) FN – False negative (Observed NO; Predicted YES) number of incorrectly classified cases, in which the company is OK.

The elements are used to establish relative indicators of effectiveness of the classification. The first, and probably

the most important, indicator is the so-called accuracy of classification, which is defined as a ratio of sums of absolute frequencies on the main diagonal to the total number of the classified ones. This can be formally expressed as follows (Eq. 10):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

As implied by the Table No. 3, the ratio in the case of our model is:

$$(379+10)/(379+4+31+10)=0.917453, \text{ i.e. } 91.7453 \%$$

The confusion matrix obtained based on a set of training data provides a slightly deflected estimate of efficiency of the developed model and therefore it is necessary to use a set of validation data. The confusion matrix for the set of validation data is provided in Table No. 4, which shows that the estimated efficiency of classification with our regression model is 90.9 %.

Table 4 Confusion matrix for the validation data

OBSERVATION	Predicted YES	Predicted NO	Percentage of correct predictions
YES	1606	53	96.8 %
NO	111	43	27.9 %

Source: authors

RECEIVER OPERATING CHARACTERISTIC

Another possibility to evaluate predictive power of our model is the so-called Receiver Operating Characteristic (ROC) curve.

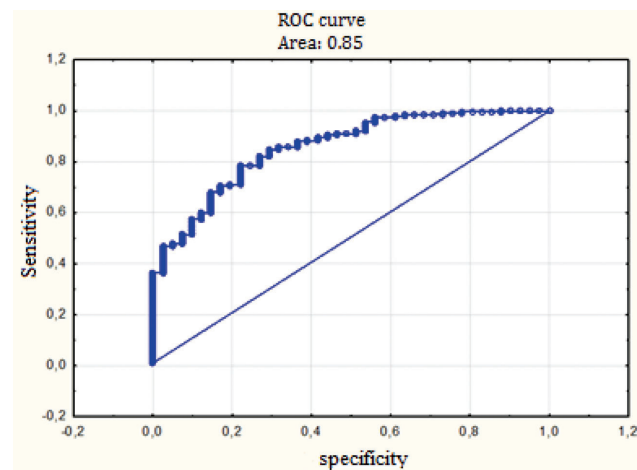


Figure 1 Diagram of the Receiver Operating Characteristic (ROC) curve for the training data set

Source: authors

The ROC curve of a model with zero prediction ability is a line segment connecting points [0;0] and [1;1].

Predictive power of the model is to a certain extent affected also by the selected threshold value and therefore the threshold shall be selected for which the respective point on the ROC curve is as close as possible to the point [0;1]. When looking at the diagram we can conclude that the ROC curve and the confusion matrix indicate the same and that the efficiency of the model is 91.7453 %.

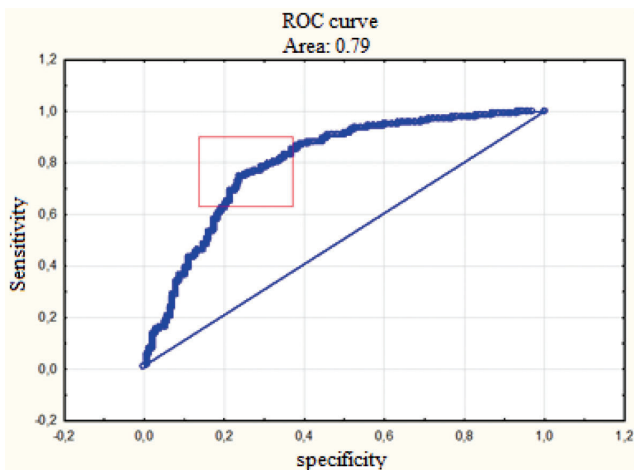


Figure 2 Diagram of the Receiver Operating Characteristic (ROC) curve for the validation data set

Source: authors

ASSESSING THE BENEFITS OF THE SOLUTIONS FOR TRANSPORT PRACTICE AND SCIENTIFIC KNOWLEDGE

The structure of the proposed model for transportation and shipping companies respects their specific features and differences from other industries of the national economy. This has been reflected in the efficiency of the developed model which is nearly 92 %. In agreement with the achieved results we can conclude that the proposed model is able to predict financial development of transportation and shipping companies in the Czech Republic [16-19].

To a certain extent the model offers a methodical procedure for its potential implementation in other countries. This is not to say that the model presented herein is applicable worldwide, however, the proposed procedure for its structure definitely is. If predictive power obtained with a model is greater than 50 % we can conclude that the model is valid. The developed methodology can be used to devise a model applicable worldwide for prediction of survival of transportation and shipping companies while taking into account that predictive power of such a model will be significantly limited by the selected set of data.

CONCLUSION

The completed research of specialized literature has shown that no method has been available to quickly predict survival of transportation and shipping companies. The solution respected the formulated hypothesis that a model taking into account specific features and characteristics of transportation and shipping companies will be more accurate and its predictive power will be higher. Our results have shown that the new model is highly efficient, it has high predictive power and therefore it is valid. When developing the model we used binary logistic regression, we analyzed a set of data from Czech transportation and shipping companies and we defined assumptions used for model development.

The selected methodology was used to find a model with high predictive power of nearly 92 % in comparison with the

originally established assumption requiring predictive power greater than 50 %. We can nearly certainly conclude that the proposed model is able to predict survival of companies in the Czech transportation and shipping sector. Outputs from the model can be used both by managements of the concerned companies, as high-level indicators for company management, and by other stakeholders, i.e. creditors, owners, competitors and others [18]-[20].

A challenge to be addressed in the future is to develop similar models for individual countries in Europe and worldwide, which need to take into account national, regional and local conditions. However, it would be much more complicated to devise a model applicable worldwide which would overcome specific geographic features of transportation and shipping companies. Such a solution would require quality selection of the data set.

REFERENCES

- [1] Mohd-Sulaiman, A. N. *Financial reporting failures, board's competency and effectiveness*. International Journal of Disclosure and Governance. Vol. 10 (2), pp. 155-174. 2013.
- [2] HO, C.-Y., McCarthy, P., Yang, Y., YE, X. *Bankruptcy in the pulp and paper industry: Market's reaction and prediction*. Empirical Economics. Vol. 45 (3), pp. 1205-1232. 2013.
- [3] Korol, T. *Early warning models against bankruptcy risk for Central European and Latin American enterprises*. Economic Modelling. Vol. 3 (1), pp. 22-30. 2013.
- [4] WU, T.-C., HSU, M.-F. *Credit risk assessment and decision making by a fusion approach*. Knowledge-Based Systems. Vol. 35, pp. 102-110. 2012.
- [5] Garcia-Gallego, A., Mures-Quintana, M.-J. *Business failure prediction models: Finding the connection between their results and the sampling method*. Economic Computation and Economic Cybernetics Studies and Research. Vol. 7. 2012.
- [6] Teng, H. S. S., Bhatia, G. S., Anwar, S. *A success versus failure prediction model for small businesses in Singapore*. American Journal of Business. Vol. 26 (1), pp. 50-64. 2011.
- [7] Sun, J., Li, H., Huang, Q.-H., He, K.-Y. *Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches*. Knowledge-Based Systems. Vol. 57, pp. 41-56. 2014.
- [8] Waszkowski, A. *Methods of classification models for enterprises insolvency prediction*. ActaScientiarumPolonorum. Oeconomia. Vol. 10 (2), pp. 95-106. 2011.
- [9] Yazdanfar, D. *Predicting bankruptcy among SMEs: Evidence from Swedish firm-level data*. International Journal of Entrepreneurship and Small Business. Vol. 14 (4), pp. 551-565. 2011.
- [10] De Andrés, J., Landajo, M., Lorca, P. *Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios*. Knowledge-Based Systems. Vol. 30, pp.67-77. 2012.
- [11] Lin, F., Yeh, C.-C., Lee, M.-Y. *A hybrid business failure prediction model using locally linear embedding and support vector machines*. Romanian Journal of Economic Forecasting. Vol. 16 (1), pp. 82-97. 2013.
- [12] Doumpos, M., Zopoudinis, C. *A multicriteria discrimination method for the prediction of financial distress: the case of Greece*. Multinational Finance Journal. Vol. 3 (2), pp. 71-101. 1999.
- [13] Hosmer, D. W., Lemeshow, S. *Applied logistic regression*. 2nd ed. New York: John Wiley; 2000.
- [14] Gujarati, D. N. *Basic Econometrics*. 4th edition. London: McGraw Hill; 2003.
- [15] Dobson, A. J. *An introduction to generalized linear Models*. 2nd edition. London: Chapman & Hall/CRC; 2002.
- [16] Meloun, M., Militký, J., Hill, M. *Computer analysis of multivariate data in the examples*. 1st edition. Prague: Academia; 2005.
- [17] Stopka, O., Kampf, R., Kolář, J., Kubasáková, I., Savage, C. H. *Draft guidelines for the allocation of public logistics centres of international importance*. Communications. Vol.16 (2), pp. 14-19. ISSN 1335-4205. 2014.
- [18] Kubasáková, I., Kampf, R., Stopka, O. *Logistics information and communication technology*. Communications. Vol.16 (2), pp. 9-13. ISSN 1335-4205. 2014.
- [19] Krile, S. *Efficient Heuristic for Non-linear Transportation Problem on the Route with Multiple Ports*, Polish Maritime Research, Gdansk, Poland, Vol. 20 (4), pp. 80-86, ISSN 1233-2585. 2013.
- [20] Krile, S., Krile, M., Prusa, P., *Non-Linear Mimimax Problem For Multi-Stop Flight Routes*, Transport, Vilnius, Vol. 30 (3), pp.361-371, ISSN 1648-4142. 2015.