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.vscteb.cz

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

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

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.

INTRODUCTION
In 2008, the economic crisis spread worldwide from the Unite 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):

$$P (X_i) = \frac{1}{1 + \exp{-(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in})}} = \frac{1}{1 + \exp{-(Di)}} \quad (1)$$

where $P (X_i)$ = probability of failure with regard to the vectors of attributes $x_i$;

$\beta_j$ = attribute coefficient $j$ , for $j = 1, \ldots, n$ and $\beta_0 = $ limited section,

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

$Di =$ “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:

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

with the probabilities $Pr(Z=1)=\pi$ and $Pr(Z=0)=1-\pi$. If we have $n$ of such mutually non-correlating quantities $Z_1, Z_2, \ldots, Z_n$, where $Pr(Z=1)=\pi$ for $j=1,2,\ldots,n$, then a random quantity $Y$ can be defined, represented by the sum of all $n$ random quantities $Z_j, j=1,2,\ldots,n$. 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):

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

We can anticipate even more general situation in which we obtain, from experiments or observations, $n$ of such independent values $y_1, y_2, \ldots, y_n$. 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 $\pi$ through the model (Eq. 3):

$$g(\pi_i) = \hat{\beta}^T x_i$$

The symbol $x_i$ represents a column vector of explanatory variables for the $i$th observation and $\hat{\beta}$ 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 \( g(.) \) is an identical function. This will result in a classical linear model in the form \( \pi = \beta'x_i \). However, it is not very suitable and one can use a distribution function where [17], [18] (Eq. 4):

\[
\pi = \int_{-\infty}^{r} f(s)ds
\]

(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):

\[
\text{logit}(\pi(x_i)) = \beta'x_i
\]

(5)

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

\[
E(y | x) = \pi(x) = \frac{\exp(\beta'x)}{1 + \exp(\beta'x)}
\]

(6)

In order to estimate unknown parameters of the logistic model, identified with the symbol of \( \beta \), 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):

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

(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):

\[
l(\hat{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i}
\]

(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):

\[
\text{logit}(\pi(x_i)) = \log \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \hat{\beta}'x_i
\]

(9)

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>
<thead>
<tr>
<th>Response</th>
<th>Group 1a</th>
<th>Group 2a</th>
<th>Group 3a</th>
<th>Group 4a</th>
<th>Group 5a</th>
<th>Group 6a</th>
<th>Group 7a</th>
<th>Group 8a</th>
<th>Group 9a</th>
<th>Group 10</th>
<th>Row Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: Observed</td>
<td>19.0</td>
<td>9.0</td>
<td>4.0</td>
<td>3.0</td>
<td>2.0</td>
<td>3.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>41</td>
</tr>
<tr>
<td>0: Expected</td>
<td>20.0</td>
<td>6.2</td>
<td>4.2</td>
<td>3.4</td>
<td>2.7</td>
<td>2.1</td>
<td>1.5</td>
<td>0.7</td>
<td>0.1</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1: Observed</td>
<td>23.0</td>
<td>33.0</td>
<td>38.0</td>
<td>39.0</td>
<td>40.0</td>
<td>39.0</td>
<td>41.0</td>
<td>42.0</td>
<td>42.0</td>
<td>46.0</td>
<td>383</td>
</tr>
<tr>
<td>1: Expected</td>
<td>22.0</td>
<td>35.8</td>
<td>37.8</td>
<td>38.6</td>
<td>39.3</td>
<td>39.9</td>
<td>40.5</td>
<td>41.3</td>
<td>41.9</td>
<td>46.0</td>
<td></td>
</tr>
<tr>
<td>All groups</td>
<td>42.0</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>46</td>
<td>424</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Stat.</th>
<th>Bankruptcy-Statistics of fitting quality (model)</th>
<th>Modeled probability that bankruptcy = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviation</td>
<td>SV 187.9935</td>
<td>0.464182</td>
</tr>
<tr>
<td>Deviance in scale</td>
<td>405</td>
<td>0.757624</td>
</tr>
<tr>
<td>Pearson’s Chi2</td>
<td>405</td>
<td>0.757624</td>
</tr>
<tr>
<td>Scaled P. Chi2</td>
<td>405</td>
<td>0.757624</td>
</tr>
<tr>
<td>AIC</td>
<td>225.9935</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>302.9385</td>
<td></td>
</tr>
<tr>
<td>Cox-Snell R2</td>
<td>0.1748</td>
<td></td>
</tr>
<tr>
<td>Nagelkerke R2</td>
<td>0.3717</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-93.9968</td>
<td></td>
</tr>
</tbody>
</table>

Source: authors


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.

![Diagram of the Receiver Operating Characteristic (ROC) curve for the training data set](image)

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](image)

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 managers 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