LOGISTIC REGRESSION AS A TOOL FOR DETERMINATION OF THE PROBABILITY OF DEFAULT FOR ENTERPRISES

Erika SPUCHLAKOVA[1], Maria KOVACOVA[2]

[1] Faculty of Operation and Economics of Transport and Communications, University of Zilina, Slovak Republic, erika.spuchlakova@fpedas.uniza.sk
[2] Faculty of Operation and Economics of Transport and Communications, University of Zilina, Slovak Republic, maria.kovacova@fpedas.uniza.sk

***Abstract**: In a rapidly changing world it is necessary to adapt to new conditions. From a day to day approaches can vary. For the proper management of the company it is essential to know the financial situation. Assessment of the company financial health can be carried out by financial analysis which provides a number of methods how to evaluate the company financial health. Analysis indicators are often included in the company assessment, in obtaining bank loans and other financial resources to ensure the functioning of the company. As company focuses on the future and its planning, it is essential to forecast the future financial situation. According to the results of company´s financial health prediction, the company decides on the extension or limitation of its business. It depends mainly on the capabilities of company´s management how they will use information obtained from financial analysis in practice. The findings of logistic regression methods were published firstly in the 60s, as an alternative to the least squares method. The essence of logistic regression is to determine the relationship between being explained (dependent) variable and explanatory (independent) variables. The basic principle of this static method is based on the regression analysis, but unlike linear regression, it can predict the probability of a phenomenon that has occurred or not. The aim of this paper is to determine the probability of bankruptcy enterprises.*

Keywords*:* Enterprise; Logistic regression; Probability of default.

JEL Classification Codes: G21, G24, G28.

1. INTRODUCTION

Logistic regression (also LOGIT) is the name for a regression model with a binary (dichotomous) dependent variable. It was designed in the 60s of the 20th century as an alternative to the least squares method. Previously concerned the most jobs by logistic regression, particularly in medicine and epidemiology. Explained (dependent) variable is e.g. the presence or absence of disease. Logistic regression model allows e.g. the risk of heart disease as a function of a number of anthropometric and biochemical parameters (gender, age, BMI, blood and pressure).

In the industry we can follow the success or failure of a product and logistic regression can help to determine which variables are significantly involved in this success. In the banking sector LOGIT is used to create a model that can estimate banks' clients applying for a loan on the basis of a series of parameters (e.g. age, sex, and educational attainment), whether they will repay the loan properly or not. Logistic regression is also an alternative to discriminant analysis and analysis of mixtures normal distribution. (Kliestik & Majerova, 2015)

Logistic regression differs from a linear regression mainly in one factor. Logit predicts the probability of an event to happen or not to happen. It uses a logit transformation to create conditions for detention between this probability and linear predictor consisting of independent variables $X_1, \ldots, X_n$. The difference between logistic and linear regression is that the logistic regression uses categorical dependent metamorphosis, whereas linear regression consists continuous explained variable. The nature of logistic regression is the modeling of the relationship between explained (dependent) variable and explanatory (independent) variables.

2. METHODOLOGY OF LOGISTIC REGRESSION

Response (dependent) variable is a binary variable, which means it can take only the values 0 and 1. Logistic regression based on generalized linear models, in which by means of so-called logit transformation logit capturing the flow, and therefore the relationship between dependent variable y and vector of independent variables x. A prerequisite of classical linear model is that explained variable y is quantitative continuous, whereas the logistic regression works with categorical response variable.

Logistic regression based on the natural logarithm of chance or hope. Explained variable y represents the probability π with which acquires a value of 1 (occurrence of reporting effect), which can be expressed as P(Y=1) = π. The probability that explained variable y acquires a value of 0 (do not occurrence of reporting effect) can be described as 1 - π.

Odds is then formulated as the ratio of these obtained probabilities. Logarithm of chance for a binary variable is indicated by the term logit. Probability of function alternative distribution has the form:

$$P(y|\pi) = \pi^y (1-\pi)^{1-y} \left[\pi/(1-\pi)\right]^y = (1-\pi)\exp\left[y\ln\left(\frac{\pi}{1-\pi}\right)\right] \qquad (1)$$

It is the fact that π ∈ (0,1), quotient $\left(\frac{\pi}{1-\pi}\right)$ can take any non-negative values from the interval (0, ∞) and the logit $\ln\left(\frac{\pi}{1-\pi}\right)$ can take any real number from the interval (-∞, ∞).

In case of using logit transformation π in a generalized linear model occurs the regression model (with the explanatory variables), which has the following form:

$$g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = x\beta \qquad (2)$$

for $x = [1, x_1, x_2, \ldots, x_k]$, $\beta = [\beta_0, \beta_1, \ldots, \beta_k]$ where $\beta_0, \beta_1, \ldots, \beta_k$ are the model parameters and π is the conditional mean of the response variable. (Hebák, 2007)

Within the logistic regression it is possible to distinguish the binary logistic regression, ordinal and nominal logistic regression.

1. Binary logistic regression - refers to the binary dependent variable, which have only two possible values, e.g. the absence and presence of the phenomenon.

2. Ordinal logistic regression - the dependent variable is ordinal variable type, acquiring more possible condition, among which there are natural arrangement, e.g. Stage of severity of disease, the type of questionnaire responses with possible answers at all, probably not, probably yes, definitely yes.
3. (Multi) nominal logistic regression - refers to the nominal dependent variable by more than two levels of the state, among which there are only differences, e.g. eye color, race.

Similarly, to the linear regression, the vector of explanatory variables for all three logistic regressions may comprise several variables, both continuous, called predictors are categorical, called factors.

2.1. BINARY LOGISTIC REGRESSION

If we consider the binary response variable Y and one or more explanatory (independent) variables, then, the binary variable takes the value 1 with probability π and the value 0 with probability $\pi - 1$. Property of probability π explanation and explanatory variables is limited interval of values (0, 1).

Explaining variable Y has Bernoulli distribution with parameter πi if the vector $x_i = [x_{i1}, x_{i2}, \ldots, x_{ik}]$, for i= 1, 2, …, n is the i-th combination values for nonrandom explanatory variables X_1, X_2, …, X_k. Probability of alternative distribution function can therefore be included by the aforementioned equation (1).

In the case that for different vectors x_i of conditional by probability of vector Y is equal, then is the explanation variable Y independent of these variables. If the probability of distribution of vector Y is different, then it is possible to talk about any type of addiction, which can be represented by the regression model. As it was mentioned previously, using the logit transformation π there is the logistic regression function in the regression function, see the equation (2). From this equation it is indicate that:

$$\frac{\pi}{1-\pi} = e^{x\beta} \tag{3}$$

Then, it is possible to express the logistic model which has the following form:

$$\pi = \frac{e^{x\beta}}{1+e^{x\beta}} = \left[1+e^{-x\beta}\right]^{-1} \tag{4}$$

Obviously, the probability π is expressed as a nonlinear function K to the explanatory variables. If we applied that K = 1, then the graph of the nonlinear function is symmetrical S-curve, which is also the graph of the distribution function of normal and logistic distribution. Asymptote of the curve are parallel to the horizontal axis and intersect the vertical axis at points 0 and 1. (Bartosova, 2015)

$$F(x) = \frac{e^{(x-a)/b}}{1+e^{(x-a)/b}}, -\infty < x < \infty \tag{5}$$

Where π is pi, and the parameter and the mean value of this distribution. If we consider that:

$$\beta_1 = \frac{1}{b} \; and \; \beta_0 = \frac{-a}{b} \qquad (6)$$

Then the distribution function can be written as:

$$F(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \qquad (7)$$

when the value of $\beta_0 + \beta_1 x$ $\beta_0 + \beta_1 x$ (logit) is 100π- percent quantile of logistic division. In the event that $\beta_1 > 0$, the graph of the distribution function is symmetrical rising s-curve. As can be seen base on equation (5) the distribution function is defined in the field of real numbers and probability π can take values in a limited range (0,1). In the event that $\beta_1 < 0$, then it is a symmetrical downward s-curve, and it isn´t the graph of the distribution function. (Bianco & Martinez, 2009)

2.2. OHLSON MODEL

Logistic regression was firstly applied for bankruptcy prediction by Ohlson (1980). He assessed 105 bankrupt and 2058 non-bankrupt American companies and proposed a model which can be applied for the prediction of bankruptcy. The model formula is following:

$$\begin{aligned} O = & -1{,}32 - 0{,}407X_1 + 6{,}03X_2 - 1{,}43X_3 - 0{,}0757X_4 - 2{,}37X_5 \\ & -1{,}83X_6 + 0{,}285X_7 - 1{,}72X_8 - 0{,}521X_9 \end{aligned} \qquad (8)$$

where:

X_1 = log Total Assests / Gross National Product price index level
X_2 = Total Liabilities / Total Assests
X_3 = Working Capital / Total Assests
X_4 = Current Liabilities / Current assets
X_5 = Net Income / Total Assests
X_6 = EBITDA / Total Liabilities
X_7 = 1 if a net loss for the last two years, 0 otherwise
X_8 = 1 if Total Liabilites > Total Assets, 0 otherwise
$X_9 = \dfrac{EAT_t - EAT_{t-1}}{|EAT_t| + |EAT_{t-1}|}$

The model contains nine explanatory variables and the probability of bankruptcy is given by the formula:

$$p = \frac{e^O}{1 + e^O} \qquad (9)$$

3. RESULTS AND DISCUSSION

The main approach for the further development of models to quantify such prediction was learning relationship between standard fixed and variable data using statistical models. In both, practice and academic studies, statistical models based on multivariate analysis, discriminant, logistic regression and neural networks were used to predict corporate bankruptcy (Wilson and Sharda 1996. Lee et al, 2005).

The corporate bankruptcy and logistic regression has a beautiful probabilistic interpretation, because the output is between 0 and 1. However, there are two main problems of logistic regression due to the nature of the problem of bankruptcy proceedings. First, Booth (1982) argues that the bankrupt company could be regarded as outliers in terms of a healthy society. In a given year, the number of corporate bankruptcies is small relative to the total number of publicly traded companies. In the event of bankruptcy of the extreme values, it is a fundamental violation of basic requirements for the distribution logistic regression.

The second problem, which occurs in the logistic regression in bankruptcy issue is the interpretation of the model correctly predicted percentage (or secret). As pointed out by Kliestik et. al. (2014), the percentage of correctly predicted is given by goodness-of-fit measure, but it can be very misleading. Percentage predicted correctly can be very misleading if the relative ratios of results are great. Furthermore, it's true, that corporate bankruptcy is usually very low because of the relative incidence of bankruptcy. So in the case of corporate bankruptcy, it is possible to obtain a relatively high percentage of correct prediction, although the bankruptcy prediction model (the least probable outcome) is very poor. Bartosova (2008) recommends that scientists should predict correctly by calculating the percentage for each result.

3.1. APPLICATION OF OHLSON MODEL ON SLOVAK COMPANIES

Based on the given model specified in previous chapter we have applied Ohlson model on the data set of Slovak companies. We have randomly chosen 500 bankrupt and 500 non-bankrupt companies and tested the prediction ability of Ohlson model which was constructed in specific conditions of United States on the data set of Slovak companies. We consider company as bankrupt if:

- negative value of earnings after taxes,
- the value of financial independence indicator is less than 0.04,
- the value of current ratio indicator is less than 1,
- company has at least two liabilities 30 days after due date from different creditors,
- the total amount of payable and not payable liabilities is higher that the value of company´s assets.

Table 1. Calculated confusion matrix of Ohlson model (Source: self-processed)

<table>
<tr><th colspan="2" rowspan="2">Ohlson model</th><th colspan="2">Predictive value</th><th rowspan="2"></th></tr>
<tr><th>0 (non-bankrupt)</th><th>1 (bankrupt)</th></tr>
<tr><td rowspan="2">Actual value</td><td>0 (non-bankrupt)</td><td>328</td><td>172</td><td>500</td></tr>
<tr><td>1 (bankrupt)</td><td>301</td><td>199</td><td>500</td></tr>
<tr><td colspan="2"></td><td>629</td><td>371</td><td>1000</td></tr>
</table>

Based on the table 1 we can calculate prediction accuracy and characteristic of the model applied on chosen data set of Slovak companies. So from gained results can be summarized that it is better to use bankruptcy prediction model which was developed in similar conditions to environment of Slovak republic or to develop new model. The prediction accuracy was only little bit more than 50%, which is quite low.

Table 2. Calculated characteristics of Ohlson models (Source: self-processed)

Model	*Ohlson model*
Type I. error	0,6020
Type II. error	0,3440
Sensitivity	0,6560
Specificity	0,3980
Model accuracy	52,70%

4. CONCLUSIONS

Prediction of corporate bankruptcy is an important and widely studied topic. Lenders and investors in companies need to know the probability of default in order to estimate the profitable business decisions. For banks, an accurate assessment of the likelihood of bankruptcy may lead to Sirens better lending practices and fair value interest rate that reflect the credit risk. However, it is necessary to predict corporate bankruptcy motions banks. For example, accounting firms may risk the courts if the auditors fail to issue an early warning, such as "concern" opinion on the issue firm. Ubiquitous in business are derivative contracts, where companies must often assess their counterparty risk. Historically, many credit or counterparty risk assessment was simply used for credit ratings issued by credit rating standard agencies. Like many investors have recently discovered this evaluation are reactive than predictive. For this reason, there is a great need to develop precise quantitative models to predict corporate bankruptcy.

ACKNOWLEDGEMENT

The contribution is an output of the scientific project VEGA 1/0428/17 Creation of new paradigms of financial management at the threshold of the 21st century in conditions of the Slovak republic.

REFERENCES

1. Adamko, P., Kliestik, T. *Proposal for a Bankruptcy Prediction Model with Modified Definition of Bankruptcy for Slovak Companies*, RSEP International Conferences on Social Issues and Economic Studies 2nd Multidisciplinary Conference, Madrid, Spain, pp. 1-7, 2016
2. Adamko, P.; Kliestik, T.; Birtus, M. *History of Credit Risk Models*, 2nd International Conference on Economics and Social Science, Advances in Education Research, Shenzhen, China, vol. 61, pp. 148-153, 2014
3. Bartosova V. *Financial analysis and planning*, Zilina, EDIS Publishers, pp. 82, ISBN 978-80-554-0001-3, 2008
4. Bartosova, V. *Optimal debt level as part of enterprise financial health prediction*, International Journal of Applied Business and Economic Research, vol. 14, no. 6, pp. 3799-3806, 2016.
5. Bianco, A. M., Martinez, E. *Robust testing in the logistic regression model*, Computational Statistics and Data Analysis, vol. 53, pp. 4095-4105, 2009
6. Bianco, A. M., Yohai, V. J. *Robust Estimation in the Logistic Model, Robust Statistics, Data Analysis, and Computer Intensive Methods*, Lecture Notes in Statistics, vol. 109, Springer Verlag, New York, pp. 17- 34, 1996

7. Booth, D. *The analysis of outlying data points by robust regression: A multivariate problem bank identification model*, Decision Sciences, vol. 13, pp. 72-81, 1982

8. Buc D., Kliestik T. *Aspects of statistics in terms of financial modelling and risk*, Proceeding of the 7th International Days of Statistics and Economics, Prague, pp. 215-224, 2013,

9. Cisko S., Kliestik T. *Financny manazment podniku II*, Zilina, EDIS Publishers, 775 p., ISBN 978-80-554-0684-8, 2013

10. Hauser, R., Booth D. *CEO bonuses as studied by robust logistic regression*, Journal of Data Science, vol. 9, pp. 293-310, 2011

11. Kincl J., Dickson A. *Jak se české banky připravují na Basel III*, available online at: http://grantthornton.cz/uploads/media_items/biii-bankovnictvi.original.pdf, 2012

12. Kliestik, T., Majerova, J. *Selected issues of selection of significant variables in the prediction models*, Financial management of firms and financial institutions: 10th international scientific conference, Ostrava, Czech Republic, pp. 537-543, 2015

13. Kliestik, T.; Lyakin, A. N.; Valaskova, K. *Stochastic Calculus and Modelling in Economics and Finance*, 2nd International Conference on Economics and Social Science (ICESS), Advances in Education Research, Shenzhen, China, vol. 61, pp. 161-167, 2014.

14. Kral, P.; Janoskova, K. *Classifying and Analysing Project Resources in Project Portfolio Optimizing Process*, 18th International Conference on Transport Means, Proceedings of the International Conference, Kaunas, Lithuania, pp. 133-136, 2014

15. Lee, K., Booth, D., Alam, P. *A comparison of supervised and unsupervised neural networks in predicting bankruptcy in Korean firms*, Expert Systems with Applications, vol. 29, pp. 1-16, 2005

16. Ohlson, J. *Financial ratios and the probabilistic prediction of bankruptcy*, Journal of Accounting Research, vol. 18, no. 1, pp. 109-131, 1980

17. O'Leary, D. E. *Using neural networks to predict corporate failure*, International Journal of Intelligent Systems in Accounting, Finance and Management, vol. 7, pp. 187-197, 1998

18. Spuchlakova, E. *Credit risk monitoring*, Management innovation and business: 2nd international conference on Management innovation and business, Bangkok, Thailand, Lecture notes in management science, vol. 44, pp. 110-115, 2014

19. Svabová, L., Durica, M. *Korelačná analýza prediktorov použitých v bankrotných predikčných modeloch na Slovensku*, Ekonomicko manažerské spektrum, vol. 10, no.1, pp. 2-11, 2016

20. Wilson, R. L., Sharda, R. *Bankruptcy prediction using neural networks*, Decision Support Systems, vol. 11, pp. 545-557, 1994