COMPARING DEA AND LOGISTIC REGRESSION IN CORPORATE FINANCIAL DISTRESS PREDICTION

VIERA MENDELOVÁ, MÁRIA STACHOVÁ
Matej Bel University in Banská Bystrica, Faculty of Economics
Department of Quantitative Methods and Information Systems
Tajovského 10, 975 90 Banská Bystrica, Slovakia
e-mail: viera.mendelov@umb.sk, maria.stachova@umb.sk

Abstract

The paper focuses on different mathematical and statistical approaches to assessing financial distress of Slovak companies. Using a selected sample of large enterprise failures in the Slovak Republic, the paper examines the capability of Data Envelopment Analysis to predict financial distress of enterprises by comparing it with logistic regression. The main goal of the paper is to investigate whether sample size of data has impact on prediction accuracy of the models considered. Both models are estimated using a database that contains financial ratios and financial status of enterprise and that was obtained from the leading Slovak corporate analytical agency CRIF – Slovak Credit Bureau. The database covers economic activities of manufacturing, construction, wholesale and retail trade, repair of motor vehicles and motorcycles.
Key words: *data envelopment analysis, logistic regression, corporate financial distress.*

1. Introduction

There are many fitted models for classifying and predicting whether a firm is a potential candidate for being financially distressed or not. It has become a subject of many analysis since well-known Altman's Z-score (Altman, 1968) and its revision (Altman, 1983), through approaches based on static classification models constructed using various statistical methods, e.g. discriminant analysis, logistic regression, decision trees (Boďa and Úradníček, 2016; Balcean and Ooghe, 2006; Brezigar-Masten and Masten, 2012; Úradníček et al., 2016), as well as studies that incorporate time dynamic into these well-known static models such (Kráľ et al., 2014; Stachová et al. 2015).

In our paper, we use a selected samples of large enterprise failures in the Slovak republic to examine the capability of Data Envelopment Analysis (DEA) and Logistic regression (LR) in assessing financial distress of enterprises. The main purpose of this study is to investigate which prediction model (DEA or LR) can produce the better results in the Slovak republic and to determine whether the sample size and structure of the sample have the impact on model prediction accuracy.

This paper is organized as follows: Section 2 presents hypothesis to be tested through our investigation, Section 3 describes the data set used in the analysis and prepares a description of both DEA and LR techniques for corporate failure classification. The results from the comparative analysis of the two techniques are summarized in Section 4. Section 5 concludes this study and discusses future research extensions.

2. Hypotheses

The general hypothesis to be tested through our investigation on the DEA and LR models is derived from the results of comparisons of these two models presented in Premachandra et al. (2009).

H1: The LR model is superior to the DEA model in terms of the overall correct evaluations,

H2: The DEA model performs extremely well in correctly identifying the bankrupt firms compared to the LR model,

H3: The LR model performs extremely well in correctly identifying the non-bankrupt firms compared to the DEA model,

H4: The LR model provides better results for large samples, compared to the DEA model, which has a better ability of correct classification for small samples,

H5: The proportion of bankrupt firms in the sample does not have a major impact on the DEA results, but in the LR model this increasing proportion improves the overall correct classification.

3. Methodology

3.1. Variables Considered

Adopting the assumption of Beaver (1966) that the financial ratios are good indicators of the financial corporate distress, six financial ratios (predictors) were used in our analysis. The data sets consist of one liquidity ratio reflecting a firms´ ability to meet its obligations (X1), one activity ratio reflecting how effectively a firm utilizes its resources (X2), one leveraging ratio expressing how a firms is sustainable and risky to lend future loans (X3) and three profitability ratios reflecting a firm´s ability to generate an acceptable rate of return (X4, X5 and X6). The following formulas for the predictor variables computation were used:

X1 – total current assets / total current liabilities,
X2 – total liabilities / total sales × 360,
X3 – total liabilities / total assets,
X4 – earnings before interest and taxes (EBIT) / total assets,
X5 – earnings after taxes (EAT) / total sales,
X6 – value newly created / total sales.

3.2. Mathematical and Statistical Techniques

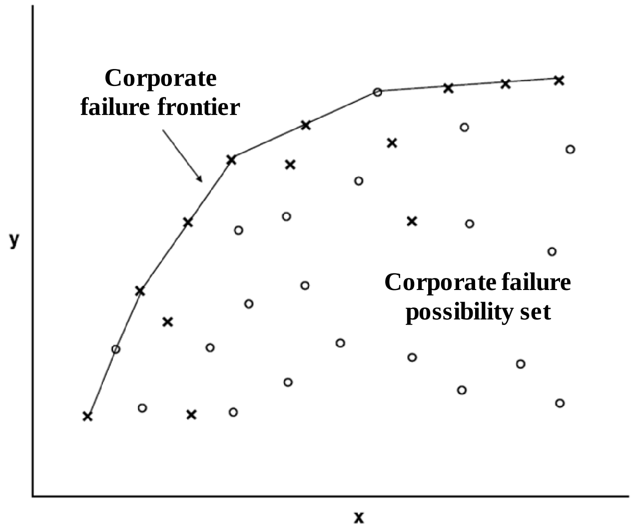
The mathematical and statistical techniques which will be used in this paper have been decided with respect to the basic methodologies followed in the original studies. Therefore, we have used the following two techniques with different characteristics and assumptions: the DEA, and the LR techniques.

3.2.1. DEA model Used for Corporate Failure Assessment

In this paper, we employ the methodology of Premachandra et al. (2009) who proposed to construct the corporate failure frontier (CFF) in the following way. Financial ratios are considered as inputs if their small values could possibly cause financial distress, and on the other hand, financial ratios are considered as outputs if their large values could possibly cause financial distress. With respect to the variables considered in our study, input variables were represented by five predictors (X1, X2, X4, X5 and X6) and output variable was represented by predictor X3.

This input-output classification identifies CFF, and indicates those firms which are about to fail. In this way, the CFF is constructed (see Figure 1) instead of the Production Possibility Frontier (PPF) that is conventionally considered in DEA.

Figure 1: Corporate failure frontier and corporate failure possibility set for one input (x) and one output (y). The symbol (○) indicates a non-default firm and the symbol (×) indicates a default firm



Source: Premachandra et al. (2009).

Since financial ratios take often negative values, the basic radial DEA models as Charnes-Cooper-Rhodes (CCR) model (Charnes et al., 1978) or Banker-Charnes-Cooper (BCC) model (Banker et al., 1984) operating on the semipositivity requirement cannot be used in these cases. For this reason, we used the additive model of Charnes et al. (1985) under the variable returns to scale conditions to discriminate healthy firms from those that are more liable for financial distress. The additive model measures efficiency of a particular firm o, $o \in \{1,...,n\}$ as follows:

$$\max_{\mathbf{s}^-,\mathbf{s}^+,\boldsymbol{\lambda}} \ \mathbf{e}'\mathbf{s}^- + \mathbf{e}'\mathbf{s}^+ \qquad \text{subject to:} \quad \begin{aligned} &\mathbf{s}^- = \mathbf{x}_o - \mathbf{X}\boldsymbol{\lambda}, \\ &\mathbf{s}^+ = \mathbf{Y}\boldsymbol{\lambda} - \mathbf{y}_o, \\ &\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{e}'\boldsymbol{\lambda} = 1, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}, \end{aligned} \tag{1}$$

where, n is the number of firms under consideration, m is the number of inputs, s is the number of outputs, $\mathbf{X}$ denotes a input matrix, $\mathbf{Y}$ denotes a output matrix, $\mathbf{e}'$ is a row vector with all elements equal to 1, $\mathbf{x}_o$ is a column vector of m inputs of the firm o, $\mathbf{y}_o$ is a column vector of s outputs of the firm o, $\mathbf{s}^-$ is a vector of m input slacks (excesses) of the firm o, $\mathbf{s}^+$ is a vector of s output slacks (shortfalls) of the firm o and $\boldsymbol{\lambda} \in R^n$ is an intensity variable vector connecting inputs and outputs.

Let $(\mathbf{s}^{-*}, \mathbf{s}^{-*}, \boldsymbol{\lambda}^*)$ be an optimal solution of (1). Then the firm o forms the CCF if and only if $\mathbf{s}^{-*} = \mathbf{0}$ and $\mathbf{s}^{+*} = \mathbf{0}$. In the context of corporate failure assessment, the firms with a high probability of their future failure tend to have a value for the objective function of the additive model (1) equal to zero, and the firms with low probability of their future failure tend to have these values greater than zero.

3.2.2. Logistic Regression Used for Corporate Failure Assessment

The second approach used in our analysis is the well-known and widely-used Logistic regression model. This method is more robust as for example discriminant analysis, because it does not assume that the independent variables are normally distributed, or they have equal variance in each group. It also does not assume linear relationship between the independent variables and depend variable and so on. The stability of logistic regression model can be negatively affected by insufficient number of data points per predictor. For more detail see (Hastie et al., 2001, Úradníček et al., 2016).

The logistic regression model can be expressed as

$$P(Y=1\,|\,x_i)=x_i^T\beta, \tag{2}$$

where Y is a binary outcome with an alternative distribution, x_i is the vector of predictor variables and β is the vector of regression coefficients.

Coefficients are computed by logarithm maximization of log-likelihood function that can be expressed for N observations as follows:

$$\sum_{i=1}^{N}\left[y_i\left(x_i^T\beta\right)-\ln\left(1+e^{x_i^T\beta}\right)\right]. \tag{3}$$

3.3. Sampling Procedure and Data Collection

The data sets for this analysis were extracted from the data base purchased from the leading Slovak corporate analytical agency CRIF – Slovak Credit Bureau, s.r.o. To take into account the differences that may exist between different sectors within the economy, only one sector was selected, i.e. economic activities 1110 – 96060 according to SK NACE Rev. 2, i.e. Manufacturing, Construction, and Wholesale and retail trade, repair of motor vehicles and motorcycles. The data set included all the four legal forms of enterprises common in the Slovak republic (i.e. v.o.s. – general partnership, k.s. – limited partnership, s.r.o. – private limited company, a.s. – joint-stock company) and related to a range of 5 fiscal periods: from 2009 – 2013. In the context of corporate financial failure classification, all variables were computed at the end of the fiscal year immediately preceding the year of corporate failure.

The original data set consists of more than 147 000 firms with 108 different financial indicators. A random sample of 2,400 firms was drawn from these 147,000 for analysis. We have detected exactly 600 failed firms satisfying all of the following necessary conditions for business failure (Boďa and Úradníček, 2016):

- its equity is negative,
- its EAT is negative.
- its current ratio attains a value lower than 1.

From these 2,400 firms, 8 sub-samples have been selected on a random basis. Each sub-sample contained 150, 300, 600 or 1,200 firms. In order to track the results of the models depending on the proportions between the failed and non-failed firms in the data set, each sub-sample contained either 5% or 25% of failed firms. The final composition of the sub-samples is presented in Table 1.

Table 1: Composition of the population and samples

Sub-samples	Total number of firms	Sample of 5% failed firms		Sample of 25% failed firms	
		Number of failed firms	Number of non-failed firms	Number of failed firms	Number of non-failed firms
Sub-sample150	150	7	143	37	113
Sub-sample300	300	15	285	75	225
Sub-sample600	600	30	570	150	450
Sub-sample1200	1,200	60	1,140	300	900

Source: the authors

4. Empirical Results and Discussion

The DEA and LR models have been constructed and the correct-classification rates have been examined separately for all sub-samples. The task of examining correct classification rates has been carried out by comparing the actual status of the firms to their predicted status according to the DEA and LR models. The classification capability of the models has been performed in terms of the Type I error, Type II error and Overall accuracy of the models.
The comparison of the classification performances of the models is based on the classification of firms into four groups (see Table 2).

Table 2: Classification of firms into four groups

Actual status	Predicted status	
	Failed	Non-failed
Failed	a	b
Non-failed	c	d

Source: the authors

The firms in groups a and d are correctly classified, while the firms in groups b and c are classified incorrectly. According to Altman (1968, p. 599), group b represents Type I error and group c represents Type II error. Let n_i, $i = a, b, c, d$ denotes the number of firms in the group i and $n = n_a + n_b + n_c + n_d$ denotes the total number of observations in the sample. Then, compute the following rates:

$$I_{CC}^{F} = \frac{n_a}{n_a + n_b}, \tag{4}$$

$$I_{IC}^{F} = \frac{n_b}{n_a + n_b}, \tag{5}$$

$$I_{IC}^{NF} = \frac{n_c}{n_c + n_d}, \tag{6}$$

$$I_{CC}^{NF} = \frac{n_d}{n_c + n_d}, \tag{7}$$

where I_{CC}^{F} refers to the percentage of the failed firms that are classified by the model as failed, i.e. the percentage of correctly classified failed firms, I_{IC}^{F} refers to the percentage of the failed firms that are classified by the model as non-failed, i.e. the percentage of incorrectly classified failed firms, I_{IC}^{NF} refers to the percentage of the non-failed firms that are classified by the model as failed, i.e. the percentage of incorrectly classified non-failed firms, and I_{CC}^{NF} refers to the percentage of the non-failed firms that are classified by the model as non-failed, i.e. the percentage of correctly classified non-failed firms. According to Altman (1968, p. 599), I_{IC}^{NF} is Type I error and I_{IC}^{NF} is Type II error.

The overall correct classification rate I_{CC} representing the overall percentage of correct classification is calculated as

$$I_{CC} = \frac{n_a + n_d}{n}, \tag{8}$$

and the overall misclassification rate I_{IC} representing the overall percentage of incorrect classification is computed as

$$I_{IC} = \frac{n_b + n_c}{n} = 1 - I_{CC}, \tag{9}$$

A higher I_{CC} (a lower I_{IC}) corresponds to a better model.

4.1. Results of the DEA and the LR Model

Results of our analysis are listed in the following tables. The Table 3 includes confusion matrices of DEA and LR models estimated on sample of 5% failed firms and the Table 4 consists of confusion matrices estimated on sample of 25% failed firms.

Table 3: Results of the DEA and the LR models for samples of 5% failed firms

Sub-sample	Actual status	Predicted status by DEA model		Predicted status by LR model	
		Failed	Non-Failed	Failed	Non-Failed
Sub-sample150	Failed	3	4	1	6
	Non-failed	14	129	1	142
Sub-sample300	Failed	3	12	3	12
	Non-Failed	13	272	3	282
Sub-sample600	Failed	3	27	1	29
	Non-Failed	17	553	3	567
Sub-sample1200	Failed	2	58	2	58
	Non-Failed	6	1,134	2	1,138

Source: the authors.

Table 4: Results of the DEA and the LR models for samples of 25% failed firms

Sub-sample	Actual status	Predicted status by DEA model		Predicted status by LR model	
		Failed	Non-Failed	Failed	Non-Failed
Sub-sample150	Failed	4	33	9	28
	Non-failed	3	110	4	109
Sub-sample300	Failed	8	67	18	57
	Non-Failed	4	221	11	214
Sub-sample600	Failed	7	143	64	86
	Non-Failed	7	443	48	402
Sub-sample1200	Failed	8	292	135	165
	Non-Failed	6	894	107	793

Source: the authors.

4.2. Comparison of the Overall Prediction Performances

The comparison of the models is carried out according to their classification performances in terms of the quantitative criteria we have mentioned at the beginning of this section. The hypotheses established in Section 2 are tested with respect to the models' quantitative criteria presented in the following Table 5.

Testing the hypothesis stated in Section 2:

H1: The LR model is superior to the DEA model in terms of the overall correct evaluations.

Results show that H1 can be accepted because the overall correct evaluations of the LR expressed in the Table 5 as an overall correct classification rate outperform the DEA model.

H2: The DEA model performs extremely well in correctly identifying the bankrupt firms compared to the LR model.

Based on results listed in the Table 5 we can see, that the LR outperform the DEA in the most cases. Thus this hypothesis cannot be accepted.

H3: The LR model performs extremely well in correctly identifying the non-bankrupt firms compared to the DEA model.

This hypothesis also cannot be accepted because results in the Table 5 shows that in the cases where models were estimated on samples of 25% failed firms the DEA achieved better results as the LR did.

H4: The LR model provides better results for large samples, compared to the DEA model, which has a better ability of correct classification for small samples.

Results show that neither one of the methods we used is dependent on the sample size.

H5: The proportion of bankrupt firms in the sample does not have a major impact on the DEA results, but in the LR model this increasing proportion improves the overall correct classification.

Our achievement are opposite to this hypothesis. The DEA results do not seem to be dependent on the proportion of bankrupt firms in the sample size, but the LR results do.

Table 5: Classification performances of the DEA and the LR models

Sub-sample	The DEA model		The LR model	
	Sample of 5% failed firms	Sample of 25% failed firms	Sample of 5% failed firms	Sample of 25% failed firms
Overall correct classification rate I_{CC}				
Sub-sample150	88.0%	76%	**95.33%**	**78.67%**
Sub-sample300	91.7%	76.3%	**95%**	**77.33%**
Sub-sample600	92.7%	75%	**94.67%**	**77.67%**
Sub-sample1200	94.7%	75.2%	**95%**	**77.33%**
Rate of correctly classified failed firms I_{CC}^{F}				
Sub-sample150	**42.86%**	10.81%	14.29%	**24.32%**
Sub-sample300	**20.00%**	10.67	**20.00%**	**24%**
Sub-sample600	**10.00%**	4.67%	3.33%	**42.67%**
Sub-sample1200	3.33%	2.67%	**3.45%**	**45%**
Rate of correctly classified non-failed firms I_{CC}^{NF}				
Sub-sample150	90.21%	**97.35%**	**99.3%**	96.46%
Sub-sample300	95.44%	**98.22%**	**98.95%**	95.11%
Sub-sample600	**97.02%**	**98.44%**	94.5%	89.33%
Sub-sample1200	99.47%	**99.33%**	**99.82%**	88.11%
Overall misclassification rate I_{IC}				
Sub-sample150	12.0%	24%	4.67%	21.33%
Sub-sample300	8.3%	23.7%	5%	22.67%
Sub-sample600	7.3%	25%	5.33%	22.33%
Sub-sample1200	5.3%	24.8%	5%	22.67%
Rate of incorrectly classified failed firms I_{IC}^{F}				
Sub-sample150	57.14%	89.19%	85.71%	75.68%
Sub-sample300	80.00%	89.33%	80%	76%
Sub-sample600	90.00%	95.33%	96.67%	57.33%
Sub-sample1200	96.67%	97.33%	96.55%	55%
Rate of incorrectly classified non-failed firms I_{IC}^{NF}				
Sub-sample150	9.79%	2.65%	0.7%	3.54%
Sub-sample300	4.56%	1.78%	1.05%	4.89%
Sub-sample600	2.98%	1.56%	5.5%	10.67%
Sub-sample1200	0.53%	0.67%	0.08%	11.89%

Source: the authors

5. Conclusion

In our study, we focus on two different approaches to assess financial distress of companies come from selected economic area in Slovak republic. We used a selected samples of large enterprise failures in the Slovak republic to examine the capability of Data Envelopment Analysis (DEA) and Logistic regression (LR) in assessing financial distress of companies. The main aim of our contribution was to investigate which prediction model (Data Envelopment Analysis or logistic regression) can produce the more accurate estimation and to determine whether the sample size and structure of the sample have the impact on model prediction accuracy. The results were confronted with hypothesis based on study Premachandra et al. (2009). Our analysis shows that from hypothesis mentioned above, only one (H1) can be accepted. In generally, we cannot say that one method is better that the other one, the accuracy and suitability of the method depends on particular data, its size and proportions.

Acknowledgements

Mária Stachová has been supported by the project VEGA 1/0647/14.

References

[1] ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. In Journal of Finance, 1968, vol. 23, iss. 4, pp. 589-609.

[2] ALTMAN, E. I. 1983. Corporate financial distress : A complete guide to predicting, avoiding, and dealing with bankruptcy. New York : Wiley, 1983.

[3] BALCAEN, S., OOGHE, H. 2006. 35 years of studies on business failure : An overview of the classic statistical methodologies and their related problems. In The British Accounting Review, 2006, vol. 38, iss 1, pp. 63–93.

[4] BANKER, R. D., CHARNES, A., COOPER, W. W. 1984. Some models for estimating technical scale inefficiencies in data envelopment analysis. In Management Science, 1984, vol. 30, iss. 9, pp. 1078-1092.

[5] BEAVER, W. H. 1966. Financial ratios as predictors of failure. In Journal of Accounting Research, 1966, vol. 4, supplement to iss. 4, pp. 71-111.

[6] BOĎA, M., ÚRADNÍČEK, V. 2016. The portability of Altman's Z-score model to predicting corporate financial distress of Slovak companies. In Technological and Economic Development of Economy, 2016, vol. 22, iss 4, pp. 532-553.

[7] BREZIGAR-MASTEN, A., MASTEN, I. 2012. CART-based selection of bankruptcy predictors for the logit model. In Expert Systems with Applications, 2012, vol. 39, iss 11, pp. 10153–10159.

[8] CHARNES, A. et al. 1985. Foundations of data envelopment analysis for Pareto Koopmans efficient empirical production functions. In Journal of Econometrics, 1985, vol. 30, iss. 1-2, pp. 91-107.

[9] CHARNES, A., COOPER, W. W., RHODES, E. 1978. Measuring the efficiency of decision making units. In European Journal of Operational Research, vol. 2, iss. 6, pp.429-444.

[10] HASTIE, T., FRIEDMAN, J. H., TIBSHIRANI, R. 2001. The elements of statistical learning, New York : Springer. 2001. ISBN-13 978-0387848570.

[11] HOLVAD, T. 2010. An examination of efficiency level variations for bus services. In International conference series on competition and ownership in land passenger transport 2001, Molde, Norway, Thredbo 7. [cit. 13-08-2016] http://hdl.ha ndle.net/2123/6320.

[12] KRÁĽ, P., STACHOVÁ, M., SOBÍŠEK, L. 2014. Utilization of repeatedly measured financial ratios in corporate financial distress prediction in Slovakia: In 17th AMSE, international scientific conference. Conference proceedings. Wrocław: Poland, 2014, pp. 156-163.

[13] PREMACHANDRA, I. M. BHABRA, G. S., SUEYOSHI, T. 2009. DEA as a tool for bankruptcy assessment : A comparative study with logistic regression technique. In European Journal of Operational Research, 2009, vol. 193, iss. 2, pp. 412-24.

[14] R CORE TEAM 2013. R : A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013, http://www.r-project.org/.

[15] STACHOVÁ, M. et al. 2015. Analysis of financial distress of Slovak companies using repeated measurement. In 18th AMSE Applications of mathematics and statistics in economics : Conference proceedings. Jindřichův Hradec 2-6 September 2015. Prague : University of Economics, 2015

[16] ÚRADNÍČEK, V. et al. 2016. Variantné metódy predikcie finančného zdravia podnikov v podmienkach dynamického ekonomického prostredia, Univerzita Mateja Bela v Banskej Bystrici : Belianum, 2016. ISBN 978-80-557-1172-0.