PREDICTING BANKRUPTCY OF POLISH MANUFACTURING ENTERPRISES – AN ALTERNATIVE MODEL BASED ON FINANCIAL RATIOS

[a]SABINA AUGUSTYN

Cracow University of Economics, Rakowicka 27, 31-510 Cracow, Poland
email: [a]sabina.augustyn@uek.krakow.pl

This paper has been financed from the funds granted to the Faculty of Management at Cracow University of Economics, as a part of the subsidy for the maintenance of research potential.

Abstract: The activity of many enterprises in the market economy is associated with the risk of bankruptcy. Bankruptcy can cause many consequences for the company itself and its environment as well, which is why many researchers keep trying to develop models for forecasting bankruptcy on the basis on various data available. This article refers to a bankruptcy forecasting model based on discriminant function analysis for industrial processing enterprises (manufacturing companies) in Poland. The model is based on the variables defined as changes in the values of financial ratios of these enterprises. A discussion of the results obtained and their comparison with the results of previous analyzes will be carried out.

Keywords: bankruptcy prediction, discriminant analysis, financial ratios, financial analysis.

1 Introduction

Bankruptcy is a phenomenon relatively common in modern economies. Bankruptcy can have major consequences for the enterprise itself and for its partners and economic environment as well. Therefore, over the years, many researchers have attempted to develop models for bankruptcy prediction. An interesting overview of the bankruptcy predicting models and methods can be found in [Aziz and Dar, 2006]. One of the oldest and still most commonly used methods of bankruptcy forecasting is the linear discriminant function. The bankruptcy prediction models are usually based on financial indicators or variables and their parameters, like mean, standard deviation, variance, logarithm, etc. [du Jardin P 2009, p.43]

One of the most important research papers on predicting the bankruptcy of companies in Poland is the project [Pociecha et. al., 2014]. An attempt was made there to build bankruptcy forecasting models for enterprises belonging to the industrial processing sector in Poland, based on data from 2005-2009. The values of 35 financial ratios were used as predictor variables in the above-mentioned analysis.

In this paper, the same database is used to construct bankruptcy prediction models, but some new variables will be introduced for this purpose. The author is aware that the data may be seen as outdated, but the choice of such database was completely purposeful. This is one of the largest datasets that were used in Poland to forecast bankruptcy an definitely the largest one when dealing with Polish manufacturing companies. The use of the new method to analyze the same dataset will bring the opportunity of a detailed comparison of the results obtained.

Predictor variables defined as yearly changes in the values of financial ratios will be used to build bankruptcy forecasting models. The author believes that information regarding not only the absolute value of financial indicators, but their relative changes may be useful in the process of predicting bankruptcy.

Based on the conducted research, the author will perform a comparative analysis of the results obtained and assess the possibility of using the proposed method in further research using the up-to-date financial data from various sectors of the economy.

2 Method

Linear discriminant analysis is a method based on the idea of using a linear function as a tool for classifying objects into one of two categories. It was first introduced in [Fisher, 1936]. The idea of discriminant analysis is to find a linear combination (transformation) of variables that best separate objects belonging to different populations.

Initially, this method was used in nature sciences, but in [Altman, 1968] the idea of predicting bankruptcy using linear discriminant function was introduced. Altman distinguished groups of bankrupt and non-bankrupt companies and used their selected financial indicators to assign a given enterprise to one of two categories. This model was called the "Z-Score model" and was based on 5 financial indicators. It allowed to estimate the probability of the company's bankruptcy within two years.

The Z-score model was based on the following financial variables:

- working capital/total assets;
- retained earnings/total assets;
- earnings before interest and taxes/total assets;
- market value of equity/book value of total liabilities;
- sales/total assets.

Over the years, many MDA models have been introduced, for example Pinches and Mingo [1973], [Beerman 1976], [Altman 1993], [Appenzaller 1998], [Morris 1998], [Altman 2000].

3 Dataset

The analysis is based on data for industrial processing enterprises (manufacturing companies) in Poland, available in EMIS Intelligence Poland database. The collected data were characterized and developed for the purposes of the research presented in [Pociecha et al., 2014].

3.1 Preliminary data

The study included companies that went bankrupt within the period of 2007 and 2010. Since the assumption was made that the attempt to predict bankruptcy will be made based on the data one year and two years ahead bankruptcy, therefore financial data for these enterprises for the years 2005 and 2009 was collected from the EMIS database. The information related to 31 variables describing the financial situation of the companies being analyzed. A single record in the database contained information on the value of all variables for a single enterprise in a given year. These records are later referred to as bankrupt or non-bankrupt and are treated as separate companies.

3.2 Missing data

The data set was scanned in search of missing data. Missing values were detected for both individual records of the companies and the financial variables being subject to analysis.

For the further research purposes, only the records that in a given year presented information on the value of at least 25 variables (80.65%) out of 31 defined were considered. The number of missing values was also determined separately in relation to all the variables divided into categories: total, bankrupt, non-bankrupt. Since the share of missing data in each category did not exceed 11%, the missing values were replaced with median values of corresponding variables.

3.3 Outliers

The next step in the analysis was to identify 35 financial indicators that were later used as variables in bankruptcy prediction models. 15 variables came directly from the EMIS database (out of 31 collected), while the next 20 were calculated based on the data collected, with respect to the theoretical background of the fundamental analysis.[1]

The set of 35 financial indicators obtained was examined in search of outliers (extreme values). The Tukey's interquartile range method of detecting outliers has been implemented. Each observation outside the range $<Q_1 - 5(Q_3 - Q_1); Q_3 + 5(Q_3 - Q_1)>$[2] has been replaced by the value of its nearest limit of the above-mentioned range. The range limits were computed separately for non-bankrupt companies, one year prior to actual bankruptcy and two years prior to bankruptcy.

3.4 Data set size

As a result, the database contained 7329 records related to 133 bankrupts and 1719 non-bankrupts. Because the records contained information about companies one and two years before going bankrupt, the database finally contained information about 182 bankrupts (2.5% of all companies) and 7147 non-bankrupts (97.5% of the total number of companies). The bankrupts population consisted of 59 enterprises for which information came from the year preceding bankruptcy and 123 enterprises for which information came from two years before filing for bankruptcy.

3.5 Distributions of financial ratios

The values of the parameters of the empirical distributions of financial ratios changed over time within each group of companies (1/ non-bankrupts, 2/ bankrupts a year prior to filing for bankruptcy 3/ bankrupts two years prior to actual bankruptcy). Moreover, the empirical distributions of financial ratios in each group of companies usually do not follow normal distribution.

3.6 Predictor variables

All the above-mentioned steps have been developed and carried out for the purposes of the analysis presented in [Pociecha et al., 2014]. And as for the present study, relative changes in the value of financial ratios in a given period compared to the previous period were determined. This action was carried out separately for each of the selected groups, i.e. for bankrupts and non-bankrupts. For example, if a value (-0.06) was determined for a given indicator in year 2008, it means that the value of a given indicator in 2008 was 8% lower than in the preceding year (i.e. 2007). The values determined this way were used as variables to build discriminant function models.

4 Assumptions

4.1 Variants of analysis

Three research variants were created:

- W1 – one-year horizon of forecasting based on data from 2005 to 2009 (for companies that went bankrupt between 2006 and 2010);
- W2 – two-year horizon of forecasting based on data from 2005 to 2008 (for companies that went bankrupt between 2007 and 2010);
- W3 – two-year horizon of forecasting based on data from 2007 only.

4.2 Selection method

When selecting the companies for research purposes two approaches were used for the study: the pair-matched sampling and the method of random sampling with replacement.

In the case of the pair-matched method for all enterprises that went bankrupt in the selected period, non-bankrupts were selected based on the same type of business activity and a similar size of the company. In addition, in the case of variants W1 and W2, the same year from which the financial data came was also taken into account. As a result, three balanced samples of enterprises were obtained.

In the case of random selection with replacement among bankrupts and non-bankrupts, a random sample was drawn of the same size for each of the test variants. These subsets were balanced, but did not include information on the type of activity, the reporting period or the size of the company.

4.3 Data split

The analysis is based on the concept of splitting the sample into two subsets: training (test) dataset and validation (holdout) dataset. This process is called cross-validation. Training sets are used to build models and estimate their parameters, while using the validation sets leads to the ability of determining forecasting properties of the models. There are different approaches to the sample distribution, see e.g. [Korol, Prusak, 2018].

In this analysis data were divided into training set and validation set in a ratio of 6:4 and 7:3.

5 Results

As a result of the analysis, 12 models were built on the basis of the following assumptions:

- 4 models per variant (W1, W2, W3);
- 6 models per sampling scheme (pair-matched, random sampling with replacement);
- 6 models per sample split (6: 4; 7: 3).

The models were evaluated on the basis of the percentage of correctly qualified cases (companies classified correctly as bankrupts or non-bankrupts). The best results are presented in the table below.

Table 1: List of discriminant function models based on changes in the value of financial indicators

Variant	Selection	Split	% of correct classif.
W1	sampling	6:4	94,12
W1	sampling	7:3	87,07
W3	sampling	6:4	83,26
W1	pair-matched	6:4	82,97
W3	sampling	7:3	82,34
W1	pair-matched	7:3	81,05
W2	sampling	6:4	77,34
W2	pair-matched	6:4	74,52
W3	pair-matched	7:3	69,23
W3	pair-matched	6:4	68,12
W2	sampling	7:3	67,05
W2	pair-matched	7:3	66,29

Source: Own calculations

As the analysis shows, the model that correctly classifies the largest part of the companies was the model based on the W1 variant, i.e. data collected a year ahead the bankruptcy was declared, and the sampling with replacement scheme. The ratio of training to validation set was 6:4. On the other hand, the model that showed the smallest ability to correctly classify companies was the model based on the W2 variant (i.e. based on data collected two years before filing for bankruptcy), and also built on the basis of a pair-matched sample and with the proportion of training and testing sets of 7:3.

Based on the results obtained, it can be stated that for the considered forecasting variants, the best results were obtained for option 1, i.e. in the case of forecasting one year before filing for bankruptcy. Of the 6 best models, as many as 4 were those based on the W1 variant. In addition, it can be seen that among the best models there no models based on the W2 variant, which

[1] The dailed information on the process of computing variables to be found in [Pociecha et al, 2014, pp.64-67]
[2] Where: Q_1 – lower quartile, Q_3 – upper quartile.

may suggest that longer forecasting horizons are not appropriate in this type of analyses.

In terms of choosing the sampling method, the random selection with replacing is definitely ahead. And in this case, out of the 6 best models, 4 were built based on this sampling plan. It can therefore be assumed that the pair-matched selection of the companies may not work properly in the case of constructing bankruptcy forecasting models, and hence the similarity criterion in terms of business activities and company size is not of great importance in the process of constructing bankruptcy prediction models.

When it comes to the use of the cross-validation method and the splitting the sample, there is no reason to unequivocally state that the proportion of the sample training and testing subsets affects the results of the prediction.

6 Discussion

The results collected here (based on the changes in the values of financial ratios) were compared with the results obtained in the process of building linear discriminant models based directly on the values of financial indicators, which were described in [Pociecha et al., 2014]. This comparison was made to determine the usefulness of the proposed approach to building discriminant functions based on the changes of the values of the same financial indicators. This list is presented below:

Table 2. The comparison of the results obtained for two types of predictor variables

Financial ratios				Changes of finacial ratios			
Variant	Selection	Split	% of correct classifications	Variant	Selection	Split	% of correct classifications
W1	SA	6:4	95,83	W1	SA	6:4	94,12
W1	SA	7:3	86,11	W1	SA	7:3	87,07
W1	PM	6:4	85,42	W3	SA	6:4	83,26
W3	SA	7:3	84,21	W1	PM	6:4	82,97
W1	PM	7:3	83,33	W3	SA	7:3	82,34
W3	SA	6:4	76	W1	PM	7:3	81,05
W2	SA	6:4	74,49	W2	SA	6:4	77,34
W2	SA	7:3	71,62	W2	PM	6:4	74,52
W3	PM	6:4	70	W3	PM	7:3	69,23
W2	PM	6:4	69,39	W3	PM	6:4	68,12
W3	PM	7:3	68,42	W2	SA	7:3	67,05
W2	PM	7:3	62,16	W2	PM	7:3	66,29

Note: SA – sampling with replacement; PM – pair-matched sample selection
Source: Own calculations and [Pociecha et al, 2014]

Based on the data presented in Table 2, it can be seen that the results of both analyzes give quite similar results. The best model was built when absolute values of financial indicators were used - 95.83% of companies were classified correctly. In the case of using variables based on relative changes in the value of financial indicators, it can be seen that the best of the developed models correctly classified 94.12% of all enterprises and it is exactly the same model, presented above.

It is worth mentioning, however, that the worst of the estimated models were classified by the surveyed enterprises in 62.16% (model based on indicators) and 66.29% (model based on changes in indicators values).

For both methods, it can also be concluded that the best way to select a sample will be random sampling with replacement, and the use of pair-matched samples in both cases gives worse results. As for the proportions of the size of teaching and validation subsets, there are no clear signals to indicate the advantage of any of the described approaches. This can also be a subject of a further analysis.

7 Conclusions

On the basis of the conducted analysis, it may be stated that constructing models of bankruptcy prediction based on a linear discriminant function can be an effective way of forecasting financial distress for many enterprises.

The above-presented results indicate that the absolute values of selected financial indicators can be used to build appropriate bankruptcy forecasting models. Relative changes in the values of the indicators can also be used as predictor variables in such analyses.

The analysis also showed that forecasting process gives the best results in the case of predicting bankruptcy one year in advance, while forecasting two years in advance does not give comparable results if the evaluation process is made on the basis of the share of correctly classified companies (both bankrupts and non-bankrupts).

Another observation made on the basis of research results is that selecting a sample should be based on random sampling with replacement rather than on purposive selection, that is the selection of non-bankrupt companies with similar size and type of activity as in the case of previously drawn bankrupts.

It is also worth emphasizing that it is not possible to clearly indicate the best proportion of the sample being divided into a training and testing subset. The results obtained for the split type of 6:4 and 7:3 did not show a significant advantage of any of these solutions.

The results presented above will make the basis for further research in this subject. The next step will be using of modified values of financial indicators to build other forecasting models, e.g. neural networks, in order to better confirm or deny the results obtained using discriminant linear models. This step will be based on the data set presented in this paper.

Another step will be to use variables based on changes in the value of financial indicators to analyze the latest available data and to examine companies without considering the category of their activity.

Literature:

1. Altman, E.I.: *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy*, New York: Wiley Finance Edition, 1993.
2. Altman, E.I.: *Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy.* The Journal of Finance: Vol. 23 Issue. 4, 1968.
3. Altman, E.I.: *Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA® Models*, http://pages.stern.nyu.edu/ealtman/Zscores.pdf, 2000.
4. Appenzaller, D.: *Mikro i makroekonomiczne przyczyny upadłości przedsiębiorstw w Polsce*. Ruch Prawniczy, Ekonomiczny i Socjologiczny: Vol.3, 1998.
5. Aziz, M.A., Dar, H.A.: *Predicting corporate bankruptcy: Where we stand?* Emerald Group Publishing Limited: Corporate Governance: Vol. 6 No.1, 2006.
6. Beerman, K.: *Prognosemoglichkeiten von Kapitalverlusten mit Hilfe von Jahresabschlusen.* Dusseldorf: IDW Verlag, 1976.
7. Du Jardin, P.: *Bankruptcy prediction models: How to choose the most relevant variables?* Bankers, Markets & Investors: issue 98, 2009.
8. Fisher, R.A.: *The Use of Multiple Measurements in Taxonomic Problems.* Annals of Eugenics: Vol. 7 Issue 2, 1936.
9. Korol, T., Prusak B.: *Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji.* III. issue. Warsaw: CeDeWu, 2018. 198 p. ISBN 978-83-8102-179-1.

10. Morris, R.: *Early Warning Indicators of Corporate Failure: A Critical Review of Previous Research and Further Empirical Evidence*, Aldershot: Ashgate Publishing Company, 1998.
11. Pinches, G., Mingo, K.A., *A Multivariate Analysis of Industrial Bond Ratings.* The Journal of Finance: Vol. 28, No. 1, 1973.
12. Pociecha, J., Pawełek, B., Baryła, M., Augustyn, S.: *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*. Cracow: Fundacja Uniwersytetu Ekonomicznego w Krakowie, 2014. 170 p. ISBN 978-83-62511-29-7.

Primary Paper Section: A

Secondary Paper Section: AH, AE, BB