Construction of a rating based on a bankruptcy prediction model

MAŁGORZATA WRZOSEK, ARKADIUSZ ZIEMBA
StatConsulting Sp. z o.o., Warsaw, Poland

Abstract

The aim of this article is to present the results of a project concerning the construction of a rating based on a bankruptcy prediction model.

The development of a tool for the effective prediction of a firms' bankruptcy has interested economists since the 1930s. The models proposed which were built during that period, and later, used new analytical techniques including the Altman model (1968) and its modifications. This model, although constructed on the basis of data describing bankrupted companies from the USA, is still used because of its universality. Many models based on "local" data were also available. Their aim is to describe the specifics of the economy in which companies operate. However, the problem in the construction of efficient bankruptcy prediction model often concerns the insufficiency or quality of the available data.

The main goal of the tasks performed was to construct a model for predicting the bankruptcy of companies, based on a sufficient and representative set of reliable data. Such prepared data was used for the purpose of the verification of the efficiency of the Altman model. The results of this project are presented later in this article.

The amount and quality of the data guaranteed the construction of reliable and correct statistical model. Information on 11 505 firms (840 of which were bankrupts) was used for the purpose of the project. The data was taken from the National Court Register and included the most up-to-date information. Financial statements with profit and loss accounts and balance sheets were also available. The information collected allowed for the indicators concerning liquidity, effectiveness and liabilities to be calculated

The analytical project resulted in the construction of highly efficient model. A rating scale was prepared allowing companies to be classified according to their risk of bankruptcy.

Available data also enabled the verification of effectiveness of the Altman model, which proved to be satisfactory, but not as good as the new model.

The whole project was conducted using information about Polish companies, however, the model constructed is versatile because of the general nature of the indicators applied.

Subsequent parts of the article cover the applied approach to the modeling of the bankruptcy of companies with reference to the methods and data previously used by other authors, a description of the data, a detailed specification of predictive variables, the results of the verification of the Altman model, the results of the construction and assessment of the quality of our own model and the ratings of the companies.

1. Introduction

Predicting the bankruptcy of companies based on financial ratios has a long history. In the 1930s, the first attempts were made in the USA to identify the significant ratios for the evaluation of the threat of bankruptcy. At that time Ramser and Foster (1931), Fitzpatrick (1932), Winakor and Smith (1935), Merwin (1942) published the results of their research based on one-dimensional analysis.

Beaver (1966) is considered to be the pioneer of bankruptcy prediction research. He applied one-dimensional discriminant analysis to 30 ratios. However, the turning point for the investigation of the problem was the use of multiple discriminant analysis for bankruptcy prediction used by Altman (1968). Altman's model (the Z-score model), which is based on 5 ratios, was published in 1968. The ratios used included 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. During the next few years Altman tried to modify his model (1984) as well as to construct alternative versions in cooperation with his colleagues (Altman, Haldeman, Narayanan, 1977). Other discriminant models were also presented (e.g. El Hennawy and Morris, 1983; Fulmer, 1984; Koh and Killough, 1990). Outside of the USA works on bankruptcy prediction using multiple discriminant analysis started to appear in the 70's. New models were published by Weinrich (1978) and Beermann (1976) for the German economy, Ko (1982) for Japanese companies or Altman and Lavallee (1981) and Legault (1987) for the Canadian economy.

In the 1980s, after Ohlson's paper (1980) had been published, logistic regression began to be used for the purpose of bankruptcy prediction. Later, some publications appeared which were dedicated to the use of AI methods (neural networks and genetic algorithms), e.g. Wilson and Sharda (1994).

Many papers dealing with bankruptcy prediction assume that the ratios used are general enough to be applied to various sectors, periods and markets. But the analyses are based on data that does not always concern homogeneous groups of companies. There are also differences in the structure of the data and its quantity as well as the period which has been analyzed prior to the bankruptcy. In one-dimensional analysis and for the purpose of the construction of multiple discriminant function the data sets often contained an equal number of insolvent and solvent companies, e.g. Fitzpatrick (20+20), Beaver (79+79), Altman (33+33), Beermann (21+21). In current approaches, the set structure is closer to reality. Ohlson, for instance, in his analysis used information about 2058 solvent companies and 105 insolvent ones.

The financial indicators used to assess the threat of bankruptcy cover all areas of financial analysis, including liquidity (e.g. Current Assets/Current Liabilities, Current Liabilities/Net Worth), leverage (e.g. Equity/Total Assets, Total Liabilities/Total Assets), profitability (e.g. Net Profit after Taxes/Sales, EBIT/Total Assets), turnover (e.g. Net Sales/Total Assets), activity (e.g. Accounts payable/Net Sales). The set of potential factors which best characterize the risk of bankruptcy can be large (for instance Beaver analyzed 30 indicators, Altman - 22). Nevertheless, in any given model only a few selected indicators are ultimately used. The set of indicators used in contemporary bankruptcy prediction models is presented in the work of Back et. al. (1996). From among 31 risk factors used in 11 models (i.a. by Beaver, Altman, Merwin) the most frequently used ones were: Working Capital/Total Assets (5 models), Current Assets/Current Liabilities (4 models), and Cash Flow/Total Debt, Operating Income/Total Assets, Quick Assets/Total Assets (3 models). Ohlson's logit model, apart from financial indicators like Total Assets/Total Liabilities, Current Assets/Current Liabilities, Operating Cash Flow/Total Liabilities, also utilizes binary variables indicating whether a loss occurred in the last two years and if the liabilities are higher than assets.

The models for economies other than the USA are also based on a similar set of indicators. One can also identify variables which occur in many models, for example, Sales Income/Total Assets (Weinrich, Springate, Legault, Altman and Lavallee, Beermann).

By using the financial reports from a couple of years prior to the bankruptcy it is possible to determine the variables which illustrate the dynamics of the financial indicators of the company. In Ko's model variables such as Inventory Turnover ratio (2 years in retrospect)/Inventory Turnover ratio (3 years in retrospect) and a net profit standard deviation from the last 4 years are used.

Analytical tools like regression or neural networks, in combination with methods of automatic variable selection, enable the models to use a very wide range of financial indicators. These indicators can also be transformed and their dynamics can be taken into account as well.

2. Acknowledgements

The data used in the project was provided by Info Veriti Polska Sp. z o. o., a leading provider of economic information in Poland, delivering on-line access to a reliable and comprehensive database of Polish and worldwide companies. Reports on the selected companies are based on a wide range of information sources such as financial statements, data on personal or capital relations between companies, statements concerning delayed payments or the press. Among the recipients of the reports generated are various departments like Marketing and Sales, HR, Finance, which use them for the purposes of marketing analysis, debt collection, contractors segmentation, the evaluation of potential cooperation with a given business partner, etc.

3. Specificity of Polish market

In Poland the problem of companies going bankrupt became relevant only in the last decade of the 20th century, after the introduction of free market economy. In the first phase of the transformation (1991-1993) the number and intensity of bankruptcies increased. The causes of these bankruptcies were frequently associated with the process of the transformation of the economy and the fact that companies were not able to adapt to the new conditions. Using the surveys carried out among bankruptcy trustees Szczerbak (2006) identified the most relevant causes of bankruptcy in the periods 1990-2000 and 2000-2004. During the first period the relevant factors were usually connected with the problems the companies had with adjusting to free market economy - such as neglecting to change the manufacturing technology in order to reduce production costs, production capacity exceeding sales or not reducing employment at the appropriate time. During the second period the factors were typical for free market economies - poor management, a negative bottom line, a lack of appropriate financial supervision and internal audit in general or high leverage (debt to equity ratio).

Models which predict the threat of bankruptcy on the basis of financial indicators for the Polish economy have been constructed since mid-nineties. Multi-dimensional discriminant analysis was used in the research conducted by Gajdka and Stos (1996, 2003), Hadasik (1998), Hołda (2001), Mączyńska (2003), among others. In subsequent years logit models as well as those based on neural networks were also created, i.e. by Michaluk (2001), Prusak (2005).

The process of transformation and a short duration of free market economy in Poland substantially affected the selection of the modeling sample. On the one hand the period available for observation was short, while on the other hand this period was unstable and the data covering longer timeframes is not homogeneous. This lack of stability also means that the models that were built on data from the early nineties are inaccurate when compared with data from later periods. The acquisition of homogeneous data is more difficult in Poland than in "older" economies. The sets analyzed are usually diversified when it comes to business status, size and sector.

The situation in Poland is that the number of bankruptcies announced is very small. This number decreased between 2003 and 2008 and was one of the lowest in Europe. According to the reports published by Creditreform, Poland has the lowest insolvency indicators (i.e. the number of bankruptcies per 10 000 companies) in Europe. In 2008 this indicator for Poland was equal to 3[1]. It is suggested that such a low bankruptcy level is due to the fact that Polish companies were well prepared to the EU accession and made good use of the opportunities that it brought. However, companies which published the reports (Coface, Creditreform) point out that the actual number of bankruptcies in Poland is much higher than the number of bankruptcies announced. In most cases the bankruptcy is not announced either because of formal errors in the bankruptcy application or because the company does not have sufficient assets to pay for the cost of the bankruptcy.

4. The approach applied to the modeling of company bankruptcy

The data was modeled using a traditional approach based on logistic regression. Logistic regression is a well recognized method for model construction in the area of credit scoring. We assume that the reader is familiar with the logistic regression approach. As an introduction to this method we recommend Hosmer D., Lemeshow S. (2000). In order to benchmark the results produced by the regression models we used a more novel approach, based on random forests. Random forests are not as popular in the area of credit scoring as traditional approaches, mainly because of their opaque and complex (practically incomprehensible) model structure. The use of random forests as a benchmark for the regression approach is motivated by the fact that this method is very robust with respect to any kind disarray in modeling and the preparation of data. This method does not always yield the best results (although quite often this is the case) but we can at least know that we are on the right path. Another advantage is the very abundant collection of potential predictable variables. The output of logistic regression provides the basis for a good interpretation of the variables entered into the model. But what about the variables that were not used? We only get one version of the truth. Slight differences in the data set could lead to a different output and interpretation (Flat – maximum effect[2]). Detailed analysis of the variable selection procedure (i.e. looking for variables entering and excluded from the model in every single step of the variable selection process) is also possible but this approach seems to be a little bit unhandy. The random forests approach uses a wide spectrum of predictable variables as a part of the model and provides a good estimate of their importance as a source of prediction. In other words, many variables are used in many places in the model. On the one hand this is a really weak point of the model because of the incomprehensibility of its structure, but on the other hand it provides an opportunity to properly

[1] During the same period the indicator for Spain was 7, for Hungary 92. It was the largest for Luxemburg - 233, while the average for Europe was about 98 and for Eastern Europe - 52.

[2] Flat-maximum effect – (or the 'curse of insensitivity') means that the predictive ability of the models is insensitive to large variations in the size of regression weights and to the number of predictors.

assess the variables, since they occur in many places within the model. This method was introduced by Breiman (2001). It combines Breiman's bagging approach and the random selection of features (Amit, Y.; Geman, D. (1997) and Ho, Tin (1995),(1997)) to grow a collection of decision trees (random forests) with a controlled level of variation. New observations which are to be classified are put down along each tree in the random forests. The trees "vote" for predefined classes by assigning the object to one of them. The forest chooses the classification that receives the most votes (among all the trees in the forest). Random forests can be based on any statistical method, but the most convenient ones are classification and regression trees.

4.1 Data description

Information about 11 505 companies was used in the project. To reach the final data set for modeling purposes, several constraints were put on the original data set. These included:

(1) Legal form. We discarded all companies which didn't provide financial statements to the National Court Register in accordance with Polish law regulations. Table 1. below presents the list of identified legal forms of the analyzed companies used for further exclusion.

Legal Form	% of population
Limited liability company	56.49
Association	13.30
Registered partnership	11.24
Cooperative	5.42
Joint-stock company	3.09
Foundation	2.37
Trade union	1.96
Association of physical culture	1.66
Agronomical organization	0.85
Limited partnership	0.70
Independent public institution of health care	0.56

Table 1: Legal form of the company

Further analysis was limited to companies with the following legal forms: joint stock company, limited liability company, registered partnership and cooperative.

(2) No turnover. We removed companies which did not report any turnover for the year that should be taken as a source of the prediction (see the definition of the target below).

(3) The level of bankruptcy in subsequent years. We ensured the same level of bankruptcy for financial statements available for modeling in every year.

(4) Sampling non-bankrupt companies, avoiding the repetition of the same company in subsequent years. Sampling without replacement was used.

(5) Time period constraints. In the project only companies that had provided financial statements for accounting years 2003 - 2006 were used.

We assumed that the process of gathering financial statements by Info Verity was free of any bias that could influence the modeling results. Additional tests were carried out to verify the hypothesis that a bivariate probit model should be constructed to avoid bias induced by the selection process. Data was divided into training and test samples in proportion 7:3. The prior probability of bankruptcy in population was estimated at 0.16%.

Table 2 describes the final data set used for modeling.

Year	Count	Count %	Target	Target Rate %
2003	1292	27.0	258	20.0
2004	1584	33.1	264	16.7
2005	1248	26.1	208	16.7
2006	660	13.8	110	16.7

Table 2: Data description

4.2 Definition of the target

We treated as bankrupt those companies for which bankruptcy procedure was started by the court in accordance with Polish bankruptcy legislation. In the case of the target definition it does not matter whether a company finally went bankrupt or not. The bankruptcy procedure can last for years without a clear outcome. On the other hand applications for instigating bankruptcy procedures (admitted by many market players) are often finally rejected by the court. For the companies for which bankruptcy procedures were instigated in the last 6 months of the calendar year we used financial statements from the previous year. For the companies where bankruptcy procedures were instigated in the first 6 month of the calendar year we used financial statement submitted one year earlier. We treated this group differently in order to guarantee a minimum six months forecasting horizon. We understand the forecasting horizon to be the time span between the moment of forecast, as determined by the financial statements used for creating the predictable variables and the moment of checking the company's status (going bankrupt or not). To simplify the terminology used in this paper whenever we use the expression “went bankrupt” we mean companies against which bankruptcy proceeding were instigated in accordance with Polish bankruptcy regulations.

4.3 Information from financial statements used for modeling purposes

We used the information available in financial statements (profit and loss accounts and balance sheet were used) as the source of potential information for modeling purposes. Financial statements are annually submitted to the National Court Register and are publicly available.

Below we present the main categories of financial indicators used as the potential predictors of the model with examples of the indicators.

I. General financial information: based directly on items from financial statements. Examples include: Total Assets, Current Assets, Liabilities, Total Revenue, Revenue on Sales.

II. Liquidity ratios: measure the availability of cash for paying the company’s debts. There are three important ratios in this category ranked according to the liquidity: Current Liquidity ratio (Current Assets/Current Liabilities), Quick ratio (Inventories + Prepayments/Current Liabilities), Operating Cash Flow ratio (Operating Cash Flow/Total Debts).

III. Profitability ratios: measure how effectively the company uses its assets and controls its expenses to generate profit. This category includes: Gross margin (Gross Profit/Net Sales), Return on Equity (ROE) (Net Income/Equity), Return on Assets (ROA) (Net Income/Total Assets), Return on Sales (ROS) (Operating Income/Net Sales). Other ratios from this group were also used.

IV. Activity ratios: measure how effectively the company’s resources are used. Examples include: Assets Turnover (Net Sales/Total Assets) or Inventory Conversion (365 days/Inventory Turnover).

V. Leverage ratios: show how high the financial leverage is in the company. This category includes among others: Debt ratio (Total Liabilities/Total Assets) or Debt Equity ratio (Long-term Debt + Value of Leases/Average Shareholders Equity).

On the basis of the financial statements eighty six predictive variables were constructed. With an additional predictor describing the sector and the legal form they comprised the set of potential predictive variables used in the selection process during the modeling phase.

4.4 Variable selection

The selection of the variables was based on: (1) exploratory data analysis (looking for one dimensional variable characteristics, checking time dependencies and stability of the predictors, etc.), (2) the predictive power of the variables (looking at p-value, standard coefficient and similar), (3) the stability of the results, (4) expert knowledge (verifying hypotheses about the usefulness of the variables in other researches, recommendations from accounting management area). Different hypotheses concerning the final set of model's predictive variables were tested. We started with extending the set of variables proposed by other authors (see pages 2 and 3) by using statistical tests of variable significance in order to include them in the models. As an alternative path, the possibility of building a model independently of the previous approaches was also explored.

4.5 The base model

Table 3 shows the results of the modeling phase. The subsequent rows of the table show the attributes of logistic regression function with their standard coefficient. Results of the hypothesis testing the significance of the parameters along with the correlation matrix are shown in Appendix.

ID	Variable Name	Variable definition	Standard coefficient
CA	Current Assets (logarithm)	= logarithm of Current Assets	+ 0.510
POS	Profit on Sale	= Profit (Losses) on Sale	– 0.366
CL	Current Liquidity	= Current assets/Current Liabilities	– 0.700
E/A	Equity/Assets ratio	= Equity/Total Assets	– 0.266
ROE	Return on Equity	= Net Profit (Loss)/Equity	+ 0.138
GROA	Gross Return on Assets	= Gross Profit (Losses)/Total Assets	– 0.185
NRFS	Net Revenue from Sales	= Net Revenue from Sales and Equivalent	– 0.360

Table 3: Variables and coefficients in the base model

CA – Current Assets are the source of day-to-day operations and can be easily liquidated in the case of bankruptcy. A high level of current assets could give an additional incentive for starting the bankruptcy procedure, because of the relatively fast cash flow expectation for shareholders and creditors alike.

POS, NFRS – these are taken directly from the balance sheet and show the effectiveness and the scale of operation separately.

CL – this indicator is extensively used in financial reporting. It measures the company's ability to pay its debts over the next year. In theory the higher current ratio the better.

E/A – measures the coverage of total assets provided by shareholders. In managerial accounting the opposite indicator is frequently used as a measure of leverage: Total Assets/Equity ratio (Total Assets divided by Shareholders Equity).

ROE – measures how much profit was generated from the available equity. To calculate ROE, we need to provide net profit for a given period (in our case one year) and the sum (level) of the company's equity at the beginning of the period (first day of the year). According to common knowledge in management accounting, the higher level of the ratio the better financial situation of the company. But high ROE ratio is not always a good predictor of company health. Some companies may have high ROE because they require no assets. The model shows a direction of ROE influence that is opposite to what could be expected. Our interpretation is as follows: High level of ROE ratio could mean that the company has achieved an abnormal leverage of equity by acting in a more risky manner (High profit = high risk). This increase the probability of bankruptcy.

GROA – Similarly to ROA (Return on Assets) it indicates how profitable the company is in relation to its assets. GROA can vary substantially within industries. It measures how effectively company could convert the money invested in its assets into gross profit.

4.6 Assessment of the model

Figure 1 presents the results of the base model quality assessment.

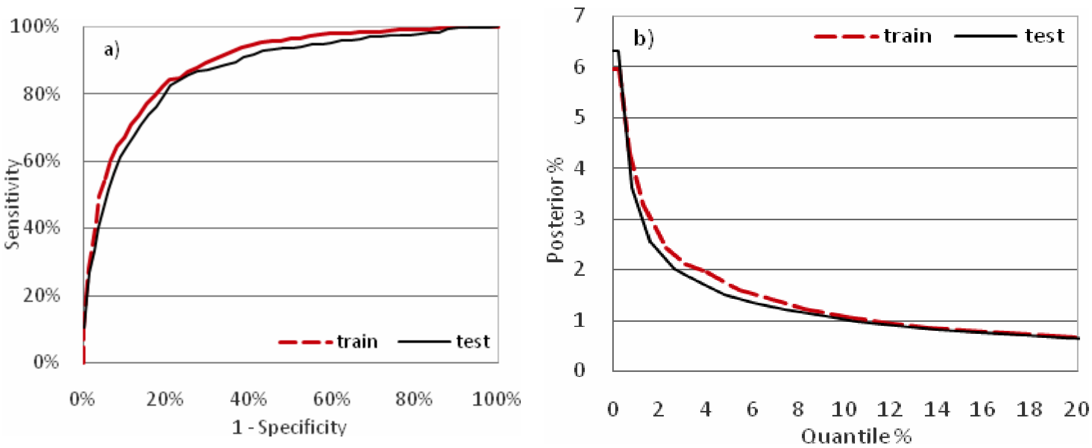

Figure 1: ROC curve (a) and Lift charts (b) for the base model[3]

The graphs show good predictive power of the model and lack of overfitting problems.

4.7 Alternative models

The base model is a representative example of winning regression models created when searching for the best results. An example of alternative logistic regression model is presented in Table 4. The results of model accuracy assessment are presented in Figure 2.

Variable	Standard Coeff
Intercept	-
L/A	0.347
C/CA	-0.350
CA	0.460
CL	-0.407
POS	-0.321
ROE	0.126
GROA	-0.170
NRFS	-0.371

[3] The Lift charts in this paper are defined on probability scale and show the leverage between the prior level of bankruptcy and the one obtained from the model. All Lift charts in this paper display only the first 20% of the population (i.e. with higher probability of bankruptcy).

Table 4: Variables and coefficients in alternative logistic regression model

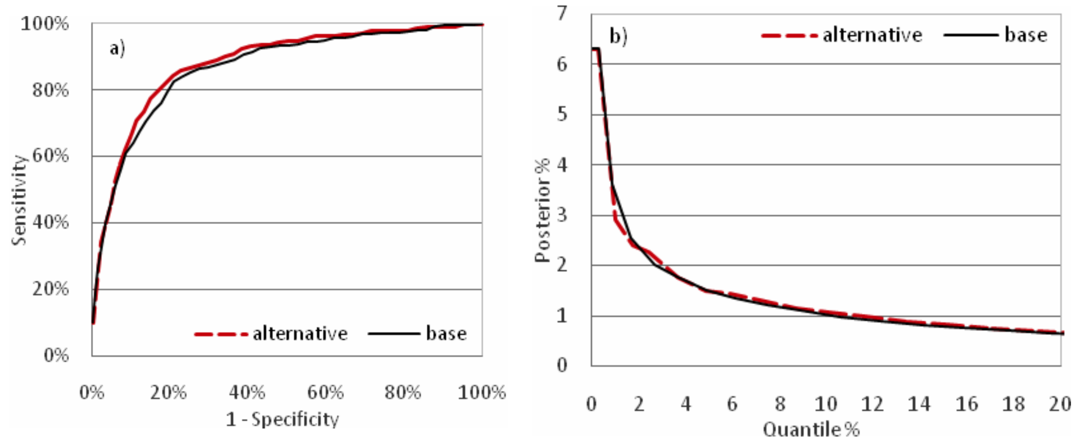

Figure 2: ROC curve (a) and Lift charts (b) for the alternative model

Two new variables (1) Liabilities to Total Assets ratio and (2) Cash to Current Liabilities ratio have been included in the model, replacing the variable E/A. Correlation between Liabilities to total assets and E/A are on the level of 0.95%. Hence, although the two models look different, the real discrepancy is not so high (flat maximum effect once again). The ROC curve looks even more promising for the alternative model but the consideration of other criteria like stability of the variables or Lift ratio compelled us to choose the previous one.

Independently, we searched for the best model provided by the random forests approach. To construct and estimate random forests we needed to settle a number of random forests parameters: number of trees (500), number of variables used as potential predictors (N=60), number of variables selected for creating splits (SQRT(N)=7,74)), maximum tree depth (20), minimal leaf size (1% = 32 observations), split evaluation statistics (Gini), measure of variable importance (Gini). The results of random forests estimation are consistent with those achieved by the logistic regression approach. The ROC curve and Lift charts estimated on the test data are presented in Figure 3. They are confronted with the base model. The conclusion is that there are not too many possibilities for improving the modeling result using the same set of predictive variables.

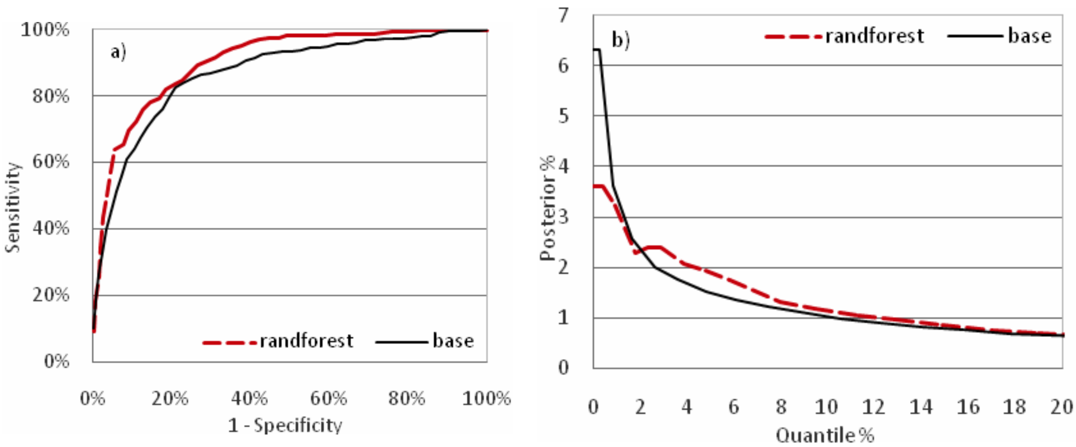

Figure 3: ROC curve (a) Lift charts (b) for random forests[4]

Table 4 in Appendix shows the summarized results describing variables that comprise the random forests model grouped by their importance. The originally obtained information about the modeling results was transformed to a more readable form by removing the variables of similar

[4] Results in this and all charts that follow are calculated for test data.

origin or meaning. To curb out the discrepancies caused by choosing an alternative measure of importance, we present results at a more aggregate level.

5. Ongoing research results (Additional Analysis)

We treated the estimation of the base model as the starting point for further research in this area. The results of this part rely on an ongoing project. We are still waiting for new data to be gathered and searching for the final solution by testing different hypotheses. However, the work results are promising enough to encourage us to include them in this paper. The hypotheses we searched for are presented in the form of the following questions:

Q_1 : How does the size of the company influence the indicators of bankruptcy in predictive modeling? We chose assets as the operational indicator of the company's size. To answer the above question we observed how the indicators change in the segments created on the basis of the size of the assets. Before deciding on the segmentation, we checked the predictor resemblance (we estimated separate models for every segments with the set of variables we had previously used for the base model) and the ROC curve results (we scored the observations in the segment using the base model and checked the fitness of the model in subsequent segments). In the first segment we put companies that did not possess any fixed assets (no current assets). Other companies (with fixed assets not equal zero) were divided into three quantiles (terciles) according to their total assets. We refer to the segments as no fixed assets, small, middle and large companies. This segmentation is not in accordance with the traditional and more formal discrimination between SME, medium and large companies based on the polish legislation, but it can be handled operationally (we did not have easy access to employment levels, which are required for the formal classification of companies). Table 5 shows the re-estimation of the base model parameters for subsequent segments.

	CA	CL	E/A	NRFS	ROE	GROA	POS
no-fix assets	0.67	-1.45	-0.20	-0.37	0.13	-0.04	-0.30
small	0.57	-0.72	-0.40	0.06	0.09	-0.18	-0.41
middle	0.34	-0.78	-0.15	-0.25	0.13	-0.36	-0.30
large	0.17	-0.32	-0.33	-0.40	0.17	-0.41	-0.23

Table 5: Re-estimation of the base model parameters for subsequent segments

Several experiments with alternative segmentation and predictive models (including the random forests approach) were carried out to support the following conclusions:

For small companies (especially with no fixed assets) liquidity indicators play a critical role. The bankruptcy is driven by the lack of liquidity. CL (Current Liquidity) ratio got the highest score in the model. The E/A (Equity/Assets) ratio showing the leverage of the company also got high score for small companies. But its relation to the company size is ambiguous . These two variables are highly correlated. The financial ratios like ROA, ROE did not turn out to be as critical as expected. We suspect that it could be the result of (1) "creative accounting", because in the case of small companies it is easier to "improve" the financial ratio and (2) relatively stronger diversity within small companies in comparison to the large ones.

For large companies liquidity is an important risk factor, but does not play a critical role. The balance shifts towards ratios measuring the effectiveness of the company. The highest score was assigned to indicators showing how profitable the company is in relation to its total assets. The most popular indicator from this group is ROA, described as Net Income divided by Total Assets. The impact of Net Revenue from Sales got a relatively high position in comparison to other predictors. High profits of companies considerably diminish the risk of bankruptcy (this can not be concluded from the model for small companies).

Q_2 : How the prediction horizon influences the predictive power of the model? To test how the predicting horizon affects scoring accuracy, additional model estimated on the financial statements from two years prior to bankruptcy was constructed. The predictive ability of this model against the base model (one year horizon) was compared. The results shown in Figure 4 indicate that the predictive power of the model diminishes when longer horizon is used for prediction but the model still has a good predictive power. We have to admit that the comparison was done during a time of relative stability in the economic environment. We noticed a rapid growth in the number of bankruptcies in the year 2009 and this could make the model more sensitive to extensions of the horizon of prediction.

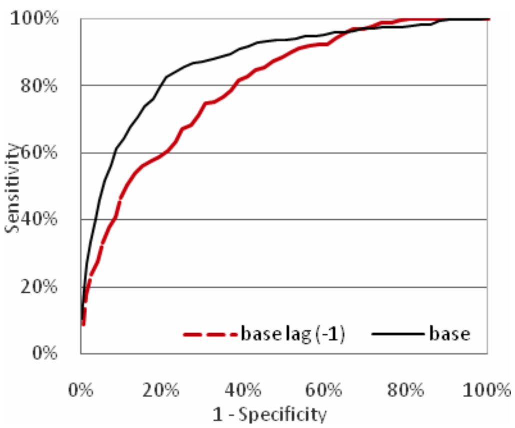

Figure 4: ROC curve for the base and new described model

Q_3: How the legal form of the company influences the indicators of bankruptcy in predictive modeling? The answer to this question is presented as the part of chapter 6 below.

Q_4: Should we benchmark financial indicators used for the model construction against their level in a specific industry? Several experiments were carried out to answer this question, but due to the ambiguous assignments of the companies to their industry according to available data (based on information that companies provided at the moment of registration) we have not managed to reach a final conclusion yet.

6. Results of the Altman model estimation for Polish market

One of the aims of the research into the assessment of bankruptcy risk was to verify the Altman's model for the Polish economy and compare the results obtained using other approaches. During the project, we used Altman's Z-score model re-estimated in 1984 for private companies not listed on the stock exchange. The analysis was carried out in two variants: using the data set used for building the base model described in chapter 4 and using the division into four categories based on the legal form of the company.

The comparison of the model over the whole data set, without considering the legal form, involved the following tasks:

- the assessment of the effectiveness of the original Z-score discriminant function built by Altman with respect to data concerning Polish companies;
- building and assessing the effectiveness of the logistic regression model consisting of five variables used in Altman's Z-score model for the data concerning Polish companies;
- the comparison of the effectiveness of the two models described above with the base model.

The table below presents the variables and coefficients used in Altman's Z-score model.

ID	Variable name	Coefficient
WC/TA	Working Capital*/Total Assets	0.717
RE/TA	Retain Earnings/Total Assets	0.847
EBIT/TA	Earnings before Interest and Taxes/Total Assets	3.10
MVE/TL	Book value of equity/Book value of Total Liabilities	0.420
S/TA	Sales/Total Assets	0.998

*Working Capital = (Current Assets-Current Liabilities)

Table 6: Variables and coefficients in Altman's Z-score model

The values of the discriminant function presented in the table were calculated for Polish companies and were subsequently used to plot the ROC curve and Lift charts[5]. In addition to using the original Altman model, a logistic regression model was also built, which contained the five indicators from the Altman model. It was assessed using the same sample as the base model. The coefficients of the assessed model are presented in the table below.

Variable	Standard Coeff
Intercept	-
WC_TA	-0.265442
RE_TA	0.2514686
EBIT_TA	-0.398451
MVE_TA	-0.461327
S_TA	-0.196571

Table 7: Logistic regression model consisting of five variables used in Altman's Z-score model

The graphs below present the ROC curve and Lift charts for the Z-score model, the regression model containing Altman's indicators and the base model described in chapter 4. The table below contains the ROC Area values for the three approaches.

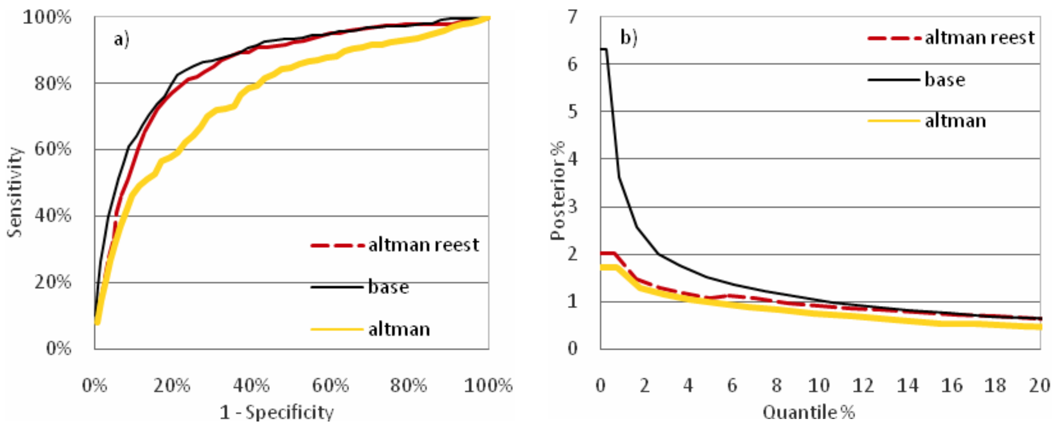

Figure 5: ROC curve (a) Lift charts (b) for the Z-score model (altman), regression model containing Altman's indicators (altman reest) and the base model (base)

Model	ROC Area
Base	0.868
Altman	0.764
Altman reest	0.847

Table 8: ROC Area values for the three models

The graphs show that the Z-score model is the least effective in the case of the data used in the comparison. Both logistic regression models (the base model and the one using Altman's indicators) have similar ROC curves. However, the comparison of the Lift charts of these two models reveals that the base model is more effective in identifying companies for which the risk of bankruptcy is the largest.

In its original version, Altman's Z-score model was constructed for companies which are listed on the stock exchange. It was subsequently modified by re-estimating the parameters for privately owned companies, however, the set of variables remained unchanged[6]. Its low effectiveness in comparison with the two other models may not suggest that it should not be applied to Polish economy, but that its effectiveness depends on the legal form of the business.. That is why in the next stage of the analysis the data set was divided according to the legal form of the company. Four categories of companies were identified - joint stock companies, limited liability companies, registered partnership and cooperative. The following models were evaluated and assessed for each category:

- Altman's Z-score model,
- logistic regression model containing five indicator's from Altman's model,
- the base model re-estimated for a given legal form of company,
- random forests model using 60 variables and 10 and 20 variables singled out as being most relevant.

The table below presents the ROC Area values obtained for different company categories and models.

The columns correspond to different legal forms of businesses: joint stock companies, registered partnerships, cooperatives and limited liability companies. The last column contains the value of ROC Area calculated for the whole set on the basis of the four models for different segments.

	Joint stock	Partnership	Coop	LLC	ROCA JOIN
Total Cases	106	147	136	1103	
Bankruptcies	29	9	6	219	
Altman’s Z-score model	0.852	0.879	0.872	0.741	0.760
Regression model containing Altman's indicators	0.896	0.899	0.876	0.821	0.847
The base model re-estimated for a given category of legal form	0.880	0.884	0.988	0.860	0.876
Random forests (~60 variables)	0.924	0.917	0.988	0.883	0.901
Random forests (best 10 variables)	0.901	0.838	0.978	0.852	0.873
Random forests (best 20 variables)	0.908	0.896	0.984	0.874	0.891

Table 9: ROC Area values for company categories and models

[6] The only change was to replace market value of equity with book value of equity.

The prediction efficiency measured as ROC Area value differs among segments for all models. The best performance for all models is in Coop segment, but the amount of bankruptcy events is very small (6). The highest value of ROCA – 0.988 – was obtained for the base model re-estimated for the Coop segment.

The worst fit for nearly all models was obtained for limited liability companies. The lowest value of ROCA – 0.741 – was obtained by the Z-score model for limited liability companies.

Among all the models, Random Forests proved to be the most effective ones. The ROCA values for these models are in nearly all the cases higher than for the remaining 3 models, independently of the segment. The random forests model built on 60 variables proved to be decisively the most effective one.

7. Company classification (rating)

The model built and the results obtained enabled us to establish a company rating based on the risk of bankruptcy. The probability of bankruptcy for a given company carries little information unless it is related to other companies. Besides, in practice, it is frequently more important to know not the given risk but how high the average risk is associated with the group that a given company belongs to and how large this group is.

Rating categories were proposed on the basis of the retrospective distribution of bankruptcies and the model statistics. The criterion for establishing the rating categories was the differentiation of bankruptcy risk in classes. In the first stage of developing the ratings, the risk classes were identified on the basis of the set used for the modeling - hence with an elevated expected risk as compared to market average risk. The table below presents the proposed risk classes together with the risk value expressed as the probability of bankruptcy.

Risk Category	Max score	Min score
6	1.0000	0.8086
5	0.8086	0.4855
4	0.4855	0.3730
3	0.3730	0.1642
2	0.1642	0.0596
1	0.0596	0.0000

Table 10: Risk classes and the probability of bankruptcy

Subsequently, the set was re-weighted to the expected market level of bankruptcy, which is equal to 0.16%. Using the re-weighted set, the risk levels for previously proposed classes were determined. The table below presents the classes created in order from the lowest to the highest risk level. The first column contains the class number. Column 2 shows the risk expressed as the bankruptcy percentage in a given class. Column 3 contains Lift values, i.e. the multiplicity of risk in a given class in relation to the risk level in the whole population. The fourth column contains the threshold risk values which separate subsequent classes. The last column shows the percentage of the population belonging to the given risk class.

Risk Category	Risk Rate	LIFT	Cut off	Population %
1	0.01%	0.0762	0.00%	50.05%

Risk Category	Risk Rate	LIFT	Cut off	Population %
2	0.08%	0.472	0.05%	30.84%
3	0.32%	1.993	0.15%	13.65%
4	0.87%	5.361	0.57%	2.26%
5	1.80%	11.037	0.96%	2.92%
6	5.98%	36.714	5.61%	0.27%

Table 11: Risk classes after re-weighting

One can see from the table that, in order to correctly assess the company's situation, one needs more information than just the value of bankruptcy risk. The risk value of 5.98% for the last class does not seem to be a clear indicator of the company's bad financial condition, yet the risk in this class is over 36 times higher than the population average and over 600 times higher than for the lowest risk class.

8. Conclusions

This paper presents the approach used to build the model predicting the threat of bankruptcy for Polish companies and the results obtained from using this model. The model was based on financial indicators determined on the basis of annual reports. The results were obtained using logistic regression and random forests methods. This approach to predicting the bankruptcy has certain weaknesses i.e. primarily, it requires the availability of adequate historic data. The arbitrariness of the time of prediction and the unreliability of financial statements can also be considered as disadvantages.

The arbitrariness of the time of prediction. Since the approach uses information included in financial statements provided annually, it is constrained by both the time of model construction and the availability of scoring. In our case (according to the Polish law) companies have 6 months to provide their financial statements. Additionally, there is no severe punishment for neglecting this requirement (in the case of the National Court Register), therefore delays are not so rare. Additional delays stem from the construction of the data set for the companies for which bankruptcy procedure was started.

Reliability of financial statements: The case of Enron (Altman, Hotchkiss, 2006) shows that financial statements do not truly reflect current financial standing of companies because of “creative accounting” (the terms "innovative" or "aggressive" are used as well). False financial information is provided to deceive market players and auditors. Altman shows that this mismatch of information diminishes the predictive power of bankruptcy models for Enron. Dechow, Sloan and Sweeney (1996) open the list of authors who try to create models for the prediction of financial information manipulation.

Although there is a high risk that companies which are about to go bankrupt provide statements based on the false information, the statistical models using these statements show good predictive power. What really matters in predictive modeling is the predictive power of variables.

In spite of the above mentioned disadvantages, an unquestionable advantage of the presented approach is the high effectiveness of the created models. It would be very difficult to improve them using currently available data. However, one can expect that the analytical tools employed could be used to obtain even more effective models, provided that more information was available. For this purpose, the available data about the companies should also contain information concerning: (1) the industry sector, (2) the dynamics of financial indicators, (3) financial flows, (4) expert evaluation of the company's condition based on financial reports (5) qualitative assessment of the company that includes the quality of management, the quality of financial reports, delays in payments, etc. The effectiveness of the presented models can also be improved by treating them as generic models and modifying them using new data, e.g. from banks, and using the Bayesian approach.

Reliance solely on publicly available information is the strongest advantage of the proposed approach for company's bankruptcy prediction. The availability of such information provides a good opportunity to overcome the crucial obstacle for using statistical models to predict defaults for companies - the lack of sufficient data. To take advantage of this opportunity, future research in this area should also cover: (1) the relevance of bankruptcy prediction to the prediction of a default. The attractiveness of bankruptcy prediction mainly comes from the consequence that bankruptcies have on the repayment of debts by companies. (2) an alternative target definition aimed at distinguishing between companies with different abilities to repay their debts. As an example we could propose (a) to base the definition of the target on the decrease of financial standing of the company (b) to base the definition of the target on a broader category of "bad companies" that also includes companies closed (i.e. which terminated their activity) due to financial problems. Although both approaches are burdened with subjectivity, they may produce results of substantial value for default prediction.

Appendix

Variable	StdErr	Wald Test	Wald Pr>ChiSq	Standard Coeff
Intercept	0.617	150.112	0.00000	-
NRFS	1.56E-009	12.869	0.00033	-0.360
CA	0.043	130.716	0.00000	0.510
ROE	0.067	25.077	0.00000	0.138
GROA	0.214	21.997	0.00000	-0.185
CL	0.100	80.035	0.00000	-0.700
E/A	0.079	55.913	0.00000	-0.266
POS	0.018	110.559	0.00000	-0.366

Table AP 1: The base model estimation results

	Intercept	NRFS	CA	ROE	GROA	CL	E/A	POS
Intercept	1.0000	0.4198	-0.9770	-0.2375	0.1723	-0.0522	0.2781	0.0226
NRFS	0.4198	1.0000	-0.4635	-0.0344	0.0859	0.0436	0.0625	-0.1047
CA	-0.9770	-0.4635	1.0000	0.1537	-0.1628	-0.1218	-0.2240	-0.0084
ROE	-0.2375	-0.0344	0.1537	1.0000	0.2293	0.0882	-0.1576	0.1046
GROA	0.1723	0.0859	-0.1628	0.2293	1.0000	-0.0315	-0.2086	-0.3285
CL	-0.0522	0.0436	-0.1218	0.0882	-0.0315	1.0000	-0.3017	-0.0950
E/A	0.2781	0.0625	-0.2240	-0.1576	-0.2086	-0.3017	1.0000	0.0133
POS	0.0226	-0.1047	-0.0084	0.1046	-0.3285	-0.0950	0.0133	1.0000

Table AP 2: Coefficient correlation matrix for the base model

	L/A	C/CA	CA	CL	E/A	NRFS	ROE	GROA	POS
L/A	1.00	-0.49	-0.08	-0.60	-0.95	-0.02	0.20	-0.48	-0.30
C/CA	-0.49	1.00	-0.13	0.78	0.36	-0.03	-0.16	0.25	0.14

	L/A	C/CA	CA	CL	E/A	NRFS	ROE	GROA	POS
CA	-0.08	-0.13	1.00	-0.04	0.13	0.35	-0.13	0.14	0.29
CL	-0.60	0.78	-0.04	1.00	0.46	-0.03	-0.22	0.32	0.22
E/A	-0.95	0.36	0.13	0.46	1.00	0.02	-0.12	0.47	0.29
NRFS	-0.02	-0.03	0.35	-0.03	0.02	1.00	-0.04	0.01	0.11
ROE	0.20	-0.16	-0.13	-0.22	-0.12	-0.04	1.00	-0.28	-0.31
GROA	-0.48	0.25	0.14	0.32	0.47	0.01	-0.28	1.00	0.52
POS	-0.30	0.14	0.29	0.22	0.29	0.11	-0.31	0.52	1.00

Table AP 3: Variables correlation matrix (Pearson) for the base model and alternative models

	Definition of the category representative
High Importance	Equity/Assets
	Profit (Loss) on Sale
	Current Assets/Current Liabilities
	Liabilities/Assets
	Gross Profit/Assets
	Short-term Investments-Cash/Short-term Liabilities
	Short-term Investments/Short-term Liabilities
	Net Profit/Sales
	Current Assets/Short-term Liabilities
Medium Importance	Earnings before Interest and Taxes/Total Assets
	Return on Equity
	Current Assets
	Liabilities Turnover
	Equity/Fixed Assets
	Short-term Receivables + Short-term Investments/Short-term Liabilities
	Revenue from Sales
	Profit on Sales/Assets
	Cash conversion cycle
	Assets
Low Importance	Interest Coverage
	Cash Assets Turnover
	Cash Total Assets
	Fixed Assets
	Assets Productivity
	Sales/Total Assets
	Receivables Turnover
	Short-term Receivables/Liabilities and Provisions for Liabilities
	Fixed Assets Productivity
	Receivables from other Entities/Current Assets
	Inventory Turnover
	Liabilities to banks/Equity
	Liabilities to banks/Liabilities

Table AP 4: Forests Importance

References:

Altman E. I., *"Corporate bond and commercial loan portfolio analysis",* Working paper series - new York University, 1997

Altman E. I., *"Handbook of corporate finance. How can ratios be used in evaluating securities?",* Wiley, Professional Banking and Finance Series, 1986

Altman E. I., *"The Success of Business Failure Prediction Models: An International Survey ",* Journal of Banking and Finance 8, 1984, pp. 171-198

Altman E. I., Hotchkiss E., *"Corporate Financial Distress and Bankruptcy*", 3rd ed., Wiley, New York, 2006

Altman E.I., *"A Further Empirical Investigation of the Bankruptcy Cost Question* ", Journal of Finance 34, No. 4 (September) 1984, pp. 1067-1089

Altman E.I., *"Financial ratios, discriminant analysis and the prediction of corporate bankruptcy ",* The Journal of Finance, September 1968, pp. 589-609

Altman E.I., Haldeman R., Narayanan P., *"ZETA analysis: a new model to identify bankruptcy risk of corporations* ", Journal of Banking and Finance, 1977, June, pp. 29-54

Altman E.I., Lavalle M., *"Business Failure Classification in Canada, Journal of Business Administration* ", Summer, 1981, pp. 147-164

Altman E.I., "*Examining Moyer's Re-Examination of Forecasting Financial Failure*", Financial Management, 1978

Amit, Y., Geman, D., "*Shape quantization and recognition with randomized trees*", Neural Computation, 1997

Back B., Laitinen T., Sere K., van Wezel M., *"Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms ",* TUCS Technical Report No. 40, September 1996

Beaver W., *"Financial ratios as predictors of failure* ", Empirical Research in Accounting: Selected Studies, supplement to vol. 5, Journal of Accounting Research, 1966, pp. 71-111

Beermann K., *" Possible Ways to Predict Capital Losses with Annual Financial Statements* ", Dusseldorf, 1976

Breiman Leo, "*Random Forests* " , Machine Learning 45 (1): 5–32, 2001

El Hennavy R., Morris R., *"The significance of base year in developing failure prediction models ",* Journal of Business Finance and Accounting, vol. 10, no. 2, 1983, pp. 209-223

Fitzpatrick P., *"A comparison of the ratios of successful industrial enterprises with those of failed companies ",* The Accountants Publishing Company, 1932

Frederikslust R. A. I., *"Predictability of Corporate Failure*", Martinus Nijhoff Social Sciences Division, Leiden, The Netherlands, 1978

Fulmer J. G. Jr., Moon J.E., Thomas A., Erwin M.J., *"A Bankruptcy Classification Model For Small Firms* ", Journal of Commercial Bank Lending, July 1984, pp. 25-37

Hadasik, D. , "*The bankruptcy of enterprises in Poland and the methods of prognosis ",* Zeszyty Naukowe AE-Poznań 153, series II, 1998, 176 p.

Ho, Tin*, "Random Decision Forest" in 3rd Int'l* Conf. on Document Analysis and Recognition ", 1995, pp. 278–282

Ho, Tin, "*The Random Subspace Method for Constructing Decision Forests*", IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998, pp. 832-844

Hołda, A., "*Prognosis of the bankruptcy of the unit in the conditions of the Polish economy with utilisation of the discriminatory ZH function* ", Rachunkowość 5,, 2001, pp. 306-310

Hosmer D., Lemeshow S. , "Applied logistic regression ", 2000

Ko C.J., *"A Delineation of Corporate Appraisal Models and Classification of Bankruptcy Firms in Japan ",* Thesis (New York University), 1982

Koh H.C., Killough L.N., *"The use of multiple discriminant analysis in the assessment of the going-concern status of an audit client* ", Journal of Business Finance and Accounting, vol. 17, no. 2, 1990, pp. 179-192

Legault Jean and Pierre Veronneau, *"CA-Score, Un Modele De Prevision De Faillite*", Research Report for the Ordre Des Comptables Agrees Du Quebec, December 1986

Legault J., *"C.A. - Score, A Warning System for Small Business Failures*", Bilanas, June 1987, pp. 29-31.

Merwin C., *"Financing small corporations: in five manufacturing industries, 1926-36 ",* National Bureau of Economic Research, 1942

Ohlson J., *" Financial ratios and the probabilistic prediction of bankruptcy ",* Journal of Accounting Research, Spring 1980, pp. 109-131

Prusak, B. *"Modern methods of predicting financial risk in companies* ", DIFIN, Warszawa, 2005, 208 p.

Ramser J., Foster L., *"A demonstration of ratio analysis ",* Bulletin No. 40, University of Illinois, Bureau of Business Research, Urbana, Illinois 1931

Szczerbak M., "*Przyczyny upadłości przedsiębiorstw w świetle opinii syndyków i nadzorców sądowych [w:] Zagrożenie upadłością* ", red. K. Kuciński, E. Mączyńska, SGH, Warszawa 2005.

Taffler R.J., "*The assessment of company solvency and performance using a statistical model",* Accounting and Business Research, No.Autumn, 1983

Taffler R.J., *"Empirical Models for monitoring of U.K. Corporations "*, Journal of Banking and Finance 8, No. 2 (June), 1984

Taffler R.J.,"*Forecasting company failure in the UK using discriminant analysis and financial ratio data"*, Journal of the Royal Statistical Society, Series A, Vol. 145 No.3, pp.342-58., 1982

Takahashi K. and E. Altman, "*Conference on the State of the Art in Bankruptcy through Classification Models"* Keio Graduate School of Business, Yokohama, Japan, June 25, 1981.

Takahashi K., K. Kurokawa and K. Watese, "*Predicting Corporate Bankruptcy through Financial Statements*", Society of Management Science of Keio University (November), 1979

Van Frederikslust R. A. I., *"Predictability of Corporate Financial Distress: Models for Prediction of Corporate Financial Distress and for Education of Debt Capacity "*, Nijhoff, 1978

Wędzki D., *"Problem wykorzystania analizy wskaźnikowej do przewidywania upadłości polskich przedsiębiorstw - studium przypadków "*, Bank i Kredyt 5/2000.

Wenrich G., *"Predicting Credit Worthiness, Directions of Credit Operations by Risk Classes "*, Weisbaden, Galder, 1978

Wilson R.L., Sharda R., *"Bankruptcy Prediction Using Neural Networks "*, Decision Support Systems, 1995 No 11, pp. 545-557

Winakor A., Smith R., *" Changes in the financial structure of unsuccessful industrial corporations ",* Bulletin No. 51, University of Illinois, Bureau of Business Research, Urbana, Illinois 1935