ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 3(322)2016

http://dx.doi.org/10.18778/0208-6018.322.09

Aneta Ptak-Chmielewska[*]

STATISTICAL MODELS FOR CORPORATE CREDIT RISK ASSESSMENT – RATING MODELS

Abstract. Taking into consideration the weakness of the models based on discrimination function (Z-score) proposed by Altman within the conditions of polish economy some attempts were taken in the 90s to adjust these models to the reality of post-communist economy. The initial interest in the models of multivariate discriminant analysis was extended by logistic regression models and then also by neural networks and decision trees. In the recent years some attempts were also taken to apply models of the event history analysis. Rating models based on developed bankruptcy risk models are basic element in credit risk management. Paper focuses on the critical assessment of statistical methods applied and points out the advantages and disadvantages of various approaches toward the estimation of models. Empirical comparative analysis were conducted based on the sample of enterprises. The possible application of statistical models in credit risk assessment of enterprises (rating models) was pointed out.

Keywords: statistical models, rating models, event history analysis

JEL: G33, C58, C52, C45, C41, C34

1. INTRODUCTION

Multivariate models for forecasting bankruptcy of the enterprises were introduced by Altman in 1968 but the works on those types of models have much shorter history in Poland.

The implementation of the western models to the market of enterprises functioning in the conditions of transitional economy as we had in Poland was not successful. It appeared that those models are not working in conditions of political and economic changes. Insufficient effects of adopting foreign models to Polish conditions contributed to the development of research concerning local models. The biggest popularity, similarly to the situation abroad, was gained by the models based on discriminant analysis.

In the 90s the activities were started to build and implement the models adjusted to the specifics of the Polish economy (papers i.e. by Hadasik (1998), Gajdka & Stos (1996), Pogodzińska & Sojak (1995)). The multivariate discriminant analysis, the regression models and neural networks models were used. Amongst the authors of the bankruptcy models who published their papers after 1990 we can enumerate: Appenzeller (2004), Hołda (2001), Michaluk

[*] Ph.D., Warsaw School of Economics.

(2000), Gruszczyński (2005), Mączyńska & Zawadzki (2006), Hamrol & Chodakowski (2008), Jagiełło (2005) and many others.

The traditional models do not take into consideration changes in time which can be significant. Such changes in time are reflected in the survival models (so called Event History Analysis), the application of which is more and more present in scientific papers (Author 2008).

From the historical perspective we can enumerate that Gajdka and Stos (1996) developed models to predict bankruptcy and presented them in 1996. Hadasik in her models used multivariate discriminant function (Hadasik 1998). Hołda (2001) also estimated models based on the sample of 80 enterprises (40 bad and 40 good) using multivariate discriminant analysis method. Models created by Prusak (2009) were built based on the sample used for financial ratio analysis of entities. Logit models were used by Stępień and Strąk (2004). The sample of bankruptcies consisted of 39 companies with legal bankruptcy applications submitted in the Court of Szczecin in years 1996–1998.

In her works Appenzeller (2004) included dynamic changes in the multivariate discriminant model. Research conducted by Mączyńska (2005) also used discriminant functions. The sample was based on 80 entities and FS data from years 1997–2002. Korol (2005) revealed the advantage of neural networks comparing to discriminant function based on 180 enterprises and FS from period 1998–2001. Strąk (2005) used decision trees convincing that it is worth moving analysis beyond the traditional discriminant approach. Dębkowska (2012) compared the discriminant analysis method with the logistic regression and decision trees based on the sample of 68 enterprises and FS from 2009. Author (2012) compared the survival analysis with the logistic regression and the discriminant analysis.

The main goal of this paper (report) was to compare the effectiveness of different statistical, econometric and data mining models in bankruptcy prediction. Those models are frequently used in rating models for credit risk assessment. In second part of this paper the use of bankruptcy models in rating models development was discussed.

2. METHODS AND MODELS

For the purpose of this paper five different models and techniques were described and discussed. Advantages and disadvantages of following methods were described and discussed:

- Discriminant analysis (Fisher multivariate linear function);
- Logistic regression;
- Survival models (semiparametric Cox regression model);
- Decision trees;
- Neural netoworks.

2.1. Discriminant analysis

The discriminant analysis is used for classification purposes into two groups: good and bad. This classification is based on the function. In this method the set of variables (interval) is used to construct a rule that distinguishes between good and bad in the best possible way.

The main purpose is correct classification into groups. The function maximizes the distance between subpopulations.

There are some limitations of this method. The variables must be normally distributed which is quite often violated. The next assumption is the equality of variances between groups. Those assumptions must be verified before the final model is estimated.

A balanced sample is required to obtain good classification and to correct errors estimation.

The discriminant functions, on which the multivariate models detection systems warn against the bankruptcy can assume various forms – they can be linear or square functions, etc. The linear discriminant function usually takes the form (Author 2012):

$$Z = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n,$$

where:
Z – dependent variable,
a_0 – intercept,
a_i, $i = 1, 2, \dots, n$ – discriminative coefficients (weights),
$X_1, X_2, \dots, X_n$ – explanatory variables (financial ratios).

The presented form of discriminant functions called Fisher discriminative function parameters a, called discriminative coefficients/ratios (weights).

After determining the form of the discriminative function the cut off value is determined which allows for classifying the entity as financially risky or not risky in a definite manner.

The average value of the discriminant function in specific groups and the cut-off value half-way between the average values are most frequently determined. If Z-value for a given enterprise is lower than Z cut-off, then this entity is classified as susceptible to bankruptcy risk, and if it is higher, then the entity is considered healthy.

The main ***advantage*** of discriminant linear function is easy understanding and easy to apply in any IT systems and applications. This model simultaneously take into consideration many variables due to weights applied. It gets one dimension from multivariable space through transformations in order to evaluate the situation on the basis of selected measure. It is possible to determine the impact of particular explanatory variables on the dependent variable (though not always). This model can be applied on a small sample and classifications are very precise in the area of analysing the bankruptcy risk of the enterprises. There is also possibility to apply for dynamic analysis. The discriminant model is available in many popular programs (Ptak-Chmielewska 2012).

The basic ***disadvantage*** of discriminant linear function is small stability. The models can get outdated due to changing economic and spatial conditions. There is no possibility to apply qualitative variables which have significant impact on the enterprise situation e.g. human factor. The strong assumption on normality of explanatory variables distribution is often violated. The assumption on equality of the matrix variance – covariance groups of individuals is also difficult to fulfill. There is necessity to know opinion probability of the population and costs of error of 1st and 2nd types, only with this information the 1st and 2nd type errors are estimated properly. Next limitation is the necessity to have couple and independent date, missings make the classification impossible. Also the linear dependency between the value of the ratio and the financial status of the entity (although in reality it is usually non-linear) makes the estimates limited. The lack of direct possibility to determine the probability of bankruptcy (if it is used to build classification models) direct application of discriminant analysis in rating models is limited. Rating models for enterprises credit risk assessment requires the probability of default estimation on the individual customer level (Ptak-Chmielewska 2012).

2.2. Logistic regression analysis

Logit models are very popular method applicable nowadays to forecast bankruptcy cases of enterprises. The logit function in binomial models assumes the form (Gruszczyński 2001, Matuszyk 2015):

$$P(Y=1)=\frac{1}{1+\exp^{-(\beta_0+\beta_1 x_1+\ldots+\beta_k x_k)}}$$

where:
$P(Y\!=\!1)$ – dependent-variable, usually determines the probability of bankruptcy,
β_0 – intercept,
β_i, $i = 1, 2, \ldots, k$ – weights (coefficients),
x_i, $i = 1, 2, \ldots, k$ – explanatory variables – financial ratios.

$P(Y\!=\!1)$ assumes the value from <0;1>, where 0 – “good” enterprise 1 “bad” enterprise. A significant issue while estimating the binomial models is do properly determine “cut off” pint. In the case of models estimated on the basis of the balanced sample the value of this point is usually equal to 0.5. the structure of the group has the influence on the value of this point (the case of good and bad enterprises).

The so called odds ratio plays the big role in interpreting the results obtained from logit analysis. This ratio is calculated as the relation of the possibility that the event happens to the probability that it will not happen.

The logit model requires that many assumptions are met. The most important are: random character of the sample, big sample, no collinearities of variable, independency of observation.

The ***advantage*** of logistic regression is no assumptions of normality of distributions of explanatory variables and no assumption of equality of the matrix of variance – covariance groups. Explanatory variables can be nominal variables. The result from <0;1> received on forms of the probability of occurrence on analysed event. This result gives the direct estimation of probability of default in rating models. Results from logistic regression are easy to interpret and easy to understand in form of odds ratios. Multivariate character of the model is guaranteed by taking into consideration at the same time many variables thanks to the application of weights. Logistic regression model is characterized by high accuracy of classification comparing to other methods also higher than in the case of linear multivariable discriminant analysis. Advantage is also availability of this method in many statistical programs.

There are some ***disadvantages*** of logistic regression model application. If the assumption of normality is met, higher accuracy of the classification is obtained by means of linear multivariable discriminant method. The explanatory variables cannot be correlated. Correlation of explanatory variables results in high volatility of the model. The model is also highly volatile in relation to significant deviations of the explanatory variables distributions from the normal distribution. Worse results of the classification in logistic regression are obtained than in the case of artificial neural networks. There is necessity to have complete data, missings make the classification impossible. Good results of the classifications are obtained for large samples with high share of bad companies, which is hard to obtain in practice (Matuszyk 2015).

2.3. Event history analysis – Cox regression model

Event History Analysis – survival analysis is described as the set of statistical techniques aiming at the description and studies of the life cycle of an individual, i.e. the frequency of some events, their sequence, distribution, the time the individual spends in different states.

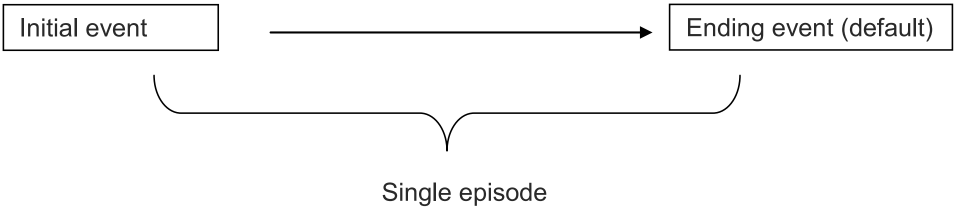
Due to the number of events that may happen we distinguish the single episode analysis and multiple episode analysis, whereas the single episode analysis is the most basic model of the event history tracking analysis.

The stochastic process, which is the subject matter of the analysis, is considered in three basic areas (Blossfeld & Rohwer 2002):

– Time that the occurrence of the distinguished states (events) is expected,
– The intensity of transitions between distinguished states,
– Number and sequence of events.

The basic analytical structure in the event history analysis is the state space and the time axis. The state space is discrete and the time measurement can be continued or discrete. The time axis itself can be defined in two ways: as a calendar time or as a relative time. The state of entry is common for all individuals of the population studied which is defined by common experience by all individuals at the moment of T_0 of an event (it is then a cohort), the occurrence of the bankruptcy event is eliminating an enterprise from a cohort of active enterprises. It is then a final event (exit state) for a single episode model.

– Interval dependent variable – Y – time to event (days or months)
– Model of intensities, the value may exceed the range [0–1]



Scheme 1. Single episode model

Source: own elaboration.

Methods of estimation distinguish between parametric and non-parametric methods. This is based on assumptions about the functional form of time distribution. If there are no such assumptions, non-parametrical methods are applied with the classical example of life table models. Non-parametric analysis gives information about changes of individual behaviours schemes in time.

In parametric approach the time between events is assumed to be a random variable with specific distribution. The most frequently used distributions are: exponential, Weibull, Gompertz. In parametric analysis regression methods are used including the influence of time on hazard rate and the inclusion of explanatory variables and heterogeneity of the population.

The combination of two approaches is named semi-parametric approach (Cox regression model). The parametric component is based on specified influence of explanatory variables on the hazard rate, but non-parametric component does not specify functional distribution of the time.

Censoring is very characteristic for event history data. If information is not available then it is censored. The most typical is right censoring when the time till event is not known but it is longer than observation period.

Take the interval variable T as the time till event occurrence since the time t_0. Distribution of variable T may be described in a few different ways apart from density and cumulative function also by survival and hazard functions.

- **survival function**

$$S(t) = P(T > t)$$

where $S(t)$ means unconditional probability that event occurs after time t, so the enterprise will survive at least till time t. This function describes the survival pattern in the population.

- **hazard function**

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

where $h(t)$ is conditional density of time to event occurrence (on condition that the event did not occur till time t), so $h(t)\Delta t$ means (approx.) probability that the event occurs in a very short period of time (t, t+Δt), on condition that the individual survived at least till time t.

The most frequently used model is semi-parametric proportional hazards Cox regression model. For Cox regression model the hazard function is given by (Frątczak et al. 2005, Matuszyk 2015):

$$h(t \mid x_1, \ldots, x_k) = h_0(t)\exp(\alpha_1 x_1 + \ldots + \alpha_k x_k)$$

where: $h_0(t)$ means base hazard, parametrically non-specified function of time and $X_1, X_2, \ldots, X_k$ – means explanatory variables (including time dependent variables).

The main advantage of the Cox model is assessment of variables influence on the process without necessity of base hazard $h_0(t)$ specification. The main disadvantage of Cox model is hazard proportionality assumption. This assumption imposes that for each pair of individuals in any time the hazard rate is fixed. This problem may be solved by including additional time dependent variables. For checking the proportionality assumption, the easy way is to include the interaction with time. The significance of these parameters confirms that the proportionality assumption is violated. In this case the model is named non-proportional hazards Cox regression model. Results of Cox model estimation are parameters describing the influence of explanatory variables on the probability of event occurrence and on the base hazard.

The main ***advantage*** of event history analysis – Cox regression model is that apart from the question about "if" we ask the question about "when" the event occurs (default). It is possible to include censored information about the customer. There is no need of fixed time observation period for default observation (like in logistic regression). Results give „dynamic" prediction of probability of the event. It is possible to include the macroeconomic changes in the model (time varying variables). In such case the model becomes non-proportional hazard rate model.

There are however some ***disadvantages***. There is strong proportionality assumptions, that must be verified before we estimate the model. All assumptions used in regression models: normality assumption, noncollinearity assumption, etc. still are in force. This model is non-resistant to missing data, all observations with any missing information will be excluded. There is necessity of the information about the exact time of the event (default) which is sometimes not available in practice (Matuszyk 2015, Author 2012, 2014a).

2.4. Data Mining – decision tree

A decision tree is a non-parametric predictive method. Observations are classified by assigning them into groups; therefore it determines that the probability of event occurrence is being calculated at the group level. Classification tree can be treated as a segmentation model with supervision (dependent variable).

A classification is a recursive partition algorithm splitting the group into two subgroups (in each successive step), to ensure uniformity of observations in these subgroups. This model does not require the earlier selection of variables (Lasek, Pęczkowski 2013).

In the preliminary analysis a large set of observations is required when building a tree as well as a relevant number of cases of variable *Y* (i.e. the number of events).

Possible unusual observation may distort the results. The main danger when using the decision tree models is the tendency to over-fitting which makes that the resulting model is unstable. This instability means that the classification rules and the estimated probability of an event do not work on the independent data.

Decision trees requires:

– A big sample

– A lot of defaults ($Y = 1$).

Outliers may distort the results. The model is not resistant to overfitting which makes the model unstable.

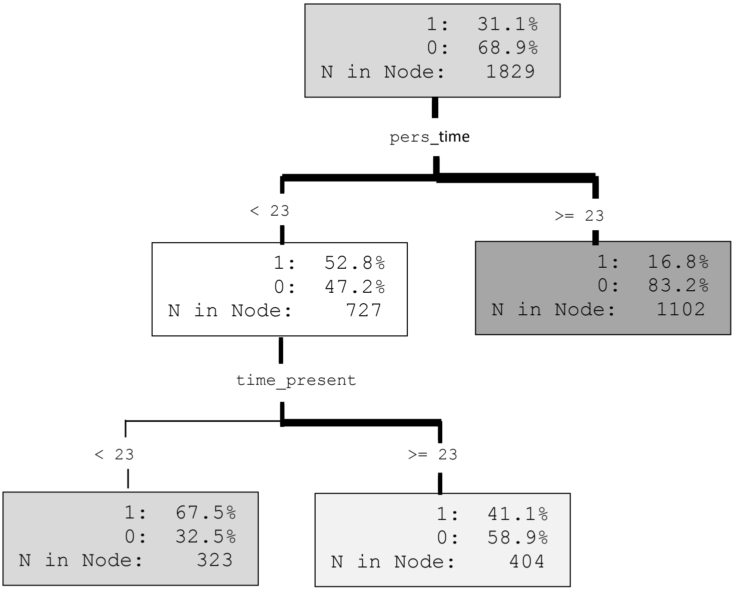


Figure 1. Decision tree scheme

Source: own elaboration.

A decision tree contains the so called root, i.e. the main element, including the entire data set, nodes and sub-segments formed by splitting the data according to the used rules. A tree branch of creates the node with further sub segments. The final division element is called a leaf which is the final segment, which is no longer splitted. Each observation of the output file is being assigned only to one end leaf. A classic decision tree model, for a binary dependent variable, contains the following items (all items are estimated on the training set):

– nodes definitions, or the principles of assigning each observation to the output to the final leaf,

– probability (posterior) for each end leaf (ratio of modelled occurrences of a binary variable in each final leaf),

– assigning level of the dependent variable in the model for each final leaf.

Decision rule can be based on maximizing the profits, minimizing the costs or minimizing the misclassification error.

Decision tree, unlike binary logistic regression, does not contain any equations or coefficients, and is based only on the data set allocation rules. The rules generated by the model can be used in prediction without the dependent variable (the result is a binary decision).

The basic ways of measuring the quality of distribution for dependent binary or discrete variables with few categories are as follows: the degree of separation achieved by the division (measured by p-value Pearson's chi-square test) or the degree of pollution reduction achieved by the separation (measured by the entropy reduction or Gini coefficient).

Stopping splitting criterion can be as follows: the level of significance p-value divisions, leaf size (minimum size of the final leaf), size distribution (node size), or the maximum size of the path splits (maximum path length).

After building the decision tree model with the selected method, the next step is cutting the tree to the correct size. It is done in stages. Firstly, one division is cut off, then all possible combinations of the trees are checked and the best of them are chosen. Then another division is cut and the best tree is checked (already shortened twice), etc. With the increase in the number of leaves, the tree value will increase at the beginning but after reaching a certain point, the growth will not be visible, or even a drop can occur. It is the optimal size of a tree.

There are advantages and disadvantages of decision trees. The main ***advantages*** are: fast adoption of the model to dynamic changes, easy interpretation and visualization of the data and resistance to missing data. There is no need of normality assumption and no assumption about equal variances between groups in decision trees model. This effects that explanatory variables can be nominal,, interval, ordinal or any other type. This model makes possible to modeling any nonlinear dependencies. It automatically selects the significant explanatory variables. There is no need of preselection.

There are some ***disadvantages*** as well. Decision trees models are very often unstable, partitioning may influence the results. At the one level only one variable is included for splitting, no multivariate modelling. The big training sample is required to stabilize the tree and avoid overfitting. The final probability is estimated on the final leaf level (the pooling method), There is no direct estimation of probability of default. The most severe disadvantage is high overfitting risk – very good classification on the training sample, poor on the testing sample (Ptak-Chmielewska 2014b).

2.5. Data Mining – neural network

An artificial neural network is built by neurons (information processing elements) and the connections between them (weights modified during the learning process). An artificial neural network is, in fact, a simplified model of the human brain (Prusak 2005).

A single neuron has multiple inputs x_n, n=1, 2, …, n, and one output. Selected explanatory variables are neuron inputs. Selection of variables is based on the chosen method, such as the factor analysis or principal components method.

For each variable a specific weight is being assigned – w_n. Once they are determined, the total neuron stimulation e is calculated that is the sum of the products of the explanatory variables and their weights (so called activation function).

The output value of the neuron depends on the total neuron stimulation, which in turn is obtained by using a suitable activation function $\varphi(y)$. The format of this function determines the type of neuron. For binary endogenous variable the activation function of the output layer is a logistic function, which narrows the range of estimates to [0;1], which makes it possible to interpret the result in terms of the event occurrence probability.

The capacity of any single artificial neuron is small due to the small computational capabilities and the ability to store a small amount of information. For this reason, the artificial neural networks, consisting of a large number of interconnected neurons, are widely used.

We can distinguish the following neural networks:

– two layer – consisting of input and output layers,

– multilayer – consisting of input and output layers, and hidden layers between them.

The most frequently used neural network is called MLP – *Multi Layer Perception* with one hidden layer.

One hidden layer is sufficient for modeling all nonlinear dependencies but interval type only.

Steps of neural network set up require: dependent variable specification, independent variables specification, partitioning, selection of architecture, training, testing.

Among ***advantages*** of neural networks can be enumerated possibility of fast adoption to dynamic changes. Information may be chaotic and truncated, which is quite difficult to model. There is no need of normality assumption and no assumption about equal variances between groups. Explanatory variables can be nominal. Parallel information is included and used in a model as in multivariate regression models. Any nonlinear dependencies and correlations may be modelled in neural network models.

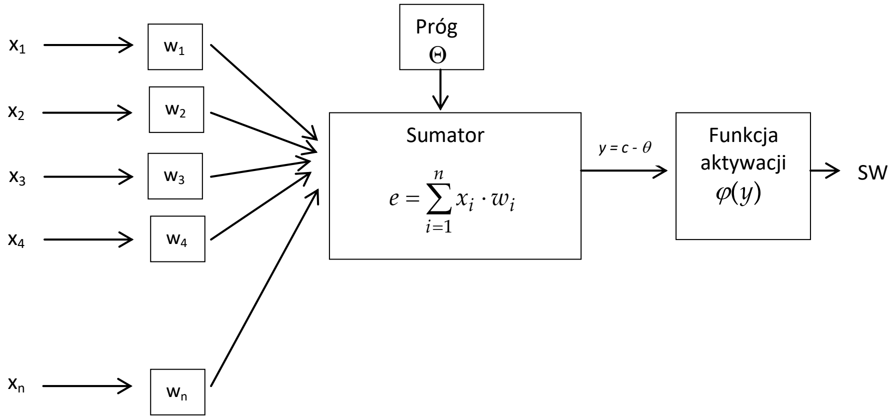


Figure 2. Neural network scheme

Source: Prusak (2005: 57).

There are however some ***disadvantages*** like long training time in case of complicated networks, and possibility of non-convergence of the model. The way of weights estimation is difficult and complicated. There is no automatic selection of variables, explanatory variables must be selected manually. There is high risk of overfitting, over trained neural networks give very good classification on training set but very poor classification on independent sample (test sample). Subjective selection of network architecture and optimization algorithms always make the risk of misspecification. The most severe disadvantage is so called *black-box*, no interpretation of parameters (Ptak-Chmielewska 2014b).

3. COMPARISON

The logistic regression is placed between the discriminant analysis and the neural network, considering the implementation difficulty. In the logistic regression the assumptions about explanatory variables are not so strict as in the case of discriminant analysis. However it **requires big samples** for precise estimation and a high quality of classification. It is not resistant to missing values.

The linear discriminant analysis model is adequate for smaller samples, smaller databases were characteristic for early 90s. Higher volumes available now in many Banks may develop their own methods using databases and more advanced statistical techniques such as neural networks, logistic regression, decision trees.

There are not information or technical limitations nowadays. The dynamic development of advanced models and techniques should be observed in rating models development in Banks. The meaning of **event history analysis** and **data mining** analysis will be growing.

4. EMPIRICAL EXAMPLE

The sample consisted of 6078 financial statements FS for 2342 Medium and Large Enterprises (above 8 mio turnover) in Poland, for which 760 defaults (bankruptcies) were identified over period of 2 years starting from date of FS.[1]

Financial Statements covered years 2002–2011, missings (<1% cases) were imputed with mean value. Extreme values for variables (ratios) were replaced with the value of 5 and 95 percentile.

For a balanced sample only defaults (760) and randomly selected 760 non-defaults were selected. The final sample consisted of 1520 individuals, with proportion of 1:1 good and bad enterprises.

20 financial ratios proposed for financial analysis of enterprise by banking analytics were used (Zaleska 2012).

The list of ratios was selected by two steps procedure:

– Step1. All correlated ratios were eliminated using Hellwig parametric method (r > 0.7),

– Step 2. Using univariate discriminant analysis the discriminative power of ratios were estimated and all ratios with AR < 0.2 were eliminated.

The final list of ratios consisted of 7 ratios.

Table 1. List of financial ratios used in estimation of the models

Label	Ratio	Definition	AR
PL_PS	quick ratio	current assets -stocks / short term liabilities	0.308
EF_ROA	profitability of assets ratio	net profit / total assets	0.357
SB_AKW	coverage of assets by equity	equity / total assets	0.478
OD_PODE	debt coverage by ebit	(net profit+ tax) / (interests+ capital payments)	0.224
AK_CRKO	working capital cycle	net working capital / net income from sales x 360	0.256
AK_UNAO	share of receivables in assets	(long and short term receivables) / total assets	0.241
AK_CRZK	liabilities cycle	average trade liabilities / net income from sales x 365	0.293

Source: own elaboration.

[1] Database was acquired from one of the Polish banks portfolio.

4.1. Discriminant function

Linear discriminant function in classification into two groups (defaults and non-defaults) can be defined as one function or can be defined as two separate function as in table 2.

The higher the value of the weight the stronger impact on classifying into group. The stronger impact for classification into default comparing to non-default has the ratio AK_UNAO (share of receivables in assets). As higher coverage of assets by equity (SB_AKW) as higher chances to be non-default.

The higher values of liabilities cycle (AK_CRZK) support classification into group=defaults comparing to classification into group=non-default.

Table 2. Results of estimation of discriminant linear model

Label	Ratio	Non-default (0)	Default (1)
	Intercept	–5.88616	–5.28586
PL_PS	quick ratio	0.45456	0.17347
EF_ROA	profitability of assets ratio	0.25585	–6.06969
SB_AKW	coverage of assets by equity	14.38695	11.36781
OD_PODE	debt coverage by ebit	–0.00190	–0.00135
AK_CRKO	working capital cycle	–0.00632	–0.00560
AK_UNAO	share of receivables in assets	10.77621	12.53975
AK_CRZK	liabilities cycle	0.02961	0.03245

Source: own elaboration using SAS 9.3.

4.2. Logistic regression

The results of estimation of logistic regression model are included in table 3. The ratio: debt coverage by ebit (OD_PODE) is not significant at the level of significance 0.05 (p-value is much higher than 0.005). AK_CRKO (working capital cycle) is not significant at the level of significance 0.05 but it is significant at the level of significance 0.1. Odds ratio is difficult to interpret because the unit is equal 1. The change in ratio in 1.0 means the change about 100%. The strongest influence is observed in share of receivables in assets (AK_UNAO) because the change in the ratio about 100% means 8 times higher risk of default, but such change is not possible.

Table 3. Results of estimation of logistic regression model

Parameter	DF	Estimation	Err	Chi-sqr	P-value	OR	OR Wald CL95%	
Intercept	1	0.5879	0.2038	8.3192	0.0039	–	–	–
PL_PS	1	–0.5763	0.1241	21.5574	<.0001	0.562	0.441	0.717
EF_ROA	1	–6.3861	0.7927	64.9069	<.0001	0.002	<0.001	0.008
SB_AKW	1	–2.6665	0.3474	58.9105	<.0001	0.069	0.035	0.137
OD_PODE	1	–0.00009	0.0013	0.0040	0.9493	1.000	0.997	1.003
AK_CRKO	1	0.00145	0.0008	2.9182	0.0876	1.001	1.000	1.003
AK_UNAO	1	2.1120	0.3563	35.1297	<.0001	8.265	4.111	16.617
AK_CRZK	1	0.00246	0.0009	6.8032	0.0091	1.002	1.001	1.004

Source: own elaboration using SAS 9.3.

4.3. Cox regression survival model

The results for Cox regression model (table 4) confirm results obtained for logistic regression model. One exception is for AK_CRKO which is now significant at the level 0.05.

The classification accuracy is comparable to accuracy of linear discriminant model. The first type error is quite small.

The observation period was censored at 24th month after date of Financial Statement.

Table 4. Results of estimation of Cox regression model

Parameter	DF	Estimation	Err	Chi-sqr	P-value	HR
PL_PS	1	–0.52295	0.09680	29.1859	<.0001	0.593
EF_ROA	1	–4.59334	0.48306	90.4179	<.0001	0.010
SB_AKW	1	–1.69140	0.24218	48.7759	<.0001	0.184
OD_PODE	1	–0.00009	0.00093	0.0109	0.9169	1.000
AK_CRKO	1	0.00150	0.00051	8.6006	0.0034	1.001
AK_UNAO	1	1.06736	0.21328	25.0442	<.0001	2.908
AK_CRZK	1	0.00109	0.00054	4.1235	0.0423	1.001

Source: own elaboration using SAS 9.3.

4.4. Decision tree

For decision tree algorithm below splitting criteria were assumed:

– *F* test with 0.2 significance level used for splitting

– Minimal leaf size 50 units.

Significance of variables included in splitting are shown in table 5. Only four variables were used in splitting. Remaining three variables (ratios) were not significant in splitting.

The most important split was due to SB_AKW (coverage of assets by equity). The splitting criteria was the value of 0.3248. After splitting there were two groups distinguished with 67.4% of defaults and with 28.3% of defaults compared to 50% at the original sample. The final leaf was reached with 14% of defaults (293 observations) and with 73.3% of defaults (277 observations) (see figure 3).

Table 5. Results of estimation of Decision tree model

Obs	NAME	NRULES	IMPORTANCE
1	SB_AKW	2	1.00000
2	EF_ROA	2	0.53064
3	OD_PODE	1	0.32108
4	AK_CRZK	1	0.31326

Source: own elaboration using SAS Enterprise Miner.

4.5. Neural network

For Neural network model MLP architecture with following criteria was used:

– One hidden layer

– Number of neurons equal number of variables (7).

– Combination function: weighted sum with intercept

– Activation function: logistic.

Results of final optimization are included in table 6.

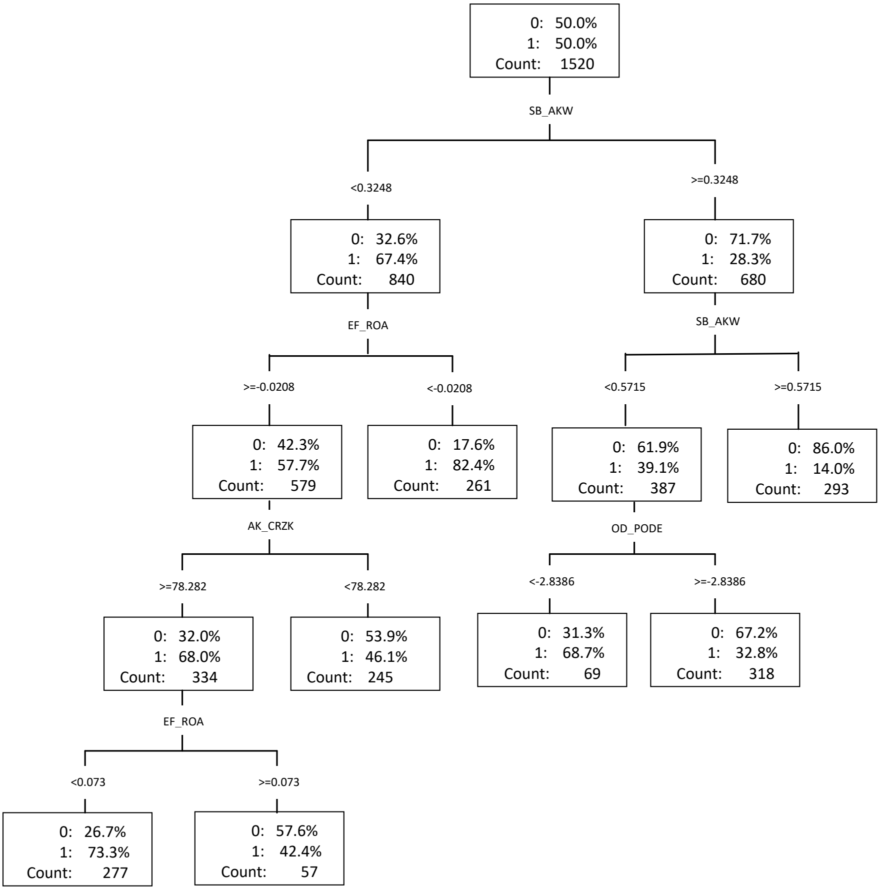


Figure 3. Decision tree – graphical presentation of the final model

Source: own elaboration using SAS Enterprise Miner.

Table 6. Results of estimation of Neural network model

Procedure NEURAL			
Optimization results			
N	Parameter	Estim	Gradient function target
1	AP_AK_CRKO_H11	–0.100862	0.007724
2	AP_AK_CRZK_H11	–0.074365	0.004461
3	AP_AK_UNAO_H11	0.257176	–0.002191
4	AP_EF_ROA_H11	–0.010791	–0.000592
5	AP_OD_PODE_H11	0.017423	–0.000635
6	AP_PL_PS_H11	–0.063980	–0.000504
7	AP_SB_AKW_H11	–0.260593	–0.000098949
8	AP_AK_CRKO_H12	1.962116	0.004406
9	AP_AK_CRZK_H12	–1.126914	0.003138
10	AP_AK_UNAO_H12	0.208302	–0.001972
11	AP_EF_ROA_H12	–1.698166	–0.000030469
12	AP_OD_PODE_H12	–0.416528	0.000129
13	AP_PL_PS_H12	–0.963936	–0.000226
14	AP_SB_AKW_H12	0.734915	0.000420
15	AP_AK_CRKO_H13	–0.954705	0.000081119
16	AP_AK_CRZK_H13	–1.942790	0.000186
17	AP_AK_UNAO_H13	–0.354209	0.000228
18	AP_EF_ROA_H13	1.441217	–0.000051045
19	AP_OD_PODE_H13	4.470955	–0.000096755
20	AP_PL_PS_H13	2.924606	–0.000225
21	AP_SB_AKW_H13	0.183680	–0.000141
22	BIAS_H11	0.734884	0.001664
23	BIAS_H12	–3.760444	0.001824
24	BIAS_H13	2.045224	–0.000544
25	H11_CZY_D1	3.747302	–0.000085812
26	H12_CZY_D1	1.189661	–0.000237
27	H13_CZY_D1	–1.034454	–0.000500
28	BIAS_CZY_D1	–0.968288	0.000585
Value of final function = 0.51238767			

Source: own elaboration using SAS Enterprise Miner.

4.6. Empirical example – comparison

Comparison was done based on Accuracy ratio (Gini Coefficient). AR does not have any particular interpretation. As higher value of this ratio as better is classification accuracy of the model. For rating model (for corporate) the satisfactory level of this ratio is between 0.6–0.7. The highest value of AR was reached by Neural Network model. Comapring accuracy we should also consider two types of classification errors:

– Ist type error – when a model classifies the default as a non-default client, and

– IInd type error – when a model classifies a good customer as potential default.

Comparing those errors of classification the smallest Ist type error was observed for discriminant function and Cox regression model. But for those models the IInd type error was the highest. The highest Ist type error was observed for Decision Tree model (see table 7).

Table 7. Results of estimation of five models – comparison

Model	Accuracy Ratio	Ist type error (%)	IInd type error (%)
Logistic Regression	0.602	25.40	29.40
Discriminant function	–	**21.68**	32.59
COX Regression	0.601	**22.47**	33.11
Decision Tree	0.584	37.32	16.82
Neural Network	0.646	25.09	26.94

Source: own elaboration.

It must be considered in the context of costs of decision. It depends on the cost of Ist and IInd type error. Normally the Ist type error is much more costly than IInd type error. Depending on the risk *appetite* of the bank the decision may be different.

5. RATING MODEL

A **rating model** is a basic tool in the credit application process – underwriting and in credit risk management system for determining enterprises creditworthiness. In recent years we observe the speed in development of information system and data warehouses. It makes possible for Banks to develop their own rating models. Most of those models are statistical or hybrid models.

Changes in Supervisory regulations (CRD IV) caused significant changes in the range of use of methods and techniques of models development. At the same time those amendments triggered changes and corrections in qualitative rating and transactions' risk assessment. This caused a lot of issues to be addressed in rating models and systems development.

The literature on Rating models in Poland is very limited. Some papers and books are focused on Scoring models (Matuszyk 2015, Gruszczyński 1999). The latest publication applies survival models in Scoring systems for individual customers (Matuszyk 2015) on Polish market. The biggest problem is on limited access to databases and internal systems information. From Polish market Wiatr (2011) describes the Rating models and systems based on a few examples.

The literature on Rating models in Europe and USA is much more developed but it is out of the scope of this paper. Polish economy is much less matured comparing to the West Europe. In our country the transformation has ended. The models well developed in West Europe cannot be applied directly on Polish market.

In his book Wiatr (2011) describes also the system functioning in one of Polish banks. The Rating model described in this book comprises of hard facts (financial ratios) and soft facts (qualitative part). Qualitative part cover): sector, company characteristics, management, cooperation with bank. Wiatr (2011) points out the specific evaluation of small and micro enterprises.

The above discussed statistical and *data mining* models can be applied for financial rating development. But this is a first part of the Rating model (see Figure 4). Very important on rating model development is also the qualitative part which is based on questionnaire. Such questionnaire covers the market situation of the company, as well as the history and condition of management in this company. This part can be more or less complicated, usually it consists of 10–20 questions.

Elaboration of rating model requires involvement of different people, different units and IT tools. As presented on Figure 5, this process takes a few steps on different levels.

Final level of approve is placed on Management Board level.

In rating models also the transaction risk is addressed. However the transaction risk is an additional element of credit risk assessment process (credit process). Wiatr (2011) describes transaction risk based on Japanese bank example and Iwanicz-Drozdowska (2012) on one of foreign banks. However those models are quite different from Polish market.

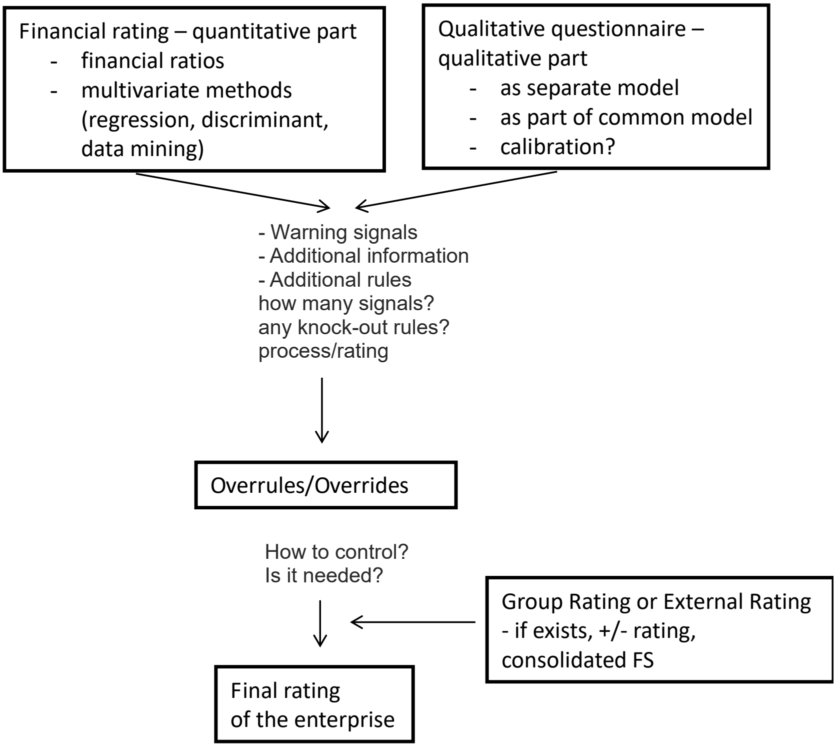

Financial rating – quantitative part
- financial ratios
- multivariate methods (regression, discriminant, data mining)
Qualitative questionnaire – qualitative part
- as separate model
- as part of common model
- calibration?
- Warning signals
- Additional information
- Additional rules
how many signals?
any knock-out rules?
process/rating
Overrules/Overrides
How to control?
Is it needed?
Group Rating or External Rating
- if exists, +/- rating, consolidated FS
Final rating of the enterprise

Figure 4. rating model composition – scheme

Source: own elaboration.

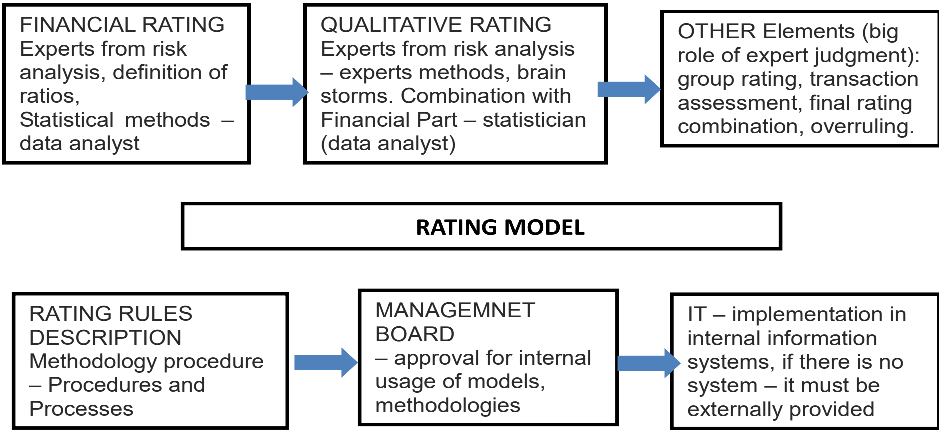
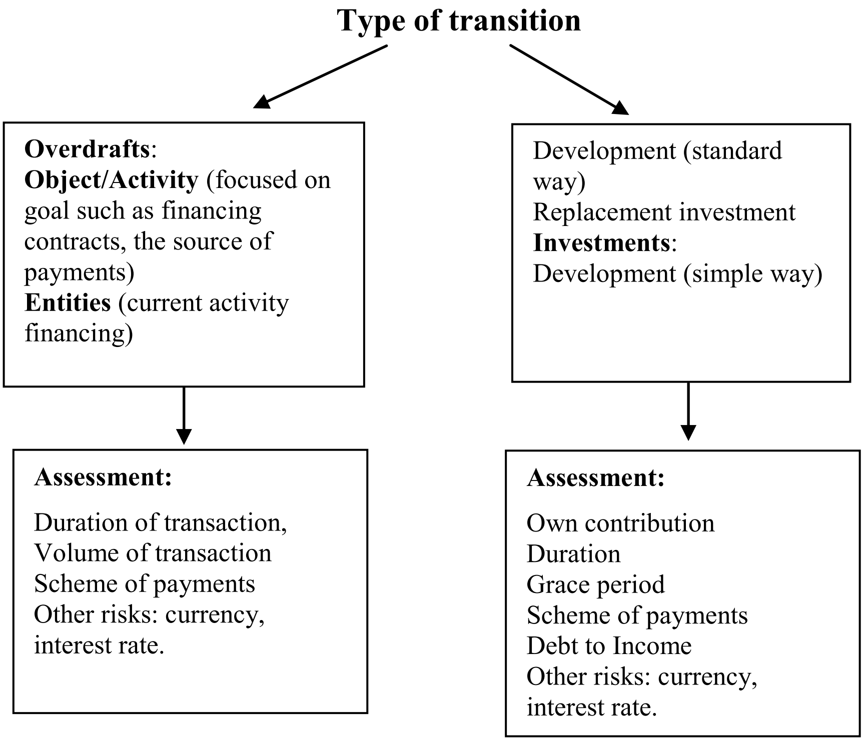

FINANCIAL RATING
Experts from risk analysis, definition of ratios,
Statistical methods – data analyst
QUALITATIVE RATING
Experts from risk analysis – experts methods, brain storms. Combination with Financial Part – statistician (data analyst)
OTHER Elements (big role of expert judgment): group rating, transaction assessment, final rating combination, overruling.
RATING MODEL
RATING RULES DESCRIPTION
Methodology procedure – Procedures and Processes
MANAGEMNET BOARD
– approval for internal usage of models, methodologies
IT – implementation in internal information systems, if there is no system – it must be externally provided

Figure 5. Rating model development – scheme

Source: own elaboration.

In Figure 6 the transaction risk assessment was presented. The process is different for simple transactions like overdrafts and for investment transactions where the risk is placed on investment (object).



Type of transition

Overdrafts:
Object/Activity (focused on goal such as financing contracts, the source of payments)
Entities (current activity financing)

Development (standard way)
Replacement investment
Investments:
Development (simple way)

Assessment:

Duration of transaction,
Volume of transaction
Scheme of payments
Other risks: currency, interest rate.

Assessment:

Own contribution
Duration
Grace period
Scheme of payments
Debt to Income
Other risks: currency, interest rate.

Figure 6. Assessment of transaction risk

Source: own elaboration.

6. CONCLUSIONS

It is important to consider all advantages and disadvantages of statistical models used in rating development. It is crucial to consider all limitations of applied statistical methods, models and techniques. The very first models applied in bankruptcy prediction were based on discriminant analysis. The linear discriminant analysis model is adequate for smaller samples, smaller databases were characteristic for early 90s. Later the logistic regression was implemented and quite often applied. Nowadays the logistic regression is applied the most frequently in Banking sector. The logistic regression is placed between the discriminant analysis and the neural network, considering the implementation difficulty. In the logistic regression the assumptions about explanatory variables are not so strict as in the case of discriminant analysis.

Higher volumes available now in many Banks may develop their own methods using databases and more advanced statistical techniques such as neural networks, logistic regression, decision trees. Regulatory requirements posed on Banks the necessity of more complex models application. There are not information or technical limitations nowadays. The dynamic development of advanced models and techniques should be observed in rating models development in Banks. The meaning of event history analysis and data mining analysis will be growing.

We shouldn't forget that rating model is much more complex issue than only statistical model based on financial ratios. It covers also qualitative rating and transaction risk assessment and should be considered in many dimensions in the context of credit process.

REFERENCES

Appenzeller D. (red.) (2004), *Upadłość przedsiębiorstw w Polsce w latach 1990–2003. Teoria i praktyka*, Zeszyty Naukowe, nr 49/2004, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.

Blossfeld H.P., Rohwer G. (2002), *Techniques of Event History Modeling. New Approaches to Causal Analysis*, Lawrence Elbaum Associates Publishers, London.

Dębkowska K. (2012), *Prognozowanie upadłości przedsiębiorstw za pomocą wybranych metod wielowymiarowej analizy statystycznej*, Zarządzanie i Finanse, vol. 10, nr 1.

Frątczak E., Sienkiewicz U., Babiker H. (2005), *Analiza historii zdarzeń. Elementy teorii, wybrane przykłady zastosowań*, OW SGH, wydanie II.

Gajdka J., Stos D. (1996), *Wykorzystanie analizy dyskryminacyjnej w ocenie kondycji finansowej przedsiębiorstw*, [w:] R. Borowiecki (red.), Restrukturyzacja w procesie przekształceń i rozwoju przedsiębiorstw, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

Gruszczyński M. (2001), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*. OW SGH. Warszawa.

Gruszczyński M. (2005), *Zalety i słabości modeli bankructwa* [w:] K. Kuciński, E. Mączyńska (red.), Zagrożenie upadłością, SGH, Warszawa.

Hadasik D. (1998), *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu, AE w Poznaniu, Poznań.

Hamrol M., Chodakowski J. (2008), *Prognozowanie zagrożenia finansowego przedsiębiorstwa. Wartość predykcyjna polskich modeli analizy dyskryminacyjnej*, „Badania Operacyjne i Decyzje", no. 3.

Hołda A. (2001), *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej*, „Rachunkowość" nr 5.

Iwanicz-Drozdowska M. (ed) (2012), *Zarządzanie ryzykiem bankowym*, Poltext, Warszawa.

Jagiełło R. (2005), *Zastosowanie analizy dyskryminacyjnej do oceny ryzyka kredytowego małych i średnich przedsiębiorstw*, [w:] G. Rytelewska (red.) Bankowość detaliczna. Potrzeby. Szanse. Zagrożenia. PWE, Warszawa.

Korol T. (2005), *Wykorzystanie sieci jednokierunkowej wielowarstwowej oraz sieci rekurencyjnej w prognozowaniu upadłości przedsiębiorstw*, [w:] K. Kuciński, E. Mączyńska (red.), Zagrożenie upadłością, SGH, Warszawa.

Kuciński K., Mączyńska E. (red.) (2005), *Zagrożenie upadłością*, SGH, Warszawa.

Lasek M., Pęczkowski M. (2013), *Enterprise Miner. Wykorzystanie narzędzi Data Mining w systemie SAS*, OW UW, Warszawa.

Matuszyk A. (2000), *Credit Scoring*, CeDeWu, Warszawa.

Matuszyk A. (2015), *Zastosowanie analizy przetrwania w ocenie ryzyka kredytowego klientów indywidualnych*, CeDeWu, Warszawa.

Mączyńska E. (2005), *Kreowanie i konstrukcja modeli dyskryminacyjnych jako narzędzi ostrzegania przed upadłością przedsiębiorstw*, [w:] K. Kuciński, E. Mączyńska (red.), Zagrożenie upadłością, Materiały i Prace Instytutu Funkcjonowania Gospodarki Narodowej, vol. XCIII, Instytut Funkcjonowania gospodarki Narodowej SGH, Warszawa.

Mączyńska E. (red.) (2010), *Meandry upadłości przedsiębiorstw*, Oficyna Wydawnicza SGH w Warszawie, Warszawa.

Mączyńska E. (red.), (2008), *Bankructwa przedsiębiorstw, wybrane aspekty instytucjonalne*, Przedsiębiorstwo współczesne, Kolegium Nauk o Przedsiębiorstwie SGH, Warszawa.

Mączyńska E., Zawadzki M. (2006), *Dyskryminacyjne modele predykcji bankructwa przedsiębiorstw*, „Ekonomista" 2/2006.

Michaluk K. (2000), *Zastosowanie metod ilościowych w procesie prognozowania upadłością przedsiębiorstwa*. Doctoral Dissertation, Wydział Zarządzania i Ekonomiki Usług. Uniwersytet Szczeciński.

Pogodzińska M., Sojak S. (1995), *Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw*, "Acta Universitatis Copernici", "Ekonomia", 25 (299), pp. 53–61.

Prusak B. (2009), *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa.

Ptak-Chmielewska A., Schab I. (2008), *Wykorzystanie modeli regresji logistycznej i hazardu do określenia determinant zaniechania zobowiązań*, w: Pociecha J. (red.) *Współczesne problemy modelowania i prognozowania zjawisk społeczno-gospodarczych*, Studia i Prace Uniwersytetu Ekonomicznego w Krakowie nr 2, Kraków.

Ptak-Chmielewska A. (2012), *Wykorzystanie modeli przeżycia i analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw*, „Ekonometria" 4(38)2012, Wrocław.

Stępień T., Strąk T. (2004), *Wielowymiarowe modele logitowe oceny zagrożenia bankructwem polskich przedsiębiorstw*, [w:] D. Zarzecki (red.), Czas na pieniądz, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin.

Ptak-Chmielewska A. (2014a), *Modele predykcji upadłości MŚP w Polsce – analiza z wykorzystaniem modelu przeżycia Coxa i modelu regresji logistycznej*, "Ekonometria" 4(46)2014, Wrocław.

Ptak-Chmielewska A. (2014b), *Modele przeżycia i metody data mining w ocenie ryzyka upadłości przedsiębiorstw*, [w:] D. Appenzeller (ed), Matematyka i informatyka na usługach ekonomii. UEK w Poznaniu pp. 50–66

Stępień T., Strąk T. (2004), *Wielowymiarowe modele logitowe oceny zagrożenia bankructwem polskich przedsiębiorstw*, [w:] D. Zarzecki (red.), Czas na pieniądz, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin.

Strąk T. (2005), *Wykorzystanie drzew klasyfikacyjnych do oceny zagrożenia bankructwem polskich przedsiębiorstw*, „Monografie i Opracowania Naukowe", SGH w Warszawie. Kolegium Zarządzania i Finansów. Finanse przedsiębiorstwa.

Wiatr M.S. (2011), *Zarzadzanie indywidualnym ryzykiem kredytowym. Elementy systemu.* OW SGH. Warszawa. Wydanie II.

Zaleska M. (2012), *Ocena ekonomiczno-finansowa przedsiębiorstwa przez analityka bankowego*, Szkoła Główna Handlowa – Oficyna Wydawnicza, Warszawa.

Aneta Ptak-Chmielewska

MODELE STATYSTYCZNE DO OCENY RYZYKA KREDYTOWEGO PRZEDSIĘBIORSTW – MODELE RATINGOWE

Streszczenie. Dostrzegając słabość modeli opartych na funkcji dyskryminacyjnej Z-score zaproponowanej przez Altmana w warunkach gospodarki polskiej podjęto w latach 90. próby dostosowania tych modeli do realiów gospodarki post-komunistycznej. Początkowe zainteresowanie modelami wielowymiarowej analizy dyskryminacyjnej poszerzono o modele regresji logistycznej a później również o sieci neuronowe i drzewa decyzyjne. W ostatnich latach podjęto również próby zastosowania modeli analizy historii zdarzeń. Modele ratingowe oparte na wypracowanych modelach upadłości stanowią kluczowy element w zarządzaniu ryzykiem kredytowym. W artykule podjęto próbę krytycznej oceny stosowanych metod statystycznych oraz wskazano na zalety i wady różnych podejść do budowy modeli. Przeprowadzono porównawczą analizę empiryczną na próbie przedsiębiorstw. Wskazano na możliwość wykorzystania modeli statystycznych do oceny ryzyka kredytowego przedsiębiorstw (modele ratingowe).

Słowa kluczowe: modele statystyczne, modele ratingowe, analiza historii zdarzeń.