Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

www.czasopisma.uni.lodz.pl/foe/
1(333) 2018

DOI: http://dx.doi.org/10.18778/0208-6018.333.05

Aneta Ptak-Chmielewska
Warsaw School of Economics, Institute of Statistics and Demography, aptak@sgh.waw.pl

Bankruptcy Risk Models for Polish SMEs – Regional Approach

Abstract: Credit risk management is a key element in bank management. For credit risk management, statistical models are used, the so-called scoring and rating models. For enterprise risk assessment, rating models are used. Rating models consist of quantitative models (based on financial ratios) and qualitative models (based on a questionnaire). For estimation of quantitative models, econometric and statistical models are used, mainly logistic regression models. In this paper, statistical models for quantitative assessment are presented, including an empirical example based on the sample of data for SMEs made available by one of Polish banks. A logistic regression model with a nominal variable – the region of activity, including territorial differences, was used. The construction of rating model was presented, including the sector and region of activity.

Keywords: credit risk, bankruptcy models, logistic regression

JEL: C52, C58, G33

1. Introduction

Credit risk management is a key element in the process of risk management in a bank. Statistical models, the so-called scoring and rating models, are used in credit risk management. In assessment of credit risk in a corporate portfolio, rating models are used. Basic parts of rating models are quantitative models (based on financial ratios) and qualitative models (based on a qualitative questionnaire).

Most of the Polish models for predicting bankruptcy were built using linear discriminant analysis. Selected models are described below.

A pioneer in the study of the credit risk of companies operating in Poland was Mączyńska (1994). Research conducted by her can be treated as an attempt to adapt Altman's model to the Polish economy.

Pogodzińska and Sojak (1995) also used discriminant analysis to predict the bankruptcy of enterprises operating in Poland. In order to build such a model, they used a data set consisting of 10 companies, including four which survived, and six that defaulted. To assess the financial situation of the companies, the authors used two indicators (explanatory variables), i.e.: quick ratio and gross profit margin. The efficacy of the model was estimated at 80%.

Gajdka and Stos (1996 also used discriminant analysis in their research. The authors built two models based on a group of 40 companies, half of which became bankrupt, while the remaining half were solvent. The effectiveness of the first model was calculated at 82.5% and of the latter at 92.5%.

Another person who conducted research focused on the prediction of corporate bankruptcy was Hadasik (Appenzeller – after the change of name). In her first study (Hadasik, 1998), she built 9 discriminatory models using a sample of companies operating between 1991 and 1997. Among these 9 models, 5 were characterised by efficiency exceeding 93%.

Hołda (2001) built a model based on a sample of 80 enterprises, 40 of which went bankrupt, and 40 remained solvent. He built the model using discriminant analysis. The effectiveness of this model was 92.5%.

Similar research was conducted by Wierzba (2000) using a sample of 48 companies, of which 24 went bankrupt or introduced an arrangement, while the other 24 remained solvent. To build the model, he used multivariate linear discriminant analysis, and chose 4 financial ratios. The effectiveness of the model was 92.0% for the first year before the bankruptcy, and 76% for two years before the bankruptcy.

Sojak and Stawicki (2001), using a sample of 58 companies, divided the studied entities into three categories: good, average and bad ones. In order to build the model, they chose 7 financial indicators and tried 3 classification functions. The effectiveness of the received model was 93.1%.

A group of models built by Prusak (2005) was also based on discriminant analysis. The data set consisted of 40 bankrupt enterprises and 40 in a good financial condition. Two models were built differing by chosen financial indicators. The effectiveness of the first one a year before the bankruptcy amounted to 100%, and was 86.08% two years before the bankruptcy. The effectiveness of the overall second, third and fourth model for the test amounted to 93.51%, 97.86% and 95.71% the year before the bankruptcy, and two years before the bankruptcy, respectively, 97.40%, 91.18% and 91.91%.

The latest proposals of research concern models for industrial enterprises (Pociecha, 2014). In this work, a review of models and ratios used in bankruptcy models was presented. New models for industrial enterprises were proposed. Estimation of those models was based on a large sample (a few thousands) of enterprises.

In the work by Jagiełło (2013), the division according to the sector of activity was proposed. Separate models were estimated for each sector (group of activities). Models were based on a sample of 80 enterprises (including 40 non-performing ones according to the Polish Accounting Standards) which were a bank's clients. That is why the definition of default was based on non-performing loans.

In bankruptcy risk models used for prediction of probability of bankruptcy usually logistic regression and discriminant analysis as well as data mining models were applied. In recent years, also event history analysis – survival models – have been used more and more often (Ptak-Chmielewska, Matuszyk, 2014). In all those models, there is a need for including nominal variables as explanatory variables. In the case of nominal variables with a lot of different categories, such as the region of activity, sector of activity, etc., the aggregation of such categories is necessary. Categories with the same level of risk can be combined (merged) diminishing the dimension and size of such a model.

The multivariate statistical method such as cluster analysis was proposed for aggregating the nominal variable categories (Frątczak, 2009). The example was illustrated based on a sample of small and medium enterprises. The sample consisted of bankrupted enterprises and enterprises in a good standing. The nominal variable *region of activity* (*province*) was included in the logistic regression model.

Research hypotheses:

H1: There is strong differentiation of bankruptcy risk according to the size and sector of activity.

H2: Information about regional differences in enterprise activity positively increases the discriminatory power of the bankruptcy risk model.

2. Data and methods

Data were supplied by one of Polish commercial banks. The sample consisted of 333 enterprises which went bankrupt in years 2009–2012 with financial statements (FS) for those enterprises for years 2008–2010. The time to their bankruptcy from the date of FS was 1–2 years. The enterprises in a good condition are 533 enterprises with FS from 2009.

The structure of the sample according to the sector of activity:

1) 288 (33.25%) trade enterprises (including 109 – 37.85% bankrupted),
2) 298 (34.41%) industrial enterprises (including 123 – 41.28% bankrupted),
3) 280 (32.34%) service enterprises (including 101 – 36.07% bankrupted).

The structure of the sample according to the size:

1) 212 (24.48%) small enterprises (including 92 – 43.40% bankrupted),
2) 654 (75.52%) medium enterprises (including 241 – 36.85% bankrupted).

2.1. Discriminant analysis

Discriminant analysis identifies the variables that properly classify observations to different groups. The main purpose of this analysis is the proper classification of observations into groups, subspaces. A discriminant function is maximisation of the distance between subpopulations.

In discriminant analysis, the classification into two groups, bankrupted or non-bankrupted companies, is based on at least two explanatory variables and one dependent variable (binary). The most typical is the linear discriminant function. The outcome of such linear combination of variables is a Score (the value of the function) which is used for the classification.

Discriminant analysis has limitations (Frątczak, 2009):

1) explanatory variables (ratios) must be normally distributed,
2) explanatory variables (ratios) must be independent (no collinearities),
3) covariances in both subpopulations must be equal.

The linear discriminant function (called Fisher's discriminant function) is as follows:

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

where:
Z – dependent variable (binary),
a_0 – intercept,
a_i, $i = 1, 2, \ldots, n$ – discriminant weights,
$X_1, X_2, \ldots, X_n$ – explanatory variables (such as financial ratios).

For classification, the cut-off point must be set up. All observations above the cut-off point (and equal) are classified into the first group and all observations below the cut-off point are classified into the other group.

Effectiveness of classification is measured by a classification table. In the classification table, two types of errors are presented: I-type error and II-type error. The first type of error measures the percentage of bankrupted enterprises classified as non-bankrupted ones, and the second type of error measures the percentage of non-bankrupted enterprises classified as bankrupted ones.

2.2. Logistic regression

The logistic regression function is S-shaped and described by the following formula:

$$P(Y=1)=\frac{1}{1+\exp^{-(\beta_0+\beta_1 x_1+\ \dots\ +b_k x_k)}},$$

where:
$P(Y = 1)$ – dependent variable,
b_0 – intercept,
b_i, $i = 1, 2, \dots, k$ – coefficients,
x_i, $i = 1, 2, \dots, k$ – explanatory variables.

The $P(Y = 1)$ takes the values from interval [0; 1]. The cut-off point is an important element in the logistic regression model. Estimation based on a balanced sample usually takes the 0.5 as the cut-off value. The structure of the sample (the percentage of bankrupted enterprises) determines the cut-off value.

Interpretation of results is usually based on odds ratios (the ratio of odds in two groups or in change of one unit in explanatory variable). Logistic regression requires a number of different assumptions to be fulfilled. The most important assumptions are: randomness of the sample, a big sample, no collinearities in explanatory variables, and independence of observations.

Altman is a precursor of multivariate methods. He presented his model in 1968. This model was a combination of ratio analysis and statistical method – multivariate discriminant analysis. The analysis of 22 financial ratios was based on a sample of 66 enterprises (33 bankrupted and 33 in a good condition). In the subsequent analysis, the ratios with the lowest predictive power were eliminated, and finally 5 financial ratios were included in the model.

In 1977, Altman with his team conducted the following research on the prediction of bankruptcy risk of enterprises. 58 bankrupted enterprises and 58 enterprises in a good condition were analysed. The model with 7 variables was estimated, but no weights or discrimination function were proposed. The proposed ZETA model had high predictive power for the period of 5 years before bankruptcy. For one-year prediction, the prediction power was 90%, for five-year prediction the predictive power was 70%.

The next version of Altman's model was developed in 1983. Altman changed the weights proposed in the first model. The misclassification error was at the level of 6%. The subsequent adjustment concerned lowering weight for the economic cycle factor and sector specificity in the value of Z-Score.

All the models developed by Altman were based on enterprises from the American market. Application of this model in different markets (e.g.: post-communist markets) does not provide satisfactory results.

In this paper, Altman's Z-Score bis model was applied:

$$Z = 0.717 \cdot X_1 + 0.847 \cdot X_2 + 3.107 \cdot X_3 + 0.420 \cdot X_4 + 0.998 \cdot X_5,$$

where:
X_1 – Working Capital/Total Assets,
X_2 – Profit Retained/Total Assets,
X_3 – EBIT/Total Assets,
X_4 – Equity/Total Liabilities,
X_5 – Sales Revenue/Total Assets,
Z – Score.

3. Empirical results

Four models were estimated using a sample of bankrupted and healthy small and medium enterprises. The first model was a basic logistic regression model using variables proposed by Altman in his Z-Score model. The next two models were proposed using the size and sector of activity and using the province as the nominal variable. For grouping the categories of province variable, the hierarchical grouping method was used. The province nominal variable represents territorial differences in bankruptcy risk. Finally, the model was estimated for small and for medium enterprises separately.

3.1. Logistic regression for variables from the Z-score model

Variables X_2 and X_4 were not significant at the 0.05 significance level. The variable X_4 was significant at the 0.1 significance level (see Table 1).

Table 1. Results of the logistic regression model for variables from the Z-Score model

Variable	Estimate	p-value
Intercept	–0.47	< 0.0001
X_1	–1.45	< 0.0001
X_2	0.06	0.5837
X_3	–2.05	< 0.0001
X_4	0.01	0.0650
X_5	0.07	< 0.0001

Source: own elaboration using SAS

Assuming the cut-off point at the 0.5 level, the percentage of correct classifications was 72.6% in total. The sensitivity of the model amounted to 42.9%, meaning the percentage of correctly classified $Y = 1$. The percentage of incorrectly classified $Y = 0$ (1 – Specificity) was 8.8% (see Table 2). The area under the curve ROC was 0.7692.

Table 2. Classification table for the logistic regression model with variables from the Z-Score model

$P = 0.5$	Model $Y = 1$	Model $Y = 0$	Total
Sample $Y = 1$	143	190	333
Sample $Y = 0$	47	486	533
Total	190	676	866

Source: own elaboration using SAS

3.2. Logistic regression with variables: the size and sector of activity

The model with the size and sector of activity was used to verify the research hypothesis concerning the differences in bankruptcy risk according to the size of enterprise and segment of enterprise's activity. Type 3 analysis confirmed that variables X_2 and the size were not significant (at the 0.05 significance level) (see Table 3).

Assuming the cut-off point at the 0.5 level, the percentage of correct classifications was 72.9% in total. The sensitivity of the model amounted to 39.9%, meaning the percentage of correctly classified $Y = 1$. The percentage of incorrectly classified $Y = 0$ (1 – Specificity) was 6.6% (see Table 4).The area under the curve ROC was 0.7779.

Table 3. Results for the logistic regression model for variables from the Z-Score model and variables: the size and sector of activity

Variable	Estimate	p-value
Intercept	–0.48	< 0.0001
X_1	–1.50	< 0.0001
X_2	0.05	0.6577
X_3	–2.14	< 0.0001
X_4	0.01	0.0431
X_5	0.08	< 0.0001
Size (small)	0.02	0.8360
Sector (trade)	–0.07	0.4977
Sector (industry)	0.31	0.0047

Source: own elaboration using SAS

Table 4. Classification table for the logistic regression model with variables from the Z-Score model and variables: the size and sector of activity

$P = 0.5$	Model $Y = 1$	Model $Y = 0$	Total
Sample $Y = 1$	133	200	333
Sample $Y = 0$	35	498	533
Total	168	698	866

Source: own elaboration using SAS

3.3. Logistic regression with the nominal variable: province

For the segmentation of provinces due to bankruptcy risk of enterprises, the hierarchical cluster analysis method was used based on the bankruptcy percentage (see Table 5). As the linkage criterion, an average linkage was applied. The distance was the highest in the case of division into three groups (see Figure 1). The provinces were clustered into three groups: low risk of bankruptcy (group 1 – 22.58% bankruptcies, 93 enterprises), medium risk (reference group 3 – 37.56% bankruptcies, 655 enterprises), and high risk (group 2 – 55.93%, 118 enterprises).

Table 5. Voivodeship – nominal variable distribution (bankruptcy was in total in 38.45% cases)

Voivodeship	Bankruptcy			Voivodeship	Bankruptcy		
	No	Yes	Total		No	Yes	Total
Dolnośląskie	39 (46.43%)	45 (53.57%)	84	Podlaskie	15 (65.22%)	8 (34.78%)	23
Kujawsko- -Pomorskie	27 (64.29%)	15 (35.71%)	42	Pomorskie	41 (62.12%)	25 (37.88%)	66
Lubelskie	16 (72.73%)	6 (27.27%)	22	Warmińsko- -Mazurskie	8 (61.54%)	5 (38.46%)	13
Lubuskie	7 (58.33%)	5 (41.67%)	12	Wielkopolskie	62 (60.19%)	41 (39.81%)	103
Mazowieckie	135 (62.79%)	80 (37.21%)	215	Zachodniopomorskie	25 (60.98%)	16 (39.02%)	41
Małopolskie	56 (78.87%)	15 (21.13%)	71	Śląskie	54 (64.29%)	30 (35.71%)	84
Opolskie	12 (63.16%)	7 (36.84%)	19	Świętokrzyskie	8 (66.67%)	4 (33.33%)	12
Podkarpackie	15 (60.00%)	10 (40.00%)	25	Łódzkie	13 (38.24%)	21 (61.76%)	34

Source: own elaboration using SAS

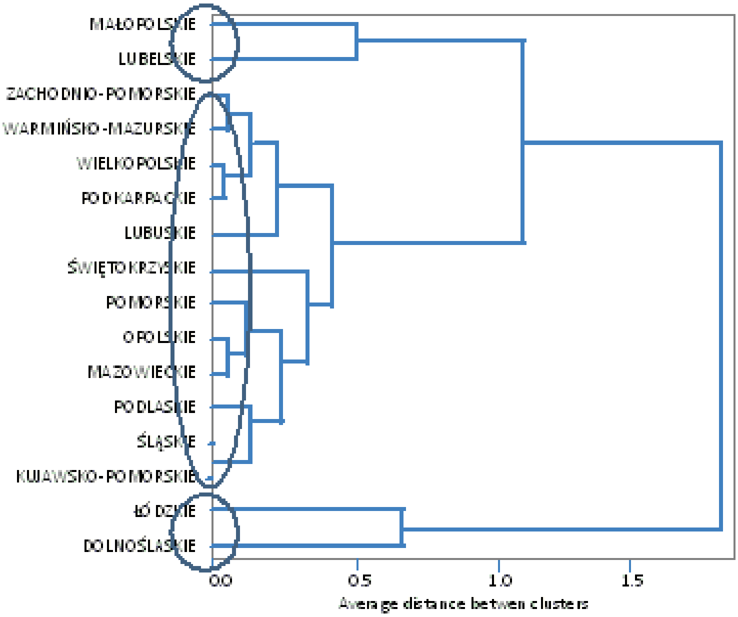

Figure 1. Voivodeship – categories grouping – hierarchical average linkage method

Source: own elaboration using SAS

Table 6. Results for the logistic regression model for variables from the Z-Score model and variables: the sector of activity and province group

Variable	Parameter	p-value
Intercept	–0.77	< 0.0001
X_1	–1.51	< 0.0001
X_2	0.07	0.5579
X_3	–2.15	< 0.0001
X_4	0.01	0.0398
X_5	0.09	< 0.0001
Sector (trade)	0.13	0.5097
Sector (industry)	0.55	0.0047
Voivodeship group 1	–0.85	0.0041
Voivodeship group 2	0.78	0.0004

Source: own elaboration using SAS

Variable X_2 is still insignificant at the level of 0.05, all the other variables are significant in this model (see Table 6). The risk of bankruptcy in trade enterprises is about 14% higher and in industrial enterprises about 74% higher compared to the reference group of enterprises in the services sector. In group 1 of provinces (Małopolskie, Lubelskie), bankruptcy risk is about 58% lower than in the reference group (group 3), while in group 2 (Łódzkie, Dolnośląskie), it is more than twice higher compared to the reference group.

AUC in this model was 0.7873.

3.4. Model for small size enterprises

There were 212 small enterprises in the sample, including 92 (43.4%) bankruptcies. Variables X_2, X_4, the Province group and Sector were insignificant at the 0.05 level (see Table 7). The Province Variable group at the 0.1 level was significant (p-value 0.064), however, there is no significant difference between group 1 and 3 and between group 2 and 3.

Table 7. Results for the logistic regression model for variables from the Z-Score model and variables: the sector of activity and the nominal variable: province – small size enterprises

Variable	Estimate	p-value
Intercept	–0.28	0.4110
X_1	–1.35	0.0023
X_2	0.04	0.8188

Variable	Estimate	p-value
X_3	–1.97	0.0002
X_4	–0.25	0.1245
X_5	0.09	0.0070
Sector (trade)	–0.05	0.9027
Sector (industry)	0.27	0.5237
Voivodeship group 1	–0.83	0.1265
Voivodeship group 2	0.78	0.1146

Source: own elaboration using SAS

3.5. Model for medium size enterprises

There were 654 medium size enterprises, including 241 (36.8%) bankruptcies. Only variable X_2 was insignificant at the 0.05 level (see Table 8).

Table 8. Results for the logistic regression model for variables from the Z-Score model and variables: the sector of activity and the nominal variable: province – medium size enterprises

Variable	Parameter	p-value
Intercept	–0.94	< 0.0001
X_1	–1.44	< 0.0001
X_2	0.19	0.3161
X_3	–2.72	< 0.0001
X_4	0.05	0.0059
X_5	0.10	0.0047
Sector (trade)	0.19	0.4080
Sector (industry)	0.69	0.0023
Voivodeship group 1	–0.87	0.0166
Voivodeship group 2	0.76	0.0027

Source: own elaboration using SAS

The division into two separate models for small and for medium size enterprises revealed the difference in the discriminatory power of the model for these two segments. Due to a small number of enterprises in the sample of small enterprises, 4 variables were insignificant but the discriminatory power of the model was very high (AUC = 0.8556). The discriminatory power of the model for medium size enterprises was much lower, and slightly lower compared to the basic (common) model (AUC = 0.7707).

4. Conclusions

To summarise the results, it can be said that:

1. Logistic regression in contrast to discriminant analysis enables to include nominal variables.
2. The cluster analysis method may be used to aggregate (merge) the categories of nominal variables to decrease the dimensionality of the variable and hence the dimensionality of the analysis. Clustering the categories for the nominal variable enables the inclusion of the nominal variable, which allows to avoid quasi-complete separation.
3. In the empirical example, the model with the clustered nominal variable: province had higher discriminatory power.
4. Bankruptcy risk of enterprises is regionally differentiated. Bankruptcy risk depends on the size of the enterprise and its sector of activity.

Both research hypotheses were confirmed.

(H1): There is strong differentiation of bankruptcy risk according to the size and sector of activity.

The size variable was not significant but two separate models revealed differences between small and medium size enterprises. Bankruptcy risk for small enterprises is higher compared to medium enterprises. The sector of activity was significant in all the models, except the model for small enterprises.

(H2): Including the information about regional differences in enterprises' activity positively increases the discriminatory power of the bankruptcy risk model. Differences in bankruptcy risk between the groups of provinces were significant. The model with the voivodeship nominal variable had higher discriminatory power compared to the model with only financial ratios and the sector and size of the enterprise.

The bankruptcy risk model may be a starting point in the development of rating models. It is the first part of such a model – the quantitative part. Rating models are used in the assessment of credit risk of enterprises in the banking system.

References

Frątczak E. (ed.) (2009), *Wielowymiarowa analiza statystyczna. Teoria i przykłady zastosowań z systemem SAS*, Szkoła Główna Handlowa, Warszawa.

Gajdka J., Stos D. (1996), *Wykorzystanie analizy dyskryminacyjnej do badania podatności przedsiębiorstwa na bankructwo*, [in:] J. Duraj (ed.), *Przedsiębiorstwo na rynku kapitałowym*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Hadasik D. (1998), *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, „Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu", no. 153, Wydawnictwo AE w Poznaniu, Poznań.

Hołda A. (2001), *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej*, "Rachunkowość", no. 5, pp. 306–310.

Jagiełło R. (2013), *Analiza dyskryminacyjna i regresja logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw*, "Materiały i Studia", no. 286.

Mączyńska E. (1994), *Ocena kondycji przedsiębiorstwa (uproszczone metody)*, "Życie Gospodarcze", no. 38, pp. 42–45.

Pociecha J. (ed.) (2014), *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*, Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków.

Pogodzińska M., Sojak S. (1995), *Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw*, "Ekonomia XXV ", vol. 299.

Prusak B. (2005), *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa.

Ptak-Chmielewska A., Matuszyk A. (2014), *Default prediction for SME using discriminant and survival models, evidence from Polish market*, "Quantitative Methods in Economics", vol. XV, no. 2, pp. 369–381.

Sojak S., Stawicki J. (2001), *Wykorzystanie metod taksonomicznych do oceny kondycji ekonomicznej przedsiębiorstw*, "Zeszyty Teoretyczne Rachunkowości", vol. 3(59), pp. 45–52.

Wierzba D. (2000), *Wczesne wykrywanie przedsiębiorstw zagrożonych upadłością na podstawie wskaźników finansowych – teoria i badania empiryczne*, "Zeszyty Naukowe", no. 9, Wydawnictwo Wyższej Szkoły Ekonomiczno-Informacyjnej w Warszawie, Warszawa.

Modele ryzyka upadłości polskich MŚP – ujęcie regionalne

Streszczenie: Zarządzanie ryzykiem kredytowym stanowi kluczowy element w zarządzaniu bankiem. Do zarządzania ryzykiem kredytowym wykorzystywane są modele statystyczne tzw. modele scoringowe i ratingowe. Do oceny ryzyka kredytowego przedsiębiorstw wykorzystuje się modele ratingowe. Składową modeli ratingowych są modele ilościowe (oparte na wskaźnikach finansowych) oraz modele jakościowe (oparte na ankiecie jakościowej). Do budowy modeli ilościowych wykorzystuje się modele statystyczne i ekonometryczne, głównie modele regresji logistycznej. W artykule omówione zostały modele statystyczne do oceny ilościowej wraz z przykładem empirycznym opartym na danych dla próby MŚP udostępnionej przez jeden z polskich banków. Wykorzystano model regresji logistycznej ze zmienną nominalną – region działalności, uwzględniający zróżnicowanie terytorialne. Pokazana została konstrukcja modelu uwzględniającego zarówno branże działalności, jak i region działalności.

Słowa kluczowe: ryzyko kredytowe, modele upadłości, regresja logistyczna

JEL: C52, C58, G33

Received: 2016-12-28; verified: 2017-04-06. Accepted: 2017-12-28