
AN EVALUATION OF THE DISCRIMINATORY POWER OF SELECTED POLISH BANKRUPTCY PREDICTION MODELS AS PART OF THE VALIDATION PROCESS

Natalia Nehrebecka

University of Warsaw, Narodowy Bank Polski, Warsaw, Poland

e-mail: nnehrebecka@wne.uw.edu.pl

ORCID: 0000-0003-3870-7231

© 2018 Natalia Nehrebecka

This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivs license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

DOI: 10.15611/fins.2018.4.05

JEL Classification: C13, G33, C45

Abstract: The purposes of this article are to present validation techniques according to their discriminatory power, while indicating the reservations about such techniques, and to check the adjustment of the existing Polish bankruptcy prediction models in the context of their discriminatory power. This is the first study that performs a validation of such models. Based on the analysis, it was found that the fifth model developed by Hadasik was characterised by a very high discriminatory power. The decision was made to base the evaluation of the discriminatory power of the modules on the Gini index, the Kolmogorov-Smirnov statistic, the H measure, the information value (IV), and the precision of the estimates of bankruptcy.

Keywords: bankruptcy prediction, discriminatory power, validation.

1. Introduction

In accordance with the Polish Financial Supervision Authority (KNF), validation is referred to as an evaluation of the effectiveness of the model used in a bank, conducted by a unit that is not involved in the process of drafting the model, usually in a way that is more comprehensive than that performed as part of monitoring. The need to perform a validation of the models of the prediction of bankruptcy in financial institutions is directly associated with worldwide practices suggested by the New Basel Capital Accord (Basel II and Basel III), whose objective is to strengthen the security and stability of the international financial market.

The performance of the model validation process requires an evaluation of its discriminatory power, with a simultaneous verification of its calibration. An assessment of its discriminatory power without an evaluation of its calibration is not sufficient in the context of the entire validation process, however for the

purpose of this article, the focus is put on a comparison of the quality of prediction. Moreover, as many publications point at the lack of a clear validation procedure and recommendations for the use of specific statistics, the selection of the relevant measures in the performance of this process is decided by the validator.

The purposes of this article are to present validation techniques according to their discriminatory power, while indicating the reservations about such techniques, and to check the adjustment of the existing Polish bankruptcy prediction models in the context of their discriminatory power. This is the first study that performs a validation of such models. It must also be noted that the subject matter of this paper is important and up-to-date. First, a verification of the quality of the models is the duty of the financial institutions in the context of the implementation of the recommendations contained in Recommendation W issued by the KNF, effective as of 30 June 2016. Moreover, the performance of a qualitative and quantitative validation assists in the risk management process.

Based on publications on this subject matter, the same set of statistics used in the evaluation of discriminatory power of rating models has been identified which includes, among others, the CAP (Cumulative Accuracy Profile) curve and the area under its AR curve (referred to as the Gini index), the ROC (Receiver Operating Characteristics) curve and the area under its AUROC curve, the Kolmogorov-Smirnov statistics, the Brier result, Information Value (IV), information entropy, and the Pietra ratio. An innovation among the validation tools is the H measure, used in evaluations of the discriminatory quality of a scoring model by Lessman, Seow, Bae-sens, and Thomas [2013], which is based on beta distribution. Due to their frequently emphasized popularity and universal use, the decision was made to evaluate the Polish bankruptcy prediction models using those measures, too.

The article has the following structure: the first chapter presents a review of the empirical literature on the validation of the PD (probability of default) model and the second chapter presents a review of Polish bankruptcy models based on discrimination techniques. The third chapter presents the methodology and the final describes the test performed.

2. A review of foreign empirical literature on the validation of the PD model

The literature discussing the process of the validation of the PD model is very broad and diverse. This chapter presents a set of results of the most recent studies on the measurement of the discriminatory power of bankruptcy prediction scoring models.

The literature puts a strong emphasis on the impact of the provisions of Basel II on the deepening of the process of risk management in the banking sector. This process involves, among others, the universal practice of the validation of the models, in particular the scoring and rating models used, which are used in institutions of that sector. All the aspects of the validations performed on the studied models make it

possible to reduce credit risk, use more accurate forecasts concerning the behavior of the portfolio in the future, and ensure better protection in the event of future economic crises.

This article focuses only on studies of the effectiveness of models, defined as the ability to properly classify the analyzed entities as bankrupt companies and sound companies. Models that properly differentiate between those two groups are considered to be positively discriminating. However, one must keep in mind that an objective evaluation of the quality of the model of bankruptcy prediction should also be based on the performance of a calibration model, as indicated, among others, by Tasche [2006], Bluemke [2014], and Bloechinger [2012]. Calibration is a test of conformity of the forecast probability of bankruptcy (PD) with the achieved probability (default rate) for the entire spectrum of available data. A model in which the number of observed bankruptcies conforms to the number of forecast bankruptcies in a given period is considered to be well calibrated. Calibration is used only for risk measures expressed in the quotient scale (e.g. the probability of bankruptcy), while discrimination is associated with nominal risk measures (e.g. rating classes).

The source that is most often cited in studies on the validation of bankruptcy models is the article by Tasche [2006], which focuses on quantitative validation complying with the requirements of the regulatory authorities supervising financial markets. Two particularly important regulations provided for in Basel II, related to quantitative validation using statistical methods, are the following: first, banks must have resilient systems that enable a validation of the accuracy and cohesion of their rating systems, their processes, and their estimation of all relevant components of risk; second, banks must demonstrate to their regulatory authorities that their use of their internal validation process makes it possible to evaluate the operation of their internal rating system and their risk estimation system in a cohesive and comprehensive manner.

A number of obligatory characteristics of the quantitative validation process have also been enumerated. Most of all, the fundamental purpose of validation is to evaluate the predictive power of the risk estimators used by banks and the use of ratings in the credit process. Secondly, validation should be an iterative process without a single specified method, instead it should comprise both quantitative and qualitative elements. Banks are the principal entities required to conduct this process, which, together with its results, should undergo an independent verification.

It was observed that there are a number of different statistical tools used to measure the discriminatory power of models, however only those that are most frequently used were focused on. The key characteristic that makes it possible to differentiate those tools is whether their use requires an estimation of the probability of insolvency for the entire portfolio (unconditional probability). If so, the tool may be used only for samples with an appropriate number of bankruptcies, otherwise it can also be used for samples that are not representative. In addition, the use of a specific tool may be strongly dependent on the intended purpose of its use.

Studies on validation indicate the use of a similar set of measures of the quality of discriminating scoring models. Those most often include the ROC (Receiver Operating Characteristic) and CAP (Cumulative Accuracy Profile) curves, which are a graphic representation of the relationship between the specificity and the sensitivity of models, along with the measures of the areas under their curves: AR (Accuracy Ratio), which is the value of the Gini index, for the curve CAP and AUROC (Area Under ROC). Moreover, the studies indicate the use of the Kolmogorov-Smirnov statistics, the Pietra ratio, and the d-Somers statistics.

Another group of indicators used in the studies of the quality of the discriminatory power of models is IV (Information Value), also referred to as the divergence or stability coefficient, and CIER (Conditional Information Entropy Ratio).

Some of the studies also use the Bayes error rate. The Bayes error rate is the probability of an incorrect classification of an observation in one group using a naive Bayes classifier, in accordance with the following formula:

$$p = \int_{x \in H_i} \sum_{C_i \neq C_{max}x} P(C_i|x)p(x)dx,$$

where x is an observation (event); C_i is the class in which the event is classified; H_i is the area that the classification function h treats as C_i . The value of the error is other than zero if the target groups are not deterministic, i.e. when there is a non-zero probability that a specific event belongs in more than one group.

Another publication with a large number of citations, which focuses on the evaluation of rating models expressed as discriminatory power, is the paper by Engelmann, Hayden, and Tasche [2003]. The article focuses on a comparison of the properties of the ROC and CAP curves and on the measures of the areas under those curves (AR and AUROC). It was noted that the calculated values are of a low statistical value; consequently, an objective evaluation of model quality requires the identification of the appropriate confidence intervals for those measures. It was also noted that it is not sufficient to compare the absolute values of the two models calibrated using the same data set. In order to perform a comparative analysis it is suggested to use a statistical test of equality of those areas. However, one must keep in mind that when designing such a test or identifying the confidence intervals, one must make an assumption concerning asymptotic normality, which is difficult to achieve in the case of samples with a small number of bankruptcies.

Performance of such a statistical flavour is based on the DeLong approach, in which the test statistics originate from the chi-square distribution with one degree of freedom and uses the calculation of variance and covariance for the areas under the ROC curve for the two different models estimated using the same sample. It can be assumed that in the case of large data sets, the assumption concerning asymptotic normality is correct.

Medema [2009] indicates that there are no universally accepted and used validation methods, even though this process is considered to be very important. Validation requires, among others, the introduction of measurable expectations concerning the force of the impact of changes in economic conditions. Usually the dynamics of those effects are not considered when building the model, and the modelling process is burdened by the fact of missing observations and by the insufficient history of creditworthiness indicators. In some cases it is not possible to build a statistical model, instead expert models are used. In the evaluation of the model, both management, qualitative, and quantitative evaluation is important. Statistical (quantitative) evaluation makes it possible to obtain objective results and to use identical assumptions for all the models being analysed. Given this approach, three stages of validation are proposed: analysis of theoretical assumptions, verification of quality, completeness and representativeness of data, and evaluation of statistical properties. The latter consists in the possibility to replicate the basic study and to analyse the stability of the parameters, in the selection of the functional form, the evaluation of the discriminatory power, the calibration, the behaviour in an out-of-sample period, and in monitoring.

The purpose of the article is to present the proposed three-stage method of the validation of models using the example of a logit model, on data coming from the Dutch commercial bank, Friesland Bank. The indicator of the measurement of the quality of the discriminatory power of the model PD is the ROC curve, the concordance coefficients, and the Brier result. The selection of these indicators was due to their common use in different studies presented in the literature on this subject matter. Based on the values of the statistics that were obtained, it was demonstrated that the logit model was characterised by a greater discriminatory power than the model based on splines.

In another study [Bluemke 2004], it was indicated that the most popular measures of the discriminatory power of rating models are AUROC, AR, the Gini index, and the Somers d statistic, which can be considered as equivalent or related to each other using easy transformations. The equivalence of the information provided by those statistics was considered to be a shortcoming; however, the ability to match selected indicators to the nature of the analysis was appreciated. It was emphasised that it is important to differentiate models that are weak due to their low discriminatory power, and those that are statistically weak due to fluctuations in the lifecycle of the model and the lifecycle of loans. This means that cycles influence both the value of the observed bankruptcy rate and the discriminatory power of the rating model and this relationship is inversely proportional: the higher the values of the observed bankruptcy rate, the lower the anticipated discriminatory power of the model. Understanding of this relationship also makes it possible to study the constancy of the discriminatory power. This claim was verified in a study based on an analysis of a single-factor credit risk model based on the CreditPro rating data from the S&P Capital IQ database available to American companies outside of the financial sector,

using the Somers d statistic. The suggested conclusion from the study is that, due to the presence of the relationship between recurring fluctuations in the model, a comparison of the absolute values of statistical measures is not sufficient for the evaluation of the discriminatory power of rating models. The estimations obtained should be compared to the threshold values (benchmark) calculated taking into account the credit lifecycle.

Bloechinger [2012], in turn, recommends new measures of discrimination and calibration that make it possible to verify the quality of prediction in the case of lacking assumptions concerning the independence of bankruptcy events. The stated reason for the need to introduce new measures is the shortcomings of the tests that have been used so far. The shortcomings include an unrealistic assumption concerning stochastic independence, use of improperly estimated asymptotic distributions due to the low frequency of bankruptcy, the need to use numerical methods in the case of some tests, the detrimental grouping phenomenon, and the possibility to calculate statistics only for one moment in time. Therefore, the recommended statistics are based on opposite values. The analysed approach was applied to the S&P rating model and the Merton model using the data available in the Bloomberg service for the non-financial companies' sector for the period starting in January 1981 and ending in December 2010. In the validation process it was indicated that both models have similar discriminatory properties, however the PD estimations in the Merton model are better calibrated.

Kruger [2015] also points to the fact that bankruptcy is infrequent and that estimates using data with a low rate of observations of insolvency are not realistic. It was suggested that the best way to validate and monitor the PD parameters in such portfolios is to use the Bayes approach because it enables a more accurate determination of the boundaries of the confidence intervals for the PD parameter. The study of the confidence intervals was performed for a generalised statistical model based on the approach adopted by Tasche [2003], which was estimated using the method of the highest credibility for a multi-period sample without any bankruptcy events, a multi-period sample with a low bankruptcy rate, and one-period samples with and without any bankruptcies.

Lingo and Winkler [2010], on the other hand, found that two popular measures of the discriminatory power of models, i.e. the Gini and AUROC coefficients, from a probability standpoint, depend on the structure of the portfolio and are stochastic, which prevents their direct use in comparisons of different portfolio. However, both measures are commonly used and their popularity may be due to the fact that they can be easily used by banks implementing desirable credit policies in order to maximise the economic benefits achieved as a result of the operation of a good scoring model. The analysis that was conducted indicates that it is counterintuitive to check the discriminatory power if the purpose of assigning a rating score is to estimate the probability of bankruptcy. In this case, high granularity and good quality of calibration are sufficient to maximise the possible discriminatory power.

Even though from a probability standpoint, it is not possible to directly compare models built based on different models using the Gini coefficient and the AUROC measure, a new method of comparative analysis was proposed which enables the selection of a model that is better with regard to its discriminatory power. Moreover, it was indicated that the analysed measures also provide information on the quality of calibration of the rating model, if it enables the estimation of the probability of insolvency. As in previous articles, it was noted that studies of discriminatory quality should not be separated from calibration, which means that the calculation of discriminatory power only is insufficient in the context of the validation process.

In the study by Orth [2010], attention was brought to a shortcoming of the measures of the quality of forecasts of insolvency in the rating models, AR and AUROC, namely the fact that insolvency is considered to be a binary variable without consideration of the time of its occurrence which prevents the use of the full information contained in censored observations. In order to solve this problem in this study, a new measure of discriminatory power of the model was proposed which is related to AR but free from its shortcomings.

In order to calculate the AR measure, one must select a present time horizon and classify the observations into one of the two groups: bankrupt or sound. This type of classification results in a loss of information because the time of the event is disregarded, and the right censored observations are skipped, which results in a reduced adequacy of the forecast for a given time horizon. Therefore it is recommended to use Harrell's C concordance index, which is used in literature on biological statistics for the purpose of the evaluation of accuracy of predictions. However this measure is still susceptible to the sample error: an analysis of this variance affects the selection of the confidence interval and on testing of hypotheses.

Even though, as a rule, various authors consider the CAP and ROC curves to be equivalent, Irwin and Irwin [2012] compared the properties of both curves and indicated that the properties of the ROC curve are better. Both curves are used for the purpose of the diagnosis of models that differentiate between the two types of status of the world and are based on the assumption that the diagnosis presents the trade-off between true and incorrect predictions, which depends on the threshold value for which a decision is made to emit an alarm signal. A well-diagnosed model is characterised by a high ratio of correct predictions, regardless of the number of incorrect predictions.

The analyses performed using those curves are similar. Both curves generate a measure of the quality of the model, which depends on the selection of the cut-off threshold. Since they take into account all the possible threshold values, both measure the model's ability to differentiate between bankrupt entities and "sound" debtors, which does not depend on the rating model used. The clear advantages of the ROC curve that cannot be transferred to the CAP curve are, first, the ease of interpretation of the AUROC measure and, second, the resistance of the shape of the ROC curve to the expected probability. Moreover, due to the fact that the ROC curve

is used more commonly than the CAP curve, the former has a theory of selection of optimum threshold value that maximises the expected benefits of a decision, also the CAP curve and the AR measure associated with it are similar to other measures that are commonly used in studies of income inequality: the Lorenz curve and the Gini index.

Irwin and Irwin recommend using the ROC curve instead of the CAP curve because, most of all, the measure of the area under the former curve has a natural interpretation as the unburdened percentage of correct decisions, which cannot be said about the measure of the area under the CAP curve, and because the ROC curve is independent of the probability of insolvency.

Prorokowski [2016] describes a number of measures based on ranks used in the validation of the discriminatory power of various parameters used by banks to evaluate the prediction of insolvency, including its probability. Even though the requirement established by the regulatory authorities in the financial markets concerning the performance of verifications of the discriminatory power of models has been in force for over a decade, there is no formal definition of the discriminatory power or predictive power of a model. In general, the operation of credit risk models is evaluated ex-post based on the results of the operation of the model on a portfolio.

The validation measures mentioned by Prorokowski include the CAP curve with AR – the measure of the area under the curve, which is the Gini index, the ROC curve, the hit ratio, the false alarm rate, the Somers d statistic, the Spearman tau, the Kolmogorov-Smirnov test, and the Kendall tau. Because the ROC and CAP curves are not comparable in time and for different portfolios, it is possible to check if the PD model has any discriminatory power. In this case the tested hypothesis is $AR = 0$.

A number of threats associated with the use of statistics based on ranks have been highlighted. First, measures AR and AUROC contain the same information. When one of the dependent variables is binary, then AR is equivalent to the Somers d statistic. In the case of a continuous variable, the Goodman-Kruskal gamma measure, the Kendall tau measure, and the Somers d measure contain the same information. Second, since the literature on this subject matter presents many versions and examples of use of the Gini index, in order to avoid ambiguity it is recommended to use the AR or Somers d measure. Third, AR and AUROC are sensitive to the object measured and consequently, should not be interpreted without detailed knowledge about the portfolio on which they are based.

Moreover, a comparative analysis of the discriminatory power of different models applied to the same dataset should not be limited to the calculation of the correlation coefficient. It is suggested to determine the confidentiality level for the difference between the two rating models. A regular comparison may be misleading because the presence of potential correlation between the models is overlooked in such circumstances. The differences between the models should be analysed statistically with a zero hypothesis that the areas under the ROC curve are equal for the two different models.

The article by Siarka [2011] focuses on the methods that enable the evaluation of the discriminatory quality of models that differentiate between a population of bankrupt entities from the sound observations. Measurement of the quality of discriminatory power is necessary wherever discriminatory methods are used. The measures used in practice include the CAP curve, again, with the measurement of the area AR under the curve, the ROC curve with the AUROC measure, and the Mann-Whitney statistic. Additionally, in order to evaluate the statistical difference between two areas under the ROC curve, the DeLong approach is used with the zero hypothesis that the two areas are not significantly different, which means that there are no grounds for claiming that a different model better discriminates observations. As described in the article by Engelmann, Hayden, and Tasche [2003], in this case, a T statistic with asymptotic distribution chi-square with one degree of freedom.

In conclusion, it was found that the validation process should include both an evaluation of the discriminatory power of the model and a calibration of its parameters. An assessment of discriminatory power without an evaluation of its calibration is not sufficient in the context of the entire validation process, however for the purpose of this article a focus is made on a comparison of the quality of prediction. The performance of a calibration would not be possible due to the short history and the lack of possibility to estimate the probability of insolvency. Moreover, as many publications point to the lack of a clear validation procedure and recommendations for the use of specific statistics, selection of the relevant measures in the performance of this process is decided by the validator.

3. Review of Polish bankruptcy models based on discrimination techniques

The subject matter of the bankruptcy of businesses has only been studied by researchers since the 1990s. As Polish companies gained their first experiences related to bankruptcies and the associated problems, the interest of researchers in the topic of bankruptcy and their intent to explain and forecast it has increased. The first Polish model using multidimensional analysis was the model developed by Mączyńska i in. 1994.

$$Z^M = 1.5 \times X_1 + 0.08 \times X_2 + 10 \times X_3 + 5 \times X_4 + 0.3 \times X_5 + 0.1 \times X_6,$$

where: X_1 – (Gross result + Amortization) / Liabilities,

X_2 – Total Assets / Liabilities,

X_3 – Operating Result / Total Assets,

X_4 – Operating Result / Sales,

X_5 – Stocks / Sales,

X_6 – Total Assets / Sales.

Another early example of the use of discrimination analysis in the context of the bankruptcy of businesses is the study by Hadasik and Sojak [1995]. The analysis was conducted on a sample of 10 businesses from the then Wrocławskie Province “in whose case there were suspicions that they would go bankrupt in 1993.” Six of those businesses did, in fact, go bankrupt, while four continued their operations. Four of the analysed businesses conducted their activities in the industrial sector and two in each of the following sectors: construction, agriculture and retail.

$$Z^{PS} = 0.644741 \times X_1 + 0.912304 \times X_2,$$

where: X_1 – (Current Assets – Stocks) / Short-term Liabilities, X_2 – Gross results / Sales.

The first important study of the bankruptcy of Polish businesses was the models prepared by Gajdka and Stos in 1996. The authors estimated a total of four models: the first two based on 40 businesses from different sectors, precisely half of which were bankrupt. The next two models had the same structure, whereby the entities were traded on a stock exchange and conducted activities in the industrial, construction, and retail sectors. The analysis conducted by the authors focused on twenty predefined financial indicators calculated one year prior to their bankruptcy in the years 1994-1995.

$$Z_{-1}^{GS} = 0.01935 \times X_{1_1} + 1.094753 \times X_{1_2} + 0.179052 \times X_{1_3} - 6.35257 \times X_{1_4} + 0.291098 \times X_{1_5},$$

$$Z_{-2}^{GS} = 0.437499 + 0.017803 \times X_{2_1} + 0.588694 \times X_{2_2} + 0.138657 \times X_{2_3} - 4.31026 \times X_{2_4} + 0.437449 \times X_{2_5},$$

$$Z_{-3}^{GS} = 0.20098985 \times X_{3_1} + 0.0013027 \times X_{3_2} + 0.7609754 \times X_{3_3} - 0.9659628 \times X_{3_4} - 0.341096 \times X_{3_5},$$

$$Z_{-4}^{GS} = 0.7732059 - 0.0856425 \times X_{4_1} + 0.000774 \times X_{4_2} + 0.9220985 \times X_{4_3} + 0.6535995 \times X_{4_4} - 0.594687 \times X_{4_5},$$

where: X_{1_1}, X_{2_1} – Current Assets / Short-term Liabilities,
 X_{1_2} – Privileged liabilities / Short-term Liabilities,
 $X_{1_3}, X_{2_3}, X_{3_1}, X_{4_1}$ – Sales / Total Assets (average),
 $X_{1_4}, X_{2_4}, X_{3_3}, X_{4_3}$ – Net results / Total Assets (average annual),
 X_{1_5} – (Net results + Amortisation) / Sales,
 X_{2_2} – Total Liabilities / Total Assets,
 X_{2_5} – (Net results + Interest) / Sales,
 X_{3_2}, X_{4_2} – Short-term Liabilities (average annual) \times 365 / cost of production sold,
 X_{3_4}, X_{4_4} – Gross Results / Sales,
 X_{3_5} – Total Liabilities / Total Assets,
 X_{4_5} – Total Liabilities / Total Assets.

In 1998 Hadasik (Appenzeller) published very interesting results of her study presented in her habilitation dissertation. She analysed a set of models based on companies that in the years 1991-1997, together with their financial reports, filed petitions for bankruptcy with the provincial court in Poznań, Piła, and Leszno, as well as based on companies that continued their operations, which were selected based on their similarity with regard to ownership structure and size. Due to the fact that the financial data of some companies was incomplete, the author decided to use a step discrimination analysis on different samples.

$$Z_{-1}^H = 2.60839 - 2.50761 \times X_5 + 0.0014115 \times X_9 - 0.009252 \times X_{12} + 0.0233545 \times X_{17},$$

$$Z_{-2}^H = 2.76843 + 0.703585 \times X_1 + 1.2966 \times X_2 - 2.21854 \times X_5 + 1.52891 \times X_7 + 0.002543 \times X_9 + 0.0186057 \times X_{12} + 0.0186057 \times X_{17},$$

$$Z_{-3}^H = 1.6238 - 1.3301 \times X_5 + 0.04094 \times X_8 - 0.0038 \times X_{12} + 2.16525 \times X_{14} + 0.0235 \times X_{17},$$

$$Z_{-4}^H = 2.36261 + 0.365425 \times X_1 - 0.765526 \times X_2 - 2.40435 \times X_5 + 1.59079 \times X_7 + 0.0023026 \times X_9 - 0.012783 \times X_{12},$$

$$Z_{-5}^H = 2.41753 - 2.62766 \times X_5 + 0.0013463 \times X_9 - 0.009225 \times X_{12} + 0.0272307 \times X_{17},$$

$$Z_{-6}^H = 2.59323 + 0.335969 \times X_1 - 0.71245 \times X_2 - 2.4716 \times X_5 + 1.46434 \times X_7 + 0.0024607 \times X_9 - 0.013894 \times X_{12} + 0.0243387 \times X_{17},$$

$$Z_{-7}^H = 2.65711 - 2.3001 \times X_5 + 0.00153 \times X_9 - 0.010416 \times X_{12} + 0.0286736 \times X_{17},$$

$$Z_{-8}^H = 1.97095 - 1.98281 \times X_5 + 0.0011843 \times X_9 + 0.180604 \times X_{11} - 0.008478 \times X_{12} + 1.53416 \times X_{14} + 0.0235729 \times X_{17},$$

$$Z_{-9}^H = 2.46506 - 2.37851 \times X_5 + 0.0014124 \times X_9 - 0.00983 \times X_{12} + 0.0297656 \times X_{17},$$

where: X_1 – Current Assets / Short-term Liabilities,

X_2 – (Current Assets – Stocks) / Short-term Liabilities,

X_5 – Total Liabilities / Total Assets,

X_7 – Working Capital / Total Assets,

X_8 – Fixed Assets / Equity,

X_9 – Net Sales / Receivables (average),

X_{11} – Sales / Total Assets,

X_{12} – Stocks \times 365 / Sales,

X_{14} – Net profit / Overall Capital,

X_{17} – Net profit / Stocks.

Another model prepared on a sample of Polish businesses is the model prepared by Wierzba in 2000. In his study the author used financial data of 24 businesses that did not face the risk of bankruptcy and of the same number of businesses that were declared bankrupt or initiated an arrangement procedure in the period starting in January 1995 and ending in April 1998. The limit point below which a business is considered to be facing the risk of bankruptcy was determined to be 0.

$$Z^W = 3.26 \times X_1 + 2.16 \times X_2 + 0.69 \times X_3 + 0.3 \times X_4,$$

where: X_1 – (Operating Result – Amortization) / Total Assets,
 X_2 – (Operating Results – Amortization) / Sales,
 X_3 – Working Capital / Total Assets,
 X_4 – Current Assets / Total Liabilities.

Another example of the use of discrimination analysis in the area of the prediction of the bankruptcy of Polish businesses in 2001 was presented by Hołda who built his model based on an analysis of 80 businesses conducting operations in sectors classified according to the statistical classification of economic activities in the European Community under no. 45 to 74, precisely half of which were businesses that had been declared bankrupt. The time interval of the model is the years 1993-1996.

$$Z^H = 0.605 + 0.681 \times X_1 - 0.0196 \times X_2 + 0.00969 \times X_3 + 0.000672 \times X_4 + 0.157 \times X_5,$$

where: X_1 – Current Assets / Current Liabilities,
 X_2 – Total Liabilities / Total Assets $\times 100\%$,
 X_3 – Net Results / Total Assets (average annual) $\times 100\%$,
 X_4 – Current Liabilities (average annual) / (Costs of Products, Goods and Materials Sold + Selling Costs + Overhead Costs) $\times 360$,
 X_5 – Sales / Total Assets (average annual).

In 2003, Gajdka and Stos published the results of their further studies of a bankruptcy prediction model. They worked on a sample of 34 businesses, 17 of which were defined as bankrupt. All the “sound” businesses were traded on the Stock Exchange for at least three more years. In refining the classification criterion, the authors defined bankruptcy as a situation where a liquidation process was initiated due to a bad economic situation, reaching a court settlement with creditors, or the declaration of a settlement with a bank. The businesses conducted operations in different sectors, including light industry, retail, services, and transport. The researchers used 20 financial indicators that were calculated based on the financial statements prepared one year before bankruptcy was declared, which in this case was the year 1994.

$$Z_{5GS} = -0.3342 - 0.0005 \times X_1 + 2.0552 \times X_2 + 1.726 \times X_3 + 0.1154 \times X_4,$$

where: X_1 – Short-term liabilities (average annual) / Production Cost of Production,

- X_2 – Net Profit / Total Assets (average annual),
 X_3 – Gross Profit / Sales,
 X_4 – Total Assets / Total Liabilities.

The value of the limit point was determined to be 0 and the area of ignorance was determined to be in the range of the value $Z_{5^{GS}}$ from -0.49 to 0.49 . A value of the $Z_{5^{GS}}$ indicator that is less than -0.49 indicates the very bad condition of the business and the high risk of bankruptcy.

A result of the continuation of the work by Appenzeller (Hadasik) was an article published in 2004 together with Szarzec. Their study was conducted on a sample of 34 publically traded companies facing the risk of bankruptcy and of the same number of similar companies of a good financial status. The risk of bankruptcy was identified based on the filing of at least one petition bankruptcy in a court or the initiation of an arrangement procedure in the years 2000-2003, regardless of their legal consequences.

$$Z_{1^{AS}} = -0.661 + 1.286 \times X_{1_1} - 1.305 \times X_{1_2} - 0.226 \times X_{1_3} + 3.015 \times X_{1_4} - 0.005 \times X_{1_5} - 0.009 \times X_{1_6},$$

$$Z_{2^{AS}} = -0.556 + 0.819 \times X_{2_1} + 2.567 \times X_{2_2} - 0.005 \times X_{2_3} - 0.0095 \times X_{2_4} + 0.0006 \times X_{2_5},$$

- where: X_{1_1}, X_{2_1} – Current Assets / Short-term Liabilities,
 X_{1_2} – (Current Assets – Stocks – Short-term Receivables) / Short-term Liabilities,
 X_{1_3} – Gross Result / Sales,
 X_{1_4} – Net Results / Total Assets (average annual),
 X_{1_5}, X_{2_3} – Stocks (average annual) / Sales $\times 365$,
 X_{1_6}, X_{2_4} – Liabilities and Provisions for Liabilities / [(Operating Result + Depreciation) $\times 12$ / fiscal period],
 X_{2_2} – Operating Result / Sales,
 X_{2_5} – Receivables Turnover + Inventory Turnover (in days).

The value of the limit point of both models was determined to be 0.

A frequently cited example of a bankruptcy prediction model is the so-called “Poznań” model, which is the result of the article by Hamrol, Czajka, and Piechocki published in 2004. The model was developed based on a sample of 100 Polish businesses, half of which faced the risk of bankruptcy.

$$Z^{HCP} = -2.368 + 3.562 \times X_1 + 1.588 \times X_2 + 4.288 \times X_3 + 6.791 \times X_4,$$

- where: X_1 – Net Financial Result / Total Assets,
 X_2 – (Working Capital – Stocks) / Short-term Liabilities,
 X_3 – Fixed Capital / Total Assets,
 X_4 – Financial Result on Sales / Sales.

Prusak, too, conducted a study of the bankruptcy of businesses that used a linear discrimination function. He collected a sample of 40 bankrupt manufacturing companies and the same number of companies that continued their operations. The financial data was taken from the financial statements published one year and two years prior to bankruptcy (1998-2002).

$$Z_{-1}^P = -1.5685 + 6.5245 \times X_{1_{-1}} + 0.14 \times X_{1_{-2}} + 0.4061 \times X_{1_{-3}} + 2.1754 \times X_{1_{-4}},$$

$$Z_{-2}^P = -1.8713 + 1.4383 \times X_{2_{-1}} + 0.1878 \times X_{2_{-2}} + 5.0229 \times X_{2_{-3}},$$

where: $X_{1_{-1}}$ – Operating Result / Total Assets (average annuals),
 $X_{1_{-2}}, X_{2_{-2}}$ – (Operating Expenses – Other Operating Expenses) / Short-term Liabilities (average annual) without Special Fund and Short-term Financial Liabilities,
 $X_{1_{-3}}$ – Current Assets / Short-term Liabilities,
 $X_{1_{-4}}$ – Operating Result / Net Sales,
 $X_{2_{-1}}$ – (Net results + Amortization) / Total Liabilities,
 $X_{2_{-3}}$ – Results on Sales / Total Assets (average annuals).

In the first model, the value of the limit point was -0.13 and in the second model, it was -0.295 . Moreover, the author defined an intermediate zone when the value of the model's indicator Z_{-1}^P was in the range of $[-0.13; 0.65]$ and for the Z_{-2}^P model $[-0.7; 0.2]$.

In his further studies, Prusak developed two more models based on a combined test and validation sample for the first pair of models, which was then subject to selection as a result of which 140 small and medium-sized businesses were identified, precisely a half of which were bankrupt.

$$Z_{-3}^P = -1.1760 + 6.9973 \times X_{3_{-1}} + 0.1191 \times X_{3_{-2}} + 0.1932 \times X_{3_{-3}},$$

$$Z_{-4}^P = -0.3758 + 3.7657 \times X_{4_{-1}} + 0.1049 \times X_{4_{-2}} - 1.6765 \times X_{4_{-3}} + 3.523 \times X_{4_{-4}},$$

where: $X_{3_{-1}}, X_{4_{-1}}$ – Result on Sales / Total Assets (average annual),
 $X_{3_{-2}}, X_{4_{-2}}$ – (Operating Costs – Other Operating Costs) / Short-term Liabilities (average annual) without Special Fund and Short-term Financial Liabilities,
 $X_{3_{-3}}$ – Current Assets / Short-term Liabilities,
 $X_{4_{-3}}$ – Short-term Liabilities / Total Assets,
 $X_{4_{-4}}$ – Operating Result / Total Assets (average annual).

One of the most valued and up-to-date examples of studies of the prediction of the bankruptcy of companies is the study by Mączyńska and Zawadzki [2006] conducted at the Institute of Economic Sciences of the Polish Academy of Sciences. The authors selected 40 entities facing the risk of bankruptcy and 40 entities not facing such a risk in the period of 1997-2002 from a sample of the 500 largest

companies traded on the Warsaw Stock Exchange. Out of a group of 45 financial indicators, the number of variables was gradually reduced, thus resulting in a set of seven models with equal limit points 0.

$$Z_{-1}^{MZ} = -9.832 + 5.577 \times X_1 + 1.427 \times X_2 + 0.154 \times X_3 + 0.31 \times X_4 + 1.937 \times X_5 + 1.598 \times X_6 + 3.203 \times X_7 + 0.436 \times X_8 + 0.192 \times X_9 + 0.14 \times X_{10} + 0.386 \times X_{11} + 1.715 \times X_{12},$$

$$Z_{-2}^{MZ} = -0.392 + 5.837 \times X_1 + 2.231 \times X_2 + 0.222 \times X_3 + 0.496 \times X_4 + 0.945 \times X_5 + 2.028 \times X_6 + 3.472 \times X_7 + 0.495 \times X_8 + 0.166 \times X_9 + 0.195 \times X_{10} + 0.03 \times X_{11},$$

$$Z_{-3}^{MZ} = -0.678 + 5.896 \times X_1 + 2.831 \times X_2 + 0.539 \times X_3 + 2.538 \times X_4 + 3.655 \times X_5 + 0.467 \times X_8 + 0.179 \times X_9 + 0.226 \times X_{10} + 0.168 \times X_{11},$$

$$Z_{-4}^{MZ} = -0.593 + 6.029 \times X_1 + 6.546 \times X_2 + 1.546 \times X_3 + 1.463 \times X_4 + 3.585 \times X_5 + 0.172 \times X_{10} + 0.114 \times X_{11},$$

$$Z_{-5}^{MZ} = -1.962 + 9.004 \times X_2 + 1.177 \times X_3 + 1.889 \times X_4 + 3.134 \times X_5 + 0.5 \times X_6 + 0.16 \times X_{10} + 0.749 \times X_{11},$$

$$Z_{-6}^{MZ} = -2.478 + 9.478 \times X_2 + 3.613 \times X_3 + 3.246 \times X_4 + 0.455 \times X_5 + 0.802 \times X_{11},$$

$$Z_{-7}^{MZ} = -1.498 + 9.498 \times X_2 + 3.566 \times X_3 + 2.903 \times X_4 + 0.452 \times X_5,$$

where: X_1 – Growth rate of sales revenues,
 X_2 – Operating Result / Total Assets,
 X_3 – Net Financial Result / Sales,
 X_4 – Accumulated three-year Gross Financial Result / Total Assets,
 X_5 – Equity / Total Assets,
 X_6 – (Equity – Share Capital) / Total Assets,
 X_7 – (Net Financial result + Amortisation) / Total Liabilities,
 X_8 – Operating Results / Financial Costs,
 X_9 – Current Assets / Short-term Liabilities,
 X_{10} – Working capital / Fixed Assets,
 X_{11} – Sales / Total Assets,
 X_{12} – The decimal logarithm of Total Assets.

The empirical studies described in the literature review above are certainly not the complete list of attempts to explain the reasons for bankruptcies of companies in Poland. Due to the topic of the article, only those models that use one-formula discrimination analysis are described, although the model of concentrations by Sojak and Stawicki [2001], the logit models by Gruszczyński [2012], and the neuron networks models by Korol and Prusak [2004] should also be mentioned. A review of Polish bankruptcy models was presented by Prusak [2005].

4. Methodology – presentation of measures used in the model validation process

Each studied entity can be described with two random variables: S and Z . The random variable S means score – a synthetic, one-dimensional indicator of creditworthiness of an entity that may have any value from the set of real numbers ($S \subset R$). The value of the variable can be determined using many statistical methods, including a discrimination function. The commonly adopted convention is that a high value of the score means high creditworthiness (“good” customers), while a low value of the score means low creditworthiness (“bad” customers). This variable can be discredited using another random variable R which means the rating of the entity ($R = \{1, 2, \dots, k\}$).

Let the random variable Z mean the observed status of the entity, which may have one of the following two values: 1 when the entity is insolvent (default) and 0 when the entity continues to be solvent (non-default). Let D mean the status of insolvency ($Z = 1$), while ND means the status of continued solvency of the entity ($Z = 0$). Both kinds of status are mutually separate, i.e.:

$$D \cap ND = \{Z = 1\} \cap \{Z = 0\} = \emptyset.$$

The unconditional probability of default is defined as:

$$p = PD = Pr(D) = Pr[(Z = 1) = 1 - Pr(Z = 0)] \in [0, 1].$$

The above probability is in the closed interval of 0 to 1; however, extreme values are assumed only in special cases. The value PD equal to 0 is assigned only on an expert basis, as in the case of borrowers that are governmental institutions, while the value equal to 1 is observed only for entities that are insolvent.

The total distribution of the two random variables (S, Z) can be described using the distribution of the random variable S that depends on the status of the entity (random variable Z). The functional forms are given in the following manner:

$$F_D(s) = Pr(S \leq s | D) = \frac{Pr(\{S \leq s\} \cap D)}{p} = \int_{-\infty}^s f_D(u) du,$$

$$F_{ND}(s) = Pr(S \leq s | ND) = \frac{Pr(\{S \leq s\} \cap ND)}{1 - p} = \int_{-\infty}^s f_{ND}(u) du.$$

In the literature on this subject matter, for the set value s , the value $F_{ND}(s)$ is referred to as the false alarm rate while the value $F_D(s)$ – as the hit rate.

Based on the claim concerning full probability, one can define the unconditional distribution of the score using the following formulas:

$$f(s) = p \times f_D(s) + (1 - p) \times f_{ND}(s),$$

$$F(s) = p \times F_D(s) + (1 - p) \times F_{ND}(s).$$

In the literature, for the set value s , $F(s)$ is sometimes referred to as the alarm rate.

Pursuant to the Bayes claim that relates unconditional behaviors, one can define the conditional probability of insolvency.

$$Pr(D|S = s) = \frac{p \times f_D(s)}{f(s)} = \frac{p \times f_D(s)}{p \times f_D(s) + (1 - p) \times f_{ND}(s)}.$$

Knowledge of the form of functions $Pr(D|S = s)$ and $F(s)$ makes it possible to define the form of the unconditional probability of insolvency and of the conditional density, using the following relationships:

$$p = \int Pr(D|S = s) \times f(s) ds,$$

$$f_D(s) = \frac{Pr(D|S = s) \times f(s)}{p},$$

$$f_{ND}(s) = \frac{(1 - Pr(D|S = s)) \times f(s)}{1 - p}.$$

All of the above relationships can also be deduced for a discrete random variable R . Higher creditworthiness is usually identified by a lower value of the rating – function $Pr(D|R = s)$ should be growing along with the growth of r . The names of the functions are also different, i.e. $Pr(R = r|D)$ and $Pr(R = r|ND)$ are referred to as the default conditional rating profile and survival conditional rating profile. The function $Pr(D|R = r)$ is defined as the *PD* curve.

The scoring model, like any other classification model, is characterised by a certain discrimination power which may be evaluated and measured. A tool that is often used by practitioners for the purpose of this analysis is the CAP curve, which enables the visual evaluation of the model. The CAP curve shows the share of the entities in the default status (expressed in percentage points) whose score is smaller than or equal to a specific orderly value of the score. Usually, instead of an orderly value of the score, the interval of $[0,1]$ is used, so that the maximum values of both axes on the figure of the curve are equal to 1. In accordance with the assumption, the fact that a higher score corresponds to higher creditworthiness, the entity with the lowest value of the score is in default and the entity with the highest value of the score is not. Therefore the relationships $CAP(0) = 0$ and $CAP(1) = 1$ are always true. In an ideal model (perfect discrimination function), in the case of the lowest value of the score, the share of entities in default would be the equal to zero, while

for any other value, the value would be equal to 1. In the other extreme case for the random model, when the conditional distributions are the same, a straight line would be observed.

Knowledge of the functional values of the unconditional distribution and of two conditional distributions of the score makes it possible to define the functional curve CAP in the following manner:

$$CAP(u) = F_D(F^{-1}(u)) \quad u \in (0,1).$$

The ratio of the size of the area between the CAP curve for the analysed model and the CAP curve for the random model to the size of the area between the CAP curve for the ideal model and the CAP curve for the random model is referred to as the accuracy ratio (AR). It assumes values in the range of 0 (random model) to 1 (ideal model) and is defined by the following formula:

$$AR = \frac{2 \times \int_0^1 CAP(u) du - 1}{1 - p}.$$

Another frequently used characteristic of the predictive power of a model is the ROC curve which illustrates specificity in relation to the sensitivity of the model. In the ideal model, the ROC curve has the shape of the reversed letter “L”. A purely random classification would be illustrated by a straight line inclined at 45°.

The functional form of the ROC curve is defined by the following formula:

$$ROC(u) = F_D(F_N^{-1}(u)) \quad u \in (0,1).$$

The numerical characteristic of the ROC curve is the size of the area under the curve which can be calculated using the integral of the specific function ROC from 0 to 1:

$$AUC = \int_0^1 ROC(u) du.$$

It was noticed that as the discrimination power of the model improves, the values AR and AUC increase and as the discrimination decreases, both of those values also decrease. In the case of a model that perfectly differentiates between both populations of entities $AUC = 1$ and $AR = 1$, while in the case of a model that has no predictive power $AUC = 1/2$, and $AR = 0$. The relationship between AR and AUC was introduced by Engelmann, Hayden, and Tasche [2003] and has the following form:

$$AR = 2 \times AUC - 1.$$

In its publication about the validation of internal rating systems, the Basel Committee on Banking Supervision recommends also using other measures for the validation of the discriminatory power of scoring models such as the test statistic of the Kolmogorov-Smirnov test, the Pietra index, entropy, the information value criterion, and the Kendall τ statistic.

The Pietra index is a value that describes the ROC curve. It is measured as the maximum area of a triangle whose two apexes are located on both ends of a diagonal curve and the third apex is a point located on the line of the ROC curve. Geometrically, it is equal to half of the maximum distance between the ROC curve and the diagonal curve. In the case of a model that perfectly discriminates both groups of entities, the value of the Pietra index is equal to 0.353 ($\sqrt{2}/4$) and in the case of a random model it is close to zero. The Kolmogorov-Smirnov test statistic has a similar definition: it is equal to the maximum distance between two distribution functions for both populations. The value of entropy comes from the information theory and reflects the degree of “uncertainty”: it is the highest when the value of probability is equal to 0.5 and the lowest for zero and one. The information value is similar to the Pietra index and refers to the difference between the distributions of the score for entities that are in default and for the remaining entities. A high information value indicates the high discriminatory power of the model. The Kendall τ value is a measure of the monotonous relationship between the probabilities of both populations.

The DeLong test is performed in order to study the equality of the areas under the ROC curve for the two models estimated on the same set. A zero hypothesis of the test means that the areas under the ROC curve do not significantly differ from each other, which means that there are no grounds for the claim that one of the models is characterised by a higher discriminatory power. For the purpose of the test, the test statistic T is calculated which has a chi square asymptotic distribution with one degree of freedom:

$$T = \frac{(U_1 - U_2)^2}{\hat{\sigma}_{U_1}^2 + \hat{\sigma}_{U_2}^2 - 2\hat{\sigma}_{U_1, U_2}},$$

where $\hat{\sigma}_{U_1, U_2}$ is the value of covariance between the AUROC values for two scoring models. Then the critical value for the assumed level of significance is taken from the chi square distribution tables, which enables verification of the hypothesis. The values U_1 and U_2 in the above formula stand for the Mann-Whitney statistic for the AUROC variable.

The H measure is an alternative way to measure the discriminatory power of the models, which deals with the problem of internal inconsistency, which is characteristic of the popular AUC measure. It is a measure of loss in the case of the erroneous classification of observations and it depends on the relative proportion of entities that are classified into each class, in turn being a characteristic that is opposite to other measures independent in this regard.

The H measure is defined by the following formula:

$$H = 1 - \frac{L}{L_{max}},$$

where L is the value of loss for a distribution of scoring points originating from a monotonous distribution, while L_{max} is the value of the maximum possible loss for this distribution. The maximum possible loss takes place when the model works randomly (with reference to the ROC curve – when it is diagonal).

The values of the H measures are in the $[0;1]$ interval and, like in the case of the previous statistics, higher values indicate a higher discriminatory power of a model.

In conclusion, the scoring model can be perceived as an example of a classification model making it possible to assign to each entity one of two mutually separate statuses: solvency (non-default) or insolvency (default). Based on the general properties of the probability theory, it is possible to define a number of relationships that are interpreted in a special manner in the context of credit risk. Each classification model is characterised by a specific discriminatory and calibration power. In the next sub-chapter the most important Polish bankruptcy models are presented, which are used in this article as an example of the application of the scoring model.

5. Description of the data

The source of the data used in the article is the database of the BISNODE for the years 2005-2015. The first step was to implement the bankruptcy models. Because the data also covered companies that are not publicly traded, it was not possible to define the indicator of market value of fixed capital for all the facilities. As a result it was not possible to define three models described by Altman. Similarly, due to the missing relevant data it was not possible to determine the value of privileged liabilities and, consequently, to calculate one model prepared by Gajdka and Stos. In the case of all remaining 31 models it was possible to estimate a synthetic indicator that, with a potentially good accuracy, distinguishes between bankrupt companies and those in a good financial situation one year prior to the potential bankruptcy.

Further in the article the calculated value of the indicator will be interpreted as the value of the *score* closer to an unspecified scoring model. This approach is based on the assumption that this relationship is (not necessarily linearly) monotonous, i.e. a smaller value of the indicator corresponds to a higher probability of the *default* status. This approach is reasonable because, theoretically, the value of the indicator is strictly related to the future status of the company.

6. Results

In the study, the AUC and AR measures, the Pietra index, the information value index, and the Kolmogorov-Smirnov test were used. The detailed results of the validation are presented in Appendix 1 and the weighted averages, with weights corresponding to the number of observations in the model in the given year, are shown in Table 1.

Table 1. Weighted average of selected measures of power validation of discriminatory models

Model	Index Pietra	Area Under Curve	Accuracy Ratio	Information Value	The test statistic of the Kolmogorov-Smirnov test
Appenzeller, Szarzec 1	0.2660	0.8039	0.6077	482.0641	0.4727
Appenzeller, Szarzec 2	0.2912	0.5171	0.0343	252.7094	0.0883
Gajdka, Stos 2	0.0754	0.1507	-0.6987	840.1429	0.5831
Gajdka, Stos 3	0.1708	0.3830	-0.2340	1940.6451	0.1843
Gajdka, Stos 4	0.1417	0.3846	-0.2309	2028.7811	0.1935
Gajdka, Stos 5	0.2791	0.8751	0.7501	476.6818	0.6477
Holda	0.2987	0.8803	0.7607	1103.7715	0.6302
Hadasik 1	0.2893	0.8946	0.7892	1118.9877	0.6640
Hadasik 2	0.2888	0.7874	0.5747	493.5577	0.4936
Hadasik 3	0.3009	0.8945	0.7891	950.3810	0.6622
Hadasik 4	0.2944	0.9038	0.8076	1072.6344	0.6806
Hadasik 5	0.2977	0.9105	0.8210	1227.2840	0.6993
Hadasik 6	0.2882	0.8916	0.7832	1144.6395	0.6662
Hadasik 7	0.2851	0.8853	0.7706	769.1622	0.6478
Hadasik 8	0.2880	0.8438	0.6876	766.2555	0.5834
Hadasik 9	0.2882	0.8897	0.7794	880.8971	0.6562
Hamrol, Czajka, Piechocki	0.2701	0.8267	0.6535	631.5631	0.5116
Mączyńska 1	0.2906	0.7859	0.5717	451.1902	0.4777
Mączyńska, Zawadzki 1	0.2868	0.8353	0.6707	697.9166	0.5700
Mączyńska, Zawadzki 2	0.2895	0.8373	0.6745	740.2041	0.5836
Mączyńska, Zawadzki 3	0.2872	0.8262	0.6525	670.1934	0.5565
Mączyńska, Zawadzki 4	0.2962	0.8270	0.6539	614.3216	0.5614
Mączyńska, Zawadzki 5	0.2971	0.8277	0.6554	643.6517	0.5443
Mączyńska, Zawadzki 6	0.2969	0.8282	0.6564	640.2911	0.5573
Mączyńska, Zawadzki 7	0.3000	0.8908	0.7816	790.8497	0.6725
Prusak 1	0.2908	0.7976	0.5951	506.7067	0.4889
Prusak 2	0.2757	0.8054	0.6107	492.3673	0.4742
Prusak 3	0.2973	0.7829	0.5658	490.4003	0.4677
Prusak 4	0.2968	0.7876	0.5752	504.4228	0.4783
Pogodzińska, Sojak	0.2343	0.7920	0.5839	750.2237	0.4437
Wierzba	0.2824	0.8186	0.6372	534.0056	0.5008

Source: own calculation.

With regard to the AUC and AR criterion, the best model is the fifth model developed by Hadasik (AUC equal to 0.91 and AR equal to 0.82). The second highest result was observed for the same model in its fourth version. Interestingly, the second model prepared by Appenzeller and Szarzec slightly better classifies entities compared to a purely random classification: in 2012, the values of AUC and AR for the Appenzeller and Szarzec model were equal to 0.503 and 0.01, respectively. The AR values for the second, third, and fourth model developed by Gajdka and Stos for each analysed year are smaller than zero, which means that a classification based on those two models brings results that are the opposite to the expected results: as the value of the score decreases, the real probability of bankruptcy decreases. In a large majority of the remaining models the value of AR is higher than 0, which indicates the high discriminatory power of those models.

The values of the Pietra index and of the test statistic of the Kolmogorov-Smirnov test, as expected, are mutually correlated and their values increase along with an increase in the absolute value of AR. This is due to the fact that the closer the value of AR to zero, the closer the score distributions for both populations of entities and the smaller the measures of the distances between those distributions. The values of the information value are the largest in the case of the fifth model developed by Hadasik, which confirms the high predictive power of this model.

7. Conclusion

The objective of the article was to present the techniques of model valuation according to the discriminatory power and to check the matching of the existing Polish models of bankruptcy prediction to the sectors of the economy in the context of the discriminatory power of those models. An answer was sought to the question of which of the models provides the highest quality of prediction for specific sectors in specific years. The studies conducted so far on Polish models were based on small and heterogeneous samples, as a result of which the results obtained cannot be treated as reliable.

Based on the analysis, it was found that the fifth model developed by Hadasik was characterised by a very high discriminatory power. The decision was made to base the evaluation of the discriminatory power of the modules on the Gini index, the Kolmogorov-Smirnov statistic, the information value (IV), and the precision of the estimates of bankruptcy. This selection of the statistics was due to their diversity and to the fact that, in accordance with the literature on the methods of validation of bankruptcy prediction models, the other statistics studies discussed in the methods contained the same information (especially the AUC and d Somers coefficients, the Gini index, and the Kolmogorov-Smirnov statistic and the Pietra index). The selection of the precision indicator is, on the other hand, a representation of the obtained contingency matrices and is important due to the nature of the models used for the purpose of evaluation of their ability to predict bankruptcy.

The study is definitely different than the studies that have been conducted so far by Polish authors due to the use of a broader range of data, their higher homogeneity with reference to the sector of the economy, and due to the use of statistical indicators. The studies conducted so far were based on a study of conformity of the estimations of the models and the real bankruptcies of companies using small and diverse samples. Thus, the present article constitutes an expansion and an implementation of the recommendations contained in those analyses concerning the use of larger and more homogenous data sets.

Based on the literature on this subject matter, validation is not complete if only the discriminatory quality of a model is checked. In order to perform an objective evaluation of a model, which must also be calibrated. Such a comprehensive evaluation can be performed by banks thanks to the observation of the credit portfolio over time. Homogeneity of the portfolio used to develop a specific model is the necessary condition and is required by supervisory bodies.

In conclusion, the performance of a validation process is important due to the link between the risk of the models and the credit risk, which in turn influences the process of risk management and the stability of the bank's operation in adverse economic conditions. However, this process requires keeping a number of assumptions that enable objective evaluation of the quality of the specific model. Most of all, as has been observed in this study, it is of key importance to ensure homogeneity of the portfolio in time and with regard to the factors affecting the risk level.

In the future, the study can be expanded to include an analysis of the changes of the values of the statistics used in time and of the conformity of those changes to the expectations associated with the phases of the business cycle. This matter has not been verified so far and a check of the stability of the model in time is required by the supervisory bodies in the validation process.

Bibliography

- Altman E.I., 1968, *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, The Journal of Finance, pp. 589-609.
- Altman E.I., 1983, *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*, John Wiley & Sons.
- Altman E.I., Haldeman R.G., Narayanan P., 1977, *ZETATM analysis. A new model to identify bankruptcy risk of corporations*, Journal of Banking and Finance, pp. 29-54.
- Altman E.I., Hotchkiss E., 2007, *Trudności finansowe a upadłość firm*, CeDeWu, Warszawa, p. 239.
- Anderson R., 2007, *The Credit Scoring Toolkit. Theory and Practice for Retail Credit Risk Management and Decision Automation*, Oxford.
- Antonowicz P., 2007, *Metody oceny i prognoza kondycji ekonomiczno-finansowej przedsiębiorstw*, Gdańsk.
- Appenzeller D., Szarzec K., 2004, *Prognozowanie zagrożenia upadłością polskich spółek publicznych*, Rynek Terminowy, nr 1, pp. 120-128.
- Bank for International Settlements. Studies on the Validation of Internal Rating Systems, May 2005, http://www.bis.org/publ/bcbs_wp14.pdf.

- Bloechlinger A., 2012, *Validation of Default Probabilities*, Cambridge.
- Bluemke O., 2014, *On the negative correlation between default rates and the discriminatory power of credit ratings*, The Journal of Fixed Income, Fall, <http://www.ijournals.com/doi/abs/10.3905/jfi.2014.24.2.019?journalCode=jfi>.
- Dyrektywa Parlamentu Europejskiego i Rady 2013/36/UE z dnia 26 czerwca 2013 r. w sprawie warunków dopuszczenia instytucji kredytowych do działalności oraz nadzoru ostrożnościowego nad instytucjami kredytowymi i firmami inwestycyjnymi, zmieniająca dyrektywę 2002/87/WE i uchylająca dyrektywy 2006/48/WE oraz 2006/49/WE.
- Engelmann B., Hayden E., Tasche D., 2003, *Testing rating accuracy*, <https://www.esma.europa.eu/file/15962/download?token=2z1UXVZH> (1.09.2016).
- Gajdka J., Stos D., 1996, *Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa spółki*, [in:] Duraj J. (ed.), *Przedsiębiorstwo na rynku kapitałowym*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Generalny Inspektorat Nadzoru Bankowego. Walidacja zaawansowanych metod wyliczania wymogów kapitałowych z tytułu ryzyka kredytowego i operacyjnego. Dokument konsultacyjny, Warszawa 2006.
- Generalny Inspektorat Nadzoru Bankowego. Konsultacje i wdrożenie postanowień Nowej Umowy Kapitałowej w sektorze bankowym w Polsce. Dokument konsultacyjny, Warszawa 2005, https://www.knf.gov.pl/Images/dokument_konsultacyjny_1_tem75-4763.pdf.
- Godlewska S., 2010, *Skuteczność polskich modeli dyskryminacyjnych w ocenie zagrożenia upadłością spółek giełdowych*, Lublin.
- Gruszczyński M., 2003, *Modele mikroekonometrii w analizie i prognozowaniu zagrożenia finansowego przedsiębiorstw*, Studia Ekonomiczne, nr 34, Wydawnictwo INE PAN, Warszawa.
- Hamrol M., Chodakowski J., 2008, *Prognozowanie zagrożenia finansowego przedsiębiorstwa. Wartość predykcyjna polskich modeli analizy dyskryminacyjnej*, Wrocław.
- Hamrol M., Czajka B., Piechocki M., 2004, *Upadłość przedsiębiorstwa – model analizy dyskryminacyjnej*, Przegląd Organizacji, nr 6, pp. 35-39.
- Hand D.J., *Measuring classifier performance: A coherent alternative to the area under the ROC curve*, Machine Learning, October 2009, <http://link.springer.com/article/10.1007/s10994-009-5119-5>.
- Hołda A., 2001, *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej ZH*, Rachunkowość, nr 5, pp. 306-310.
- Irwin J., Irwin T., 2012, *Appraising credit ratings: Does the CAP fit better than the ROC?*, International Monetary Fund, <https://www.imf.org/external/pubs/ft/wp/2012/wp12122.pdf>.
- Jagiello R., 2013, *Analiza dyskryminacyjna i analiza logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw*, Materiały i Studia, Warszawa.
- Kisielińska J., Waszkowski A., *Polskie modele do prognozowania bankructwa przedsiębiorstw i ich weryfikacja*, http://www.wne.sggw.pl/czasopisma/pdf/EIOGZ_2010_nr82_s17.pdf.
- Komisja Nadzoru Finansowego. Rekomendacja J dotycząca zasad gromadzenia i przetwarzania przez banki danych o nieruchomościach, Warszawa 2012.
- Komisja Nadzoru Finansowego. Rekomendacja T dotycząca dobrych praktyk w zakresie zarządzania ryzykiem detalicznych ekspozycji kredytowych, Warszawa 2013.
- Korol T., Prusak B., 2005, *Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji*, CeDeWu.pl Wydawnictwo Fachowe, Warszawa.
- Kruger M., *Validation and monitoring of PD models for low default portfolios using PROC MCM*, https://www.researchgate.net/publication/282074223_Validation_and_monitoring_of_PD_models_for_low_default_portfolios_using_PROC_MCMC.
- Lessman S., Seow H.-V., Baesens B., Thomas C.L., 2013, *Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update*, http://www.business-school.ed.ac.uk/waf/crc_archive/2013/42.pdf.

- Lingo M., 2008, *Discriminatory power: an obsolete validation criterion?*, DefaulRisk.com, February, http://www.defaultrisk.com/pp_test_42.htm.
- Mazurczak A., Turek-Radwan M., 2013, *Skuteczność modeli predykcji bankructwa opracowanych w polskich ośrodkach naukowych. Metody i techniki diagnozowania w doskonaleniu organizacji*, Kraków.
- Mączyńska E., 1994, *Ocena kondycji przedsiębiorstwa (uproszczone metody)*, *Życie Gospodarcze*, nr 38, pp. 42-45.
- Mączyńska E., Zawadzki M., 2006, *Dyskryminacyjne modele predykcji bankructwa przedsiębiorstw*, *Ekonomista*.
- Medema L., Koning R.H., Lensink R., Medema M., 2009, *A practical approach to validating a PD model*, *Journal of Banking & Finance*, 33(4), pp. 701-708.
- Orth W., 2012, *The Predictive Accuracy of Credit Ratings: Measurement and Statistical Inference*, *International Journal of Forecasting*, January-March. <http://www.sciencedirect.com/science/article/pii/S0169207011001014>.
- Pociecha J. (ed.), 2014, *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*, Kraków.
- Pogodzińska M., Sojak S., 1995, *Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw*, AUNC, *Ekonomia XXV*, z. 299, Toruń.
- Prorokowski L., 2016, *Rank-order statistics for validating discriminative power of credit risk models*, *Bank i Kredyt*.
- Prusak B., 2005, *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Warszawa.
- Rezac F., Rezac M., 2011, *How to measure the quality of credit scoring models*, *Czech Journal of Economics and Finance*.
- Rozporządzenie Parlamentu Europejskiego i Rady UE nr 575/2013 z dnia 26 czerwca 2013 r. w sprawie wymogów ostrożnościowych dla instytucji kredytowych i firm inwestycyjnych, zmieniające rozporządzenie UE nr 648/2012.
- Siarka P., 2011, *Quality measures of scoring models*, *Journal of Risk Management in Financial Institutions*, London.
- Siedlecka U., 1996, *Prognozowanie ostrzegawcze w gospodarce*, Warszawa.
- Sojak S., Stawicki J., 2001, *Wykorzystanie metod taksonomicznych do oceny kondycji ekonomicznej przedsiębiorstw*, [in:] Bednarski L. (ed.), *Zeszyty Teoretyczne Rachunkowości*, vol. 3(59), Warszawa, pp. 56-67.
- Tasche D., 2006, *Validation of internal rating systems and PD estimates*, [in:] *The Analytics of Risk Model Validation*, Elsevier, Cambridge, UK.
- Tasche D., 2016, *Validation of internal ratings and PD estimates*, <https://arxiv.org/pdf/physics/0606071.pdf>.
- Wojnar J., 2014, *Ocena skuteczności modeli analizy dyskryminacyjnej do prognozowania zagrożenia finansowego spółek giełdowych*, Tarnów.
- Wysiński P., 2013, *Zastosowanie scoringu kredytowego w zarządzaniu ryzykiem kredytowym*, *Biznes międzynarodowy w gospodarce globalnej*. *International Business and Global Economy*, Gdańsk.

OCENA SIŁY DYSKRYMINACYJNEJ WYBRANYCH POLSKICH MODELI PREDYKCJI BANKRUCTWA W RAMACH PROCESU WALIDACJI

Streszczenie: Celem artykułu było przedstawienie technik walidacji ze względu na moc dyskryminacyjną, jednocześnie ze wskazaniem zastrzeżenia powyższych technik, oraz sprawdzenie dopasowania istniejących polskich modeli predykcji bankructwa w kontekście zdolności dyskryminacyjnych. Jest to pierwsze badanie, które przeprowadza walidację takich modeli. Na podstawie analizy uzyskano, że piąty model opracowany przez Hadasik charakteryzował się bardzo wysoką zdolnością dyskryminacyjną. Ocenę siły dyskryminacyjnej modeli zdecydowano się oprzeć na współczynniku Giniego, statystyce Kołmogorowa-Smirnova, mierze H, wartości informacji IV oraz na precyzji oszacowań bankructwa.

Słowa kluczowe: prognozowanie bankructwa, moc dyskryminacyjna, walidacja.