STATISTICS IN TRANSITION new series, June 2026
Vol. 27, No. 2, pp. xxx–xxx, https://doi.org/10.59139/stattrans-2026-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Effectiveness of bankruptcy prediction models constructed for differently selected diagnostic variables

Bernard Kokczyński[1], **Dorota Witkowska**[2]

Abstract

The purpose of the article is to compare the effectiveness of discriminant models for predicting corporate bankruptcy, constructed using different methods of diagnostic variables selection. We compared several methods, such as arbitrary selection, the forward stepwise method applied after the initial selection of variables by means of a significance test of differences between group averages (further called the two-step method), the Hellwig method, the *tstatistics* and the backward stepwise method. We also assessed the models' accuracy in terms of the synthetic measure. It was constructed by applying eight measurements of the classification effectiveness, such as the values of Wilks' lambda statistic and AUC together with the percentage of correctly identified companies, i.e. total, bankrupts and non-bankrupts in training and testing sets. The results show that the backward stepwise method and the two-step method generate models with the highest accuracy of classification. In addition, the study found that Wilks' lambda statistic is not a good approximation of the classification abilities of bankruptcy models. The contribution of our paper is a comparative methodological study, focusing on the impact of alternative diagnostic variable selection techniques on the linear discriminant function accuracy used to bankruptcy prediction.

Key words: diagnostic variable selection, linear discriminant function, bankruptcy prediction, Euclidean distance.

1. Introduction

Shi and Li (2019) list the currently popular bankruptcy prediction methods, i.e.: logistic regression, classification trees and artificial neural networks. In contrast, Kopczyński (2022) emphasizes that linear discriminant analysis still retains its relevance and application in predicting corporate bankruptcy, despite the development of

[1] Department of Corporate Financial Management, Faculty of Management, University of Lodz, Poland. E-mail: bernard.kokczynski@gmail.com. ORCID: https://orcid.org/my-orcid?orcid=0000-0002-9379-0376.
[2] Department of Corporate Financial Management, Faculty of Management, University of Lodz, Poland. E-mail: dorota.witkowska@uni.lodz.pl ORCID: 0000-0001-9538-9589.

newer methods. However, regardless of the bankruptcy prediction method used, the key issue remains the choice of diagnostic variables[3].

According to Grzybowska and Karwański (2023), explanatory variables are selected to reflect expected influences based on theory, previous research, and local context in temporal and spatial dimensions. Gruszczyński (2012) points out that the process of selecting diagnostic variables in practice is often limited to looking for such variables that intuitively appear to be causal for the explained variable, which is then subject to empirical verification. Witkowska (2023) emphasizes that independent variables should comprehensively describe the most important aspects of the phenomenon under analysis, providing general information about individual units rather than unique data. Many economic variables are characterized by high mutual correlation, which leads to information redundancy. In such a case, there is a need to reduce the number of diagnostic variables, which can also be caused by a limited sample size or a large number of estimated parameters.

According to Woo Ahn (2016), the problem of selecting the optimal set of variables is one of the most active areas of research in statistics. This phenomenon is reflected in numerous studies on the development of new methods of variable selection. For example, Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Tibshirani (1996), Smoothly Clipped Absolute Deviation (SCAD) proposed by Fan and Li (2001), Iterative BIC-Based Variable Selection for Model-Based Clustering (IBVS-MBC) proposed by Raftery and Dean (2006), Hierarchical Distance-Based Variable Grouping and Selection (HDB-VGS) proposed by Korzeniewski (2016), Supervised Factor Models (SFM) using constraint optimization proposed by Wang et al. (2023), One Covariate at a Time Multiple Testing (OCTMT) proposed by Chudik et al. (2024), to name a few.

Despite the development of modern machine learning and penalized regression methods, linear discriminant analysis is still widely used in applied bankruptcy studies, particularly in Central and Eastern European research and practice.

The literature provides a number of proposals for the selection of diagnostic variables. Altman (1968) sought to determine the relative contribution of each independent variable to the model, assessed the correlations between explanatory variables, checked the predictive accuracy of the models, and subjected them to judgment as an analyst. Gajdka and Stos (1996) selected discriminating variables by analyzing the correlations between the value of the dependent variable Y and the values of 20 arbitrarily chosen financial indicators, selecting those most correlated with the value of Y. Mączyńska and Zawadzki (2006) created an initial set of 45 diagnostic variables based on the literature and assumptions about the mechanism of financial deterioration of companies. The

[3] In the article, phrases such as discriminant, diagnostic, independent or explanatory variables will be used interchangeably.

selection process included an evaluation of the discriminatory ability of the variables on the basis of: the test of differences of the financial indicators characterizing the groups, the Mahalanobis distance, Wilks' lambda statistic, the accuracy of the classification obtained from the application of single-indicator discriminant models, and the value of the correlation coefficient between the variables. The final selection of indicators also considered expert judgment. Herman (2018) used three variable selection techniques: the method called by him *tstatistics*, the selection of independent variables based on the Spearman correlation coefficient between them and the dependent variable, and the stepwise forward selection of variables, taking into account the variables that reduce Wilks' lambda statistics[4] the most. In the collective work of Valaskov et al. (2023), the results of the significance test of differences between group averages were used to select discriminant variables.

The aim of the paper is to investigate how different variable selection strategies affect model quality which is understood as effective classification of companies. The empirical study compares linear discriminant models constructed on the basis of diagnostic variables distinguished using various techniques for their selection, and indicating which variable selection techniques lead to the most effective linear discriminant models from the point of view of multivariate evaluation of the model validity.

In assessing their effectiveness, eight commonly used accuracy measures were considered, i.e. Wilks' lambda statistic, the value of AUC, the percentage of properly recognized bankrupts, non-bankrupts and all analyzed enterprises in the training and testing samples. The model with the lowest value of Euclidean distance from the defined pattern was considered the best. In other words, the aim of analysis is to point out which method of variable selection specified the most effective bankruptcy prediction models. An additional purpose of the study is to find out whether Wilks' lambda statistic is a good approximation of the discriminatory power of the models.

2. Research methods

In the study a linear discriminant function was used to identify bankrupt and going concern enterprises. Verification of the ability to differentiate between groups of a discriminant model is usually done by using Wilks' lambda statistic, which is calculated as the ratio of the determinant of the within-group variance and covariance matrix to the determinant of the total variance and covariance matrix (Pociecha et al., 2014; Herman, 2018). Many researchers point out that Wilks' lambda statistic can be used as a measure to evaluate the discriminatory ability of classification models. Kopczyński (2022) emphasizes that the value of this statistic makes it possible to assess the effectiveness of the

[4] Before applying the selection techniques, the author eliminated explanatory variables that were highly correlated among themselves (i.e. Spearman correlation coefficient > 0.90).

entire model in separating groups. In a similar vein is Herman (2018), who uses a stepwise selection of forward variables, including in the model those variables that most significantly cause Wilks' lambda value to fall. Mączyńska and Zawadzki (2006) point out that a lower Wilks' lambda value means better discriminatory ability of variables. Similar conclusions are found in the work of Pociecha et al. (2014), where a range of Wilks' lambda values from 0 (best classification ability) to 1 (no discriminatory power) is defined.

In order to verify the discriminatory power of the model, it is testified whether Wilks' lambda is significantly less than unity, i.e. a pair of hypotheses is posed:

$$H_0: \lambda = 1$$
$$H_0: \lambda < 1$$

the test statistic has χ^2 distribution and it is of the form (Jagiełło, 2013):

$$\chi^2 = -\left(N - \frac{k+p}{2} - 1\right) \ln\left(\hat{\lambda}\right) \tag{1}$$

where: $\hat{\lambda}$ is the value of Wilks' lambda statistic estimated from the sample.

Evaluation of the discriminant model efficiency is based on an analysis of prediction accuracy, which can be carried out on a training sample, used to estimate model parameters, or on a test sample, not used in the estimation process, to assess the accuracy of classification. Ptak-Chmielewska (2012) emphasizes that classification using the model distinguishes between an overall classification error and two types of partial errors if the consequences of misclassification into one group are "more costly" than for the other group. When considering the problem of bankruptcy prediction, the classification error of the first type E_1 occurs if the method used misclassifies a bankrupt as a non-bankrupt (2), and the error of the second type E_2 occurs if the model recognizes a non-bankrupt as a bankrupt (3). The overall classification error E is the sum of errors of the first and second type with respect to the total number of observations, reflecting the total percentage of misclassified cases in the analyzed model (4). Errors[5] are most often expressed as a percentage:

$$E_1 = \frac{Fb}{Tb+Fb} \cdot 100 \tag{2}$$
$$E_2 = \frac{Fn}{Tn+Fn} \cdot 100 \tag{3}$$
$$E = \frac{Fb+Fn}{Tb+Tn+Fb+Fn} \cdot 100 \tag{4}$$

where: Fb, Fn - the number of misidentified bankrupts and non-bankrupts, respectively, Tb, Tn – the number of correctly identified bankrupts and non-bankrupts,

[5] Most analyses of the quality of decision rules are based on a comparison of the classification errors of different classifiers. Classification error is not an ideal indicator of a classifier's predictive value. This measure does not consider different a priori probabilities of objects belonging to classes and the problem of unbalanced classes. Misztal (2014) points out that these problems can be partially solved by taking into account different weights or costs of misclassifications, if they can be estimated.

respectively. Thus, the classification efficiency is indicated by the percentage of correctly recognized objects, which completes the classification error to 100%.

Alternatively, the analysis of the classifier quality can be broadened by using additional measures of quality. In this case, the most commonly used are sensitivity (the ability of a classifier to correctly identify bankrupts) and specificity (the ability of a classifier to correctly recognize non-bankrupts). Misztal (2014) emphasizes that these measures are crucial in constructing Receiver Operating Characteristic (ROC) curves. The area under the ROC curve, referred to as AUC, is an integrated measure of a classifier's ability to distinguish between bankrupts and non-bankrupts. An AUC value of 1 indicates a perfect model, while an AUC value of 0.5 indicates a model operating at random. Kopczyński (2022) limits the comparison of two ROC curves solely to comparing the AUC values, omitting their graphical representation. In this case, determining the optimal AUC value becomes crucial. Harańczyk (2010) points out that an AUC value at a certain level is not unambiguously good or bad. It depends on the field and the specifics of the problem under consideration. According to Kopczyński (2022), on the other hand, the AUC values above 0.7, in the context of bankruptcy prediction models, are considered a sign of satisfactory discriminatory ability of the model, while values above 0.6 indicate sufficient discriminatory ability.

2.1. Techniques for selecting discriminating variables for models

The model construction started by applying an arbitrary selection of variables. In the first model, one financial indicator was selected from each of the following groups: liquidity indicators, operating efficiency indicators, debt indicators and profitability indicators. The second model was constructed using a tool based on the generative artificial intelligence Chat GPT (version 3.5), which selected variables based on an implemented list of primary diagnostic variables and an analysis of available literature sources. In both cases, variable selection was based on their "popularity" in the literature, rather than on the statistical properties of the variables derived from the data analysis.

The next technique used for the selection of independent variables, was a two-step method consisting of a test of significance of differences between group means, performed with the SPSS program[6], by one-way ANOVA analysis of variance and consisting of stepwise forward selection of variables. In other words, the following null hypothesis was verified:

$$H_0: E(X_1) = E(X_2) \tag{5}$$

where: $E(X_1), E(X_2)$ denotes the expected values of the random variables derived from the first and second groups, respectively. The test statistic expresses the ratio of the

[6] The study used the SPSS program Statistics version 29.0.0.

between-group variance to the within-group variance, which in the case of two groups can be written as (Malarska, 2005):

$$F = \frac{\sum_{j=1}^{2}(\bar{x}_j - \bar{x})^2 n_j}{\frac{1}{n-2}\sum_{i=1}^{n_j}\sum_{j=1}^{2}(x_{ij} - \bar{x}_j)^2} \tag{6}$$

where: $\sum_{j=1}^{2}(\bar{x}_j - \bar{x})^2 n_j$ is the intergroup variability expressing the variation of the two group averages around the overall average, $\sum_{i=1}^{n_j}\sum_{j=1}^{2}(x_{ij} - \bar{x}_j)^2$ is the within-group variability expressing the variation of individual observations around group averages, $\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$ are group averages calculated for each of the distinguished groups (j=1, 2), $\bar{x} = \frac{1}{n}\sum_{i=1}^{n_j}\sum_{j=1}^{2} x_{ij}$ is the overall average calculated from all observations, n_j is the number of observations in each group, $n = \sum_{j=1}^{2} n_j$ is the number of total observations, x_{ij} is the observation of a distinguished feature in the *i*-th object located in the *j*-th group.

In the first stage of the model building, the variables most differentiating the two groups of companies (bankrupt and non-bankrupt) were identified based on Wilks' lambda and Fisher-Snedecor statistics implemented in the SPSS package. The study assumed a significance level of 0.05. In the second stage, models were constructed using the stepwise progressive method. In this variable selection procedure, at each step, the variable that most contributes to improving the model's classification performance was added to the model, starting with the most important determinant of the phenomenon under study (as in Pociecha et al. (2014) and Herman (2018)).

The further method used in the study, was so-called Hellwig's method of selecting diagnostic variables extensively described in Witkowska's work (2023). One of the formal procedures for selecting diagnostic features, in the analysis of multivariate objects, is based on the **R** correlation matrix, which analyses potential diagnostic features:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1K} \\ r_{21} & 1 & r_{23} & \cdots & r_{2K} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{K1} & r_{K2} & r_{K3} & \cdots & 1 \end{bmatrix} \tag{7}$$

where: for each pair of variables x_i and x_j (i, j = 1, 2, …, K), r_{ij} is Pearson's linear correlation coefficient.

Witkowska (2023) emphasizes that the criterion for the selection of diagnostic variables is based on the formally determined or arbitrarily adopted critical value of the correlation coefficient r^*, which in this study was adopted at the level of 0.8. On this basis, clusters, i.e. subsets of features for which the absolute values of correlation coefficients do not exceed the critical value, are distinguished from the matrix. Within the clusters, one so-called central feature and a number of so-called satellite features are determined. The features in the clusters are called systemic, while the remaining features are referred to as isolated. Central and isolated features may form a set of diagnostic features.

In order to distinguish individual features, the sum of the elements of each column of the correlation matrix **R** is calculated. Then the column p for which the sum of absolute values is the largest is searched for, and the feature corresponding to this column is considered the central feature. In the highlighted column, elements that satisfy the inequality are identified (Witkowska, 2023):

$$\left|r_{pj}\right| \geq r^* \tag{8}$$

The features corresponding to the highlighted rows are considered satellite features. The highlighted columns and rows are then removed from the correlation matrix **R**, resulting in a reduced correlation matrix. This procedure is repeated until the set of features is exhausted, resulting in the identification of more cluster points and the creation of new reduced correlation matrices.

The next method of the variable selection used in the study is the method called by Herman (2018) *t-statistics*. According to this procedure, 5 variables were selected that had the highest absolute values of the Student's t-statistic, in a test comparing the mean value of indicators in the study groups, for independent samples (5). The test statistic is calculated for samples with heterogeneous variances as (Wiktorowicz et al., 2020):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{9}$$

where: $\bar{x}_1; \bar{x}_2$ are averages from samples, S_1^2 ; S_2^2 are the variances from samples, $n_1; n_2$ are sample sizes. The statistic (9) has *t*-Student distribution with $\left(\frac{1}{2} + \frac{S_1^2 S_2^2}{S_1^2 + S_2^2}\right)(n_1 + n_2 - 2)$ degrees of freedom. The rejection of the null hypothesis occurs if *p-value* < 0.05. This approach was supplemented by verification of information redundancy, carried out using Pearson's correlation matrix, where 0.8 was taken as the critical value.

The last method used to select diagnostic variables for discriminant models, which was implemented in SPSS software, was the backward stepwise method. Once all variables are entered, those that meet the elimination criteria are removed until all variables meeting the criteria are exhausted. The process begins with the full set of explanatory variables assumed initially in the theoretical model. Then, in each step, the variable most weakly related to the phenomenon under study is eliminated, and the process continues until only significant variables remain. Once removed, the variable is not included in the model again in further stages of the analysis (Wiktorowicz et al., 2020). In SPSS, independent variables were included or removed from the model based on the value of the Fisher-Snedecor statistic (6).

2.2. Models Evaluation

In the study several methods of discriminant model accuracy assessment are applied, such as: Wilks' lambda value, AUC value, classification efficiency of bankrupts, non-bankrupts and overall in the training and the test samples, assessed as the percentage of properly recognized firms. Values of these measurements were used to construct the synthetic measures in the k-dimensional space as Euclidean distance from the pattern, which is defined as follows:

$$\left[\sum_{j=1}^{k}\left(z_{ij}-z_{0j}\right)^{2}\right]^{\frac{1}{2}} \tag{10}$$

where k is the number of variables used for the models' assessment (k = 2, 4, 6 or 8), z_{ij} and z_{0j} are the standardized[7] variables in the i-th (i =1, 2,…, 8) model and the pattern, respectively. Equal weighting of all accuracy measures was adopted in order to avoid introducing subjective preferences and to ensure a neutral evaluation of model performance.

The Euclidean-distance-based synthetic measure is intended as a decision-support tool rather than a purely statistical construct. It allows for a simultaneous, multivariate assessment of the model performance when several accuracy measures are considered jointly. Such an approach reflects the practical situation faced by analysts, who must evaluate classification models under multiple, often conflicting, criteria instead of relying on a single performance indicator.

In the study two types of patterns were defined – the ideal model and the best hypothetical model. Pattern values for the ideal model were assumed as extremal values showing the perfect discrimination strength or excellent classification ability of the model, whereas the best hypothetical model was characterized by the lowest Wilks' lambda value, highest classification efficiency and highest AUC value among all considered models. Each model was assigned ranks in each category, where the highest (i.e. rank 1) was given to the model for which Euclidean distance from the pattern was the lowest.

To find out which accuracy measurements give similar information, all models were ranked from the best to the worst according to each measurement. Then the Spearman correlation coefficients between the ranks of models ordered due to Wilks' lambda statistics and ordered according to other characteristics of the models were calculated.

3. Data and Results

The analyses were based on data concerning 416 Polish non-financial non-public enterprises. The companies formed a choice-based, matched, and balanced sample.

[7] Standardization formula: $z_{ij} = (x_{ij} - \bar{x}_j)/S_j)$ where: x_{ij} denotes the value of the model assessment measurement, $\bar{x}_j$ and S_j are the average value and standard deviation of the j-the variable (j = 1, 2,…, k), respectively.

Bankrupts were defined as companies that had filed a bankruptcy petition with the court in the years 2019-2022. Non-bankrupts were defined as companies that had not filed a bankruptcy petition with the court, they operated in the same industry and had a similar size of annual revenues as companies considered bankrupt.

Firms included in the research belong to three economic sectors in equal proportions: trade, manufacturing and services. The source of the data used in the study was the Emerging Markets Information Service (EMIS). The extracted financial data of the companies from their financial statements was used to calculate 56 financial indicators, which are the most commonly used in the literature in the context of bankruptcy prediction. These were indicators of liquidity (denoted by the letter P), debt (- Z), operating efficiency (- S), profitability (- R), dynamics (-D), and company size (- W). Based on these, a preliminary set of discriminant features was created from which variable selection was carried out for the construction of discriminant models. Models were estimated based on a training sample in which there were 332 objects (80%), and verified on a test sample with 84 objects (20%). Objects for the test sample were selected by random drawing, using a random number generator in Excel[8]. Companies considered bankrupt in each industry were drawn, and their counterparts in the non-bankrupt group were paired. In order to maintain the representativeness of economic sector participation in the study, the drawing was carried out separately in trade, manufacturing and services. The results of the sampling from each industry were then summed to form a balanced teaching and testing sample, which reflects the industry structure of the available database.

In the course of the study, all potential 56 diagnostic variables were analyzed for their use in building bankruptcy prediction models. For this purpose, the methods for selecting independent variables discussed in subsection 2.1 were used. Table 1 provides a description of each set of diagnostic variables, along with information on the method used to select them.

Among the variables presented in Table 1, the most frequently occurring variable was W03 (logarithm of assets), which appeared in six different sets. The frequently occurring variables were P02 (quick ratio), Z02 (debt-to-equity ratio), R10 (return on average current assets), W02 (logarithm of asset structure) in three sets, and Z04 (share of equity in total assets), R02 (operating profitability to total assets), R04 (net margin), R09 (average gross profit to assets), S01 (asset turnover ratio), S07 (conversion of receivables), which were present in two sets. Variables such as: Z03 (long-term debt to equity), Z05 (proportion of current liabilities to total assets), R01 (EBITDA), R05 (return on equity), R07 (operating profit to assets), R12 (operating margin after depreciation), R13 (return on current assets), S03 (ratio of inventory to operating

[8] The use of Excel was limited to generating random numbers and does not affect the randomness or validity of the sampling procedure.

expenses), S04 (ratio of inventory to sales revenue), S09 (inventory turnover in days), S18 (averaged net cash conversion cycle), S19 (averaged coverage of current liabilities by operating expenses), S20 (averaged asset turnover ratio), W01 (asset structure), D01 (revenue dynamics), D02 (equity dynamics) appeared in only one of the sets.

Table 1. Sets of financial variables selected using various selection methods

Selection method	Symbol of the variable set	List of variables in each set
arbitrary own choice	A	P02, R04, S01, Z02
arbitrary GPT Chat selection	B	P02, R04, S07, W03, Z02
two-step method (variant I)	C	R02, R09, W03
two-step method (variant II)	D	R10, W03
Hellwig's selection method diagnostic variables (central variables)	E	P02, R09, R10, R12, R13, S01, S04, S18, S19
Hellwig's selection method diagnostic variables (isolated variables)	F	D01, D02, R01, R02, R05, S03, S07, S09, W01, W02, W03, Z02, Z03, Z04
tstatistics	G	R07, W02, W03, Z04, Z05
backward stepwise method	H	R10, S20, W02, W03

Source: own work.

Note: D - dynamics, P - liquidity, R - profitability, S - operating efficiency, W - company size, Z - debt.

Table 2. Model estimation results

Discriminant function	Wilks' lambda	AUC
A = 0.001·P02 + 0.007·Z02 + 0·R04 + 0.093·S01 -0.435	0.981	0.500
B = 0·P02 - 0.002·Z02 + 0·R04 + 0·S07 + 1.068·W03 -3.413	0.867	0.723
C = 0.217·R02 + 1.026·W03 + 0.032 ·R09 -3.322	0.870	0.722
D = 1.043·W03 + 0.003·R10 -3.357	0.863	0.730
E = 0·S18 + 0.271·R09 – 0.001·P02 + 0.028·R13 + 0·R14 – 0.018·R16 – 0.181 ·S04 + 0·S19 + 0.002·R10 + 0.243	0.952	0.715
F = -0.002·Z02 + 0.078·Z03 + 0·R01 + 0.129·R02 + 0·R05 – 0.061·S03 + 0·S07 – 0.051·W01 + 0.358·W02 + 0·D01 + 0.034·Z04 + 0.779·W03 – 0.001·D02 + 0 ·S09 -2.133	0.798	0.763
G = 0.873·W03 + 0.074·Z04 + 0.286·W02 + 0.245·R07 + 0.045·Z05 - 2.606	0.856	0.718
H = 0.309·W02 + 0.926·W03 + 0.003·R10 + 0.081·S20 - 2.971	0.831	0.730

Source: own work.

Note: The zero values of some model parameters are due to the rounding to three decimal places used.

On the basis of selected financial variables, discriminant models were estimated. Parameter estimates together with the values of Wilks' lambda statistic and AUC are presented in Table 2. The lowest Wilks' lambda value was obtained by the model F

(0.798), and the highest by the model A (0.981). All Wilks' lambda values for the presented models are relatively high, which may indicate their limited discriminatory power for classifying objects. The obtained AUC values (except for the model A) were greater than 0.7, which indicates the satisfactory discriminatory power of the models.

Table 3. Classification performance

Model	Classification effectiveness in training set			Classification effectiveness in test set		
	bankrupts	non-bankrupts	total	bankrupts	non-bankrupts	total
A	32.70%	76.20%	54.45%	37.50%	70.00%	53.75%
B	65.50%	72.60%	69.05%	70.00%	72.50%	71.25%
C	66.70%	77.40%	72.05%	70.00%	75.00%	72.50%
D	68.50%	77.40%	72.95%	70.00%	80.00%	75.00%
E	41.10%	91.10%	66.10%	42.50%	92.50%	67.50%
F	65.50%	76.80%	71.15%	72.50%	67.50%	70.00%
G	64.90%	73.80%	69.35%	67.50%	65.00%	66.25%
H	63.70%	79.20%	71.45%	67.50%	75.00%	71.25%

Source: own work.

Table 4. Assigned tied ranks associated with model features

Model	Classification effectiveness in training set			Classification effectiveness in test set			Wilks' lambda	AUC
	bank-rupts	non-bankrupts	total	bank-rupts	non-bankrupts	total		
A	8	6	8	8	6	8	8	8
B	3.5	8	6	3	5	3.5	5	4
C	2	3.5	2	3	3.5	2	6	5
D	1	3.5	1	3	2	1	4	2.5
E	7	1	7	7	1	6	7	7
F	3.5	5	4	1	7	5	1	1
G	5	7	5	5.5	8	7	3	6
H	6	2	3	5.5	3.5	3.5	2	2.5

Source: own work.

The classification efficiency of the models (Table 3) was verified using data from the test sample, which includes data not used in the estimation of the parameters of the discriminant function, and compared with the classification results obtained for the training sample. It was noted that higher classification performance of bankrupts was obtained in the test sample when comparing it to that obtained for the training sample. An inverse relationship is found in the context of classifying non-bankrupts (except for models D and E). In the training and test samples, the highest overall classification efficiency was distinguished by model D. The most effective in classifying bankrupts was model F, and non-bankrupts was model E (Table 4).

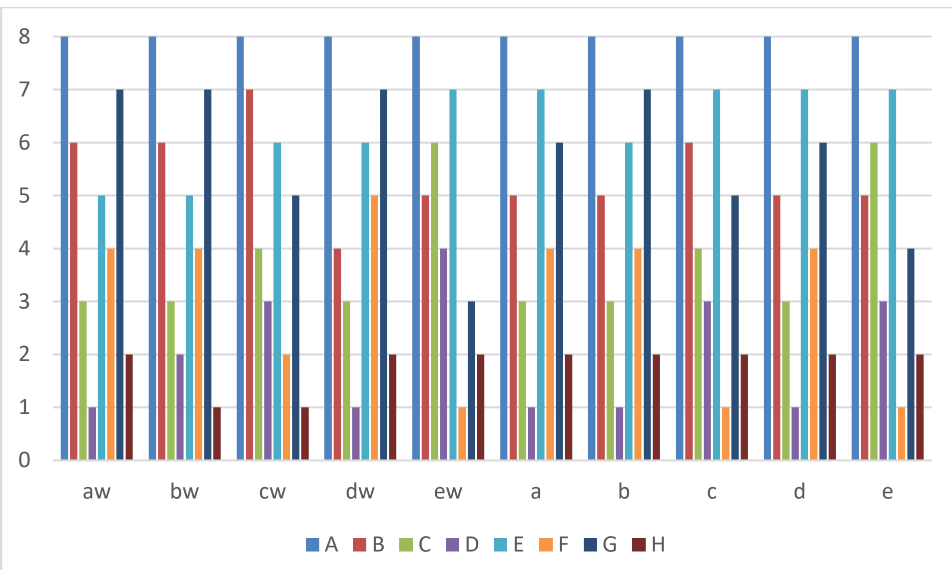
The assigned tied ranks were used to calculate the Spearman correlation coefficients between positions of models obtained due to values of Wilks' lambda statistic and rankings made according to other characteristics of the models (Table 5). It is noticeable that strong correlation is visible only for rankings made due to AUC.

Table 5. Spearman's rank correlation coefficients for classification model features

Classification effectiveness in training set			Classification effectiveness in test set			AUC
bankrupts	non-bankrupts	total	bankrupts	non-bankrupts	total	
0.3631	-0.0417	0.5238	0.6131	-0.3512	0.2560	0.8393

Source: own work.

Figure 1. Comparison of models' ranking positions based on Euclidean distance



Source: own work.
Note: symbols aw, bw, cw, dw and ew denote Euclidean distance from the pattern defined as hypothetical model, i.e. the best among the considered set of models, whereas a – e from the pattern being the ideal model.

In order to determine the best method of selecting diagnostic variables, the Euclidean distances (10) were calculated for each model (A – H) based on distinguished model correctness measures, and described two patterns. It was assumed that Euclidean distances are evaluated using: (a) all accuracy measurements – 8 measures, (b) without overall classification effectiveness – 6 measures, (c) and (d) excluding classification accuracy observed in the test or training set respectively – 4 measures and (e) considering only values of Wilks' lambda statistics and AUC – 2 measures. In other words, ten Euclidean distances were evaluated (Figure 1). As it is visible, regardless of the pattern and number of measurements taken into consideration, model A, which was constructed using arbitrary selected diagnostic variables, ranks last, whereas the first position was held by models D (5 times), F (3 times) or H (2 times). Therefore, using the majority voting rule, model D was found to be the best. However, it is worth mentioning that model H ranked only first or second in all the rankings. Thus, to create the final rank of all models, the sum of ranks was calculated, which ordered models as follows: H (with the sum of ranks 18), D (20), F (30), C (38), B (54), G (57), E (63) and A (80).

4. Conclusion

Based on the provided analysis, models H, D and F seemed to be the best in terms of simultaneously considering several accuracy measures. Thus, we claim that the backward stepwise method, the two-step method (i.e. combination of the test of significance of differences between group averages and stepwise forward selection of variables) as well as the Hellwig method of selection of isolated variables are found to be the most fruitful methods for selecting diagnostic variables for bankruptcy prediction models. It is also worth noticing that model F (variables were chosen by the Hellwig method) contains the highest number of variables which describe all aspects of the company performance, which can contribute to greater stability over time for this model (in comparison to the models with a small number of variables).

The highest value of Euclidean distance was obtained by model A, thus the arbitrary choice turned out to be the least effective method of variable selection. This means that in the selection of diagnostic variables, statistical properties of data must be considered.

It should be noted that the reported classification results are based on a single train–test split. Alternative splits or resampling procedures could lead to changes in the absolute values of accuracy measures. However, given the consistency of rankings across multiple evaluation variants, the relative performance of the variable selection methods is unlikely to change substantially.

In the study, the value of Wilks' lambda statistic does not seem to point out the classification abilities of bankruptcy prediction models well since Spearman's rank correlation evaluated between the rankings made due to the values of Wilks' lambda statistic and according to the measures of classification accuracy was rather weak. Thus, classification efficiency analysis remains an indispensable tool in assessing the quality of discriminatory models.

The study is subject to several limitations. First, the analysis is based on a single-country sample (i.e. Polish enterprises), which restricts the external validity of the results. Second, the sample size and the balanced design may influence classification outcomes. Future research could extend the analysis to other countries, larger datasets, and alternative sampling schemes, as well as compare the presented approach with modern classification methods.

References

Aczel, A., (2000). Statystyka w zarządzaniu [Statistics for management]. *Wydawnictwo Naukowe PWN*, Warsaw.

Altman, E., (1968). Financial Ratios, Discriminant Analysis and the Prediction of the Corporate Bankruptcy. *The Journal of Finance*, Vol. 23, pp. 589–609.

Chudik, A., Pesaran, H., Sharifvaghef, M., (2024). Variable selection in high dimensional linear regressions with parameter instability. *Journal of Econometrics,* Vol. 246, pp. 1–25.

Fan, J., Li, R., (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, Vol. 96/456, pp. 1348–1360.

Gajdka, J., Stos, D., (1996). Wykorzystanie analizy dyskryminacyjnej do badania podatności przedsiębiorstwa na bankructwo [Application of discriminant analysis to examine corporate bankruptcy risk], In: Duraj, J. (ed.), Przedsiębiorstwo na rynku kapitałowym [Enterprise on the capital market]. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź, pp. 138–148.

Gruszczyński, M., (2012). Mikroekonometria, Modele i metody analizy danych indywidualnych [Microeconometrics: Models and methods of individual data analysis], 2nd ed., *Wolters Kluwer Polska*, Warsaw.

Grzybowska, U., Karwański, M., (2023). Selekcja zmiennych metodami statystycznymi i uczenia maszynowego. Porównanie podejść na przykładzie danych finansowych, *[Variable selection using statistical and machine learning methods: Comparison of approaches based on financial data], Metody Ilościowe w Badaniach Ekonomicznych [Quantitative Methods in Economics],* Vol. XXIV/4, pp. 229–241.

Harańczyk G., (2010). Krzywe ROC, czyli ocena jakości klasyfikatora i poszukiwanie optymalnego punktu odcięcia [ROC curves: Evaluation of classifier quality and search for the optimal cut-off point], *StatSoft Polska*, available at: 25.06.2024, https://media.statsoft.pl/_old_dnn/downloads/krzywe_roc_czyli_o cena_jakosci.pdf

Herman, S., (2018). Analiza porównawcza zdolności predykcyjnej wybranych metod prognozowania upadłości przedsiębiorstw [Comparative analysis of predictive ability of selected corporate bankruptcy forecasting methods]. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* [*Legal, Economic and Sociological Movement*], Vol. LXXX/3, pp. 199–216.

IBM., (2023). *Tabela badań Levene'a*, available at: 26.12.24, https://www.ibm.com/docs/ pl/spss-statistics/29.0.0?topic=variances-levene-test-table.

Jagiełło, R., (2013). Analiza dyskryminacyjna i regresja logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw [Discriminant analysis and logistic regression in the process of assessing corporate creditworthiness]. *Materiały i Studia* [*Materials and Studies*], No. 286, pp. 1–116.

Kopczyński, P., (2022). Prognozowanie upadłości przedsiębiorstw [Corporate bankruptcy forecasting]. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Korzeniewski, J., (2016). New method of variable selection for binary data cluster analysis. *Statistics in Transition new series*, Vol. 17/2, pp. 1–10.

Mączyńska, E., Zawadzki, M., (2006). Dyskryminacyjne modele predykcji upadłości przedsiębiorstw [Discriminant models for corporate bankruptcy prediction]. *Ekonomista* [*Economist*], No. 2, pp. 205–235.

Malarska, A., (2005). Statystyczna analiza danych wspomagana programem SPSS [Statistical data analysis supported by the SPSS software], *SPSS Polska*, Cracow.

Misztal, M., (2014). Wybrane metody oceny jakości klasyfikatorów - przegląd i przykłady zastosowań [Selected methods for classifier quality assessment – review and application examples], In: Jajuga, K., Walesiak, M. (eds.), Taksonomia 23. Klasyfikacja i analiza danych – teoria i zastosowania [Taxonomy 23. Classification and data analysis – theory and applications], Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu [Research Papers of Wrocław University of Economics]. *Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu*, Wrocław, pp. 156–166.

Nowakowski, M., (2019). The ANOVA method as a popular research tool. *Studia i Prace WNEiZ US*, 55, pp. 67–77.

Pociecha, J., Pawełek, B., Baryła, M., Augustyn, S., (2014). Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej [Statistical methods of bankruptcy forecasting in changing economic conditions], *Fundacja Uniwersytetu Ekonomicznego w Krakowie*, Cracow.

Ptak-Chmielewska, A., (2012). Wykorzystanie modeli przeżycia i analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw [Application of survival models and discriminant analysis in assessing corporate bankruptcy risk]. *Ekonometria* [*Econometrics*], Vol. 4(38), pp. 157–172.

Raftery, E., Dean, N., (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, 101/473, pp. 168–178.

Shi, Y., Li, X., (2019). An overview of bankruptcy prediction models for corporate firms: A Systematic literature review. *Intangible Capital*, 15/2, pp. 114–127.

Tibshirani, R., (1996). Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society*, Ser. B, Vol. 58/1, pp. 267–288.

Valaskova, K., Gajdosikova, D., Belas, J., (2023). Bankruptcy Prediction in the Post-Pandemic Period: A Case Study of Visegrad Group Countries. *Oeconomia Copernicana*, 14(1), pp. 253–293.

Wang, Z., Zhu, Z., Yu, C., (2023). Variable Selection in Macroeconomic Forecasting with Many Predictors. *Econometrics and Statistics*, pp. 1–18.

Wiktorowicz, J., Grzelak, M., Grzeszkiewicz-Radulska, K., (2020). Analiza Statystyczna z IBM SPSS Statistics [Statistical analysis with IBM SPSS Statistics]. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Witkowska, D., (2023). Wybrane metody ilościowe w finansach [Selected quantitative methods in finance]. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Woo Ahn, K., (2016). Modern variable selection techniques. *Datum Newsletter Division of Biostatistics*, Vol. 22/1.