The use of different approaches for credit rating prediction and their comparison

Použití různých přístupů k predikci kreditního ratingu a jejich srovnání

Martina Novotná [1]

Abstract

The objective of the paper is to apply three selected approaches for credit rating prediction, linear discriminant analysis, multinomial logistic regression and decision trees. The models are based on data of European companies' financial indicators and MORE rating. A special attention is paid to comparison of models and interpretation of results. The analysis is also focused on the identification of variables with the most significant impact on credit rating assessment. The estimated models can be used for credit rating prediction and can serve as a useful tool in the investment decision process.

Key words

Credit rating, decision trees, discriminant analysis, logistic regression, modelling, prediction.

JEL Classification: C35, C38, G24

1. Introduction

Credit rating models can be used as a guideline when evaluating unrated firms. One of the best known models in this area is applied by E. I. Altman (1968), whose default model is often used as a tool in financial analysis of a company. This model has the ability to identify companies with the possible financial problems and was proposed on the basis of multivariate discriminant analysis. The other research in this area is primarily focused on bond rating and bond rating models. In addition to discriminant analysis, regression analysis became one of the most used methods to estimate rating in the primary research. An approach of the multivariate discriminant analysis was introduced by Pinches and Mingo (1973), Ang and Patel (1978), Belkaoui (1980) and Altman and Katz (1976). Subsequent research was concentrated on comparison of particular statistical methods; e.g. Kaplan and Urwitz (1979) compare ordered probit analysis with ordinary least square regression, Wingler and Watts (1980) compare ordered probit analysis with multiple discriminant analysis. Recent studies come from the theoretical framework mentioned above and extend statistical methods for new non-conservative approaches such as neural networks introduced by e.g. Surkan and Singleton (1990) or Dutta and Shekhar (1988). Other studies examine the impact of financial variables on credit rating for a given country or region, e.g. Gray, Mirkovic and Ragunathan (2005). Contribution of own credit models is evaluated by some authors, see for example J. Rerolle and C. Rimaud (2009). As they confirm, research in credit risk area and credit models has in comparison with certified rating important value added, because it enables to react on changes and new information sooner than in the case of complete dependency.

[1] Ing. Martina Novotná, Ph.D., Technical University Ostrava, Department of Finance, martina.novotna@vsb.cz.

2. Statistical approaches for credit rating prediction

The three following methods are used in this paper, multinomial logistic regression, discriminant analysis and decision trees. All techniques are suitable for the problem of credit rating prediction, where there are more than two categories of dependent variable (such as five rating categories). The overall description of these methods is in the chapters below.

2.1 Multivariate discriminant analysis

Discriminant analysis is a common statistical method used for classification tasks a suitable method for credit rating modelling. The analysis can be used for two major objectives: i) description of group separation and ii) prediction or allocation of observations to groups. Discriminant functions are linear combinations of variables that best separate groups, for example the *k* groups of multivariate observations. For the following part of this paragraph, the explanation and definitions were taken from Rencher (2002) and Huberty and Olejnik (2006).

For k groups with n_i observations in the ith group, we transform each observation vector y_{ij} to obtain $z_{ij} = a'y_{ij}$, $i = 1, 2,..., k$; $j = 1, 2,...,n_i$, and find the means $\overline{z}_i = a'\overline{y}_i$, where $\overline{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. We seek the vector **a** that maximally separates $\overline{z}_1, \overline{z}_2, ... \overline{z}_k$. The separation criterion among $\overline{z_1}, \overline{z_2} ..., \overline{z_k}$ can be expressed in term of matrices,

$$\lambda = \frac{a'Ha}{a'Ea} \tag{1}$$

where matrix H has a between sum of squares on the diagonal for each of the *p* variables, and matrix E has a within sum of squares for each variable on the diagonal. Another expression of the separation criterion is

$$\lambda = \frac{\mathrm{SSH}\ (z)}{\mathrm{SSE}\ (z)} \tag{2}$$

where SSH (*z*) and SSE (*z*) are the between and within sums of squares for z. The main task of the discriminant analysis is to find a set of weights (*a* values) for the outcome variables to determine a linear composite:

$$Z = a_1Y_1 + a_2Y_2 + \cdots + a_pY_p \tag{3}$$

so that the ratio (2) is maximized. The discriminant analysis follows by assessing the relative contribution of the $y's$ to separation of several groups and testing the significance of a subset of the discriminant function coefficients. The discriminant criterion (1) is maximized by λ_1, the largest eigenvalue of $E^{-1}H$; the remaining eigenvalues correspond to other discriminant dimensions. The test of significance is usually based on the Wilks' lambda.

2.2 Multinomial logistic regression

The multinomial logistic regression is a modification of binary logistic, where only two possible outcomes can occur. The definitions and derivations used in this chapter were extracted from Hosmer and Lemeshow (2000).

The model for dichotomous outcome variable is based on logistic distribution and we use the quantity $\pi(x) = E(Y|x)$ to represent the conditional mean of Y given x when the logistic distribution is used,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}. \tag{6}$$

The central idea of logistic regression is a transformation of $\pi(x)$, so-called logit transformation. This transformation is given by the following equation, showing the case of two dependent (binary) variables.

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{7}$$

The logit, g(x), is linear in its parameters and may range from $-\infty$ $to + \infty$, depending on the range of x. To estimate the logistic regression model, we find the values of parameters β_0 and β_1 which maximize the probability of obtaining the observed set of data. Thus, we must first construct the likelihood function which expresses the probability of the observed data as a function of the unknown parameters. The likelihood function for binary dependent variable can be defined as the log likelihood,

$$L(\beta) = ln[l(\beta)] = \sum_{i=1}^{n}\{y_i ln[\pi(x_i)] + (1 + y_i) ln[1 - \pi(x_i)]\}\ . \tag{8}$$

The multinomial logistic regression is then a modification of the binary alternative. Our analysis of credit rating to five categories requires four logit functions and determination of the so-called reference or baseline category, which is then compared with other logits. A general expression for the conditional probability in the five category model is

$$P(Y = j|\mathbf{x}) = \frac{e^{g_j(x)}}{\sum_{k=1}^{5} e^{g_k(x)}}\ , \tag{9}$$

where $P(Y = j|\mathbf{x})$ = π_j(x) for k = 1, 2, 3, 4, 5 and $g_5(\mathrm{x}) = 0$ for the baseline category five.

2.3 Decision trees

Using decision trees enables to create a tree-based classification model and the rules can be used for prediction purposes. Decision trees can graphically represent alternative choices that can be made and enable the decision maker to identify the most suitable option in a particular circumstance[2]. A decision problem can be presented in the form of a matrix (table) or a tree. Decision trees conventions are described for example in Mian (2011, pp. 168). Rokach and Maimon (2008) states that a decision tree is a predictive model and can be used both for decision and classification problems. Classification trees can be used to classify an object or an instance (such as companies) to a predefined set of classes (rating groups). The companies are firstly classified according to the most relevant variable, then into subgroups according to other variable, and so on (Witzany, pp. 46, 2010). As Witzany (2010) says, the decision rules should maximize a divergence measure of the difference in default risk between the two subsets. The splitting is repeated until no group can be split into two subgroups which are statistically different. According to Wei-Yin (2008), there are three major tasks of a classification tree: (i) how to partition the data at each step, (ii) when to stop partitioning and (iii) how to predict the value of y for each x in partition. Common algorithms for decision tree induction include ID3, C4.5, CART, CHAID and QUEST (Rokach and Maimon, 2008).

Since two algorithms, CART and CHAID will be used for rating prediction; the following text is focused on their description. We assume, that N_j is the number of class j training samples and $\pi(j)$ is the prior probability of class j. At each node t, $N_j(t)$ is the number of class j training samples in t and $p(j|t) = \pi(j)N_j(t)/N_j$ is the estimated probability that an observation in t belongs to class j (Wei-Yin, 2008).

[2] http://businesscasestudies.co.uk/business-theory/strategy/decision-tree-analysis.html

CART (or CRT) refers to classification and regression trees. This algorithm splits the data into segments that are as homogenous as possible with respect to the dependent variable. A terminal node in which all cases have the same value for the dependent variable is a homogenous (pure) node. The extent to which a node does not represent a homogenous subset of cases is an indication of impurity. For categorical dependent variables such as rating, the Gini index, twoing or ordered twoing can be used as impurity measures[3]. If the Gini index is used as a node impurity criterion, then

$$i(t) = 1 - \sum_{j=1}^{J} p^2\,(j|t). \qquad (10)$$

Based on Wei-Yin (2008), we suppose that a split divides the data in t into a left node t_L and a right node t_R. CART selects the split that maximizes the decrease in impurity $i(t) - p_L i(t_L) - p_R i(t_R)$, where p_L and p_R are the proportions of data in t_L and t_R. CART enables to consider misclassification costs and provide prior probability distribution.

CHAID refers to chi-squared automatic interaction detection. At each step, CHAID selects the independent (predictor) variable that has the strongest interaction with the dependent variable. If categories of each predictor are not significantly different with respect to the dependent variable, they are merged[4]. For each input attribute a_i, CHAID finds the pair of values in V_i that is least significantly different with respect to the target attribute. The significant difference is measured by the p value obtained from a statistical test. The statistical test used depends on the type of target attribute. An F test is used if the target attribute is continuous; a Pearson chi-squared test if it is nominal; and a likelihood ratio test if it is ordinal (Rokach and Maimon, 2008). For each selected pair of values, the p value obtained is compared with a certain merge threshold. If it is greater, it merges the values and searches for an additional potential pair to be merged. It is repeated until no significant pairs are found. The best input attribute to be used for splitting the current node is then selected, such that each child node is made of a group of homogeneous values of the selected attribute. This procedure stops also when one of the following conditions is fulfilled: (i) maximum tree depth is reached; (ii) minimum number of cases in a node for being a parent is reached, so it cannot be split any further; (iii) minimum number of cases in a node for being a child node is reached (Rokach and Maimon, 2008).

3. Estimation of credit rating models

The analysis is based on the sample of 4 802 companies from eight countries from Central and Eastern Europe: the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia. Only very large and large companies[5] from mining, manufacturing and construction industries with MORE[6] rating were selected. The sample contains data from the period 2002 – 2008 of which approximately 70 % will be used as experimental sample. The table below (Table 1) shows an uneven representation of rating grades.

[3] PASW Decision trees 18
[4] PASW Decision trees 18
[5] Companies in Amadeus database are considered to be very large (or large) when they have operating revenue equal or greater than 100 (10) million euro, or total assets equal or higher than 200 (20) million euro, or number of employees is equal or greater than 1000 (150).
[6] MORE ratings classify companies similarly as rating agencies (Bureau van Dijk Electronic Publishing, 2008). The MORE rating is calculated using a unique model that references the company's financial data to create an indication of the company's financial risk level.

Table 1: Description of experimental sample

MORE Rating	Number of cases	Marginal percentage	Rating grade
AA	162	5.4%	5
A	721	24.1%	4
BBB	1529	51.1%	3
BB	507	16.9%	2
B	73	2.4%	1
Total	**2992**	**100%**	

The list of independent variables considered in the analysis of credit rating is shown in the following table (Table 2).

Table 2: Financial variables

Category	Economic rationale	Financial variables	
Size	Adequate protection	Total assets	LogTA
Profitability	Ability to earn a satisfactory returns	Return on total assets Return on equity	ROA ROE
Capitalization	Measure of capital structure and leverage	Long term debt to total assets Equity to total assets	LogLTDTA EQTA
Liquidity	The flow of financial resources	Current ratio	LogCURR
Interest coverage	Ability to service the financial charges	Interest cover	LogINTCOV

3.1 Discriminant analysis

The discriminant analysis is carried out by means of two methods, the simultaneous method resulting in the model containing all independent variables, and the stepwise method considering only variables with the greatest discriminating ability. Tests of equality of group means show that all variables contribute to the model. Based on Wilks' lambda, the variables ROA (.483), LogINTCOV (.524) and EQTA (.557) are those with the greatest ability at discriminating between groups. Since we discriminate between five groups, the analysis results in five discriminant functions, all of them are statistically significant. The stepwise approach gives the model containing five variables with the greatest discriminating ability: ROA, LogINTCOV, EQTA, LogCURR and ROE (the rank of each variable reflects its discriminating ability).

The coefficients of classification functions (Fisher's linear discriminant functions) of this model are shown in the Table 3. To classify individual cases, the values of five functions must be calculated and the group corresponding to the function with the highest value is selected.

Table 3: Classification function coefficients (MDA)

Variable	1	2	3	4	5
EQTA	.380	.504	.640	.743	.805
ROE	.312	.311	.271	.192	0.046
ROA	-.567	-.374	-0.080	.408	1.246
LogCURR	-11.752	-10.157	-6.906	-1.433	3.330
LogINTCOV	2.279	3.720	6.036	9.879	11.675
Constant	-11.295	-13.962	-21.938	-39.267	65.155

The classification functions are weighted more heavily in favour of classifying group three, because the groups are not equally sized. Results of classification are presented in the table below (Table 4) and show overall classification ability of two models.

Table 4: Classification table (discriminant analysis)

Method	Experimental sample	Test sample
Simultaneous method	85.8%	70.6%
Stepwise method	85.7%	70.4%

3.2 Logistic regression

Multinomial logistic regression allows one to find the coefficients of predictors included in the model by using maximum likelihood method. In direct logistic regression, all predictors enter the equation simultaneously, while the stepwise method allows one to identify the most important variables for classification that should be included in the final model. In the case of five grouping categories, four logit functions are estimated. The rating group 5 denotes to the highest quality rating category (AA) and is used as a reference value. We form four logits, $g_1(x), g_2(x), g_3(x), g_4(x),$ comparing Y=1, Y=2, Y=3, Y=4 to the reference value. To fit the logistic regression model in equation (9) we estimate the unknown parameters using the maximum likelihood method. Based on the model fitting criteria (-2 Log Likelihood), both models are significantly different with ones with the constant only. The statistical significance of each of the coefficients is evaluated using the Wald test[7]. Two coefficients are not statistically significant in all cases, LogTA in first two equations $g_1(x), g_2(x)$ and EQTA in the next two equations $g_3(x), g_4(x)$.

To form the logistic regression model, the conditional probability is expressed for each category in the model. For each case, we calculate probabilities $\pi(1), \pi(2), \pi(3), \pi(4), \pi(5)$ and select the rating category with the highest probability value. The stepwise approach model includes six following variables, ROA, LogINTCOV, EQTA, LogCURR, ROE and LogTA. Parameter estimates of logit functions are presented in the Table 5. The "null" model would classify all objects into the modal category, 3, and would be correct in 51.1% of the time. The overall classification ability of estimated models, both on experimental and test sample, is demonstrated in the Table 6.

Table 5: Parameter estimates (logistic regression)

Variable	1	2	3	4
EQTA	-.377	-.167	.012	.030
ROE	.586	.559	.460	.140
ROA	-2.906	-2.256	-1.619	-.579
LogCURR	-29.815	-21.051	-11.408	-3.575
LogINTCOV	-13.979	-12.616	-9.910	-4.255
LogTA	.152	-.962	-1.425	-1.650
Constant	50.034	48.926	39.171	23.657

Table 6: Classification table (MLR)

Method	Experimental sample	Test sample
Full factors method	88.0%	72.4%
Stepwise method	87.9%	64.5%

7 A squared parameter estimate divided by its squared standard error is a Ch-square statistic.

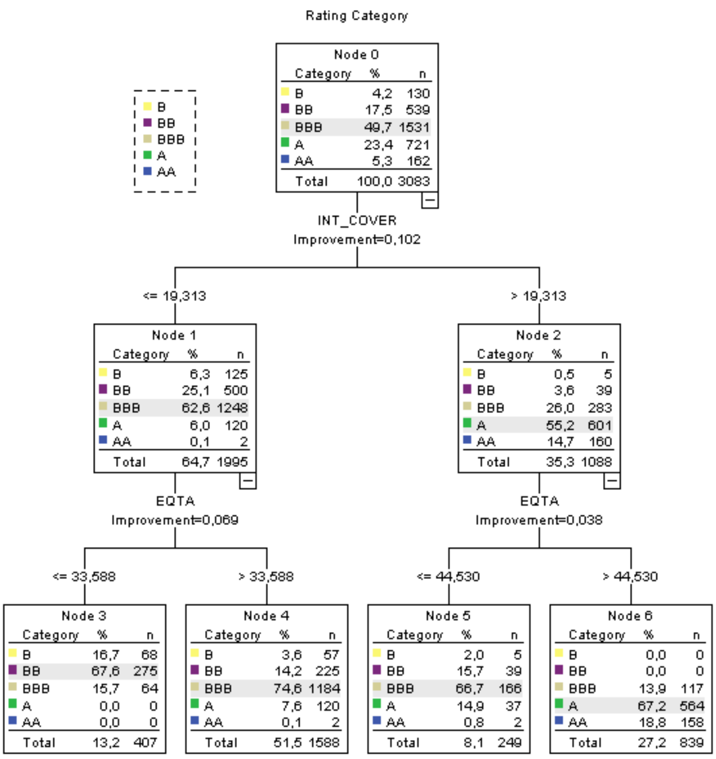
3.3 Decision trees

Seven variables were selected as the inputs for the decision trees and two techniques were employed, CHAID and CART. The algorithm enables to set growth limits as the number of levels in the tree and a minimum number of cases for parent and child nodes. Increasing the minimum values tends to produce trees with fewer nodes, as can be seen in the table below. For example in CART models, the values of 100 cases for parent nodes and 50 cases for child nodes results in a tree with 23 nodes and 5 depths, while the values of 400 cases for parent nodes and 200 cases for child nodes results in a tree with 7 nodes and 2 depths. The result of CHAID model is a tree of 13 nodes and 2 depths (400-200).

The CHAID method results in the model containing three financial variables, INT_COV, CURR, EQTA. As there are five rating categories, the predicted category is the credit rating category with more than 20 % of cases in the node. The significance values for splitting nodes and merging categories is based on the method of Pearson statistics. The model proves that INT_COV is the best predictor for credit rating. For example, for INT_COV from the interval (5.865, 11.833), 77.1 % of companies fall within the rating category BBB, while for INT_COV of ≤ 2.186, 41.9 % companies are from the worse category BB. When the value of INT_COV is from the interval (37.182, 71. 817), 63.6 % of companies belong to category A. The overall classification ability of the model is 62.6 % based on cross validation of cases which means that the risk of misclassifying a company is approximately 37 %. The model provides better prediction of middle groups when compared with AA or B.

In the model derived by CART method, all seven variables are included. The best three predictors are INT_COV, EQTA, ROA. The method used to measure impurity is based on Gini coefficient calculated as squared probabilities of membership for each category of the dependent variable. The simper model (CRT 2) correctly classifies 71 % of cases, however is not able to correctly classify groups B and AA. The graphical presentation of the CRT 2 tree model is in the figure below (Figure 1). The model with 23 nodes and 5 depths (CRT 1) is too large to present it in this paper. Since this model's overall classification ability is 77.3 %, it is better than the simpler model. The improvement is evident in the prediction of all groups, especially AA.

Figure 1: CRT 2 decision tree



4. Conclusion

The paper is focused on the use of three methods for the estimation of credit rating models. The methods are discriminant analysis, logistic regression and decision trees. All the methods have been used on financial data of European companies with MORE rating. The approaches used for credit rating modelling in this paper allow identifying variables with the highest effect on credit rating assessment. The analysis suggests that variables such as return on assets, interest coverage, and the ratio of equity to total assets are those with the highest impact on rating and thus the best discriminating power among rating groups. These results are confirmed by all three approaches used in this paper. The models estimated by means of discriminant analysis contain seven (direct method), or five (stepwise method) independent variables, while logistic regression models include seven (full method), or six (stepwise method) independent variables. The models estimated with the use of decision trees include either three (CHAID) variables or all seven variables (CART). The classification ability of the decision tree models based on the experimental sample is low relatively to other models. For this reason, only discriminant and logistic models have been validated on the test sample. Their classification ability is relatively high and achieves approximately 70 %. All methods used in this paper are suitable for credit rating modelling. Discriminant analysis and logistic regression models are relatively simple to use and achieve a high classification ability. The advantage of the method of decision tress is its graphical representation and a simple identification of the most significant variables.

Acknowledgements

The paper is based on research activities sponsored through SGS research project No. SP2012/19. The support is greatly acknowledged.

References

[1] Altman E.I., Katz S., 1976. Statistical bond rating classification using financial and accounting data. *M. Schiff and G. Sorter, eds.: Topical Research in Accounting* , NYU Press, New York.

[2] Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. In *Journal of Finance*, 4, pp. 589 – 609.

[3] Ang J.S., Patel K., 1975. Bond rating methods: Comparison and validation. *The Journal of Finance*, 30, pp. 631 – 640.

[4] Belkaoui A., 1980. *Industrial bond ratings: A discriminant analysis approach*. Working paper (University of Ottawa).

[5] Dutta, S, Shekhar S., 1988. Bond ratings: A non-conservative application of neural network. *IEEE International Conference on Networks, 2, San Diego, California,* pp. 443-450.

[6] Gray, S, Mirkovic, A, Ragunathan, V., 2006. The determinants of credit ratings: Australian evidence. *Australian Journal of Management*, 31, pp. 333-354.

[7] Haimes Y. Yacov, 2011. *Risk modeling, assessment, and management*. 3rd ed. New Jersey: John Willey & Sons.

[8] Hosmer D.W., Lemeshow S., 2000. *Applied logistic regression*. John Wiley & Sons Inc.: New York.

[9] Huberty C.J., Olejnik S., 2006. *Applied MANOVA and Discriminant Analysis*. John Wiley & Sons Inc.: New York.

[10] Kaplan R, Urwitz G., 1979. Statistical models of bond ratings: A methodological inquiry. *Journal of Business*, 52, pp. 231-261.

[11] Mian, M. A, 2011. *Project economics and decision analysis, Volume 2: Probabilistic models*. 2nd ed. Tulsa: PenWell Corporation.

[12] Novotná, M., 2011. Application of selected scoring models on corporate credit rating. *In 29th International Conference on Mathematical Methods in Economics 2011 - part I*, pp. 511 – 516. Praha: Proffesional Publishing.

[13] Pinches G., Mingo K.A., 1973. A multivariate analysis of industrial bond ratings. *The Journal of Finance*, 28, pp. 1 – 18.

[14] Rencher, A.C., 2002. *Methods of Multivariate Analysis*. New York: John Wiley & Sons.

[15] Rerolle J, Rimaud C., 2009. Does independent credit research add value? *Alternative Intelligence Quotient* 2009, 32, pp. 31 – 37.

[16] Rokach, L. and Maimon, O., 2008. *Data mining with decision trees. Theory and applications*. Singapore: World Scientific Publishing.

[17] Wei-Yin, L., 2008. Classification and regression tree methods. In *Encyklopedia of Statistics in Quality and Reliability*, pp. 315-323. Chichester: Wiley VCH.

[18] Wingler T.R., Watts J.M., 1980. An analysis of bond rating changes in the electric utility industry. *The Financial review*, 15, pp. 1 – 89.

[19] Witzany, J., 2010. *Credit risk management and modelling*. Praha: Oeconomica.

Summary

Cílem příspěvku je použití a srovnání tří vybraných metod k odhadu kreditních ratingových modelů, lineární diskriminační analýzy, multinomické logistické regrese a metody rozhodovacích stromů. Modely jsou odhadovány na základě finančních ukazatelů a MORE ratingu evropských firem. Analýza je zaměřena jak na odhad predikčních modelů, tak na identifikaci proměnných s největším vlivem na kreditní rating. Výsledky jednotlivých metod potvrzují podobné závěry, nicméně klasifikační úspěšnost modelů je vyšší v případě diskriminačních a logistických modelů, které umožňují relativně jednoduchým způsobem predikovat příslušnost k ratingové skupině. Model odvozený na základě metody rozhodovacích stromů je méně klasifikačně úspěšný, ale jeho výhoda spočívá v grafickém znázornění rozhodovacího pravidla a jednoduchém způsobu identifikace proměnných s největším vlivem.