TECHNICAL ASSUMPTIONS OF MODELS PREDICTING CORPORATE FINANCIAL DISTRESS

Dagmar Čámská

Abstract

This paper focuses on technical assumptions of models predicting corporate financial distress. A use of the prediction models for evaluating related business units is a part of corporate risk management. Increasing number of insolvency proposals since 2008 has caused an increasing need for a reliable production and it starts a new wave of a prediction models' construction. Availability of statistical programs and friendly use of their packages enables the model's creation. Models' construction is easier than ever before but the models also have to provide reliable and useful results. Approaches used the most nowadays are discriminant analysis and logistic regression. These methods have many similar characteristics but also some differences which appear also in an area of requirements and assumptions. These limitations should be known and respected otherwise the created model is not enough robust and it does not provide good results for the evaluation of out-sample units. The technical statistical requirements of the discriminant analysis and the logistic regression are compared. This paper should not only provide the description, comparison but also a solution if it is more appropriate to use the discriminant analysis or the logistic regression in the model creation phase.

Key words: discriminant analysis, logistic regression, explanatory power, bankruptcy models

JEL Code: G33, M2, C51

Introduction

Models predicting financial distress provide a quick and inexpensive answer about the corporate financial situation. These models are one of tools which predict a possible crisis (Krause, 2013). This possible crisis can even lead to a corporate bankruptcy which affect many other related counterparties on a microeconomic level as well as on a macroeconomic level. The microeconomic level is connected with particular companies as suppliers or customers and employees of the ailing evaluated company. The macroeconomic level is meant as the government and its interventions in the economy. Especially the Czech labour market has the limited flexibility (Pavelka and Löster, 2013) which has to be taken into account by government policies in a case of social impacts of corporate bankruptcies.

Prediction of corporate financial distress is a serious research task since 1960's. There have been created many tools and methods. The scientific discussion about appropriate explanatory power of these models is usually reopened by serious shocks and changes. One of the latest has started because of the last global economic crisis. This environment raises a wave of new predicting tools, models and methods. Approaches by Karas and Režňáková (2014) and Hálek (2013) have been introduced recently in the Czech Republic. Technical models' construction is easier than ever before because there are available plenty statistical programs whose packages are friendly nowadays. On the other hand the new created models should provide reliable and useful results which support a decision making process. It means that the new models would not be applicable only for in-sample data but also and especially for out-sample data. Out-sample data are the data which were not used for the model's creation at the beginning. It is obvious that the model should work for in-sample data as prerequisite.

Previously stated ideas can be sum it up as following. It is not difficult to create the model predicting financial distress nowadays but it is difficult to create the model correctly. Models have to meet some basic requirements. Requirements from the point of view of users and possible applicability are summarized for example by Čámská (2014). There is also a level of the model's explanatory power and therefore models are tested and evaluated by further and further researches using current data and focused on the different fields of the economic activity (for example Čámská, 2013a or Klečka, Scholleová, 2010 and many others). This paper is untraditionally focused on models' prerequisites from the mathematical-statistical point of view. These prerequisites should be taken into account by an author during a phase of the model's creation. For the purpose of this paper these prerequisites are called as technical assumptions of models predicting corporate financial distress. Assumptions are derived from the method/technique which is used for the model's creation. Therefore the following chapter is focused on a discovering the most used techniques for the creation of the model predicting financial distress.

1 Techniques used for the model's creation

Literature review offers a wide as well as a narrow view on the possible techniques used for the creation of the models predicting financial distress. It depends if the models are focused on the prediction of default probability connected with Basel banking rules or if they are focused just on the prediction and another factor is which data are available (complete, incomplete, financial, nonfinancial, comparable, incomparable etc.). De Laurentis, Maino and Molteni (2010) work with three basic groups – experts-based approaches, statistical-based models and heuristic & numerical approaches. Most of publicly known models are based on the quantitative data, especially derived from financial accounting statements, and therefore this paper works further with statistical-based models. Still these models can be created with the use of plenty mathematical-statistical approaches. McKee (2000) divides these approaches into following categories according to technical bases:

- Univariate ratio models
- Multiple discriminant analysis
- Linear probability models
- Multivariate conditional probability models
- Logit and Probit
- Recursive partitioning models
- Survival Analysis (proportional hazard model)
- Expert systems
- Mathematical programming
- Neural networks
- Rough set approach.

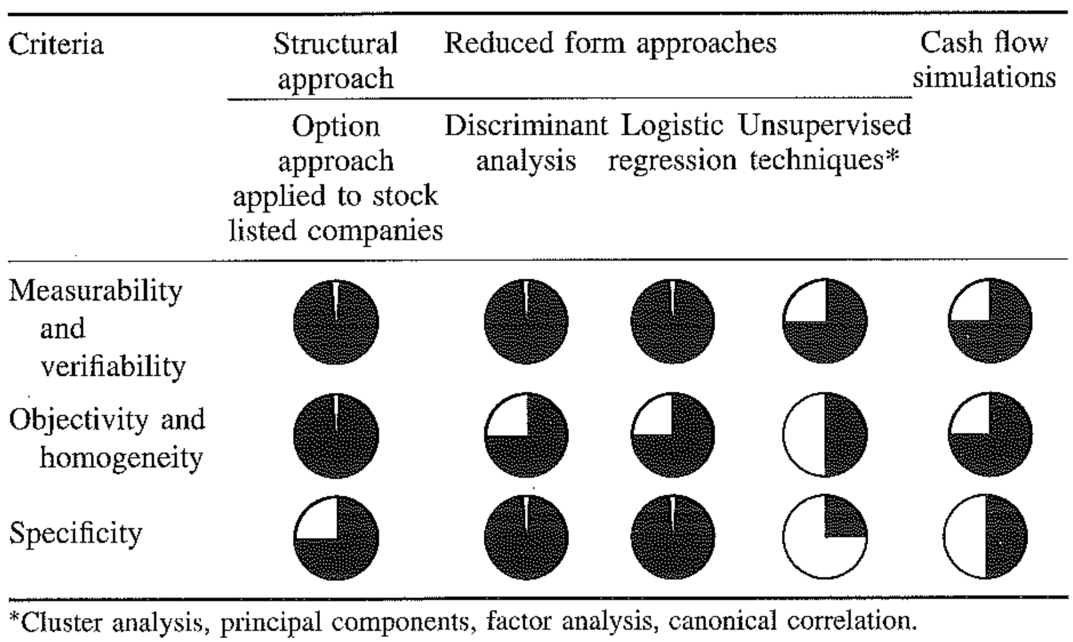
Each of aforementioned techniques have its advantages and disadvantages. In a favour of discriminant analysis and models as Probit and Logit it is that their results can be interpreted by a person without high mathematical-statistical education or knowledge. Recursive partitioning models result intuitively in decision trees which unfortunately have many levels and it makes the interpretation difficult. On the other hand neural networks look like black boxes and therefore they are for their users in transparent and unreliable. De Laurentis, Maino and Molteni (2010) use another dividing which may be listed as:

- Structural approach applied to stock listed companies
- Reduced form approaches as discriminant analysis, logistic regression and unsupervised techniques

- Cash flow simulations.

De Laurentis, Maino and Molteni (2010) also provide a comparison of these techniques which shows advantages and disadvantages in the case of criteria measurability and verifiability, objectivity and homogeneity and specificity. This comparison is displayed by figure 1 where a black circle means that the criterion is fully met and a white circle means that the criterion is not fulfilled at all.

Fig. 1: Comparison of statistical-based approaches



Source: De Laurentis, Maino and Molteni (2010, p. 77)

The range of this paper is limited and therefore only the most used techniques will be discussed and compared further. Discriminant analysis and Logistic regression are the most popular techniques in the Czech Republic and another countries of Central Europe how it is clear from Čámská (2012) and Čámská (2013b). These methods have many similar characteristics and they fulfil the same level of criteria mentioned in figure 1. On the other hand they also have some differences which appear also in an area of requirements and assumptions. Following chapters are dedicated to these requirements and assumptions.

2 Discriminant analysis

Discriminant analysis is an example of a technique which works on a basis of multivariate ratio model. These ratios can be derived from the perspective of profitability, liquidity, leverage, size or type of economic activity etc. Discriminant analysis assumes the existence of two or more populations. In the case of models predicting financial distress these populations are financial ailing companies (insolvent) and successful financial healthy companies (solvent). The aim of discriminant analysis is to maximise homogeneity inside groups and minimize areas which overlap. This can be written by a following formula

$$\min\left\{\sum_{i=1}^{p}\left(x_{i,k}-\overline{x_{i,solvent/insolvent}}\right)^2\right\} \qquad (1),$$

where $x_{i,solvent/insolvent}$ present the mean value of variables in each followed group. The method has some requirement when we want to avoid model instability and inaccuracy. According to De Laurentis, Maino and Molteni (2010)[1] these statistical requirement are:

- independent variables are normally distributed;
- absence of heteroscedasticity, that is, the matrix C has to have similar values on the diagonal;
- low independent variables multi-colinearity, that is, matrix C has to have homogenous and preferably low values off the diagonal, not statistically significant;
- homogeneous independent variables variance around groups' centroids, that is, matrix C to be (roughly) the same for firms in both solvent and insolvent groups.

The fulfilling of the first three conditions could be solved by adopting quadratic discriminant analysis instead of linear discriminant analysis. The last condition is a serious problem in the reality because insolvent companies have higher variance than solvent companies. Linear discriminant analysis has generally strong assumptions which is uneasy to fulfil in the case of using financial ratios. Morris (1998) says that variables of financial ratios have skewed distribution. In a practice many authors and models' designers try to have an equal number of solvent and insolvent companies. From the mathematical-statistical point of view it is not necessary. The numbers of cases in both groups do not have to be equal and even not comparable. Rencher and Christensen recommend (2012) that the number of cases in

[1] De Laurentis, Maino and Molteni (2010) summarize these requirements according to Landau, Everitt (2004). A handbook of statistical analysis using SPSS-PASW, Giri (2004). Multivariate Statistical Analysis: Revised and Expanded, Stevens (2002). Applied Multivariate Statistics for Social Sciences and Lyn (2009). Consumer Credit Models – Pricing, Profit and Portfolios.

the smallest group has to exceed the number of used variables. Generally small data samples can lead to heterogeneity and the final discriminatory function would be not enough robust.

3 Logistic regression

Logistic regression is a type of regression function which is suitable for a relation in which the dependent variable has a dichotomous character. This variable can be coded as binary. In the case of models predicting financial distress it means that financial ailing companies are coded as 0 or 1 and financial healthy companies conversely. Logistic regression has some serious advantages as no assumed linear relation between dependent and independent variables and it can work with presence of heteroscedasticity. On the other hand there are also requirements which are sum it up by Fernandes (2005) following:

- Each y_i follows a Bernoulli distribution with parameter $\pi(x^k{}_i)$.
- The error terms are independent.
- No relevant variables are omitted, no irrelevant variables are included, and the functional form is correct.
- There is a linear relationship between the logit of the independent variables and the dependent.
- There is no significant correlation between the independent variables (no multicolinearity).

Logit model works with variables $x_1, x_2, \ldots, x_p$ and coefficients $ß_0, ß_1, ß_2, \ldots, ß_p$ which create a function g which can be written for the company i following

$$g(\pi_i)=\beta_0+\beta_1\times x_{i1}+\beta_2\times x_{i2}+,....,+\beta_p\times x_{ip}\text{, where i = 1,…, n.} \qquad (2)$$

Π_i is the probability that the state of the world (solvent or insolvent) occurs. If the dependent variable has the Bernoulli distribution it is possible to prove that the previous formula can be rewritten as

$$g(\pi_i)=\log\frac{\pi_i}{1-\pi_i}=\beta_0+\sum_{j=1}^{p}\beta_j\times x_{ij}\text{, where i = 1, …, n.} \qquad (3)$$

This formula can be transformed further (details Kennedy, 2008) and we get formula 4 which shows that the probability of the state of the world is equal to a ratio

$$\pi_i=\frac{1}{1+e^{-\left(\beta_0+\sum_{j=1}^{p}\beta_j\times x_{ij}\right)}}\text{, where i = 1, …, n.} \qquad (4)$$

Literature contains several versions of this formula and exactly four types mentioned further are correct. These types can be obtained with a use of easy mathematical transformations. First it depends if the author chooses the basic state of the world – solvent or insolvent. Let assume that the basic state of the world is solvent and it is coded as Y = 1. It is possible to substitute $\left(\beta_0 + \sum_{j=1}^{p} \beta_j \times x_{ij}\right)$ as z and therefore the probability of solvent company is written as

$$\pi(Y=1) = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \ . \tag{5}$$

Only two states of the world can occur – solvent and insolvent and this implies that the probability of solvent case plus the probability of insolvent case has to be equal to 1 because the other state of the world cannot occur. If formula 5 is relevant for the solvent case then we can get formula 6 relevant for the insolvent case.

$$\pi(Y=0) = \frac{1}{1+e^{z}} = \frac{e^{-z}}{1+e^{-z}} \, . \tag{6}$$

An alternative of logit model is probit model which does not arises from the logistic function but from the cumulated normal function. These models are very similar. The reason why almost all models are constructed as logit is historical because logit model is not so difficult for computation. This reason can be omitted nowadays when there are so many statistical programs available.

3 Application

Logistic regression is a worldwide popular tool for the creation of models predicting financial distress because its requirements can be fulfilled easily than requirements of linear discriminant analysis. It is especially the case of nonlinear relation and working without strict homoscedasticity assumption. Although logistic regression has its advantages transition economies (Čámská, 2012 or Čámská, 2013b) prefer discriminant analysis. It implies that created models are not enough robust and it partly decreases their explanatory power. From the statistical point of view it is better to prefer logistic regression than discriminant analysis. If a model's designer want to create the model with high explanatory power the person does not have to take into account only quality of data but also technical requirements of the created model. The model's designer should know the models' limitations otherwise the person is not able to create models with sufficient accuracy.

Conclusion

This paper focused on technical assumptions of models predicting corporate financial distress. Concretely it worked with requirements of linear discriminant analysis and logistic regression. It showed the requirements for the both techniques which were compared. The paper provides a solution that it is better for the model's designer to prefer logistic regression instead of linear discriminant analysis whose requirement cannot be fulfilled in the case of used financial data.

These models based on the statistical approaches work only on probabilistic roots and they will be never able to classify all evaluated cases correctly. This limitation will never disappear and it has to be accepted by designers as well as users.

Acknowledgment

The author would like to thank the people who allowed him to write this paper and continue his teaching and research work.

References

Čámská, D. (2012). *National view of bankruptcy models*. In Loster, T., Pavelka T. (Eds.), *6th International Days of Statistic and Economics at VŠE (pp. 268-278)*. Slaný: Melandrium.

Čámská, D. (2013a). *Predicting financial distress of companies in the agricultural sector.* In Eglite, A. (Eds.), *International Scientific Conference on Economic Science for Rural Development at Latvia Univ Agr (pp. 257-262)*. Jelgava: Latvia Univ Agriculture.

Čámská, D. (2013b). *Predicting financial distress of companies operating in construction industry.* In *Proceedings of the 8th International Conference ACCOUNTING AND MANAGEMENT INFORMATION SYSTEMS AMIS 2013 in Bucharest (pp. 642-657).* Bucharest: The Bucharest University of Economic Studies.

Čámská, D. (2014). *Requirements for models predicting corporate financial distress.* In Löster, T., Pavelka T. (Eds.), *8th International Days of Statistic and Economics at VŠE (pp. 316-323)*. Slaný: Melandrium.

De Laurentis, G., Maino, R., & Molteni, L. (2010). *Developing, validating and using internal ratings. Methodologies and case studies*. Chichester: John Wiley & Sons.

Fernandes, J.E. (2005). Corporate credit risk modelling: Quantitative rating system and probability of default estimation. Retrieved from images.to.camcom.it/f/tofinanza/I_/I_01.pdf

Hálek, V. (2013). *Predikce finanční tísně podniku na základě vlastního bankrotního modelu CCB*. Hradec Králové: Gaudeamus.

Karas, M. (2014). *Možnosti využití bankrotního modelu k měření úvěrového rizika podniku.* In Jedlička, P. (Ed.), *Sborník recenzovaných příspěvků z mezinárodní vědecké konference Hradecké ekonomické dny 2014, Díl I., Ekonomický rozvoj a management regionů in Hradec Králové (pp. 435-442)*. Hradec Králové: Gaudeamus.

Kennedy, P. A. (2008). *Guide to Econometrics*. Malden: Blackwell Publishing.

Klečka, J., Scholleová, H. (2010). Bankruptcy models enunciation for Czech glass making firms. *Economics and management, 15*, 954-959.

Krause, J. (2013). Risk management in companies and the importance of selected measures for overcoming the crisis. *WSEAS Transactions on Business and Economics, 10*(3), 133–141. Retrieved from http://www.wseas.org/multimedia/journals/economics/2013/045707-101.pdf

McKee, T.E. (2000). Developing a Bankruptcy Prediction Model via Rough Sets Theory. *International Journal of Intelligent Systems in Accounting Finance & Management, 9*(3), 159-173.

Morris, R. (1998). Bankruptcy prediction model: Just how useful are they? *Credit management, May*, 43-45.

Pavelka, T., Löster, T. (2013). *Flexibility of the Czech Labour Market from a Perspective of the Employment Protection Index.* In Löster, T., Pavelka T. (Eds.), *7th International Days of Statistic and Economics at VŠE (pp. 1090-1099)*. Slaný: Melandrium.

Rencher, A.C., Christensen, W.F. (2012). *Methods of Multivariate Analysis.* John Wiley & Sons.

Contact

Dagmar Čámská

University of Economics, Prague, Faculty of Business Administration, Department of Strategy

nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic

dagmar.camska@seznam.cz