Available online at www.sciencedirect.com

ScienceDirect

Procedia Economics and Finance 23 (2015) 850 – 855

Procedia Economics and Finance

www.elsevier.com/locate/procedia

2nd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM, 30-31 October 2014, Prague, Czech Republic

Logit and Probit Model used For Prediction of Financial Health of Company

Tomáš Klieštik[a], Katarína Kočišová[a]*, Mária Mišanková[a]

[a] *University of Žilina, Faculty of Operation and Economics of Transport and Communications, Department of Economics, Univerzitná 1, 010 26 Žilina, Slovakia*

Abstract

Corporate failure can exist in various types and dimensions, and has different effects on stakeholders according to magnitude of the failure and its type. The rise of corporate failure in different types brought about the use of different definitions and different concepts connoting failure. Over the past 35 years, the topic of "business failure prediction" has developed to a major research domain in corporate finance. Many academic studies have been dedicated to the search for the best corporate failure prediction, based on publicly available data and statistical techniques. This article will be focus on techniques used for prediction of bankruptcy such as logit and probit analysis.

Selection and/ peer-review under responsibility of Academic World Research and Education Center

Keywords: logit analysis; probit analysis; prediction; financial health; company;

1. Introduction

Martin (1977) first applied the logit regression to construction of early warnings of bank failure. Ohlson (1980), we can mark him as pioneer in economic area in application of logit analysis – multivariate conditional probability model to business failure prediction. He introduced a logistic regression approach to develop a bankruptcy prediction model to assess the probability of corporate failure. He did not agree with discriminant analysis for example because of requirement of identical variance-covariance matrices for both groups – failed, non-failed and because of

* Ing. Katarína Kočišová. Tel.: +421-41-513-3227
E-mail address: katarina.kocisova@fpedas.uniza.sk

Selection and/ peer-review under responsibility of Academic World Research and Education Center
doi:10.1016/S2212-5671(15)00485-2

requirement of normally distributed predictors. Ohlson found out that output of the MDA (ordinal ranking) provides nothing about the probability of default. In applied MDA models is used size and industry as matching criteria and he criticized that the use of variables as predictors rather than the use the variables for matching. To solve this problem it is used conditional logit analysis for predicting the probability of default so it means that logit requires less restrictive statistical assumptions and offer better empirical discrimination. His sample included 105 publicly traded industrial companies that become bankruptcy during the period 1970 – 1976. This model examined liquidity, profitability, leverage and solvency and the result is that the Ohlson´s model was able to identify about 88% of 105 bankrupt companies approximately one year before bankruptcy (Zvaríková & Majerová, 2014).

Bliss (1934) came as the first with idea of probit analysis in 1934 published in the journal Science. He used probit analysis to solve the problem of finding effective pesticide for spraying insect that eating grapes. In 1947 John Finney (1947) described probit analysis in more detail in the book Probit Anylsis. The first, who employed a probit model to estimate financial distress prediction, was Zmijewski (1984). Zmijewski (1984) for his research selected 96 failed and 3880 non-failed companies in the period 1972 – 1978. He examined three independent variables: the ratio of net income to total assets, the ratio of total liabilities to total assets, the ratio of current assets to current liabilities.

2. Logit

Logistic regression is standard mathematical-statistic method for decades. It is applied almost in cases that explaining, dependent variable is not is not continuous but binary – dichotomous or alternative (if the explaining variable is continuous we use classical linear regression), so it means it can take only two values. Variable y that we try to explain using ratio (explanatory variables), it takes discrete value. If the company prospers it is the value 1 and if the company declared bankrupt it is the value 0. For this reason, we cannot use “classic” regression but adjusted regression analysis, called Logit and Probit models (Spuchľaková & Cúg, 2014).

Logit analysis is characterized by prediction of probability of the event that either occur or not. Calculated probability is thus equal to either 1 or 0. It is necessary to realize logit transformation within the logistic regression to establish this condition. This logit transformation is based on “ratio of chances and hopes”. Given transformation allows the ideal relationship between dependent variable y and a vector of independent variables x. If values of independent variable are very low the probability of the variable y close to zero and if the values of independent are high the probability of y close to one. Logistic regression uses categorically explained variable (Kollár, 2014).

We meet in literature with different types of classification variables. According to one of approaches it is distinguished variables, where the main criterion is the type of relationship between values (Gregová, 2007):

- **Nominal** variable is classified only quantitative. This means that its value either belongs or does not belong into certain category and we do not know determine the order of these categories. Individual values of nominal variable are expressed by words or numerical codes (so it make easier further computer processing). A typical example is marital status (single, married, divorced, widowed), place of birth (Bratislava, Rome, Wien, and so on), nationality (Slovak, Czech) and so on. A special case of nominal variable is dichotomous variable which acquiring only two possible values (gender, smoker and non-smoker, etc.).
- **Ordinal** variable meet all conditions that are required from nominal variable and in addition we can determine the order of its values. It is not possible to determine how much do one value higher or lower. We classify here for example the level of education (primary, secondary without graduation, secondary with graduation, and so on), the degree of customer satisfaction with certain product, etc.
- **Interval** variable allows to find out not only the order but also quantifies difference between of its two values. It does not allow to determine their ratio because it does not have “rational zero” in its scale of values. An example is the monthly income of household or level of blood cholesterol. In literature commonly reported example is temperature in Celsius (or Fahrenheit), where 0°C does not mean the absence of the temperature (temperature 30°C is higher about 15°C than 15°C but not two times higher).
- **Ratios** variable have defined “rational zero” and therefore it make sense to talk about how many times is one value higher (lower) than the other. If we measure the temperature on the Kelvin scale, so we can determine not only the fact that the temperature 100 degrees is about 50 higher than 50 degrees, but also the fact that the temperature is exactly two times higher. Another example is weight, number of household members, age, etc.

Nominal and ordinal variables are complexly marked as qualitative, interval and ratio variables and known as quantitative (cardinal, numerical). Quantitative variables can take form of discrete values so it means only integer values, or continuous values so it means discretionary value from certain interval (Micháliková, Spuchľáková & Cúg, 2014).

The aim of logistic regression (it is similar to the linear regression) is expressed dependence of magnitude Y on the variable x_k. It does not use a linear dependence. Observed data are interleaved by logistic curve instead of line and its prescript is (Cisko & Klieštik, 2013):

$$\Pi = \frac{e^{\beta+\beta_1 x_1+\ldots+\beta_k x_k}}{1+e^{\beta+\beta_1 x_1+\ldots+\beta_k x_k}} = \frac{1}{1+e^{-(\beta+\beta_1 x_1+\ldots+\beta_k x_k)}} \tag{1}$$

where:

Π - probability of default,
x_k - the value of k-th financial indicator,
β_k - coefficients of individual indicators.

Variables of this logistic function x_k are chosen financial indicators while coefficients β_k is necessary to estimate – these coefficients are estimated by using of the method maximal likelihood. In this case it cannot to stated that these coefficients are weights of individual financial indicators and their finally interpretation is then more complicated than the linear models. For determine the probability of default it is not necessary to use calibration of model, it is sufficient to calculate median of known probabilities of all subjects (observed companies) which are in individual group. Logistic function may be increasing and also decreasing but its value is always between zero and one (Valášková, Gavláková & Dengov, 2014).

Logistic regression eliminates disadvantages of discriminant analysis – does not assume normal distribution of independent variables and homogeneity of variation-covariance matrices (Kollár & Cisko, 2014).

We can see in following table that we can use not only generalized logit model. Starting with the simple binary logit model, research progressed during the 1960s and 1970s to the multinomial logit (MNL) and nested logit models, the latter becoming the most popular of the generalized logit models. Also mixed logit models and its variants have now supplanted simpler models in many areas of economics, marketing, management, transportation, health, housing, energy, research and environmental science (Train, 2003; Jones & Hensher, 2008).

Here introduce the paper, and put a nomenclature if necessary, in a box with the same font size as the rest of the paper. The paragraphs continue from here and are only separated by headings, subheadings, images and formulae. The section headings are arranged by numbers, bold and 10 pt. Here follows further instructions for authors.

Table 1. Different logit models and their strengths and challenges

	Classical MNL	Nested Logit	Mixed Logit	Latent Class-MNL
Major Strenght	- Closed-form solution - Provides one set of globally optimal parameter estimates - Simple calculation - Widely understood and used in practise - Easy to interpret parameter estimates - Easy to calculate probability outcomes	- Closed-form solution - Provides one set of globally optimal parameters - Relatively easy to interpret - Relatively easy to calculate probability outcomes - Prartially corrects fo IID condition - Incorporates firm-specific observed and unobserved heterogeneity to some extent (especially the covariance	- Allows for complete relaxation of IID condition - Avoids violation of the IIA condition - High level of behavioural definition and richness allowed in model specification - Includes additional estimates for random parameters, heterogeity in means and decompositions in variances (these influences are effectively treated as "while	- Clossed-form solution - Semi-parametric specification - Like mixed logit, this model form is free form many limiting statstical assumptions, such as homogeneity in variances and normality assumptions - Incorporates firm-specific observed and unobserved heterogeneity through "latent class" constructs

	Classical MNL	Nested Logit	Mixed Logit	Latent Class-MNL
	- Less demanding data quality requirements	extension)	noise" in basic models)	- Less complex interpretation than mixed logit
Major Challenges	- Highly restrictive error assumptions(IID condition) - Violates the IIA assupmption - Ignores firm-specific observed and unobserved heterogeneity which can lead to inferior model specification and spurious interpretation of model outputs - Parameters are point estimates with little behavioural definition - Often provide good aggregate fits but can be misleading given simple form of the model - Tends to be less behaviourally resposive to changes in attribute levels	- Only partially corrects for IID condition - Analytically very closely related to basic MNL model (thus shares many of the limitations of MNL) - Does not capture potential sources of correlation across nests - Judgement required in determining which alternatives can be appropriately partitioned into nests (nested logit requires well separated nests to reflect their correlation)	- Open-form solution (requires analytical integration and use of simalted maximum likelihood to estimate model parameters) - Lack of a single set of globally optimal parameter estimates (i.e. due to the requirement for simalted maximum likelihood) - Assumptions must be imposed for the distribution of unobserved influences - Complex interpretation - Model estimation can be time consuming due to computational intensity - High quality data contraints	- Lacks flexibility in specification of firm-specific unobserved - Model estimation can be time consuming due to computational intensity - Assumption that manifest variables within latent classes are independent can be unrealistic - High quality data constraints

3. Probit

Probit analysis is alternative of logit method. The main difference is that assume normal distribution of random variables (independent variables in model). The difference lies in fact that logistic function has harder “fat tails”. There are no significant differences in practise, only in the case that sample contains numerous observations with extreme values. Parameters estimates obtained by logit and probit models cannot be compared directly because the logarithmic distribution has variance equal $\pi^2/3$, therefore the estimates obtained by logit model have to be multiplied by $3^{1/2}/\pi$ in order to be comparable with estimates obtained in the probit model (Lehútová, 2011).

4. Conclusion

Logit and probit models are very similar to each other. Distribution function in the logit model on the contrary of probit model has “flatter tails” (distribution has more observations appearing at the end of the distribution function). Distribution function of probit model has steeper slope so we can see on the figure 1. Both distribution function are almost linear between π = 0,2 and π = 0,8 (Majerčák & Majerčáková, 2013).

Parameter estimation in logit and probit models vary significantly. Logarithmic distribution has variance equal to π2/3, β estimations comes from logit model have to be multiplied by 31/2/π in order to be comparable with estimations acquired from probit model which has variance equal to one (Kicová & Kramárová, 2013).

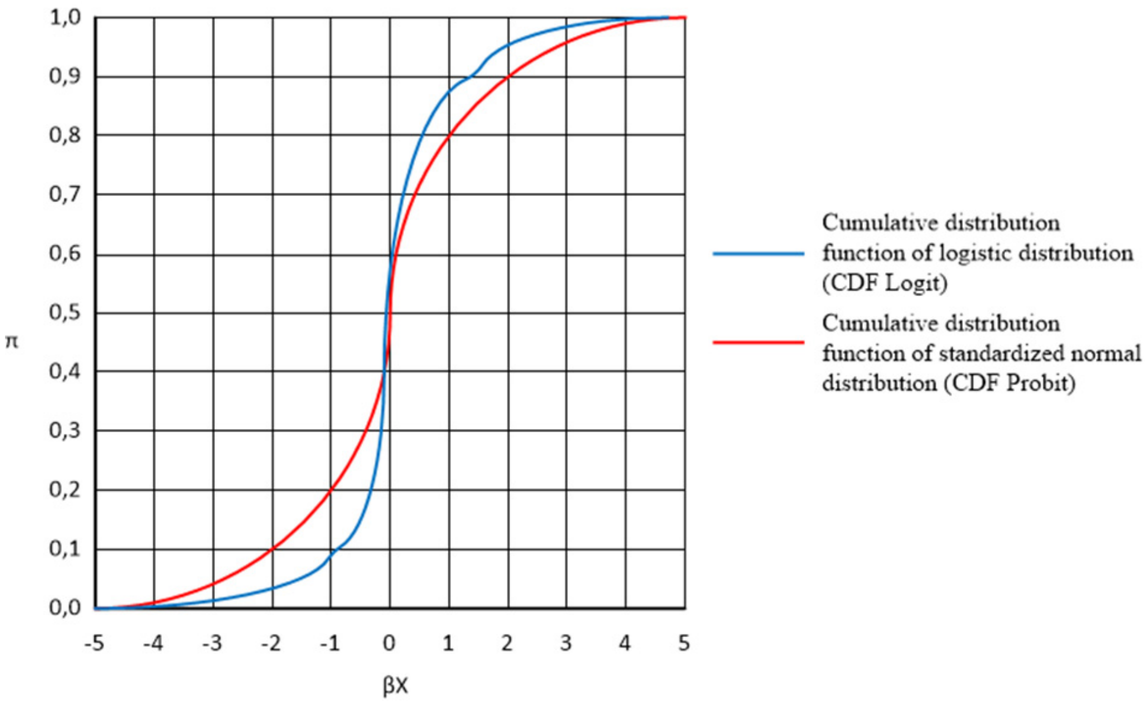

Fig. 1. Compare distribution function of Logit and Probit models

Estimation of probability of default in these models is dependent variable which variance is independent variable. Independent variables are expressed by various financial indicators or other indicators which representing external variables reflected financial environment.

Logit model has two practical advantages on the contrary of probit model, despite their mutual similarity (German, 2008):

- **Simplicity** – equation of logistic distribution function is very simply while normal cumulative distribution function contains unquantified integral.
- **Interpretability** – inverse linear transformation of logit model can be interpreted directly as logarithm of *chances*, while inverse transformation of probit has not direct interpretation.

Acknowledgements

The contribution is an output of the science project VEGA 1/0656/14- Research of Possibilities of Credit Default Models Application in Conditions of the SR as a Tool for Objective Quantification of Businesses Credit Risks.

References

Bliss, C. I. (1934). The Method of Probits. *Science*, 79, 9-38.
Cisko, Š., & Klieštik, T. (2013). *Finančný manažment podniku II*. Žilina: EDIS Publisher.
Finney, D. J. (1947). *Probit Analysis: A Statistical treatment of the Sig Moid Response Curve.* Cambridge University Press.
German, M. (2008). *Modelovanie vzťahov na trhu práce pomocou metód maximálne vierohodnosti*, Diploma thesis, Univerzita Komenského v Bratislave.
Gregová, E. (2007). *Malé a stredné podniky v podmienkach transformujúcich sa ekonomík SR a RF*. Žilina: EDIS Publishers, University of Žilina.
Jones, S., & Hensher, D. A. (2008). *Advanced in Credit Risk Modelling and Corporate Bankruptcy Prediction.* Cambridge University Press.
Kicová, E., & Kramárová, K. (2013). Innovations in outsourcing. *The 7th International Days of Statistics and Economics,* 588-600.
Kollár, B. (2014). Credit Value at Risk and Options of Its Measurement. *2nd International Conference on Economics and Social Science (ICESS 2014), Information Engineering Research Institute, Advances in Education Research*, 61, 143-147.
Kollár, B., & Cisko, Š. (2014). Credit risk quantification with the use of CreditRisk+. *International Conference on Management, Education,*

Business, and Information Science, 43-46.
Lehutová, K. (2011). Application of Corporate Metrics method to measure risk in logistics. *International Scientific Conference Computer systems aided science industry and transport*, 2209-2213.
Majerčák, P., & Majerčáková, M. (2013). The enterprise valuation and categories of the value. *Financial management of firms and financial institutions,* 469-475.
Martin, D. (1977). Early Warning of Bank Failure: A Logit Regression Approach. *Journal of Banking and Finance,* 1, 249-276.
Michalíková, F. K., Spuchľáková, E., & Cúg, J. (2014). A Comparative Anatomy of Credit Risk Models. *2nd International Conference on Economics and Social Science (ICESS 2014), Information Engineering Research Institute, Advances in Education Research*, 61, 69-74.
Ohlson, J. A. (1980). Financial Ratios and Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research,* 18, 109-131.
Spuchľáková, E., & Cúg, J. (2014). Lost Given Default and the Credit risk. *International Conference on Management, Education, Business, and Information Science,* 12-15.
Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.
Valášková, K., Gavlaková, P., & Dengov, V. (2014). Assessing credit risk by Moody's KMV model. *2nd International Conference on Economics and Social Science (ICESS 2014), Information Engineering Research Institute, Advances in Education Research*, 61, 40-44.
Zmijewski, M. (1984). *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*. *Journal of Accounting Research,* 22, 59-82.
Zvaríková, K., & Majerová, J. (2014). Financial literacy in the Slovak republic. *Journal Procedia – Social and Behavioral Sciences*, 110, 1106-1115.