2009 International Conference on Information Management and Engineering

Using Bayesian Networks for Bankruptcy Prediction: Empirical Evidence from Iranian Companies

Arezoo Aghaie
Faculty of Management and Accounting
Azad University-Mobarakeh Branch
Mobarakeh, Iran
E-mail: arezooaghaie2001@yahoo.com

Ali Saeedi
Department of Accounting
University of Isfahan
Isfahan, Iran
E-mail: alisaeediv@yahoo.com

Abstract - Financial distress and bankruptcy of companies may cause the resources to be wasted and the investment opportunities to be faded. Bankruptcy prediction by providing necessary warnings can make the companies aware of this problem so they can take appropriate measures with these warnings. The aim of this study is model development for financial distress prediction of listed companies in Tehran stocks exchange (TSE) using Bayesian networks (BNs). The sample consists of 72 bankrupt firms and 72 non bankrupt ones from 1997 to 2007 and bankrupt firms are those firms that subject to Business Law par. 141.

In order to develop a bankruptcy prediction model, we consider 20 predictor variables including liquidity ratios, leverage ratios, profitability ratios and other factors like firm's size and auditor's opinion and then we use two methods for choosing variables. The first method is based upon conditional correlation between variables and the second method based upon conditional likelihood. Then three models for predicting financial distress are developed using naïve bayes model and regression model and the result of three models are compared. The accuracy in predicting bankruptcy of the first naïve bayes model's performance that is based upon conditional correlation is 90% and the accuracy of the second naïve bayes model is 93% and finally the accuracy of the logistic regression that was built for comparing to naïve bayes models is 90%.

Collectively the results show that it is possible to predict financial distress using Bayesian models. Also, because this prediction is based on the information provided in financial statements of companies, it can be an evidence that the financial statements of companies have information content. With respect to the remainder variables in developed models in this research we find firms that have lower profitability and have more long term liabilities and have lower liquidity are more in risk of financial distress. To reduce financial distress risk, firms should use more conservative methods which lead to decrease in debts and reduce their costs. Further analyses show that the discretization into two, three and four states cause the model's performance to increase but increasing states into five states causes the model's performance to decrease.

Keywords: Bankruptcy Prediction; Financial Distress; Bayesian Networks; Naïve Bayes; Discretization of Continuous Variables; Logistic regression

1. INTRODUCTION

In today's dynamic economic environment, the number and the magnitude of bankruptcy filings are increasing significantly. Even auditors, who have good knowledge about firms' financial position, often fail to make an accurate judgment on firms' going-concern conditions (Sun & Shenoy, 2006). The prediction of financial distress is an important and challenging issue that has served as the impetus for many academic studies over the past three decades (Bevear, 1966).

The aim of this study is to model Financial Distress prediction of listed companies in Tehran Stocks Exchange (TSE) using Bayesian Networks (BNs). The sample consists of 72 bankrupt firms and 72 non bankrupt ones from 1997 to 2007 and bankrupt firms are those firms that subjected to Business Law par. 141.

In order to develop a bankruptcy prediction model, we consider 20 predictor variables including liquidity ratios, leverage ratios, profitability ratios and other factors like firm's size and auditor's opinion and then we use two methods for choosing variables. The first method is based upon conditional correlation between variables and the second method based upon conditional likelihood. Then, three models for predicting financial distress are developed using naïve bayes models and regression model and the result of three models are compared.

The accuracy in predicting bankruptcy of the first naïve bayes model's performance that is based upon conditional correlation is 90% and the accuracy of the second naïve bayes model is 93% and finally the accuracy of

978-0-7695-3595-1/09 $25.00 © 2009 IEEE
DOI 10.1109/ICIME.2009.91

the logistic regression that was built to compare to naïve bayes models is 90%.

The second naïve bayes model's performance that we develop based upon conditional likelihood, in Comparison to the Sun and Shenoy's (2006) model, is better and the logistic regression model's performance that is based upon conditional likelihood, is better too.

Collectively, the results show that it is possible to predict financial distress using Bayesian models. Also, because this prediction is based on the information provided in financial statements of companies, it can be an evidence that the financial statements of companies have information content. With respect to the remainder variables in developed models in this research we find firms that have lower profitability and have more long term liabilities and have lower liquidity are more in risk of financial distress. To reduce financial distress risk, firms should use more conservative methods which lead to decrease in debts and reduce their costs. Further analyses show that the discretization into two, three and four states cause the model's performance to increase but increasing states into five states causes the model's performance to decrease.

2. LITERATURE REVIEW

For a number of years, there were considerable research by accountants and finance people trying to find a business ratio that would serve as the sole predictor of corporate bankruptcy. Beaver (1966) conducted a very comprehensive study using a variety of financial ratios. His conclusion was that the cash flow to debt ratio was the single best predictor. Beaver's univariate analysis led the way to a multivariate analysis by Altman (1968), who used Multiple Discriminant Analysis (MDA) in his effort to find a bankruptcy prediction model. He selected 33 publicly-traded manufacturing bankrupt companies between 1946 and 1965 and matched them to 33 firms on a random basis for a stratified sample (assets and industry). The results of the MDA exercise yielded an equation; he called it Z-score, which correctly classified 94% of the bankrupt companies and 97% of the non-bankrupt companies one year prior to bankruptcy. The ratios used in the Altman (1968) model are: working capital over total assets, retained earning over total assets, earnings before interest and taxes over total assets, market value of equity over book value of total liabilities, and sales over total assets (Leano, 2005). Beginning in the 1980s more advanced estimation methods, such as logit (Ohlson, 1980) and probit (Zmijewski 1984), were employed.

In 1990, the NNs technique has jumped into the field of corporate bankruptcy prediction and it has been a very popular technique ever since. Odom and Sharda (1990) were the first to apply NNs to bankruptcy prediction problem. And other studies using the NNs techniques are Altman, et al (1994), Atiya (2001), Cadden (1991), Coats & Fant (1992), Tam & Kiang (1992) and Wilson & Sharda (1994). The NNs technique dominates the literature on business failure in 1990s, and still most frequently used in corporate bankruptcy prediction (Yeong, 2004).

Later on, Sarkar and Sriram (2001) developed Bayesian Network (BN) models for early warning of bank failures. They found that both a naïve BN model and a composite attribute BN model have comparable performance to the well-known induced decision tree classification algorithm. Some other techniques, such as Rough Set Theory (McKee 1998), Discrete Hazard models (Shumway, 2001), and Genetic Programming (McKee and Lensberg 2002), have also been introduced to the bankruptcy prediction area (Sun and Shenoy, 2006).

3. BAYESIAN NETWORK MODELS

Bayesian Networks are gaining an increasing popularity as modeling tools for complex problems involving probabilistic reasoning under certainty. Bayesian networks (BN) are probabilistic graphical models that represent a set of random variables for a given problem, and the probabilistic relationships between them. The structure of a BN is represented by a direct acyclic graph (DAG), in which the nodes represent variables and the edges express the dependencies between variables (Pearl 1988).

3.1. UNDERLYING CONCEPTS AND THEORY

Bayesian networks are based upon probability theory and the basic measure of our belief in a proposition (say A) will be the function P (A). The basic concept in the Bayesian treatment of certainties in Bayesian Network; is conditional probability which gives a measure of how our beliefs in certain propositions are changed by the introduction of related knowledge (Kyprianidou, 2002). Bayes rule can be expressed as follows:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

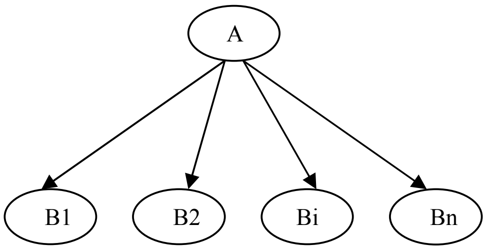
3.2. ADVANTAGES AND DISADVANTAGES OF BAYESIAN NETWORKS

The major advantage of Bayesian Network is that the output is explicitly a probability, which can be easily interpreted. A decade ago the calculations needed to prograte the probabilities through the nodes of complex Bayesian Networks were prohibitively time consuming but the increased capacity of modern computers has attracted interest in the implementation of Bayesian Networks. An important advantage of a Bayesian Network is the availability of a graphical model framework of a problem, which is useful for both people and computers. One problem of Bayesian Network centers on the quality and extends of the prior beliefs used in Bayesian inference processing. A Bayesian Network is only as useful as this prior knowledge is reliable since the quality of these prior beliefs will distort to the entire network and invalidate the results. Perhaps the most significant disadvantage of an approach involving Bayesian Networks is the fact that there is no universally accepted method for constructing a network from data (Kyprianidou, 2002).

3.3 A Naïve Bayes Bayesian Network Model

A naive Bayesian network is a very simple structure in which all random variables representing observable data have a single, common parent node—the class variable. The naive Bayesian classifier has been used extensively for classification because of its simplicity, and because it embodies the strong independence assumption that, given the value of the class, the attributes are independent of each other (Ceruti, 2002). Figure 1 presents a graphical representation of a naïve Bayesian network model.

Figure 1: A Naïve Bayes BN Model



In a naïve Bayes model, the node of interest has to be the root node, which means, it has no parent nodes. In a bankruptcy prediction context, in Figure 1, A represents the bankruptcy variable. $B1$, $B2$ …, Bn represent n bankruptcy predictor variables. The naïve Bayes model assumes the following conditional independence:

$Bi \perp \{B1, B2, ..., Bi-1, Bi+1, ...Bn\} \mid A$,

for i =1, 2, …, n.

The above assumption says that predictors, $B1$, $B2$ …, Bn are conditionally mutually independent given the state of bankruptcy (Sun and Shenoy, 2006).

4. HYPOTHESIS DEVELOPEMENT

In this research we test the hypothesis of possibility of using Bayesian networks for predicting financial distress of firms listed in Tehran Stock Exchange. The research hypothesis is as follows:

H0: It is possible to predict financial distress of firms listed in Tehran Stock Exchange using Bayesian networks.

H1: It isn't possible to predict financial distress of firms listed in Tehran Stock Exchange using Bayesian networks.

5. SAMPLE AND DATA

Sample firms used in this study are companies that listed in Tehran Stocks Exchange (TSE) across various industries during the period 1997 – 2007. We do not impose any selection restriction on the size or industry characteristics when forming bankrupt and non-bankrupt samples. At first we identify bankrupt firms: Those firms that are subjected to Business Law par. 141. The process of identifying bankrupt firms results in 72 bankrupt firms during the period 1997-2007. Then 72 non bankrupt firms are selected during the period 1997-2007 randomly.

Through our own analysis and reviewing past research 20 variables are identified as potential bankruptcy predictors. These variables are included financial ratios measuring firm's liquidity, leverage, turnover, profitability and firm's size and other factors like auditors' opinions. All variables for bankrupt firms are calculated in the year that firms are subjected to Business Law par. 141. And all variables for non bankrupt firms are calculated in the base year. The variables in this study are shown in table 1.

6. RESEARCH PROCESS AND RESEARCH RESULTS

6.1. First method for Variable Selection in Naïve Bayes Models

There exists a large pool of bankruptcy predictors. An appropriate selection of a subset of variables is necessary for developing a useful naïve Bayes model. Variable selection is really important on account of irrelevant and redundant features may confuse the learning algorithm and obscure the predictability of truly effective variables.

So, a small number of predictive variables are preferred over a very large number of variables including irrelevant and redundant ones. One of the purposes in this paper is to provide a good method to guide the selection of variables in naïve Bayes models. Hence, we compare two different methods for selecting the variables. The first method is depending upon correlations and conditional correlations among variables.

First, we obtain the correlations among all variables, including 20 potential predictors and the variable of interest, firm's bankruptcy status. Variables that have significant correlations are assumed to be dependent and therefore connected. The correlations are obtained using the entire sample of 144 firms, including 72 non-bankruptcies and 72 bankruptcies. At the first stage, four predictors (x5, x8, x11, x12) are connected with bankruptcy status, since they have dependency with bankruptcy. These variables are first-order variables. Then, we identify second-order variables, the variables that have affect on first-order variables. Conceptually, second-order variables are those that have significant correlations with first-order variables. To select a given first-order variable's second-order variables, we follow the similar method used to select first-order variables. The major difference is that now we consider each first-order variable instead of bankruptcy as a root variable. For example the conditional correlation between x5 and x17, x7, x9, x19, x4 is significant so they connect to x5 in the model.

After obtaining the conditional correlation between all variables and selecting the first-order and second-order variables, x2, x3, x10, x15, x18 and x20 are eliminated. Figure 2 shows the first naïve bayes model.

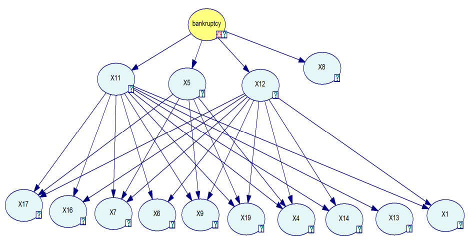


Figure 2: The Structure of the first Naïve Bayes Model

The naïve Bayes model is typically used with discrete-valued data. Prior research has used bracket median method (Sarkar and Sriram 2001) and extended Pearson-Tukey (EP-T) method (Sun and Shenoy, 2004) for

Variables	Definition
X1	Natural log of (Total Assets/ GNP Index)
X2	(Current Assets – Current Liabilities)/Total Assets
X3	Current Assets/ Current Liabilities
X4	Operating Cash Flows /Total Liabilities
X5	Current Assets/Total Assets
X6	Cash/Total Assets
X7	Total Liabilities/Total Assets
X8	Long Term Debts/Total Assets
X9	Sales/Total Assets
X10	Current Assets/Sales
X11	Earnings before Interest and Taxes/Total Assets
X12	Net income/Total assets
X13	One if net income was negative for the last two years, else zero
X14	Retained Earnings/Total Assets
X15	(Net income in year t – Net income in t–1)/(Absolute net income in year t + Absolute net income in year t–1)
X16	Natural log of total assets
X17	Zero if auditors' opinions is unqualified otherwise one
X18	Net income/ sales
X19	Retained Earnings/ total owner's equity
X20	Quick assets / total assets

Table 1: Definitions of Potential Predictor Variables

discretization, which divides the continuous cumulative probability distribution into *n* equally probable intervals. We used Uniform Widths method to convert continuous variables into discrete. During the discretization process, one problem that researchers face is to decide the number of states for discretization. We start from two states to five states and test the model with all samples in the research.

When continuous variables are discretized into 2 states, the model's accuracy in predicting bankruptcy is 83%, and its accuracy in predicting non-bankruptcy is 74%. When the number of discretization states increases to 3, the model's accuracy in predicting bankruptcy is 84% and its accuracy in predicting non-bankruptcy is 84%. When the number of states increases to 4, the model's accuracy in predicting bankruptcy is 90%, which is statistically indifferent to the model's performance with 2 or 3 states and its accuracy in predicting non-bankruptcy is 89%. When we increase the number of states for discretization further, the model's performance continues to drop.

6.2. Second method for Variable Selection in Naïve Bayes Models

In the second method, the variables are selected based upon conditional likelihood. First, we obtain the correlations among all variables, including 20 potential predictors and the variable of interest, firm's bankruptcy status. Variables that have significant correlations are assumed to be dependent and therefore connected. The correlations are obtained using the entire sample of 144 firms, including 72 non-bankruptcies and 72 bankruptcies. At first stage four predictors (x5, x7, x11, x20) are connected with bankruptcy, since they have dependency with bankruptcy. These variables are first-order variables. Then, we identify second-order variables. After selecting the first-order and second-order variables, x1, x6, x13, x14, x15, x16, x18 and x19 are eliminated. Figure 3 shows the second naïve bayes model.

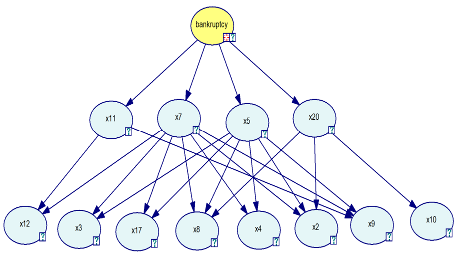


Figure 3: The Structure of the second Naïve Bayes Model

When continuous variables are discretized into 2 states, the model's accuracy in predicting bankruptcy is 87%, and its accuracy in predicting non-bankruptcy is 80%. When the number of discretization states increases to 3, the model's accuracy in predicting bankruptcy is 92% and its accuracy in predicting non-bankruptcy is 87%. When the number of states increases to 4, the model's accuracy in predicting bankruptcy is 94%, which is statistically indifferent to the model's performance with 2 or 3 states and its accuracy in predicting non-bankruptcy is 92%. When we increase the number of states for discretization further, the model's performance continues to drop. The rates show that this model's performance is better than the first one.

6.3. Naïve Bayes vs. Logistic Regression

In this section, we compare the performance of the naïve Bayes models in Figure 2 and 3 with that of logistic regression, a widely used bankruptcy prediction tool. The study sample is all 144 firms that we had in this study. Using the same four variables presented in Figure 3, logistic regression has an average prediction rate of 90.2%. The estimation of logistic regression is as follows:

$$Y=-4.956+8.483*x5+5.643*x7-25.597*x11-8.958*x20$$

Where $y = \ln \pi/1-\pi$

That $\pi = P(y_i=1) = P(\text{bankruptcy})$

7. SUMMARY AND CONCLUSIONS

In this study, we examine several important methodological issues related to the use of naïve Bayes Bayesian Network (BN) models to predict bankruptcy. First, we provide two different methods that guide the selection of predictor variables from a pool of potential variables. Under the first method, only variables that have significant correlations with the variable of interest, the status of bankruptcy, are selected. As a result, 4 variables are selected from a pool of 20 potential predictors. Then we selected second order variables. The first naïve BN consisting of these selected variables have an average prediction accuracy of 90% for the bankruptcy sample and 89% for the non-bankruptcy sample.

Second, we investigate the impact on a naïve Bayes model's performance of the number of states into which continuous variables are discretized. We find that the model's performance is the best with the continuous variables being discretized into 4 states. When the number of states is increased to 5 or more, the model's performance deteriorates. Then, we use conditional likelihood method for selecting variables and run the second naïve bayes model. This model has an average prediction accuracy of 94% for the bankruptcy sample and 92% for the non-bankruptcy sample. Finally, we run a logistic regression model and our result show that the naïve bayes model that based upon conditional likelihood has the best performance of predicting bankruptcy among these three models. The results show that it is possible to predict financial distress using Bayesian models. Also, because this prediction is based on the information that is existed in financial statements of companies, it can be evidence that the financial statements of companies have information content. With respect to the remainder variables in developed models in this research we find firms that have lower profitability and have more long term liabilities and have lower liquidity are more in Risk of financial distress. To reduce financial distress risk, firms should use more conservative methods which lead to decrease in debts and reduce their costs.

8. RESEARCH RECOMMENDATIONS

1. We recommend investors to use these naïve bayes models for evaluating firms and for deciding about investment in companies' stocks.
2. We recommend Tehran Stock Exchange to use these naïve bayes models for listing firms and evaluating them.
3. We recommend banks and financial institutes to use these naïve bayes models for loan making dicisions.

REFERENCES

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

Beaver, W.H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, Issue 4, pp. 71-111.

Fulmer, J.G & Moon, J.E & Gavin, T.A & Erwin, J.M. (1984). A Bankruptcy Classification Model for Small Firms. *The Journal of Commercial Bank Lending*, pp. 25-37.

Grice, J.S & Ingram, R.W. (2001). Tests of the Generalizability Of Altman's Bankruptcy Prediction Model. *Journal of Business Research,* pp. 53-61.

Hector Leaño, Clinton Martin, John McBride, Jennifer Nguyen and Padmaja Pallom. (2005). Bankruptcy Prediction Model Using Discriminant Analysis on Financial Ratios Derived from Corporate Balance Sheet.

Hu, Yu-chiang. Ansell, Jake. (2006). Developing Financial Distress prediction models. Working paper. Management school and Economics. University of Edinburgh, UK.

Kyprianidou, C. (2002). Analyzing Basic Genetics Using Bayesian Networks. (Ph.D) dissertation. Msc Actuarial science cases business school city university.

McKee, T.E, and T. Lensberg. (2002). Genetic Programming and Rough Sets: A Hybrid Approach to Bankruptcy Classification. *European Journal of Operational Research,* pp. 436–451.

Odam, M. & Sharda, R. (1990). Neural Network for Bankruptcy Prediction. *Probus Publishing Company*, pp. 177-185.

Ohlson, J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, Vol. 18, pp. 109-131.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. Network of Plausible Inference.

Sarkar, S & Sriram, R. (2001). Bayesian Models for Early Warning of Bank Failures. *Management Science*, pp. 1457-1475.

Sun, L & Shenoy, P. (2007). Using Bayesian Networks for Bankruptcy Prediction. *European Journal of Operational Research*, Vol. 180, Issue. 2, pp. 738-753.

Yeong, H. (2004). Financial Data Modeling and Analysis for Bankruptcy Prediction. Working paper. University of Technology. Sydney.

Zimijewski, M.E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting research,* pp. 59-82.