Bankruptcy Prediction Using Bayesian, Hazard, Mixed Logit and Rough Bayesian Models: A Comparative Analysis

Arindam Chaudhuri[1]

[1] Faculty of Post Graduate Studies and Research in Engineering and Technology, Marwadi Education Foundation's Group of Institutions, Rajkot, India

Correspondence: Arindam Chaudhuri, Faculty of Post Graduate Studies and Research in Engineering and Technology, Marwadi Education Foundation's Group of Institutions, Rajkot, India. E-mail: arindam_chau@yahoo.co.in

Abstract

Bankruptcy prediction has been a topic of active research for business and corporate institutions in recent times. The problem has been tackled using various models viz. Statistical, Market Based and Computational Intelligence in the past. In this work, we analyze bankruptcy using both parametric and nonparametric prediction techniques. This investigation concentrates on the impact of choice of cut off points, sampling procedures and business cycle on accuracy of bankruptcy prediction models. Misclassification can result in erroneous predictions leading to prohibitive costs to investors and economy. To test the impact of choice of cut off points and sampling procedures, four bankruptcy prediction models are examined viz. Bayesian, Hazard, Mixed Logit and Rough Bayesian techniques. To evaluate the relative performance of models, a sample of firms from Lynn M. LoPucki Bankruptcy Research Database in US is used. The choice of cut off point and sampling procedures are found to affect rankings of various models. The results indicate that empirical cut off point estimated from training sample resulted in lowest misclassification costs for all the models. Although Hazard and Mixed Logit models resulted in lower costs of misclassification in randomly selected samples, Mixed Logit model did not perform well across varying business cycles. Hazard model has highest predictive power. However, higher predictive power of Rough Bayesian and Bayesian modes when ratio of cost of Type I to cost of Type II errors is high is relatively consistent across all sampling methods. This advantage of Bayesian models may make them more attractive in current economic environment. This study also compares the performance of bankruptcy prediction models by identifying conditions under which a model performs better. It applies to a varied range of user groups including auditors, shareholders, employees, suppliers, rating agencies and creditors' concerns with respect to assessing failure risk.

Keywords: bankruptcy prediction, Bayesian networks, Hazard model, Mixed Logit model, Rough Bayesian networks

1. Introduction

Bankruptcy prediction (Altman, 1993; Bellovary et al., 2007; Pate, 2002) is an important and challenging topic for business and corporate institutions. Prediction of corporate bankruptcy is a phenomenon of increasing interest to investors or creditors, borrowing organizations and governments alike. It is a key goal of any management. Timely identification of organizations' impending failure is desirable. Bankruptcy is the condition in which an organization cannot meet its debt obligations and petitions federal district court for either reorganization of its debts or liquidation of its assets. In action, the property of debtor is taken over by receiver or trustee in bankruptcy for benefit of creditors. An effective prediction in time is valued priceless for business in order to evaluate risks or prevent bankruptcy (Altman et al., 1977; Altman, 1993). A fair amount of research has therefore focused on bankruptcy prediction (Agarwal et al., 2008; Altman, 1968, 1993, 2007; Altman et al., 1977; Altman et al., 1994; Beaver et al., 2005; Begley et al., 1996; Chava et al., 2004; Grice et al., 2001; Hensher et al., 2007; Hillegeist et al., 2004; Hsieh, 1993; Jones, 1987; Katz et al., 1985; McKee, 2003; Mensah, 1984; Michael et al., 1999; Ohlson, 1980; Ravi et al., 2007; Robertson, et al., 1991; Sarkar et al., 2001; Shumway, 2001; Sun et al., 2007; Tam, 1991; Weiss et al., 2004; West, 1985; Wilson et al., 1994; Zavgren, 1983; Zmijewski, 1984). There may be early warning signs of impending financial distress and this would allow the manager to act in a

pre-emptive manner to mitigate the situation from worsening. Signs of potential financial distress are evident long before bankruptcy occurs (Altman, 1993). Financial distress begins when an organization is unable to meet its scheduled payments or when projection of future cash flows points to an inability to do so in near future. The causes leading to business failure and subsequent bankruptcy (Bellovary et al., 2007; Grice et al., 2001; Zavgren, 1993) can be divided into economic, financial, neglect, fraud, disaster and others. Economic factors include industry weakness and poor location. Financial factors include excessive debt and insufficient capital. Research shows that financial difficulties are the result of managerial error and misjudgment. When errors and misjudgments proliferate, it could be a sign of managerial neglect. Corporate fraud became a public concern during late nineties. However, no models are yet available that could detect and flag corporate fraud. Disaster is sometimes the cause of corporate failure. It includes human error and malice.

Bankruptcy filing is not exclusive to any specific economy. Globalization can feed waves of economic distress across societies and national economies after original economy witnesses its deleterious impact. Countries like Japan, Belgium, Thailand, Greece, Hungary etc. are developing their own bankruptcy prediction models to deter disastrous consequences of ultimate financial distress. Predicting corporate failure using past financial data is a traditional and modern topic of financial business (Altman, 1968; Altman et al., 1977; Altman, 1993; Altman et al., 1994; Beatty et al., 2002; Beaver, 1966; Hensher et al., 2007; Jones et al., 2004; Merton, 1974; Neophytou et al., 2004; Platt et al., 2004; Robertson et al., 1991; Thomson, 1991; Zavgren, 1983). The solution to this problem is discriminant function from variable space in which observations are defined into binary set. In United States, defaults and bankruptcies have increased markedly over past decade. In fact, 17 out of 20 largest bankruptcy filing in United States happened during this period. An appreciable volume of work in literature has been devoted to forecasting corporate failure. The methodologies employed have been based mainly on various Statistical and Computational Intelligence models. During this distress period, three important Statistical models viz. Bayesian (Sarkar et al., 2001), Hazard (Shumway, 2001) and Mixed Logit (Jones et al., 2004) have been successfully applied to bankruptcy prediction. All these models have theoretical advantages over existing prediction models.

In this study, we address the issue of empirical testing of comparative power of the above mentioned three approaches along with Rough Bayesian model (Pawlak, 1982, 1991; Pawlak et al., 1994; Sarkar et al., 2001; Sun et al., 2007; Verikas et al., 2010; Yao, 2007, 2008, 2010; Yao et al., 1992; Yao et al., 1990). We examine the role of cut off point choice and sampling procedures in evaluating relative performance of bankruptcy prediction models. Specifically, following four important questions are addressed: (i) Does rankings of bankruptcy prediction models depend on arbitrary choice of cut off point? (ii) What is the procedure to determine the optimal cut off point? (iii) Does performance outcomes of bankruptcy prediction models depend on sampling procedures? (iv) How does the cost of Type I and Type II errors influence performance of bankruptcy prediction models? The first two questions focus on the critical role of cut off point as it affects bankruptcy prediction models. The third question extends bankruptcy inquiry to acknowledge the sensitivity of change in economic cycle. The fourth question provides insight to the fundamental question concerning choice of failure prediction models which are critical in optimal allocation of resources. The relevance of this research is twofold viz. (i) The current financial crisis is a result primarily of failure in accounting for credit risk and consequently failure of accounting for default risk by borrowers. This crisis thus demonstrates the need for better methods of evaluating risk of bankruptcy. (ii) The prudential banking regulation Basel II proposes banks use internal models to assess their risks and in particular their credit risks i.e. default risk and stockholders' equity necessary to cover risk of default.

This investigation is built upon prior and concurrent research in three different streams in the bankruptcy prediction literature. First, we build on studies by Jones et al. (2004) and Sarkar et al. (2001), Shumway (2001) and Hensher et al. (2007) who proposed the use of advanced probability modeling in prediction of corporate bankruptcy. These studies indicate that Bayesian, Hazard and Mixed Logit models have valuable applications in financial distress research. We attempt to extend these findings focusing on sensitivity of performance of their models to cut off point selection, to sampling procedure and to the ratio of cost of Type I to Type II error. The second stream of research demonstrates the importance of industry, change in accounting regulation and change in ability of financial statement data in forecasting accuracy of bankruptcy prediction models (Beaver et al., 2005; Chava et al., 2004; Hillegeist, 2004). The basic notion of their research is that besides industry effect, aspects of accounting systems such as going concern and conservatism principle among others limits the performance of any accounting based bankruptcy prediction model. Given their results, we study the impact of overall economic business cycle on testing bankruptcy prediction models. The third stream of research focuses on criterion used to compare bankruptcy prediction models such as classification accuracy, rates of misclassifications and cost of

misclassification (Begley et al., 1996; Grice et al., 2001; Jones et al., 2004; Sun et al., 2007). In this work, we address usefulness as measured by rankings of the four models to three criterions.

This study leads to the following important results. The choice of cut off points and sampling procedures were found to affect rankings of various models. The results indicate that empirical cut off point estimated from training sample resulted in lowest misclassification costs for Rough Bayesian model. Although Hazard, Mixed Logit and Rough Bayesian models resulted in lower costs of misclassification in randomly selected samples, Mixed Logit model did not perform as well across varying business cycles. In general, Rough Bayesian model has highest predictive power. However, the higher predictive power of the Bayesian model when ratio of cost of Type I to cost of Type II errors is high, is relatively consistent across all sampling methods.

The significance of this study to private and academic sectors by identifying under what conditions it is most appropriate to use which model in predicting bankruptcies. Auditors often fail to make accurate judgments on organizations' going concern conditions, notwithstanding their knowledge of organization (Hopwood et al., 1994; McKee, 2003). An appropriate choice of a bankruptcy model can help auditor in recognizing his disclaimer or qualification as to going concern nature of business (Altman et al., 1974). Bankruptcy prediction models can also serve audit researchers in better understanding auditor's biases (Sarkar et al., 2001). Second, in addition to auditors, creditors, stock-holders and senior management are all interested in bankruptcy prediction because it affects all of them (Wilson et al., 1994). An early warning model will allow management to take corrective actions before it is too late (Whalen et al., 1988). In fact, research in this area assumes greater significance because poor credit risk model might lead to sub optimal capital allocation (Agarwal et al., 2008). Regulators can intervene early so that mitigating actions can be taken to reduce expected costs of failure (Thomson, 1991). Finally, an appropriate model can help price distress security as shown by Katz et al. (1985) who found that abnormal return can be earned by identifying changes in scores calculated from bankruptcy model.

This paper is organized as follows. In section 2, the related work in bankruptcy prediction and its methodological implications are discussed. The experimental framework is presented in the next section. In section 4, the different data models for analysis are described. This is followed by the experimental results. Finally, conclusions are given in section 6.

2. Related Work

Beaver (1966) used univariate approach and found that net income to total debt had highest predictive ability. Although univariate analysis seems easy and intuitive to implement, the ambiguity of explanatory power does not provide clear signals (Jones, 1987; Zavgren, 1983). Altman (1968) used Linear Discriminant Analysis (LDA) which assumes equivalent covariance matrices across groups. In 1977, he proposed ZETA analysis which relaxes assumption of equivalent distribution of two group's covariance matrices. However, ZETA analysis is still criticized for assumption of multivariate normal distribution of independent variables (Jones, 1987; Tam, 1991). Tam (1991) found that most of the financial variables are not normally distributed. Unlike LDA and ZETA which are used as discrimination tool, Logit (Martin, 1977; Ohlson, 1980) and Probit (Zmijewski, 1984) are models designed for estimation of probability. These two models require assumptions only on residuals' distribution thus avoiding criticisms aimed at LDA and ZETA analysis. As parametric models dominated the research focus at beginning, studies using non parametric models started to develop in late 1980s. The most commonly used was Artificial Neural Networks (ANN) (Tam, 1991). Over past two decades, number of studies related to ANN in bankruptcy prediction is 39, compared with 11 in LDA, 19 in Logit and 3 in Probit (Bellovary et al., 2007). However, criticism of non parametric models is that significance of variables is not testable.

In the past decade, three important models viz. Bayesian (Sarkar et al., 2001), Hazard (Shumway, 2001) and Mixed Logit (Jones et al., 2004) have been applied to bankruptcy prediction. These models have subtle theoretical advantages over previous ones. Bayesian model applies well known Bayesian equation. Prior knowledge, practical estimates and subjective preference are easy to incorporate simultaneously or separately into the model. These prior recognitions are then adjusted by objective estimates from historical and empirical evidence. Subsequently posterior probability is obtained. The noticeable characteristics of this model are transparent. It is intuitive and easy to understand. The improvement of Hazard model over Logit is that former explicitly models bankruptcy not as process that happens at a point in time, an assumption made by all previous models, but as process that lasts for period of firm's life. Shumway (2001) argued that Hazard model is preferable as it incorporates time varying covariates. The advancement of Mixed Logit model over Logit is that it takes into account both observed information and unobserved information. Under this setting, there are two means to model unobserved information. The former is random parameter specification and later is error components approach.

Bankruptcy prediction models are usually evaluated from sample chosen in particular time period. Since distribution of accounting variables is dynamic (Mensah, 1984; Webster et al., 2005) models are likely to be sample specific (Agarwal et al., 2008). Grice et al. (2001) re-estimated Zmijewski's (1984) and Ohlson's (1980) models using time periods other than those used to originally develop models and found that accuracy of each model declined from period of 1988–91 to 1992–99. Begley et al. (1996) came to similar conclusion that Altman (1968) and Ohlson (1980) models did not perform well in 1980s even when coefficients were re-estimated. Robertson et al. (1991) cautioned that one could not assume that predictive power of models can transcend to industries. For example, LDA originally developed for manufacturing industry is used by practitioners as one of the important indicators of credit worthiness across many industries without re-estimating coefficients. Grice et al. (2001) found that Ollison's (1980) model was sensitive to industry classification while Zmijewski's (1984) model was not. Zmijewski (1984) indentified two sampling errors in evaluating bankruptcy prediction models. The first is choice based sample bias. The sampling procedure for bankruptcy prediction analysis initially identifies two groups of observations viz. bankrupt and non bankrupt. This procedure violates assumption of exogenous random sampling since probability that a firm enters sample depends on observed status. He found that predictive power is upwardly biased when sample selected has probability of bankruptcy which deviates from population probability of bankruptcy. He further suggested that weighted exogenous sample maximum likelihood can be used to adjust bias. Platt et al. (2002) tested choice based bias and results were consistent with Zmijewski's (1984) findings. The second sampling error refers to sample selection bias. Sampling procedures usually eliminate observations with incomplete data. Zmijewski (1984) used a bivariate normal approach to estimate correlation between bankruptcy and observation with missing data. He found that organizations with missing data would have higher probability of bankruptcy. In other words, sample selection bias understates bankruptcy probability. Beaver et al. (2005) pointed out that importance of intangible assets has increase over time due to technology based assets generated through research and development expenditure. In addition financial derivative markets experienced an explosion in 1990s. While financial derivatives are mainly used as substitute for leverage such a marked increase would lead to underestimate leverage ratios of organizations. Begley et al. (1996) also state that leverage variables play an important role. Beaver et al. (2005) argued that degree of discretion on financial statements is increasing. These three developments have direct effect on financial ratios which then undermine predictive power of bankruptcy prediction models whose inputs are mainly financial variables.

Changes in regulation may also have an impact on accuracy of accounting based prediction models. Since 1973 many new standards have been established for various perspectives such as Statement Number 87 for pensions, Number 106 for post retirement benefits, Number 107 for financial instruments and Number 115 for debt and equity. Any new standard will have an effect in long run on financial statement (2005). The use of bankruptcy filing has become strategic consideration since changes to bankruptcy laws in late 1970s (1996).

The criterion mainly used to compare bankruptcy prediction models is classification accuracy, rates and cost of misclassification. Classification accuracy is used by many researchers (Grice et al., 2001; Jones et al., 2004; Sun et al., 2007; Wiginton, 1980) because of its intuitive nature. Bankruptcy prediction models are applied in holdout sample or original sample that is used to estimate parameters of model and then sample is separated into groups according to organization's observed status. The rate of successful prediction is calculated within each group. The Lachenbruch validation approach is another validation method used by Altman et al. (1977). It is also known as Jackknife method where one observation is withheld from estimation sample and its status is predicted. Then same procedure is repeated until each observation has been predicted and individual classification accuracy is accumulated over entire sample.

Unlike classification accuracy rate of misc1assification counts observations that are incorrectly classified. Two types of error are defined accordingly. Type I and II errors refer to incorrect predictions of bankrupt and non bankrupt organizations respectively. Numerically, rate of errors is equal to one minus rate of classification accuracy from respective group. Ohlson (1980) demonstrated that rate of misclassification varies across different cut off points. To find optimal cut off point, he summed rates of Type I and II errors and optimal solution is found by minimizing sum. The cost of Type I error is loss of principle when debtor defaults. The cost of Type II error refers to opportunity cost that is difference between interest revenue generated from loans that should have been issued and risk free rate of return. Altman et al. (1977) examined 26 largest United States commercial banks and 67 smaller regional banks. They found that cost of Type I error is 35 times that of Type II error. In contrast, Weiss et al. (2004) suggested a way to specify cost of misclassification in relation to individual organization's characteristics by incorporating models by Hull et al. (2000) and Merton (1974).

3. Experimental Framework

In this research, the data is captured from bankrupt organizations identified using Lynn M. LoPucki's Bankruptcy Research Database. The database includes bankrupt organizations with $100 million or more in assets measured in 1980 dollars which filed Chapter 11 in United States starting from 1980. The database records date of filing and last available financial statement before filing so that back casting problem can be easily solved.
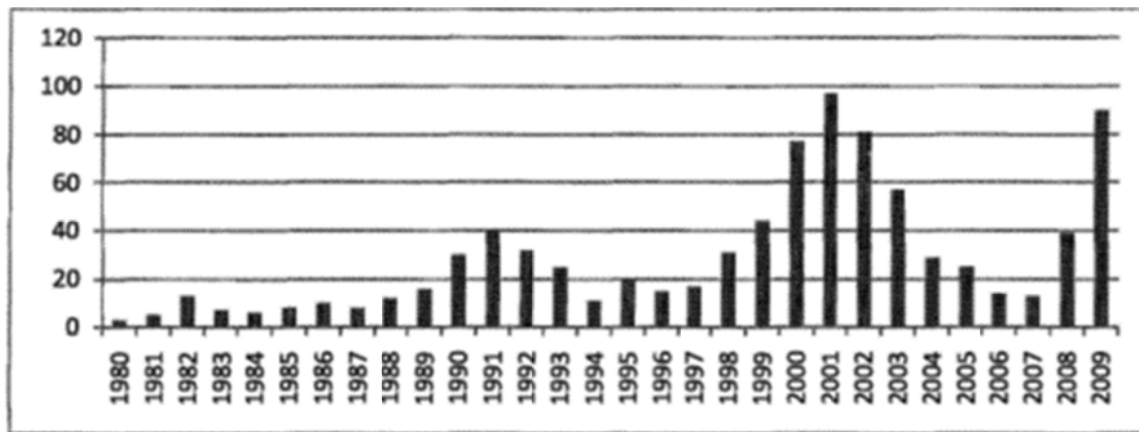

Figure 1. Number of Filings with $100M or more (Chapter 11)

[Reproduced from the data obtained from LoPucki's Database]

Figure 1 provides number of bankruptcy filings in each year. The number of bankruptcies jumps significantly at beginning of new millennia when high tech bubble burst. The bankruptcies rise from an average 20 cases in 1990s to 97 cases in 2001. Most recently following financial crisis, bankruptcy cases soared to very high level again; a total of 90 cases in 2009. It appears that there is a cycle roughly covering every decade. The sampling period is from 1980 to 2007 when latest data is available in Compustat. Organizations that had missing values in any of independent variable are eliminated from sample. As a result final sample comprises 224 bankrupt firms. Due to lack of bivariate logit distribution, the process of elimination of organizations with missing data results in sample selection bias that cannot be adjusted using Zmijewski's method. Non bankrupt organizations are collected from Compustat. An organization with active status and Compustat code for the month of fiscal year end that is different from zero is considered as non bankrupt. The organization year is sampled according to business cycle. Business cycles are defined in accordance with National Bureau of Economic Research (NBER). A business cycle consists of a peak and a trough. Expansion starts from a trough and ends at a peak while contraction starts from a peak and ends at a trough. There are five peaks and four toughs that can be identified since 1980 as shown in Table 1 below. Eight sub cycles in terms of expansion or contraction can be determined as shown in Table 2. An organization whose fiscal year ends within a sub cycle is selected for that sub cycle and it is eliminated from selection in later periods. The sampling period is from 1980 to 2007 and total sample size for non bankrupt organizations is 2561, after firms with missing values for independent variables are eliminated. Table 2 provides breakdown of the number of non bankruptcies and bankruptcies in each business cycle.

Table 1. Peaks and troughs since 1980 to 2007

Peak	Trough
January 1980	July 1980
July 1981	November 1982
July 1990	March 1991
March 2001	November 2001
December 2007	

Table 2. Breakdown of business cycle

No. of Business Cycle	No. of Sub Cycle	Sub Cycle	Starting Period	Ending Period	No. of Bankruptcy	No. of Non Bankruptcy
1	1	Contraction	Feb 1980	July 1980	1	48
	2	Expansion	Aug 1980	July 1981	2	32
2	3	Contraction	Aug 1981	Nov 1982	6	174
	4	Expansion	Dec 1982	July 1990	31	161
3	5	Contraction	Aug 1990	March 1991	7	21
	6	Expansion	April 1991	March 2001	91	1107
4	7	Contraction	April 2001	Nov 2001	27	42
	8	Expansion	Dec 2001	Dec 2007	59	976
Total					**224**	**2561**

The predictive ability of bankruptcy models is usually considered to be sensitive to number and combination of independent variables in the literature. Bellovary et al. (2007) summarized that number of variables in literature ranges from 1 to 57 and total of 752 different factors are used. However, models using only two variables can have predictive accuracies ranging from 86% to 100% which is comparable to models using higher number of independent variables. Beaver et al. (2005) also argued that the effect of selection of independent variables may have only marginal influence because statement variables are correlated. They further found that informative power of statement variables is actually decreasing due to change in application of financial tools and standards. This loss of informative power in accounting information can be compensated with usage of market variables. In this work six variables are used which are identified by Beaver et al. (2005) as independent variables in all three models. Statement variables are collected from Compustat and market variables are sampled from CRSP. The six variables are defined by Beaver et al. (2005) as follows:

(a) ROA: Net income divided by total assets. This variable captures ability that an organization generates income from its assets. In bankruptcy prediction, it is considered to be critical element used to measure ability of an organization to repay its interest or debts. (Compustat code: net income = DATA172, total asset = DATA6)

(b) ETL: EBITDA divided by total liabilities. EBITDA is earning before interest, tax, depreciation and amortization. It is calculated as sales minus cost of goods sold minus selling, general and administration expense. ETL is indicator to measure liquidity of an organization to generate cash in order to meet interest and principal requirement, especially in short term. (Compustat code: sales = DATA12, COGS = DATA41, selling, general and administration expense = DATA189)

(c) LTA: Total liabilities divided by total assets which is measure of organization's capital structure. (Compustat code: total liabilities = DATAl81)

(d) LERET: The cumulative residual return defined as difference between cumulative monthly return for the organization less cumulative monthly return on market index of NYSE, AMEX and NASDAQ firms. Share price return is recognized as leading indicator in economics. A large decline in an organization's return may signal financial difficulties.

(e) LSIGMA: The standard deviation of residual return from regression of twelve monthly returns of the organization on monthly returns of market index. This variable reflects market perception of organization's performance.

(f) LRASIZE: The logarithm of ratio of market capitalization of organization divided by market capitalization of market index. Market capitalization is calculated as number of share outstanding times stock price.

LERET and LSIGMA are computed for 12 month period ending with third month after fiscal year end. LRSIZE is computed as of end of third month after fiscal year end.

Table 3. Descriptive statistics of independent variables

Bankruptcy				
Variable	Mean	SD	Min	Max
ROA	-0.0198	0.2257	-4.7533	0.3619
ETL	0.1967	0.3450	-4.5566	2.5477
LTA	0.6598	0.2447	0.0159	2.2261
LERET	-0.0974	0.6328	-4.0301	4.1521
LSIGMA	0.1285	0.0831	0.0225	0.7704
LRSIZE	-9.8788	1.6228	-15.5839	-4.2751
Non Bankruptcy				
Variable	Mean	SD	Min	Max
ROA	0.0032	0.3961	-9.6133	20.3319
ETL	0.9032	17.2584	-64.0345	1018.8
LTA	0.5019	0.2622	0.0001	4.8621
LERET	0.0917	0.5642	-3.2852	10.856
LSIGMA	0.1199	0.0934	0.0005	2.976
LRSIZE	-10.0175	2.0678	-17.0762	-3.2056

4. Data Analysis

In order to establish the validity of different bankruptcy predictive models the following parameters are used in this work.

(a) x_i is the status of organization i in a given year for $i = 1, \ldots \ldots, n$.

(b) $x_i = \begin{cases} 0 & \text{if organization is not bankrupt} \\ 1 & \text{if organization is bankrupt} \end{cases}$

(c) $\mathbf{X}_i$ is 7×1 vector of independent variables for organization i. First element is 1.

(d) β_j is coefficient of jth element of variable j of $\mathbf{X}_i$ where β_0 is the intercept.

(e) β is 7×1 vector of coefficients.

In the remaining subsections, analysis of bankruptcy prediction is performed using Bayesian, Hazard, Mixed Logit and Rough Bayesian models.

4.1 Bayesian Model

Bayesian model developed from Bayes' Theorem (Duda et al., 1973; Good, 1965) is represented with two distinct interpretations. It is a way of understanding how the probability that a theory is true is affected by a new piece of evidence. In Bayesian interpretation, it expresses how a subjective degree of belief should rationally change to account for evidence. It has applications in wide variety of contexts ranging from marine biology to the development of Bayesian spam blockers for email systems. In the philosophy of science, it has been used to try to clarify relationship between theory and evidence. Many insights in the philosophy of science involving confirmation, falsification, relation between science and pseudoscience and other topics can be made more precise and sometimes extended or corrected by using Bayes' Theorem. The application of Bayes' Theorem to update beliefs is called Bayesian Inference. The Bayesian equation for this work is expressed as follows:

$$P\left(x_i = 1 \middle| y_{i,1}, \ldots \ldots, y_{i,6}\right) = \frac{P(y_{i,1}, \ldots \ldots, y_{i,6} | x_i = 1) P(x_i = 1)}{P(y_{i,1}, \ldots \ldots, y_{i,6})} \tag{1}$$

Sarkar et al. (2001) found that the naive model which assumes independence across predictive variables is comparable in performance with composite model which assumes dependence across some or all predictive variables as shown in Equation (1). The naive model used in this work is as follows:

$$P\left(x_i = 1 \middle| y_{i,1}, \ldots \ldots, y_{i,6}\right) = \frac{\prod_{j=1}^{6} P(y_{i,j} | x_i = 1) P(x_i = 1)}{\prod_{j=1}^{6} P(y_{i,j})}; \forall i = 1, \ldots \ldots, n \tag{2}$$

The variable y_j is continuous and it is to be discretized before inputs are given to the model. Sun et al. (2007) found that the optimal level of discretization is either 2 or 3. Here EPT method is used (Sun et al., 2007) to discretize each input variable into 3 levels which approximates continuous distribution by dividing outcome space into levels with probability of occurrence 18.5%, 63% and 18.5% respectively. The two points for discretization are determined in each training sample and are also used to discretize the holdout sample. The bankruptcy rate of training sample is used as estimator of probability of bankruptcy for individual organization i in holdout sample.

4.2 Hazard Model

Hazard models are a class of survival models in statistics. Survival models relate time that passes before some event occurs to one or more covariates that are associated with that quantity. In hazard model, the unique effect of a unit increase in a covariate is multiplicative with respect to hazard rate. For example, taking a drug may halve one's hazard rate for a stroke occurring or changing material from which a manufactured component is constructed may double its hazard rate for failure. Shumway (2001) proved that multi-period Logit model is equivalent to a discrete time hazard model. The multi-period Logit model is estimated from data of each organization in each year as if each organization year is an independent observation. The hazard function for one observation is expressed as follows:

$$h(t_i) = P_{t_i}(x_i = 1); \forall i = 1, \ldots \ldots, n \tag{3}$$

Here $h(t_i)$ is the probability that organization i will be bankruptcy in period from t_i to $t_i + 1$.

The survival function of the organization is given by:

$$S(t_i) = \prod_{t_i=1}^{T_i}[1 - P_{t_i}(x_i = 1)] \tag{4}$$

It is the probability that an organization survives to time T_i. The Logit function is assumed to be the functional form of hazard function so Equation (4) can be rewritten as follows:

$$h(t_i) = P_{t_i}(x_i = 1) = \frac{e^{(X'_{i,t_i}\alpha)}}{1+e^{(X'_{i,t_i}\alpha)}} \tag{5}$$

Again, the likelihood function is given by:

$$l = \prod_{i=1}^{n}\left(P_{T_i}(x_i = 1)^{x_i}[1 - P_{T_i}(x_i = 1)]^{(1-x_i)} \prod_{t_i=1}^{T_i-1}[1 - P_{T_i}(x_i = 1)]\right) \tag{6}$$

The Equation (6) measures discrete time multi-period model. T_i is the last organization year of organization i in our sample. The only difference in comparing with Shumway's model is that the organization's age is omitted from independent variables in an attempt to maintain consistency of independent variables across three models.

4.3 Mixed Logit Model

Mixed Logit model (Johnson et al., 2007) is a general statistical model for examining discrete choices. The motivation for mixed logit model arises from limitations of standard logit model. The standard logit model has three primary limitations taken care of by mixed logit model. Mixed Logit obviates three limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns and correlation in unobserved factors over time. Mixed logit can also utilize any distribution for random coefficients. Mixed Logit model can approximate to any degree of accuracy any true random utility model of discrete choice, given an appropriate specification of variables and distribution of coefficients. In Mixed Logit model utility associated for each observation i can be expressed as follows (Revelt et al., 1998):

$$Ut_i = X'_i\alpha + \epsilon_i + \gamma_i \tag{7}$$

In Equation (7), ϵ_i is the error component that is correlated among alternatives and heteroskedastic for each individual organization and γ_i is the random term with mean zero which is independently and identically distributed over alternatives and individual organization. If Logit function is assumed to be the functional form of probability of bankruptcy and given ϵ_i, the conditional Logit model is identical to traditional Logit model.

$$\text{Now, } P(x_i = 1|\epsilon_i) = \frac{e^{(X'_i\alpha+\epsilon_i)}}{1+e^{(X'_i\alpha+\epsilon_i)}} \tag{8}$$

Also ϵ_i can be represented as:

$$\epsilon_i = \mathbf{r}_i' \beta + \mathbf{s}_i' \gamma \tag{9}$$

Here, $\mathbf{r}_i'$ is vector of subset or full set of observed independent variables and β is random vector with mean zero whose density function is $\sum$ which can be chosen as any distribution with mean zero. The term $\mathbf{r}_i'\beta$ induces heteroskedasticity or correlation or both across unobserved utility components of alternative status of an observation by specifying $\sum$ in a desired manner. This term is known as error components approach which treats unobserved information as a separate error component in random component. Also $\mathbf{s}_i'$ is subset or full set of $\mathbf{X}_i'$ and γ is vector of random parameters or coefficients. Random coefficients allow heterogeneity across individual firms with respect to their sensitivity to observed exogenous variables. Denoting m^{th} element of γ as γ_m then γ can be further represented as:

$$\gamma_m = a_m + b_m V \tag{10}$$

Here, V is random variable following any distribution with mean zero and variance 1. The term $\mathbf{s}_i'\gamma$ is known as random parameter approach. In this work, we adopt random parameter approach in accordance with Jones et al. (2004). The parameters a_m and b_m must be estimated in addition to β. V is defined to be $N(0,1)$. Thus, γ_m is also normally distributed with mean a_m and variance b_m^2. Then, unconditional probability of bankruptcy can be expressed as expected value of indicator function $I(x_i) = 1$ if $x_i = 1$ otherwise zero.

$$P(x_i = 1) = E[I(x_i = 1)] \tag{11}$$

According to law of iterated expectation Equation (11) can be transformed as follows:

$$P(x_i = 1) = E\big[E[I(x_i = 1)|\epsilon_i]\big] = \int P(x_i = 1|\epsilon_i)\, g(\epsilon_i) d\epsilon_i \tag{12}$$

In this work, LSIGMA and ROA are selected as subset s whose coefficients are random. The likelihood function is given by:

$$l = \prod_{i=1}^{n}[\int P(x_i = 1|\epsilon_i)g(\epsilon_i)d\epsilon_i]^{\gamma_i}[1 - \int P(x_i = 1|\epsilon_i)g(\epsilon_i)d\epsilon_i]^{(1-\gamma_i)} \tag{13}$$

The estimates of Hazard and Mixed Logit models may be inconsistent and biased because of choice based sample (Zmijewski, 1984). He summarized three ways to estimate models with choice based sample viz. (a) weighted exogenous sample maximum likelihood (WESML) (b) conditional maximum likelihood (CML) and (c) full information concentrated maximum likelihood (FICML). WESML is computationally least complex so it is used to estimate Hazard and Mixed Logit models in this work. The log likelihood function of Hazard model using WESML is adjusted and is as follows:

$$L_H = \frac{P_P(x=1)}{P_S(x=1)} \sum_{i=1}^{n} x_i \left(log\big[P_{t_i}(x_i = 1)\big] + \sum_{t_i=1}^{T_i-1} log\big[1 - P_{t_i}(x_i = 1)\big] \right) +$$

$$\frac{P_P(x=0)}{P_S(x=0)} \sum_{i=1}^{n}(1 - x_i)\left(\sum_{t_i=1}^{T_i-1} log\big[1 - P_{t_i}(x_i = 1)\big] \right) \tag{14}$$

The log likelihood function of Mixed Logit model using WESML is adjusted and is given by:

$$L_H = \frac{P_P(x=1)}{P_S(x=1)} \sum_{i=1}^{n} x_i \left(log[\int P(x_i = 1|\epsilon_i)g(\epsilon_i)d\epsilon_i] \right) +$$

$$\frac{P_P(x=0)}{P_S(x=0)} \sum_{i=1}^{n}(1 - x_i)\left(log[1 - \int P(x_i = 1|\epsilon_i)g(\epsilon_i)d\epsilon_i] \right) \tag{15}$$

In Equations (14) and (15), we have $P_P(x = 1)$ as population bankruptcy rate (0.847% as reported by Zmijewski 1984); $P_P(x = 0) = 1 - P_P(x = 1)$; $P_S(x = 1)$ is sample bankruptcy rate and $P_S(x = 0) = 1 - P_S(x = 1)$.

4.4 Rough Bayesian Model

Rough Bayesian model is developed using probabilistic Rough Sets and Bayesian models. We first review basic formulations of probabilistic Rough Sets and then present Rough Bayesian model in following subsections.

4.4.1 Decision Theoretic Rough Sets

Let us consider an equivalence relation $E \subseteq U \times U$ on U i.e. E is reflexive, symmetric and transitive. Two objects in U satisfy E iff they have same values on all attributes. The pair $aprs = (U, E)$ is called an

approximation space. The equivalence relation E induces a partition of U denoted by U/E. The building blocks of Rough Sets are equivalence classes of E. For an object $x \in U$ equivalence class containing x is given by $[x] = \{y \in U | xEy\}$. For subset $C \subseteq U$ we divide universe U into five disjoint regions viz. positive region $POS(C)$, positive boundary region $BND^+(C)$, boundary region $BND(C)$, negative boundary region $BND^-(C)$ and negative region $NEG(C)$ (Pawlak, 1982, 1991; Pawlak et al., 1994):

$$POS(C) = \{x \in U | [x] \subseteq C\}$$
$$BND^+(C) = \{x \in U | [x] \cap C \neq \emptyset \wedge [x] \subseteq C\}$$
$$BND(C) = \{x \in U | [x] \cap C \neq \emptyset \wedge [x] \nsubseteq C\} \qquad (16)$$
$$BND^-(C) = \{x \in U | [x] \cap C = \emptyset \wedge [x] \nsubseteq C\}$$
$$NEG(C) = \{x \in U | [x] \cap C = \emptyset\}$$

It can be argued with *certainty* that any object $x \in POS(C)$ belongs to C and that any object $x \in NEG(C)$ does not belong to C. It cannot be decided with certainty whether or not an object $BND^+(C), BND(C)$ and $BND^-(C)$ belongs to C. The qualitative categorization in Rough Sets proposed by Pawlak may be too restrictive to be practically useful. Probabilistic Rough Sets enables some tolerance of uncertainty in which Pawlak's Rough Sets are generalized by considering degrees of overlap between equivalence classes and the set to be approximated i.e. $[x]$ and C in equation system (16). The conditional probability of an object belonging to C given that object is in $[x]$ estimated using cardinalities is given by:

$$Prob(C|[x]) = \frac{|C \cap [x]|}{|[x]|} \qquad (17)$$

In Equation (17), $|\cdot|$ denotes cardinality of set. Pawlak and Skowron (1994) suggested that conditional probability is actually rough membership function. According to above definitions the five regions can be equivalently defined as:

$$POS(C) = \{x \in U | Prob(C|[x]) = 1\}$$
$$BND^+(C) = \{x \in U | 0 \leq Prob(C|[x]) < 1\}$$
$$BND(C) = \{x \in U | 0 < Prob(C|[x]) < 1\} \qquad (18)$$
$$BND^-(C) = \{x \in U | 0 < Prob(C|[x]) \leq 1\}$$
$$NEG(C) = \{x \in U | Prob(C|[x]) = 0\}$$

The equation system (18) is defined by using two extreme values 0 and 1. These are of qualitative nature and magnitude of $Prob(C|[x])$ is not taken into account. The main result of decision theoretic Rough Sets is parameterized probabilistic approximations. This can be represented by replacing values 1 and 0 in equation system (18) by pair of threshold values α and β with $\alpha > \beta$. The (α, β) probabilistic positive, positive boundary, boundary, negative boundary and negative regions are defined by the following expressions:

$$POS_{(\alpha,\beta)}(C) = \{x \in U | Prob(C|[x]) \geq \alpha\}$$
$$BND^+_{(\alpha,\beta)}(C) = \{x \in U | \beta \leq Prob(C|[x]) < \alpha\}$$
$$BND_{(\alpha,\beta)}(C) = \{x \in U | \beta < Prob(C|[x]) < \alpha\} \qquad (19)$$
$$BND^-_{(\alpha,\beta)}(C) = \{x \in U | \beta < Prob(C|[x]) \leq \alpha\}$$
$$NEG_{(\alpha,\beta)}(C) = \{x \in U | Prob(C|[x]) \leq \beta\}$$

The five probabilistic regions lead to five way decisions (Yao, 2007, 2008, 2010; Yao et al., 1992; Yao et al., 1990). We accept an object x to be member of C if probability is greater than α. We reject x to be member of C if probability is less than β. We neither accept nor reject x to be member of C if probability lies between α and β instead we make decision of deferment. The threshold values α and β can be interpreted in terms of cost or risk of five way classification. The values can be systematically computed based on minimizing overall risk of classification.

4.4.2 Classification Based on Bayes Model

The conditional probabilities are not always directly derivable from data. In such cases alternative ways are needed to be considered to calculate their values. A commonly used method is to apply Bayes Theorem such that:

$$Prob(C|[x]) = \frac{Prob(C)Prob([x]|C)}{Prob([x])} \tag{20}$$

$$\text{Again, } Prob([x]) = Prob([x]|C)Prob(C) + Prob([x]|C^c)Prob(C^c) \tag{21}$$

$Prob(C|[x])$ is the posteriori probability of class C given $[x]$; $Prob(C)$ is the apriori probability of class C and $Prob([x]|C)$ is the likelihood of $[x]$ with respect to C. A monotonically increasing function of conditional probability may be defined to construct an equivalent classifier. This observation can lead to significant analytical and computational simplifications. The probability $Prob([x])$ in Equation (21) can be eliminated by taking odds form of Bayes Theorem which is given as follows:

$$Odd\big(Prob(C|[x])\big) = \frac{Prob(C|[x])}{Prob(C^c[x])} = \frac{Prob([x]|C)}{Prob([x]|C^c)} \cdot \frac{Prob(C)}{Prob(C^c)} = \frac{Prob([x]|C)}{Prob([x]|C^c)} Odd(Prob(C)) \tag{22}$$

A threshold value on probability can be interpreted as another threshold value on odds. For positive region we have:

$$Prob(C|[x]) \geq \alpha \Leftrightarrow \frac{Prob(C|[x])}{Prob(C^c[x])} \geq \frac{\alpha}{1-\alpha} \Leftrightarrow \frac{Prob([x]|C)}{Prob([x]|C^c)} \cdot \frac{Prob(C)}{Prob(C^c)} \geq \frac{\alpha}{1-\alpha} \tag{23}$$

By applying logarithms to both sides of the above equation we get:

$$log\frac{Prob([x]|C)}{Prob([x]|C^c)} + log\frac{Prob(C)}{Prob(C^c)} \geq log\frac{\alpha}{1-\alpha} \tag{24}$$

Similar expressions can be obtained for negative and boundary regions. Thus, the five regions can now be represented as follows:

$$POS^B_{(\alpha',\beta')}(C) = \{x \in U|\ log\frac{Prob([x]|C)}{Prob([x]|C^c)} \geq \alpha'\}$$

$$BND^{B^+}_{(\alpha,\beta)}(C) = \{x \in U|\ \beta' \leq log\frac{Prob([x]|C)}{Prob([x]|C^c)} < \alpha'\}$$

$$BND^B_{(\alpha',\beta')}(C) = \{x \in U|\ \beta' < log\frac{Prob([x]|C)}{Prob([x]|C^c)} < \alpha'\} \tag{25}$$

$$BND^{B^-}_{(\alpha,\beta)}(C) = \{x \in U|\ \beta' < log\frac{Prob([x]|C)}{Prob([x]|C^c)} \leq \alpha'\}$$

$$NEG^B_{(\alpha',\beta')}(C) = \{x \in U|\ log\frac{Prob([x]|C)}{Prob([x]|C^c)} \leq \beta'\}$$

In equation system (25),

$$\alpha' = log\frac{Prob(C^c)}{Prob(C)} + log\frac{\alpha}{1-\alpha} \tag{26}$$

$$\beta' = log\frac{Prob(C^c)}{Prob(C)} + log\frac{\beta}{1-\beta} \tag{27}$$

This interpretation simplifies calculation by eliminating $Prob([x])$. The detailed estimations of related probabilities need to be further addressed.

4.4.3 Bayesian Model for Estimating Probabilities

The naive Rough Bayesian model provides a practical way to estimate conditional probability based on naive Bayesian classification (Duda et al., 1973). The rough set model proposed by Pawlak (1982, 1991) has information about a set of objects that are represented through an information table with finite set of attributes. Formally, an information table can be expressed as follows:

$$S = (U, At, \{V_a|a \in At\}, \{I_a|a \in At\}) \tag{28}$$

Here, U is finite nonempty set of objects called universe; At is finite nonempty set of attributes; V_a is nonempty

set of values for $a \in At$; $I_a : U \to V_a$ is an information function. The information function I_a maps an object in U to a value of V_a for an attribute $a \in At$ i.e. $I_a(x) \in V_a$. Each object x is described by $\bigwedge_{a \in At} a = I_a(x)$ where $v_a \in V_a$ and atomic formula $a = I_a(x)$ indicates that value of an object on attribute a is $I_a(x)$. For simplicity we express description of $[x]$ as feature vector viz. $Descp([x]) = (v_1, \ldots \ldots, v_n)$ with respect to set of attributes $\{a_1, \ldots \ldots, a_n\}$ where $I_{a_i}(x) = v_i$. Now, conditional probability $Prob(C|Descp[x])$ can be re-expressed by apriori probability $Prob(C)$, likelihood $Prob(Descp[x]|C)$ and probability $Prob(Descp[x])$ where $Prob(Descp[x]|C)$ is joint probabilities of $Prob(v_1, \ldots \ldots, v_n|C)$ and $Prob(Descp[x])$ is joint probability of $Prob(v_1, \ldots \ldots, v_n)$. In practice, it is difficult to analyze interactions between components of $[x]$ especially when n is large. A solution to this problem is to calculate likelihood based on naive conditional independence assumption (Duda et al., 1973; Pawlak, 1982, 1991). That is we assume each component v_i of $[x]$ to be conditionally independent of every other component $v_j \forall j \neq i$. For Bayesian interpretation of five regions based on equation (24), we can add the following naive conditional independence assumptions:

$$Prob(Descp[x]|C) = Prob(v_1, \ldots \ldots, v_n|C) = \prod_{i=1}^{n} Prob(v_i|C) \tag{29}$$

$$Prob(Descp[x]|C^c) = Prob(v_1, \ldots \ldots, v_n|C^c) = \prod_{i=1}^{n} Prob(v_i|C^c) \tag{30}$$

The Equations (29) and (30) can be re-expressed as follows:

$$log \frac{Prob(Descp[x]|C)}{Prob(Descp[x]|C^c)} \geq log \frac{Prob(C)}{Prob(C^c)} + log \frac{\alpha}{1-\alpha} \tag{31}$$

$$\Leftrightarrow \sum_{i=1}^{n} log \frac{Prob(v_i|C)}{Prob(v_i|C^c)} \geq log \frac{Prob(C)}{Prob(C^c)} + log \frac{\alpha}{1-\alpha}$$

Here, $Prob(C)$ and $Prob(v_i|C)$ can be estimated from the frequencies of training data by substituting the following values:

$$Prob(C) = \frac{|C|}{|U|} \tag{32}$$

$$Prob(v_i|C) = \frac{|m(a_i,v_i) \cap C|}{|C|} \tag{33}$$

Here, $m(a_i, v_i)$ is called meaning set and is defined as $m(a_i, v_i) = \{x \in U | I_{a_i}(x) = v_i\}$ i.e. set of objects whose attribute value equals v_i with respect to attribute a_i. Similarly, $Prob(C^c)$ and $Prob(v_i|C^c)$ can be estimated. We can rewrite equation system (25) as:

$$POS_{(\alpha',\beta')}^{B}(C) = \{x \in U | \sum_{i=1}^{n} log \frac{Prob(v_i|C)}{Prob(v_i|C^c)} \geq \alpha'\}$$

$$BND_{(\alpha,\beta)}^{B^+}(C) = \{x \in U | \beta' \leq \sum_{i=1}^{n} log \frac{Prob(v_i|C)}{Prob(v_i|C^c)} < \alpha'\}$$

$$BND_{(\alpha',\beta')}^{B}(C) = \{x \in U | \beta' < \sum_{i=1}^{n} log \frac{Prob(v_i|C)}{Prob(v_i|C^c)} < \alpha'\} \tag{34}$$

$$BND_{(\alpha,\beta)}^{B^-}(C) = \{x \in U | \beta' < \sum_{i=1}^{n} log \frac{Prob(v_i|C)}{Prob(v_i|C^c)} \leq \alpha'\}$$

$$NEG_{(\alpha',\beta')}^{B}(C) = \{x \in U | \sum_{i=1}^{n} log \frac{Prob(v_i|C)}{Prob(v_i|C^c)} \leq \beta'\}$$

5. Experimental Results

This section illustrates the results obtained towards bankruptcy prediction using Bayesian, Hazard, Mixed Logit and Rough Bayesian models. A comparative analysis among different stated methods is also highlighted. All the bankruptcy models are implemented in MATLAB on Pentium IV PC with 1 GB RAM. In the remaining subsections, we discuss the experimental results in terms of cut off point, Type I and Type II errors, optimal cut

off point, randomly selected samples and samples in different business cycles and sub cycles.

5.1 Cut off Point, Type I and Type II Errors

If an organization's fitted probability is greater than cut off point, then it is classified as bankrupt and vice versa. If an organization is bankrupt and is classified as non bankrupt, the error is Type I error and reverse is Type II error. To show relationship between cut off point, Type I and Type II errors, we first randomly and equally divide total sample into two samples viz. training and holdout sample. The training sample includes 112 bankruptcies and 1280 non bankruptcies. The hold out sample comprises 112 bankruptcies, 1281 non bankruptcies. The training sample is used to evaluate prior probabilities or estimators for the four models and hold out sample is used for classification purpose to calculate two types of error.
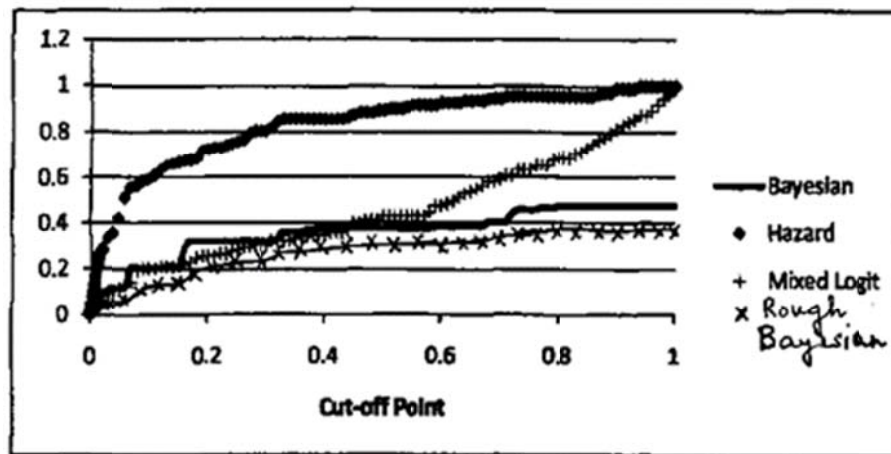

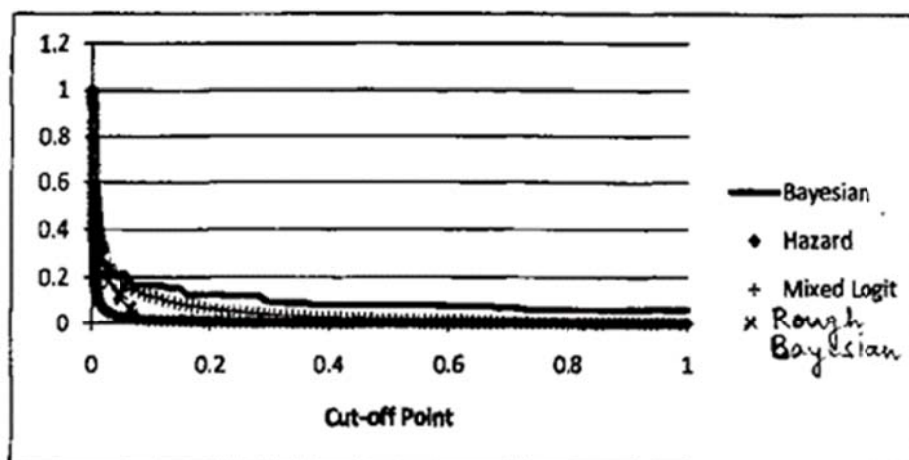Figure 2. Rate of Type I error

Figure 3. Rate of Type II error

The rate of errors is calculated as number of organizations that are misclassified over number of organizations in the group. Figures 2 and 3 exhibit relationships between cut off point and two types of errors as well as tradeoff between errors for the four models. As cut off point move towards 1, Type I error increases. In contrast, Type II error decreases as cut off point increases. As probability generated using Bayesian model is zero or greater than one, the rate of Type I error can be far less than one. The rate of Type II error can also be tilted above zero even if cut off point is chosen as one.

If classification error or classification accuracy is the validation method applied to compare bankruptcy models, Figures 2 and 3 indicate that different conclusions are obtained by choosing different cut off points. There are

some important methods to choose cut off point viz. cut off point of 50% suggested by Johnson et al. (2007) where prior probability and costs are difficult to incorporate and another one suggested by Grice et al. (2001), Zmijewski (1984) and Ohlson (1980); industry failure cut off rate of 0.41% suggested by Martin (1977) assuming that apriori probability for group membership is equal to sample probability and cut off point of 3.8% suggested by Begley et al. (1996) obtained from Ohlson (1980) that minimizes sum of Type I and Type II errors.

The results based on different cut off points are illustrated in Table 4. Within each model, Type I or Type II is the rate of Type I or Type II errors and Type I + Type II is the sum of Type I and Type II errors. When cut off point is 50%, Type I error is the mainly for validation then Bayesian model outperforms others. However, when sum of Type I and Type II errors is used, Hazard, Mixed Logit and Rough Bayesian models dominates for cut off points of 0.41% and 3.8%. It is to be noted that these results are arbitrary and does not reflect true quality of models. The above cut off point choices has two limitations. First, an inappropriate assumption of equal costs of misclassification exists if sum of two errors is used as validation method. Altman et al. (1977) estimated that cost of Type I error is 35 times greater than that of Type II error. This is because when Type I error occur, creditors lose their total principle, in contrast to the opportunity cost resulting from occurrence of Type II error. Secondly, the differences of numbers of organizations in each group are generally ignored. The number of bankrupt organizations is far less than non bankrupt organizations. According to Zmijewski (1984) frequency of financial failure has not exceeded 0.75% since 1934.

Table 4. Predictive results with different cut off points

cut off point	Bayesian			Hazard		
	Type I	Type II	Type I + II	Type I	Type II	Type I + II
50%	0.3750	0.0796	0.4546	0.8928	0.0007	0.8936
0.41%	0.0267	0.5359	0.5627	0.0803	0.2546	0.3350
3.80%	0.1160	0.2164	0.3324	0.3571	0.0359	0.3930

cut off point	Mixed Logit			Rough Bayesian		
	Type I	Type II	Type I + II	Type I	Type II	Type I + II
50%	0.4285	0.0148	0.4434	0.4086	0.0125	0.4211
0.41%	0.0089	0.5796	0.5886	0.0069	0.4596	0.4665
3.80%	0.1071	0.2187	0.3258	0.1272	0.2166	0.3438

5.2 Optimal Cut off Point

An appropriate cut off point which minimizes the cost of misclassification is given by following equation:

$$\min_\alpha C = \sum_{i=1}^n \left[C_{I,i} I(\hat{P}_i \le \alpha)\right]^{x_i} \cdot \left[C_{II,i} I(\hat{P}_i > \alpha)\right]^{(1-x_i)} \tag{35}$$

Here, C is total cost of misclassification, $C_{I,i}$ and $C_{II,i}$ are the cost of Type I and Type II error for organization i respectively, α is the cut off point, $\hat{P}_i$ is probability of bankruptcy for organization i, x_i is the observed status of organization I, $I(\cdot)$ is indicator function that returns 1 if logical function is satisfied, otherwise returns 0. Weiss et al. (2004) provided a methodology to estimate level of individual costs. However this estimation will dramatically increase computational time. So we use a broad specification of costs as in Altman et al. (1977). Equation (36) provides minimization of simplified total cost.

$$\min_\alpha C = C_I P_I(\alpha) Q_I + C_{II} P_{II}(\alpha) Q_{II} \tag{36}$$

Here, P_I and P_{II} are rates of Type I and Type II error respectively, a function of cut off point α, Q_I and Q_{II} are rates of bankrupt and non bankrupt firms in sample. Equation (36) cannot be solved in a closed form. Thus, optimal cut off point is not known in advance for holdout sample. A proxy of optimal cut off point can be found using various methods. One technique will be using cut off points discussed in literature.

Table 5. Random selection and test results

$\frac{C_I}{C_{II}}$; C; α	Bayesian Model				Hazard Model			
	0.41%	3.8%	50%	Optimal	0.41%	3.8%	50%	Optimal
1	0.5274	0.2332	0.0913	0.0776*	0.4157	0.0828	0.0713	0.0660*
10	0.5438	0.3228	0.4595	0.2977*	0.4450	0.3337	0.6982	0.2672*
20	0.5620	0.4224*	0.8687	0.4345*	0.4776	0.6125	1.3948	0.3591*
35	0.5893*	0.5718*	1.4824	0.6289	0.5264	1.0306	2.4396	0.4340*
50	0.6166	0.7212	2.0962	0.5771*	0.5752	1.4488	3.4845	0.5049*
75	0.6621*	0.9702	3.1191	0.6535*	0.6565*	2.1457	5.2259	0.6332*
100	0.7076*	1.2192	4.1420	0.7169*	0.7379*	2.8427	6.9674	0.7381*

$\frac{C_I}{C_{II}}$; C; α	Mixed Logit Model				Rough Bayesian Model			
	0.41%	3.8%	50%	Optimal	0.41%	3.8%	50%	optimal
1	0.2621	0.0676*	0.0744	0.0668*	0.2565	0.0575*	0.0696	0.0666*
10	0.3090	0.3982	0.7263	0.2578*	0.2799	0.4080	0.7565	0.2476*
20	0.3612*	0.7655	1.4506	0.3633*	0.3057*	0.7856	1.5555	0.3586*
35	0.4395*	1.3165	2.5371	0.4504*	0.4066*	1.3367	2.6936	0.4279*
50	0.5177*	1.8675	3.6236	0.5320*	0.5075*	1.9624	3.8796	0.5055*
75	0.6481*	2.7858	5.4345	0.6896*	0.6286*	2.9690	5.7569	0.6777*
100	0.7785*	3.7040	7.2453	0.7677*	0.7984*	3.9040	7.9090	0.7075*

[Optimal is estimated optimal cut off point evaluated with training sample; *indicates that cost of misclassification is statistically least in one $\frac{C_I}{C_{II}}$ under 5% significant level. For certain $\frac{C_I}{C_{II}}$ multiplier, e.g. 20 in Bayesian Model, there are two*indicating that these two costs under cut off point 3.8% and optimal are not significantly different while they both are significantly less than other two costs under cut off point 0.41% and 50%.]

Another technique is to estimate cut off point using training sample. To determine an empirical cut off point estimated using training sample is optimal proxy in holdout sample, the random selection sample is run 30 times. The training sample is used to evaluate apriori probabilities and to compute an empirical cut off point. The holdout sample is used to generate predictive results. Three other cut off points are selected for comparative purpose viz. 0.41%, 3.8% and 50%. Pair sample *t* tests are conducted under each pair of cut off points within each model. The results are given in Table 5 using various cut off points. The first column contains ratios of cost of Type I to cost of Type II errors. Instead of using only one specification of Type I and Type II costs, we consider several ratio specifications in order to have robust check. The results suggest that empirical optimal cut off point calculated with training sample is the best proxy for the true optimal cut off point for holdout sample as it dominates most specifications of Type I costs over Type II costs by producing least cost of misclassification. Such results are consistent for all four models. Following estimated optimal cut off point, industry bankruptcy rate of 0.41% is considered as second best cut off point. It is not appropriate to use it when ratio of Type I cost and Type II cost is low, since most of organizations will be classified as bankrupt reducing cost of Type I error while simultaneously increasing cost of Type II errors. A cut off point of 50% is generally used when equivalent costs of Type I error and Type II error are assumed but empirical results suggest that it will produce more cost of misclassification even though two types of costs are assumed to be equal.

5.3 Randomly Selected Samples

As discussed in subsection 5.1, different choices of cut off points can lead to different conclusion about comparative power of prediction of various models. As a result, tests should be conducted under conditions that are free of distortion posed by an arbitrary cut off point while maintaining true characteristics of models. The results in subsection 5.2, suggest that cut off points estimated from training sample are more preferable than other specifications. We use this empirical cut off point to compare four new models. The 30 random samples in subsection 5.2 are used again. Pair sample *t* tests are performed to generate total costs of misclassification for each pair of models. The results are presented in Table 6 which is reproduction of Table 5 using optimal cut off point.

Table 6. Costs of misclassification

$\dfrac{C_I}{C_{II}}$	Bayesian Model	Hazard Model	Mixed Logit Model	Rough Bayesian Model
1	0.0776	0.0660*	0.0668*	0.0666*
10	0.2977	0.2672*	0.2578*	0.2476*
20	0.4345	0.3591*	0.3633*	0.3586*
35	0.6289	0.4340*	0.4504*	0.4279*
50	0.5771	0.5049*	0.5320*	0.5055*
75	0.6535*	0.6332*	0.6896*	0.6777*
100	0.7169*	0.7381*	0.7677*	0.7075*

[*indicates least cost at 5% significant level (two or three*in one row indicates that they are not statistically different).]

Hazard, Mixed Logit and Rough Bayesian models are effectively no different in terms of predictive power for all specification of Type I and Type II cost. In general, Rough Bayesian model has highest predictive power. When the ratio of cost of Type I to cost of Type II errors reaches a very high level, Bayesian model is comparable to the other two. A possible shortcoming of randomly selected samples can result in situations where ex post results are used to explain and predict prior events creating a noisy sample.

5.4 Samples in Different Business Cycles and Sub Cycles

Since random selection produces results that may be tarnished with noisy sample, we conduct our analysis around business cycles. There are 4 business cycles in our sample. A comparative study is conducted on adjacent business cycles. Priors, estimates and optimal cut off point are evaluated from preceding cycle and cost of misclassification is calculated with subsequent cycle. Since first business cycle is short and sample size is fairly small, we combine first two cycles. Tables 7 and 8 illustrate results on business cycle and on sub cycles using same methodology.

Table 7. Results estimated across business cycles

	Bayesian Model		Hazard Model		Mixed Logit Model		Rough Bayesian Model	
	Panel A							
$\dfrac{C_I}{C_{II}}$	**1st and 2nd Cycle → 3rd Cycle**							
	Optimal	Cost	optimal	cost	optimal	Cost	optimal	cost
1	5.4524	0.1435	0.0294	**0.0897**	0.0751	0.1174	0.0778	0.1272
10	0.0115	0.3792	0.0004	0.3689	0.0099	0.2797	0.0100	**0.2596**
20	0.0115	0.4037	0.0004	0.4097	0.0041	0.3556	0.0086	**0.3550**
35	0.0115	0.4404	0.0004	0.4710	0.0031	0.4584	0.0073	**0.4403**
50	0.0115	**0.4771**	0.0004	0.5322	0.0031	0.5195	0.0072	0.5055
75	0.0115	**0.5383**	0.0001	0.7616	0.0031	0.6215	0.0070	0.6073
100	9.67e-05	**0.7234**	0.0001	0.8636	0.0031	0.7235	0.0069	0.7357
	Panel B							
$\dfrac{C_I}{C_{II}}$	**3rd Cycle → 4th Cycle**							
	Optimal	Cost	optimal	cost	optimal	Cost	optimal	cost
1	1.5010	0.0715	0.2742	**0.0596**	0.4229	0.0597	0.4030	0.0599
10	0.1890	0.2798	0.0129	**0.2034**	0.0806	0.3541	0.1666	0.2579
20	0.0286	0.5081	0.0089	**0.2748**	0.0806	0.4990	0.0555	0.4596
35	0.0239	0.5579	0.0077	**0.3761**	0.0618	0.5715	0.0199	0.3999
50	0.0239	0.5987	0.0060	**0.4683**	0.0290	0.5416	0.0240	0.5276
75	0.0239	0.6666	0.0060	**0.5813**	0.0290	0.6096	0.0240	0.6286
100	0.0239	0.7346	0.0060	**0.6943**	0.0290	0.6775	0.0240	0.7227

[Numbers in bold are the lowest cost within four models for each $\dfrac{C_I}{C_{II}}$]

The Tables 7 and 8 indicate optimal cut off point that is decreasing as cost of Type I errors increases. However, decrease is mitigated by rise in Type II error. From Tables 7 and 8, it is evident that as the ratio $\frac{c_I}{c_{II}}$ is high, the cut off point decreases to a very small number. With respect to total cost, the results are in contradiction to those given in randomly selected samples. Panel A of Table 7 shows that Bayesian model outperforms others except when the ratio $\frac{c_I}{c_{II}}$ is lower than 35. Mixed Logit model generates less misclassification cost overall as shown in Panel B of Table 7. In general, Rough Bayesian model underperforms others. Such results are consistent for tests on sub cycles as shown in Table 7. Mixed Logit model dominates most specification of the ratio $\frac{c_I}{c_{II}}$. However, Bayesian model has better performance when the ratio $\frac{c_I}{c_{II}}$ is high. Based on the results obtained from randomly selected samples and cross business cycle samples, Rough Bayesian model has highest predictive power. The predictive power of Bayesian model in higher range of Type I costs is in general higher than other three. This conclusion is consistent with random sample and across business cycle. Figures 4 and 5 utilize the randomly selected sample and provide estimated probability for organizations with bankrupt or non bankrupt status. They demonstrate outperformance of Bayesian model when Type I costs are high. In both figures, horizontal axis represents probability and vertical axis represents number of organizations. As illustrated in Figure 4, patterns are quite similar between Hazard, Mixed Logit and Rough Bayesian models whose probability mass lies between 0 and 0.1. However, for Bayesian model, a majority of organizations have probabilities higher than 1. Figure 5 shows that there are more than 1000 non bankrupt organizations having probability lying between 0 to 0.01 for Hazard, Mixed Logit and Rough Bayesian models. For Bayesian model only 660 organizations have probabilities in lowest range. Figures 4 and 5 represent two main differences between Bayesian model and Hazard, Mixed Logit and Rough Bayesian models. In general, probabilities generated from Hazard, Mixed Logit and Rough Bayesian models are lower than those from Bayesian model. The probabilities generated from the Hazard, Mixed Logit and Rough Bayesian models are more concentrated in lowest range. When Type I costs are high, empirical cut off points will be forced to be a very small number. If cut off points are small for Hazard, Mixed Logit and Rough Bayesian models more organizations need to be classified because most of the organizations' probabilities are concentrated within lowest range. This fact will result in either higher total costs of Type I or Type II errors.

Table 8. Results estimated across Sub Cycles

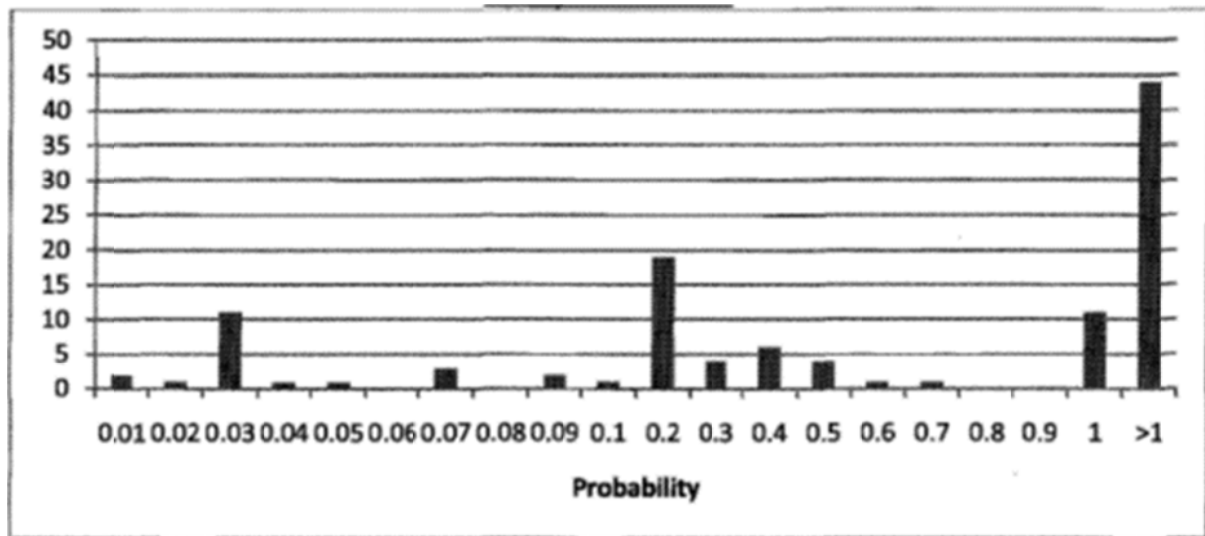|	Bayesian Model		Hazard Model		Mixed Logit Model		Rough Bayesian Model	
	Panel A							
$\frac{c_I}{c_{II}}$	6th Sub Cycle $\rightarrow$ 7th Sub Cycle							
	optimal	Cost	optimal	cost	optimal	Cost	optimal	cost
1	0.4035	0.2463	0.2637	0.2676	0.3706	**0.1739**	0.3675	0.1740
10	0.1965	1.1739	0.0154	**0.5352**	0.1137	0.8985	0.1055	0.7589
20	0.0829	1.7971	0.0085	0.6901	0.0629	0.3478	0.0735	**0.3376**
35	0.0193	0.5942	0.0073	1.1267	0.0629	0.5652	0.0735	**0.5255**
50	0.0193	**0.8115**	0.0059	1.5492	0.0285	0.8695	0.0282	0.8282
75	0.0193	**1.1739**	0.0059	2.2535	0.0285	1.2318	0.0282	1.2122
100	0.0193	**1.5362**	0.0059	2.9577	0.0285	1.5942	0.0282	1.5575
	Panel B							
$\frac{c_I}{c_{II}}$	7th Sub Cycle $\rightarrow$ 8th Cycle							
	optimal	Cost	optimal	cost	optimal	Cost	optimal	cost
1	0.7672	0.1159	0.0004	**0.1004**	0.0337	0.2405	0.0340	0.1279
10	0.0381	0.6772	0.0003	**0.2067**	0.0003	0.3922	0.0002	0.4040
20	0.0381	0.6869	0.0003	0.2743	0.0003	0.4405	0.0002	**0.2640**
35	0.0381	0.7014	0.0003	0.3758	0.0003	0.5130	0.0002	**0.3676**
50	0.0381	0.7159	0.0003	**0.4772**	0.0003	0.5855	0.0002	0.4899
75	0.0381	0.7400	0.0003	**0.6463**	0.0003	0.7062	0.0002	0.7272
100	0.0381	**0.7642**	0.0003	0.8154	0.0003	0.8270	0.0002	0.8086

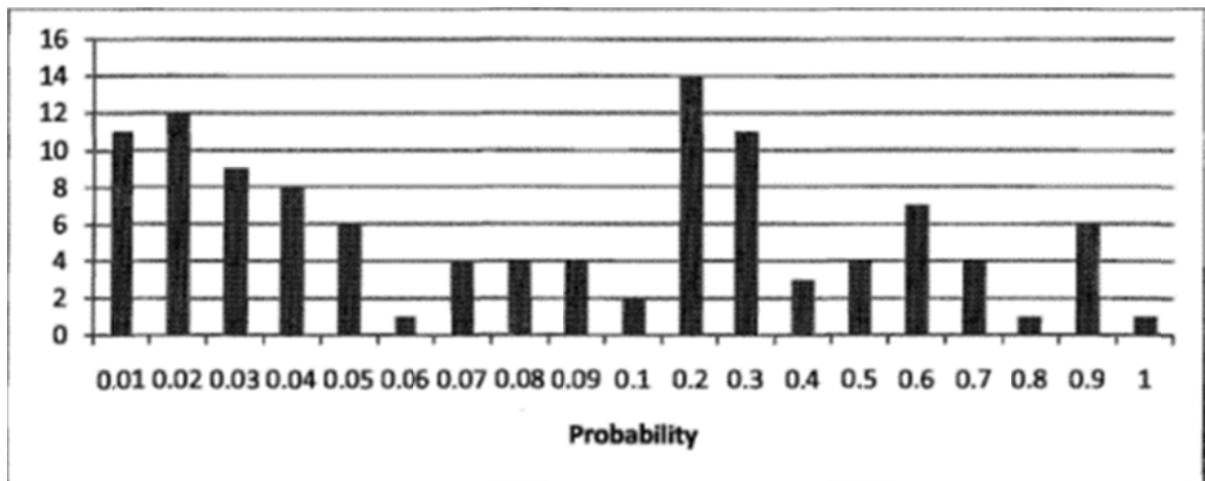Figure 4(a). Bankrupt organizations' probabilities (Bayesian model)

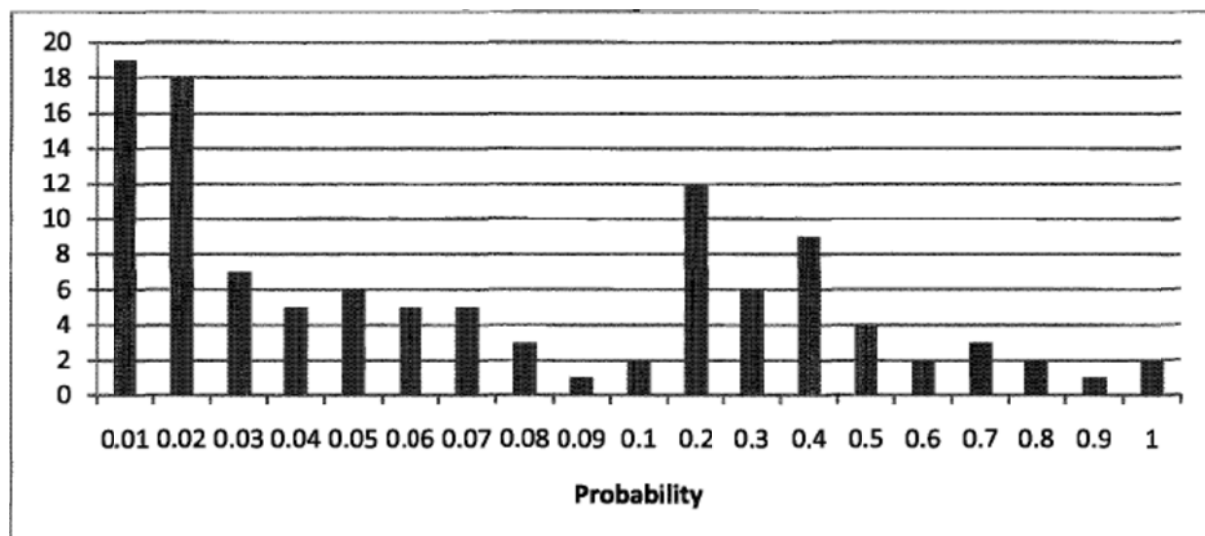Figure 4(b). Bankrupt organizations' probabilities (Hazard model)

Figure 4(c). Bankrupt organizations' probabilities (Mixed Logit model)
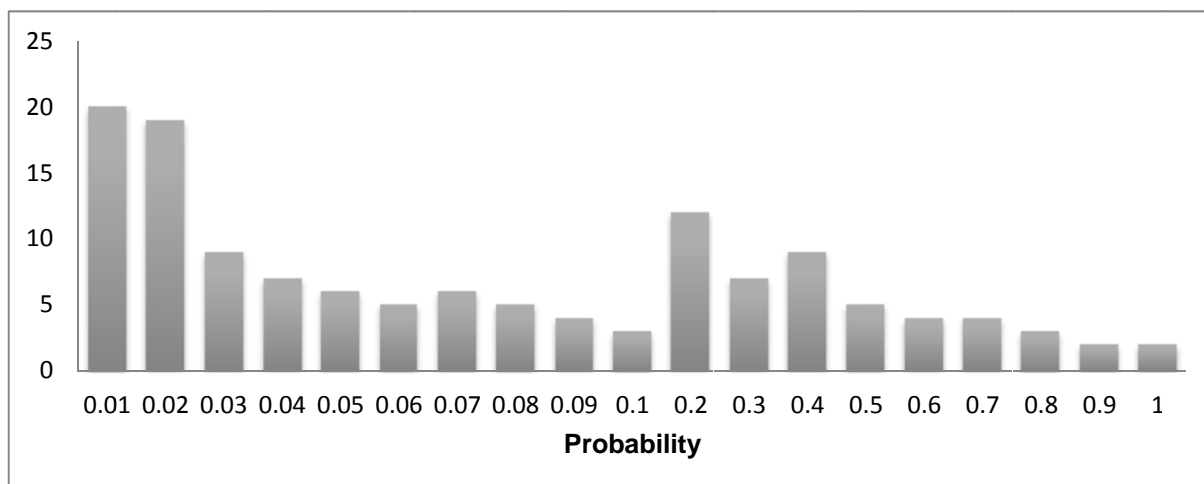
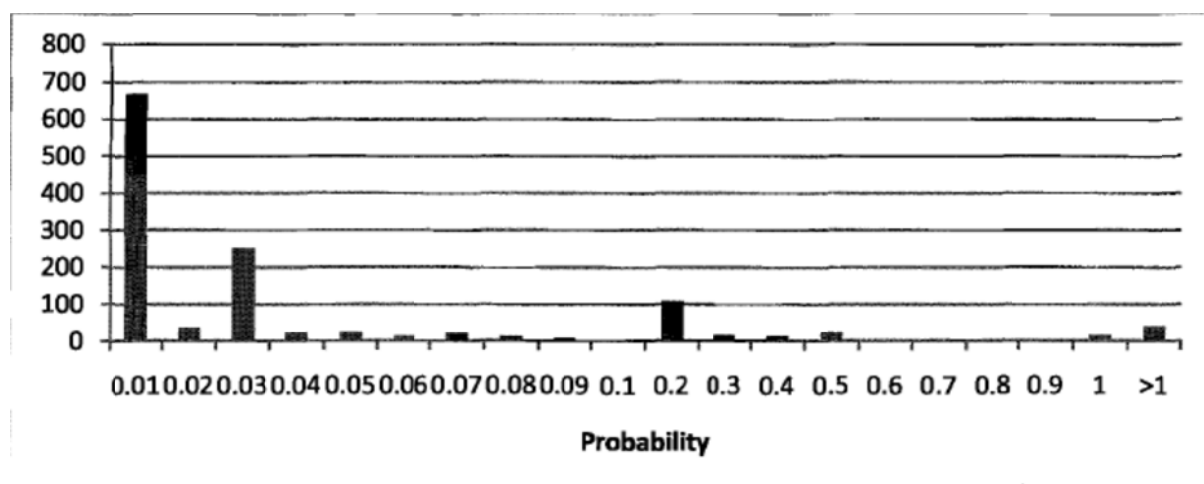Figure 4(d). Bankrupt organizations' probabilities (Rough Bayesian model)

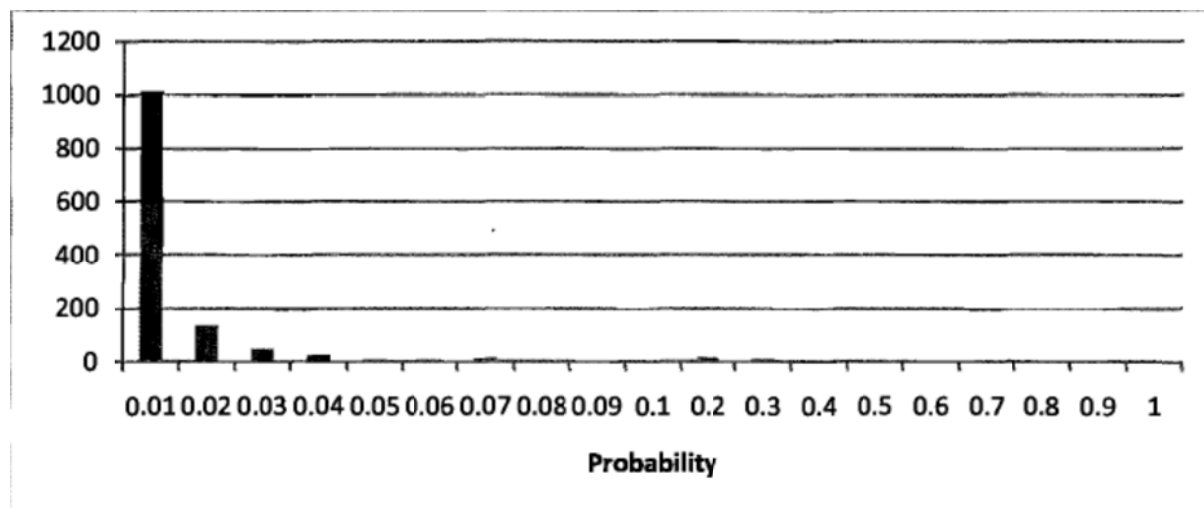Figure 5(a). Non bankrupt organizations' probabilities (Bayesian model)

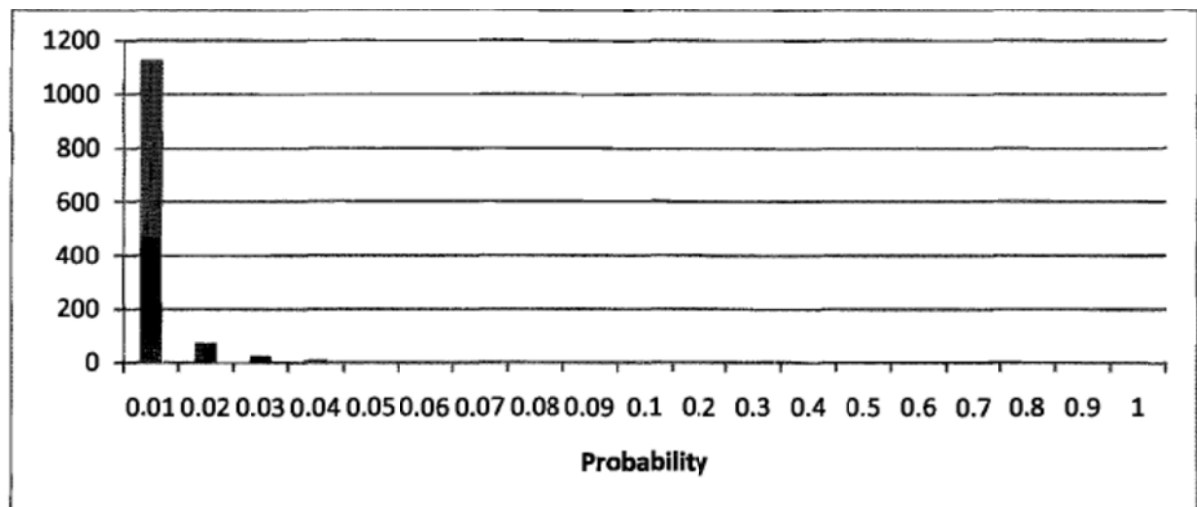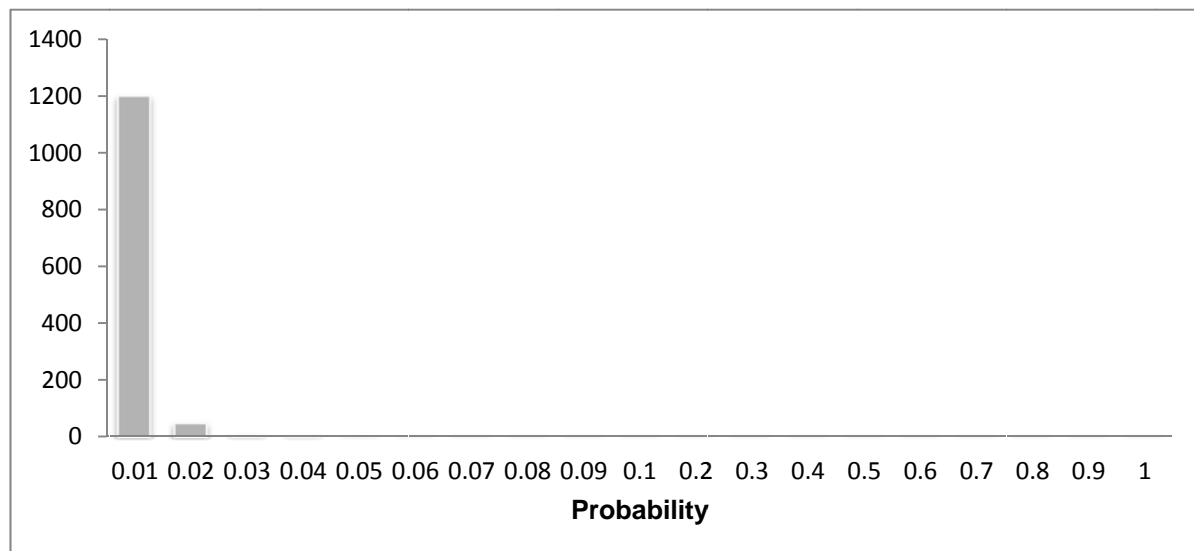Figure 5(b). Non bankrupt organizations' probabilities (Hazard model)

Figure 5(c). Non bankrupt organizations' probabilities (Mixed Logit model)

Figure 5(d). Non bankrupt organizations' probabilities (Rough Bayesian model)

6. Conclusions

In this work, a comparative analysis of corporate bankruptcy prediction is made through Bayesian, Hazard, Mixed Logit and Rough Bayesian models based on their empirical performance. To make fairness in comparisons, factors that diminish or inflate predictive accuracy are controlled or adjusted. In addition to back casting, choice based sample bias, change of financial ratio etc as identified by previous investigations, it is found that selection of cut off points and sampling procedure influences predictive results. To reduce the level of arbitrariness in selecting cut off point, we use optimal cut off point estimated by using training sample as a proxy for true optimal cut off point in holdout sample. This proxy results in least cost of misclassification when compared to cut off points used in previous works. Furthermore using this empirical optimal cut off point, the results indicate that Hazard, Mixed Logit and Rough Bayesian models are statistically superior to Bayesian model based on randomly selected samples when the ratio of cost of Type I to cost of Type II errors is low. When the ratio of cost of Type I to cost of Type II error is high, Bayesian model appears to be equivalent to the other models in randomly selected sample. Furthermore, with cross business cycle sample and cross sub cycle sample, results indicate that Bayesian model dominates other models when the ratio of Type I and Type II costs is high. Taking into account for results generated from the sampling methods, Rough Bayesian model performs relatively better in general. However, higher predictive power of Bayesian model when the ratio of cost of Type I to cost of Type II errors is high is consistent across all sampling methods. Such an advantage of Bayesian model may make

it more attractive in current economic environment where many large organizations have come under financial distress including cases such as Enron, Lehman and GM. These organizations were financed with significant amount of debt leading to high level of Type I error cost. However, they were generally considered not to be candidates for bankruptcy. Therefore, credit spread of such large organizations will be low leading to a low level of Type II error cost which is the opportunity cost for a bank issuing a loan. This investigation leaves several issues which can be considered as subject of future studies. Unlike this work, where we took a methodological approach which focuses on sensitivity of performance of Bayesian, Hazard, Mixed Logit and Rough Bayesian models, one can use an empirical approach to further analyze which model is sensitive to what factors or factors which may influence a model the most. In addition, rather than using a wide specification of cost, a model may be developed that more accurately estimates cost of both errors on an individual level helping to better compare bankruptcy prediction models. Finally, two major differences that characterize United States and Canadian markets is level of litigation risk and standard setting approach. An interesting study from international perspective is to determine differential impact of financial ratios and corporate governance quality in predicting bankruptcies.

References

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market based and accounting based bankruptcy prediction models. *Journal of Banking and Finance, 32*(8), 1541-1551. http://dx.doi.org/10.1016/j.jbankfin.2007.07.014

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589-609. http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E. I., & McGough, T. P. (1974). Evaluation of a company as a going concern. *The Journal of Accountancy, 138*(6), 50-57.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETA Analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance, 1*(1), 29-54. http://dx.doi.org/10.1016/0378-4266(77)90017-6

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (The Italian experience). *Journal of Banking and Finance, 18*(3), 505-529. http://dx.doi.org/10.1016/0378-4266(94)90007-8

Altman, E. I. (1993). *Corporate financial distress and bankruptcy*. John Wiley and Sons.

Altman, E. I. (2007). *Revisiting credit scoring models in a Basel 2 environment*. Lecture at National Taiwan University. Retrieved from www.fin.ntu.edu.tw/~hwangdar/94emba19.ppt

Beatty, A., Ke, B., & Petroni, K. (2002). Earnings management to avoid earnings declines across publicly and privately held banks. *Accounting Review, 77*(3), 547-570. http://dx.doi.org/10.2308/accr.2002.77.3.547

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*(Supplement), 71-111. http://dx.doi.org/10.2307/2490171

Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies, 10*(1), 93-122. http://dx.doi.org/10.1007/s11142-004-6341-9

Begley, J., Ming, J., & Watts, S. (1996). Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's Models. *Review of Accounting Studies, 1*(4), 267-284. http://dx.doi.org/10.1007/BF00570833

Bellovary, J. L., Giacomino, D., & Akers, M. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education, 33*(Winter), 1-43.

Chava, S., & Jarrow, R. (2004). Bankruptcy prediction with industry effect, market versus accounting variables and reduced form of credit risk models. *Review of Finance, 8*(4), 537-569. http://dx.doi.org/10.1093/rof/8.4.537

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. Wiley, New York.

Good, I. J. (1965). *The estimation of probabilities*: *An essay on modern bayesian methods*. MIT Press.

Grice, J. S., & Dugan, M. T. (2001). The limitations of bankruptcy prediction models: Some cautions for the researcher. *Review of Quantitative Finance and Accounting, 17*(2), 151-166. http://dx.doi.org/10.1023/A:1017973604789

Hensher, D. A., & Jones, S. (2007). Forecasting corporate bankruptcy: Optimizing the performance of the mixed logit model. *Abacus, 43*(3), 241-264. http://dx.doi.org/10.1111/j.1467-6281.2007.00228.x

Hillegeist, S., Cram, D., Keating, E., & Lundstedt, K. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies, 9*(1), 5-34. http://dx.doi.org/10.1023/B:RAST.0000013627.90884.b7

Hopwood, W. S., McKeown, J. C., & Mutchler, J. P. (1994). A reexamination of auditor versus model accuracy within the context of the going concern opinion decision. *Contemporary Accounting Research, 10*(Spring), 409-431. http://dx.doi.org/10.1111/j.1911-3846.1994.tb00400.x

Hsieh, S. J. (1993). A note on the optimal cut off point in bankruptcy prediction models. *Journal of Business Finance and Accounting, 20*(3), 457-464. http://dx.doi.org/10.1111/j.1468-5957.1993.tb00268.x

Hull, J., & White, A. (2000). Valuing credit default swaps I: No counterparty default risk. *Journal of Derivatives, 8*(1), 29-40. http://dx.doi.org/10.3905/jod.2000.319115

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Upper Saddle River, New Jersey, Pearson Prentice Hall.

Jones, F. L. (1987). Current techniques in bankruptcy prediction. *Journal of Accounting Literature, 6*, 131-164.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review, 79*(4), 1011-1038. http://dx.doi.org/10.2308/accr.2004.79.4.1011

Katz, S., Lilien, S., & Nelson, B. (1985). Stock market behavior around bankruptcy model distress and recovery predictions. *Financial Analysts Journal, 41*(1), 70-73. http://dx.doi.org/10.2469/faj.v41.n1.70

Martin, D. (1977). Early warning of bank failure. *Journal of Banking and Finance, 1*(3), 249-276. http://dx.doi.org/10.1016/0378-4266(77)90022-X

McKee, T. E. (2003). Rough sets bankruptcy prediction models versus auditor signaling rates. *Journal of Forecasting, 22*, 569-586. http://dx.doi.org/10.1002/for.875

Mensah, Y. M. (1984). An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. *Journal of Accounting Research, 22*(1), 380-395. http://dx.doi.org/10.2307/2490719

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance, 29*(2), 449-470.

Michael, S., Georgios, D., Nikolaos, M., & Constantin, Z. (1999). A fuzzy knowledge based decision aiding method for the assessment of financial risks: The case of corporate bankruptcy prediction. In *European Symposium on Intelligent Techniques*, *Crete*, *Greece*.

Neophytou, E., & Molinero, C. M. (2004). Predicting corporate failure in the United Kingdom: A multidimensional scaling approach. *Journal of Business Finance and Accounting, 31*(5-6), 677-710. http://dx.doi.org/10.1111/j.0306-686X.2004.00553.x

Ohlson, J. A. (1980). Financial ratios and probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109-131. http://dx.doi.org/10.2307/2490395

Pate, C. (2002). *Business chapter 11 bankruptcies*: *Recent levels and implications for 2002*. Pricewaterhouse Coopers.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences, 11*(5), 341-356. http://dx.doi.org/10.1007/BF01001956

Pawlak, Z. (1991). *Rough sets*, *Theoretical aspects of reasoning about data*. Dordrecht: Kluwer Academic Publishers.

Pawlak, Z., & Skowron, A. (1994). Rough membership functions. In R. R. Yager, M. Fedrizzi, & J. Kacprzyk, (Eds.), *Advances in the Dempster Shafer Theory of Evidence* (pp. 251-271). John Wiley and Sons, New York.

Platt, H. D., & Platt, M. B. (2002). Predicting corporate financial distress: Reflections on choice based sample bias. *Journal of Economics and Finance, 26*(2), 184-199. http://dx.doi.org/10.1007/BF02755985

Ravi, K. P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European Journal of Operational Research, 180*(1), 1-28. http://dx.doi.org/10.1016/j.ejor.2006.08.043

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics, 80*(4), 647-657. http://dx.doi.org/10.1162/003465398557735

Robertson, J., & Mills, R. (1991). The uses and abuses of corporate prediction models. *Management Accounting, 69*, 20-22.

Sarkar, S., & Sriram, R. S. (2001). Bayesian models for early warning of bank failures. *Management Science, 47*(11), 1457-1475. http://dx.doi.org/10.1287/mnsc.47.11.1457.10253

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business, 74*(1), 101-124. http://dx.doi.org/10.1086/209665

Sun, L., & Shenoy, P. P. (2007). Using bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research, 180*(2), 738-753. http://dx.doi.org/10.1016/j.ejor.2006.04.019

Tam, K. Y. (1991). Neural network models and the prediction of bank bankruptcy. *Omega, 19*(5), 429-445. http://dx.doi.org/10.1016/0305-0483(91)90060-7

Thomson, J. B. (1991). Predicting bank failures in the 1980s. *Economic Review*, *Federal Reserve Bank of Cleveland, 27*(Quarter 1), 9-20.

Verikas, A., Kalsyte, Z., Bacauskiene, M., & Gelzinis, A. (2010). Hybrid and ensemble based soft computing techniques. *Soft Computing, 14*(9), 995-1010. http://dx.doi.org/10.1007/s00500-009-0490-5

Webster, E., & Thornton, D. B. (2005). Earnings quality under rules versus principles based accounting standards: A test of the skinner hypothesis. *Canadian Accounting Perspectives, 4*(2), 167-192. http://dx.doi.org/10.1506/6TPA-D5HH-GKMB-9GJV

Weiss, L. A., & Capkun, V. (2004). *The impact of incorporating the cost of errors into bankruptcy prediction models*. Working Paper.

West, R. C. (1985). A factor analytic approach to bank condition. *Journal of Banking and Finance, 9*(2), 253-266. http://dx.doi.org/10.1016/0378-4266(85)90021-4

Whalen, G., & Thomson, J. B. (1988). Using financial data to identify changes in bank condition. *Economic Review*, *Federal Reserve Bank of Cleveland, 24*(2), 17-26.

Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis, 15*(3), 757-770. http://dx.doi.org/10.2307/2330408

Wilson, R. L., & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems, 11*(5), 545-557. http://dx.doi.org/10.1016/0167-9236(94)90024-8

Yao, Y. Y. (2007). Decision theoretic rough set models. *Proceedings of RSKT 2007*, *Lecture Notes in Artificial Intelligence*, *LNAI* 4481, 1-12.

Yao, Y. Y. (2008). Probabilistic rough set approximations. *International Journal of Approximate Reasoning*, 49(2), 255-271. http://dx.doi.org/10.1016/j.ijar.2007.05.019

Yao, Y. Y. (2010). Three way decisions with probabilistic rough sets. *Information Sciences, 180*(3), 341-353. http://dx.doi.org/10.1016/j.ins.2009.09.021

Yao, Y. Y., & Wong, S. K. M. (1992). A decision theoretic framework for approximating concepts. *International Journal of Man Machine Studies, 37*(6), 793-809. http://dx.doi.org/10.1016/0020-7373(92)90069-W

Yao, Y. Y., Wong, S. K. M., & Lingras, P. A. (1990). Decision theoretic rough set model. In Z. W. Ras, M. Zemankova, & M. L. Emrich (Eds.), *Methodologies for Intelligent Systems 5* (pp. 17-24), New York, North Holland.

Zavgren, C. (1983). The prediction of corporate failure: The state of the art. *Journal of Accounting Literature, 2*(1), 1-37.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research, 22*(Supplement), 59-82. http://dx.doi.org/10.2307/2490859