

**Using a Bayesian Model for Bankruptcy Prediction:
A Comparative Approach**

by

Zhanpeng He

Submitted in partial fulfillment
of the requirements for the degree of

Master of Science in Management (Finance)

Faculty of Business, Brock University
St. Catharines, Ontario, Canada

© Zhanpeng He 2011

Abstract

The purpose of this study is to examine the impact of the choice of cut-off points, sampling procedures, and the business cycle on the accuracy of bankruptcy prediction models. Misclassification can result in erroneous predictions leading to prohibitive costs to firms, investors and the economy. To test the impact of the choice of cut-off points and sampling procedures, three bankruptcy prediction models are assessed- Bayesian, Hazard and Mixed Logit. A salient feature of the study is that the analysis includes both parametric and nonparametric bankruptcy prediction models. A sample of firms from Lynn M. LoPucki Bankruptcy Research Database in the U. S. was used to evaluate the relative performance of the three models. The choice of a cut-off point and sampling procedures were found to affect the rankings of the various models. In general, the results indicate that the empirical cut-off point estimated from the training sample resulted in the lowest misclassification costs for all three models. Although the Hazard and Mixed Logit models resulted in lower costs of misclassification in the randomly selected samples, the Mixed Logit model did not perform as well across varying business-cycles. In general, the Hazard model has the highest predictive power. However, the higher predictive power of the Bayesian model, when the ratio of the cost of Type I errors to the cost of Type II errors is high, is relatively consistent across all sampling methods. Such an advantage of the Bayesian model may make it more attractive in the current economic environment. This study extends recent research comparing the performance of bankruptcy prediction models by identifying under what conditions a model performs better. It also allays a range of user groups, including auditors, shareholders, employees, suppliers, rating agencies, and creditors' concerns with respect to assessing failure risk.

Table of Contents

1. Introduction	1
2. Literature Review	5
2.1 Development of Bankruptcy Prediction Models.....	5
2.2 Limitations of Bankruptcy Prediction Models.....	6
2.3 Validation Methods for Comparison of Bankruptcy Prediction Models	9
3. Econometric Models	11
3.1 Bayesian Model	12
3.2 Hazard Model.....	13
3.3 Mixed Logit model	14
4. Sample Selection and Descriptive Statistics	16
5. Empirical Results	21
5.1 Cut-off Point and Type I, Type II Errors	21
5.2 Optimal Cut-off Point	23
5.3 Randomly Selected Samples.....	27
5.4 Samples in Different Business Cycles and Sub Cycles.	28
6. Conclusion.....	33
7. Reference.....	35

1. Introduction

The prevention of firm bankruptcy is a key goal of management. There may be early warning signs of impending financial distress and this would allow the manager to act in a preemptive manner to mitigate the situation from worsening. In the United States, defaults and bankruptcies have increased markedly over the past decade, in fact, 17 of the 20 largest bankruptcy filing in the U.S over this period. A substantial amount of effort in the academic literature has been devoted to forecasting corporate failure. The methodologies employed have been based mainly on various statistical models. During this unprecedented distress period, three models- Bayesian (Sarkar & Sriram, 2001), Hazard (Shumway, 2001) and Mixed Logit (Jones & Hensher, 2004)- have been applied to bankruptcy prediction. All these models have theoretical advantages over existing prediction models. All three models have been applied successfully to finance and accounting e.g. Beatty et al (2002). In this study, we address the issue of empirical testing of the comparative power of the three approaches; this study also examines the role of cut-off point choice and sampling procedures in evaluating the relative performance of bankruptcy prediction models. Specifically, we address the following four questions: (1) Do the rankings of bankruptcy prediction models depend on the arbitrary choice of the cut-off point? (2) Procedure to determine the optimal cut-off point. The first two research questions focus on the critical role of the cut-off point as it affects bankruptcy prediction models. (3) Do performance outcomes of bankruptcy prediction models depend on sampling procedures? This question extends the bankruptcy inquiry to acknowledge the sensitivity of the change in the economic cycle. (4) How the cost of type I errors and Type II errors influence the performance of bankruptcy prediction models. The answers to these questions provide insight to the fundamental question concerning the choice of failure prediction models which are critical in the optimal allocation of

resources.

The relevance of our study is twofold. (1) The current financial crisis is a result primarily of a failure in accounting for credit risk, and consequently a failure of accounting for default risk by the borrowers. This crisis thus demonstrated the need for better methods of evaluating the risk of bankruptcy. (2) 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 the stockholders' equity necessary to cover the risk of default.

Our study builds on prior and concurrent research in three different streams in the bankruptcy prediction literature. First, we build on studies by Jones and Sarkar and Sriram (2001), Shumway (2001) and Hensher (2004) who proposed the use of advanced probability modeling in the 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 the sensitivity of the performance of their models to the cut-off point selection, to the sampling procedure and to the ratio of cost of type I error to type II error.

The second stream of research demonstrates the importance of the industry, change in accounting regulation, and change in ability of financial statement data in the forecasting accuracy of bankruptcy prediction models (Beaver et al. 2005; Chava and Jarrow 2004; Hillegeist et al. 2004). The basic notion of their research is that besides the industry effect, aspects of the accounting systems such as the going concern principle and the conservatism principle among others will limit the performance of any accounting-based bankruptcy prediction model. Given their results, we investigate the

impact of the overall economic business cycle on testing bankruptcy prediction models.

The third stream of research focuses on the criterion used to compare bankruptcy prediction models such as classification accuracy, rates of misclassifications and cost of misclassification- see Sun and Shenoy (2007), Jones and Hensher (2004), Grice and Dugan (2001), and Begley et al. (1996). In this research, we attempt to address the usefulness, as measured by rankings, of the three statistical models to the three criterions.

Our sample consists of 224 bankrupt firms which were collected from Lynn M. LoPucki's Bankruptcy Research Database and 2561 non bankrupt firms. The Lynn M. LoPucki's Bankruptcy Research Database includes bankrupt firms with \$100 million or more in assets measured in 1980 dollars, which filed Chapter 11 in the US from 1980. One of the advantages of Lynn M. LoPucki's database is that it records the date of filing and the last available financial statement before filing, so that the "backcasting" (a term used by Jones et al. 2004) identified by Ohlson (1980) can be easily solved.

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

The findings of this study contribute to the 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 firms' going-concern conditions, notwithstanding their knowledge of the firm (Hopwood et al. 1994; McKee 2003). An appropriate choice of a bankruptcy model can aid the auditor in recognizing the auditor's disclaimer or qualification as to the going-concern nature of a business (Altman and McGough, 1974). Bankruptcy prediction models can also serve audit researchers in better understanding auditor's biases (Sarkar and Sriram, 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 and Sharda, 1994). An early warning model will allow management to take corrective actions before it is too late (Whalen and Thomson, 1988). In fact, research in this area assumes greater significance because a poor credit risk model might lead to sub-optimal capital allocation (Agarwal and Taffler, 2008). Regulators can intervene early so that mitigating actions can be taken to reduce the expected costs of failure (Thomson, 1991). Finally, an appropriate model can help price distress security as shown in Katz, Lilien and Nelson (1985) who found that abnormal return can be earned by identifying changes in scores calculated from the bankruptcy model.

The rest of the paper is as follow. Section 2 discusses the academic literature and its methodological implications for the modeling of bankruptcy. Section 3 describes the econometric models and research design that are used. Sample selection and descriptive statistics are discussed in section 4. Empirical results are presented in section 5. Section 6 concludes with a summary of the main findings and implications for future research.

2. Literature Review

2.1 Development of Bankruptcy Prediction Models

Beaver (1966) used the univariate approach and found that net income to total debt had the highest predictive ability. Although the univariate analysis seems easy and intuitive to implement, the ambiguity of the explanatory power does not provide clear signals (Jones, 1987; Zavgren, 1983). Altman (1968) used linear discriminant analysis (MDA) which assumes equivalent covariance matrices across groups. In 1977, he proposed ZETA analysis which relaxes the assumption of equivalent distribution of two groups' covariance matrices, however ZETA analysis is still criticized for the 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 MDA and ZETA which are used as discrimination tool, Logit (Martin, 1977; Ohlson, 1980) and Probit (Zmijewski, 1984) are models designed for the estimation of probability. These two models require assumptions only on the residuals' distribution, thus avoiding the criticisms aimed at DA and ZETA analysis.

As parametric models dominated the research focus at the beginning, studies using non-parametric models started to develop in the late 1980s. The most commonly used was Neural Network (e.g. Tam, 1991). Over the past two decades, the number of studies related to Neural Network in bankruptcy prediction is 39, compared with 11 in MDA, 19 in Logit and 3 in Probit (Bellovary et al., 2007). However the criticism of the non-parametric models is that the significance of variables is not testable.

Recently, three models - Bayesian (Sarkar & Sriram, 2001), Hazard (Shumway, 2001) and Mixed Logit (Jones & Hensher, 2004)- have been applied to bankruptcy

prediction. They are deemed to be theoretical advantageous over previous models. The Bayesian model applies the 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 the posterior probability is obtained. The noticeable characteristics of this model are transparent. It is intuitive and easy to understand.

The improvement of the Hazard model over Logit is that the former explicitly models bankruptcy not as a process that happens at a point in time, an assumption made by all previous models, but as a process that lasts for a period of the firm's life. Shumway (2001) argued that the Hazard model is theoretically preferable as it incorporates time-varying covariates.

The advancement of the Mixed Logit model over Logit is that it takes into account both observed information (represented by financial variables that can be acquired from financial statements or any other open source) and unobserved information (represented by predefined variables that follow pre-defined distributions). Under this setting, there are two means to model the unobserved information. The first one is the random parameter specification; the second one is the error components approach.

2.2 Limitations of Bankruptcy Prediction Models

Bankruptcy prediction models are usually evaluated from a sample chosen in a particular time period. Since the distribution of accounting variables are dynamic (Mensah, 1984), models are likely to be sample specific (Agarwal and Taffler, 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 the models, and found that the accuracy of each model declined from the period of 1988-1991 to the period of 1992-1999. Begley et al. (1996) came to the similar conclusion that Altman (1968) and Ohlson (1980) models did not perform well in the 1980s, even when the coefficients were re-estimated.

Robertson et al. (1991) cautioned that one could not assume that the predictive power of models can transcend to industries. For example, MDA, originally developed for the manufacturing industry, is used by practitioners as one of the important indicators of credit worthiness across many industries without re-estimating the coefficients. Grice et al. (2001) found that Ohlson's (1980) model was sensitive to industry classification while Zmijewski's (1984) model was not.

Zmijewski (1984) identified 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, bankrupt and non-bankrupt. This procedure violates the assumption of exogenous random sampling since the probability that a firm enters the sample depends on the observed status. He found that the predictive power is upwardly biased when the sample selected has a probability of bankruptcy, which deviates from the population probability of bankruptcy. He further suggested that the weighted exogenous sample maximum likelihood (WESML) can be used to adjust the bias. Platt et al. (2002) tested the choice-based bias and results were consistent with Zmijewski's (1984) findings.

The second sampling error refers to the sample selection bias. Sampling procedures usually eliminate the observations with incomplete data. Zmijewski (1984) used a

bivariate normal approach to estimate the correlation between bankruptcy and observation with missing data. He found that firms with missing data would have a higher probability of bankruptcy. In other words, sample selection bias understates bankruptcy probability.

Beaver et al. (2005) pointed out that, the importance of intangible assets has increase over time due to technology-based assets generated through R&D expenditure. In addition, financial derivative markets experienced an explosion, in the 1990s. While financial derivatives are mainly used as a substitute for leverage, such a marked increase would lead to underestimate the leverage ratios of firms. Begley et al. (1996) also state that the leverage variables play an important role. Beaver et al. (2005) argued that degree of discretion on financial statements is increasing. These three developments have a direct effect on financial ratios, which then undermine the predictive power of bankruptcy prediction models whose inputs are mainly financial variables.

Changes in regulation may also have an impact on the accuracy of accounting-based prediction models. Since 1973, many new standards have been established for various perspectives, such as Statement No.87 for pensions, No. 106 for post retirement benefits, No. 107 for financial instruments and No.115 for debt and equity. Any new standard will have an effect in the long run on financial statement. (Beaver et al., 2005). The use of bankruptcy filing has become strategic consideration since the changes to the bankruptcy laws in the late 1970s (Begley et al., 1996).

2.3 Validation Methods for Comparison of Bankruptcy Prediction Models

The criteria mostly used to compare bankruptcy prediction models are classification accuracy, rates of misclassification and cost of misclassification. Classification accuracy is used by many researchers (Jones and Hensher 2004; Grice and Dugan 2001; Sun and Shenoy 2007; Wiginton 1980), because of its intuitive nature. Bankruptcy prediction models are applied in the hold-out sample or the original sample that is used to estimate parameters of the model, and then the sample is separated into groups according to the firm'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 the Jackknife method, where one observation is withheld from the estimation sample and its status is predicted, then the same procedure is repeated until each observation has been predicted and individual classification accuracy is accumulated over the entire sample.

Unlike classification accuracy, the rate of misclassification counts the observations that are incorrectly classified. Two types of error are defined accordingly. Type I (Type II) errors refer to the incorrect predictions of bankrupt (non-bankrupt) firms. Numerically, the rate of errors is equal to one minus the rate of classification accuracy from a respective group. Ohlson (1980) demonstrated that the rate of misclassification varies across different cut-off points. To find the optimal cut-off point, he summed the rates of the type I and II errors and the optimal solution is found by minimizing the sum.

The cost of a type I error is the loss of the principle when the debtor defaults. The cost of a type II error refers to the opportunity cost that is the difference between interest revenue generated from loans that should have been issued and the risk-free rate of return.

(Altman et al., 1977). Altman et al. (1977) examined 26 largest U.S. commercial banks and 67 smaller, regional banks. They found that cost of a type I error is 35 times that of a type II error. In contrast, Weiss and Capkun (2004), suggested a way to specify the cost of misclassification in relation to individual firm's characteristics by incorporating Hull & White (2000) spread model and Merton (1974) model.

My study differs from existing research in three different ways. First, we take into account the choice of the cut-off point with cost of misclassification, which is the major concern for probabilistic models. The literature suggests that, the choice of the cut-off point used to predict bankruptcy is somewhat arbitrary e.g. 0.41% by Martin (1977); 50% by Grice et al. (2001), Zmijewski (1984) and Ohlson (1980); 3.8% by Begley et al. (1996). In this study, a cut-off point that presents the fair comparison will be estimated.

Second we compare the performance of bankruptcy models under rigorous sampling procedures. The prediction accuracy will be upward bias if one sample is used to estimate parameters meanwhile the same sample is also used to estimate the bankruptcy probability, as the model is already fitted into the data. However, the degree to which such bias is produced varies across different models and is difficult to estimate, so an adjustment method must be account for. One method to solve this problem is to use a hold-out sample for testing (Hsieh 1993; Grice and Dugan 2001). However more than half of studies do not use a hold-out sample: 87 cases without hold-out versus 77 cases with hold-out (Bellovary et al., 2007). One should be cautioned that, even if a hold-out sample is used, prediction accuracy may still be biased upwardly when the training sample and the hold-out sample are not substantially different in the underlying economic conditions (Grice and Dugan, 2001). A more rigorous test should be conducted in two periods that are likely to exhibit economic changes.

In addition, Choice-based sampling bias was common to many of the previous studies (e.g. Altman 1968; Jones and Hensher 2004). Zmijewski (1984) found that sampling bias produces inconsistent and biased estimators and that the prediction accuracy is unreasonably high. In this study, we use weighted exogenous sample maximum likelihood (WESML) to adjust choice-based sample bias as Zmijewski (1984) did.

Third we use the cost of misclassification as a criterion to compare the accuracy of bankruptcy models. Wiginton (1980) cautioned that the use of classification accuracy rates should be utilized in a manner that aids classification decisions with cost of misclassification. We also control for the “backcasting” problem. As pointed out by Ohlson (1980), “backcasting” happens when a firm files for bankruptcy before the public have access to the most recent annual report. If data is collected from such a report, accuracy is biased upwardly. To correct such bias, financial statements representing the fiscal year ending with a minimum of three months before bankruptcy filing should be used for data collection. In the next section, we discuss the bankruptcy models.

3. Econometric Models

The notation used throughout the following discussions is:

y_i is the status of firm i in a given year, for $i=1,2,\dots,n$. $y_i = \begin{cases} 0 & \text{if firm is not bankrupt} \\ 1 & \text{if firm is bankrupt} \end{cases}$

$x_{i,j}$ is the independent variable j for firm i , for $i=1,2,\dots,n; j=1,2,\dots,6$. In this study,

j =ROA, ETL, LTA, LERET, LSIGMA and LRSIZE respectively. The definitions of the independent variables are presented in section 4.

\mathbf{X}_i is a 7×1 vector of the independent variables for firm i ($i=1,2,\dots,n$). First element is 1.

β_j is the coefficient of the j th element of variable j of \mathbf{X}_i where β_0 is the intercept

β is a 7×1 vector of coefficients.

3.1 Bayesian Model

The Bayesian equation for this study is expressed as follows:

$$P(y_i = 1 | x_{i,1}, \dots, x_{i,6}) = \frac{P(x_{i,1}, \dots, x_{i,6} | y_i = 1)P(y_i = 1)}{P(x_{i,1}, \dots, x_{i,6})}, \text{ for } i = 1, \dots, n \quad (1)$$

Sarkar et al. (2001) found that the naïve model which assumes independence across predictive variables is comparable in performance with the composite model which assumes non-independence across some or all predictive variables as shown in (1). So the naïve model is used in this study, as follows:

$$P(y_i = 1 | x_{i,1}, \dots, x_{i,6}) = \frac{\prod_{j=1}^6 P(x_{i,j} | y_i = 1)P(y_i = 1)}{\prod_{j=1}^6 P(x_{i,j})}, \text{ for } i = 1, \dots, n \quad (2)$$

The variable x_j which is continuous needs to be discretized before inputted to the model. Sun et al. (2007) found that the optimal level of discretization is either 2 or 3. We use EP-T method as Sun et al. (2007) did to discretize each input variable into 3 levels, which approximates the continuous distribution by dividing the outcome space into levels with probability of occurrence of 18.5%, 63% and 18.5% by respectively. The two points for discretizing purpose are determined in each training sample. And the two points are also used to discretize the holdout sample. The bankruptcy rate of the training sample is used as the estimator of the probability of bankruptcy for individual firm i in the holdout sample.

3.2 Hazard Model

Shumway (2001) proved that multiperiod Logit model is equivalent to a discrete-time hazard model. The Multiperiod Logit model is estimated from data of each firm in each year as if each firm year is an independent observation. So the hazard function for one observation can be expressed as follows:

$$h(t_i) = P_{t_i}(y_i = 1), \text{ for } i = 1, 2, \dots, n \quad (3)$$

Where $h(t_i)$ is the probability that firm i will go into bankruptcy in period of t_i to $t_i + 1$.

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

The survival function is the probability that a firm survives to time T_i . The Logit function is assumed to be the functional form of the hazard function, so (3) can be re-written as

$$h(t_i) = P_{t_i}(y_i = 1) = \frac{\exp(\mathbf{X}'_{i,t_i}\beta)}{1 + \exp(\mathbf{X}'_{i,t_i}\beta)} \quad (5)$$

The likelihood function is shown in equation (6):

$$\ell = \prod_{i=1}^n \left(P_{T_i}(y_i = 1)^{y_i} \cdot [1 - P_{T_i}(y_i = 1)]^{(1-y_i)} \cdot \prod_{t_i=1}^{T_i-1} [1 - P_{t_i}(y_i = 1)] \right) \quad (6)$$

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

3.3 Mixed Logit model

In the Mixed Logit model, the utility associated for each observation i can be expressed as

$$U_i = \mathbf{X}_i' \boldsymbol{\beta} + \eta_i + \varepsilon_i \quad (7)$$

In (7), η_i is the error component that can be correlated among alternatives and be heteroskedastic for each individual firm, and ε_i is the random term with mean 0, which is independently and identically distributed over alternatives and individual firm. If the Logit function is assumed to be the functional form of the probability of bankruptcy and η_i is given, the conditional Logit model is identical to traditional Logit model.

$$P(y_i = 1 | \eta_i) = \frac{\exp(\mathbf{X}_i' \boldsymbol{\beta} + \eta_i)}{1 + \exp(\mathbf{X}_i' \boldsymbol{\beta} + \eta_i)} \quad (8)$$

In addition η_i can be specified as follows:

$$\eta_i = \mathbf{z}_i' \boldsymbol{\mu} + \mathbf{w}_i' \boldsymbol{\gamma} \quad (9)$$

Where \mathbf{z}_i is a vector of a subset or full set of observed independent variables and $\boldsymbol{\mu}$ is a random vector with mean 0, whose density function is Ω . Ω can be chosen as any distribution with mean 0. The combination of $\mathbf{z}_i' \boldsymbol{\mu}$ can induce heteroskedasticity or correlation or both across unobserved utility components of the alternative status of an observation by specifying Ω in a desired manner. This combination is known as the error components approach, which treats the unobserved information as a separate error component in the random component.

\mathbf{w}_i is a subset or full set of \mathbf{X}_i' and $\boldsymbol{\gamma}$ is a vector of random parameters (or random coefficients). Random coefficients allow heterogeneity across individual firms with respect to their sensitivity to observed exogenous variables. Denoting the m th

element of γ as γ_m , then γ can be further specified as

$$\gamma_m = b_m + s_m \cdot K \quad (10)$$

Where K is a random variable specified to follow any distribution with mean 0 and variance 1. The combination of $w_i \gamma$ is known as a random parameter approach.

In this study, we only adopt the random parameter approach in accordance with Jones et al. (2004). So b_m and s_m are parameters that must be estimated in addition to β . K is defined to be $N(0,1)$. Thus γ_m is also normally distributed with mean b_m and variance s_m^2 .

Then the unconditional probability of bankruptcy can be expressed as the expected value of the indicator function $I(y_i)$, which is 1 if $y_i = 1$ otherwise 0.

$$P(y_i = 1) = E[I(y_i = 1)] \quad (11)$$

According to the law of iterated expectation, (11) can be transformed as

$$P(y_i = 1) = E\{E[I(y_i = 1) | \eta_i]\} = \int P(y_i = 1 | \eta_i) f(\eta_i) d\eta_i \quad (12)$$

The integral in (12) is calculated using the Matlab function “quadgk” which integrates from minus infinity to positive infinity.

In this study, LSIGMA and ROA are selected as the subset w whose coefficients are random. So the likelihood function is

$$\ell = \prod_{i=1}^n \left[\int P(y_i = 1 | \eta_i) f(\eta_i) d\eta_i \right]^{y_i} \cdot \left[1 - \int P(y_i = 1 | \eta_i) f(\eta_i) d\eta_i \right]^{(1-y_i)} \quad (13)$$

The estimates of the Hazard and Mixed Logit models may be inconsistent and biased, because of the choice-based sample (Zmijewski 1984). He summarized three ways to estimate models with choice-based sample: weighted exogenous sample

maximum likelihood (WESML), conditional maximum likelihood (CML) and full information concentrated maximum likelihood (FICML). WESML is computationally the least complex, so it is used to estimate the Hazard and Mixed Logit models in this study.

The log-likelihood function of the Hazard model using WESML is adjusted and is

$$L_H = \frac{P_p(y=1)}{P_s(y=1)} \sum_{i=1}^n y_i \cdot \left(\log[P_t(y_i=1)] + \sum_{t_i=1}^{T_i-1} \log[1 - P_t(y_i=1)] \right) + \frac{P_p(y=0)}{P_s(y=0)} \sum_{i=1}^n (1 - y_i) \cdot \left(\sum_{t_i=1}^{T_i} \log[1 - P_t(y_i=1)] \right) \quad (14)$$

The log-likelihood function of the Mixed Logit model using WESML is adjusted and is

$$L_M = \frac{P_p(y=1)}{P_s(y=1)} \sum_{i=1}^n y_i \cdot \log \left[\int P(y_i=1 | \eta_i) f(\eta_i) d\eta_i \right] + \frac{P_p(y=0)}{P_s(y=0)} \sum_{i=1}^n (1 - y_i) \cdot \log \left[1 - \int P(y_i=1 | \eta_i) f(\eta_i) d\eta_i \right] \quad (15)$$

where for both (14), (15),

$P_p(y=1)$ is the population bankruptcy rate (0.847%, reported by Zmijewski 1984)

$P_p(y=0) = 1 - P_p(y=1)$

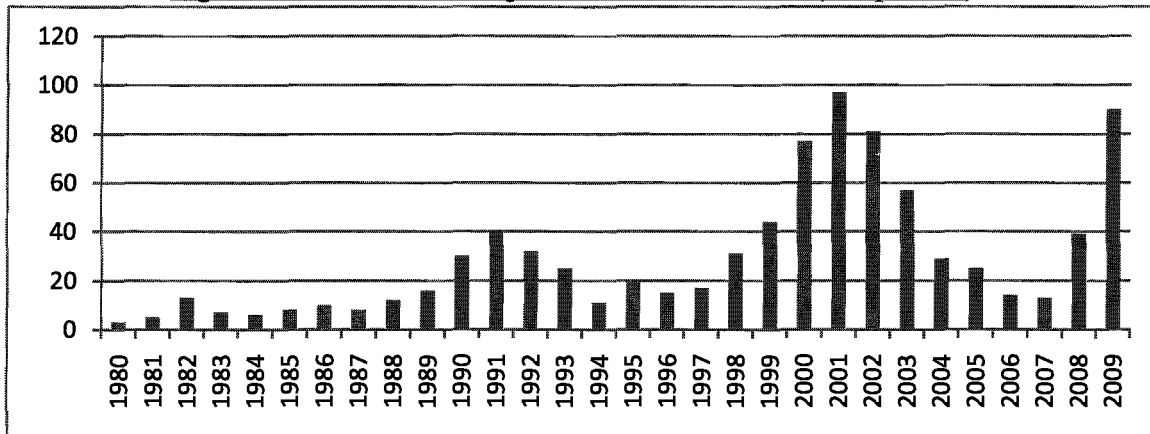
$P_s(y=1)$ is the sample bankruptcy rate

$P_s(y=0) = 1 - P_s(y=1)$

4. Sample Selection and Descriptive Statistics

Bankrupt firms were identified using Lynn M. LoPucki's Bankruptcy Research Database. The database includes bankrupt firms with \$100 million or more in assets measured in 1980 dollars, which filed Chapter 11 in the US starting from 1980. The database records the date of filing and last available financial statement before filing, so that the “backcasting” problem can be easily solved.

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



*Re-produced from the data obtained from LoPucki's Database

Figure 1 provides the number of bankruptcy filings in each year. The number of bankruptcies jumps significantly at the turning of new millennia when the high-tech bubble burst. The bankruptcies rise from an average 20 cases in 1990s to 97 cases in 2001. Most recently, following the financial crisis, bankruptcy cases soared to a 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 the latest data is available in Compustat. Firms that had missing values in any of the independent variable were eliminated from the sample. As a result, the final sample comprises 224 bankrupt firms. Due to the lack of a bivariate logit distribution, the process of elimination of firms with missing data results in sample selection bias that cannot be adjusted using Zmijewski's method.

Non-bankrupt firms are collected from Compustat. A firm with "active" status and a Compustat code for the month of fiscal-year end that is different from zero is considered as nonbankrupt. The firm year is sampled according to the business cycle. Business cycles are defined in accordance with the 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 troughs that can be identified since 1980, as follow*:

<u>Peak</u>	<u>Trough</u>
Jan,1980	Jul,1980
Jul,1981	Nov,1982
Jul,1990	Mar,1991
Mar,2001	Nov,2001
Dec,2007	

*Excerpted from: <http://www.nber.org/cycles/cyclesmain.html>

Eight sub-cycles (expansion/contraction) can be determined as shown in Table 1. A firm whose fiscal year ends within a sub-cycle is selected for that sub-cycle, and it will be eliminated from selection in later periods. The sampling period is from 1980 to 2007 and the total sample size for non-bankrupt firms is 2561, after firms with missing values for independent variables are eliminated. Table 1 provides the breakdown of the number of non-bankruptcies and bankruptcies in each business cycle.

Table 1. 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	Jul,1980	1	48
	2	Expansion	Aug,1980	Jul,1981	2	32
2	3	Contraction	Aug,1981	Nov,1982	6	174
	4	Expansion	Dec,1982	Jul,1990	31	161
3	5	Contraction	Aug,1990	Mar,1991	7	21
	6	Expansion	Apr,1991	Mar,2001	91	1107
4	7	Contraction	Apr,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 the number and the combination of independent variables in the literature. Bellovary et al. (2007) summarized that the number of variables in the literature ranges from 1 to 57 and

that a 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 the models using a higher number of independent variables (Bellovary et al. 2007). Beaver et al. (2005) also argued that the effect of the selection of independent variables may have only marginal influence, because statement variables are correlated. They further found out that the informative power of statement variables is actually decreasing due to the change in the application of financial tools and the standards. This loss of informative power in accounting information can be compensated with the usage of market variables (Beaver et al. 2005). In this study, we use the six variables 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:

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

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 the indicator to measure the liquidity of a firm to generate cash in order to meet the interest and principal requirement, especially in the short term. (Compustat code: sales=DATA12, COGS=DATA41, selling, general and administration expense=DATA189)

LTA: total liabilities divided by total assets, which is the measure of company's

capital structure. (Compustat code: total liabilities=DATA181)

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

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

LRSIZE: the logarithm of the ratio of the market capitalization of the firm divided by the market capitalization of the market index. Market capitalization is calculated as number of share outstanding times stock price.

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

Table 2. Descriptive Statistics of Independent Variables

Bankruptcy					Non-Bankruptcy				
Variable	Mean	Std	Min	Max	Variable	Mean	Std	Min	Max
ROA	-0.0198	0.2257	-4.7533	0.3619	ROA	0.0032	0.3961	-9.6133	20.3319
ETL	0.1967	0.345	-4.5566	2.5477	ETL	0.9032	17.2584	-64.0345	1018.8
LTA	0.6598	0.2447	0.0159	2.2261	LTA	0.5019	0.2622	0.0001	4.8621
LERET	-0.0974	0.6328	-4.0301	4.1521	LERET	0.0917	0.5642	-3.2852	10.856
LSIGMA	0.1285	0.0831	0.0225	0.7704	LSIGMA	0.1199	0.0934	0.0005	2.976
LRSIZE	-9.8788	1.6228	-15.5839	-4.2751	LRSIZE	-10.0175	2.0678	-17.0762	-3.2056

5. Empirical Results

5.1 Cut-off Point and Type I, Type II Errors

If a firm's fitted probability is greater than the cut-off point, the firm is classified to be bankrupt and vice versa. If a firm, whose observed status is bankrupt, is classified to be non-bankrupt, such error is referred as a type I error and the reverse is a type II error. To show the relationship between the cut-off point and type I and type II errors, we first randomly and equally divide the total sample into two samples namely training and hold-out 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 three models and the hold-out sample is used for classification purpose to calculate the two types of error.

The rate of errors (Type I or II) is calculated as the number of firms that are misclassified over number of firms in the group. Figures 2 and 3 exhibit the relationship between the cut-off point and the two types of errors as well as the tradeoff between the two errors for the three models. As a cut-off point moves toward 1, the type I error increases. In contrast, type II error decreases as the cut-off point increases. Because the fitted probability as generated using the Bayesian model can be zero or greater than one, the rate of type I errors can be far less than one. The rate of type II error can also be tilted above zero, even if the cut-off point is chosen to be one.

Figure 2. Rate of Type I Error

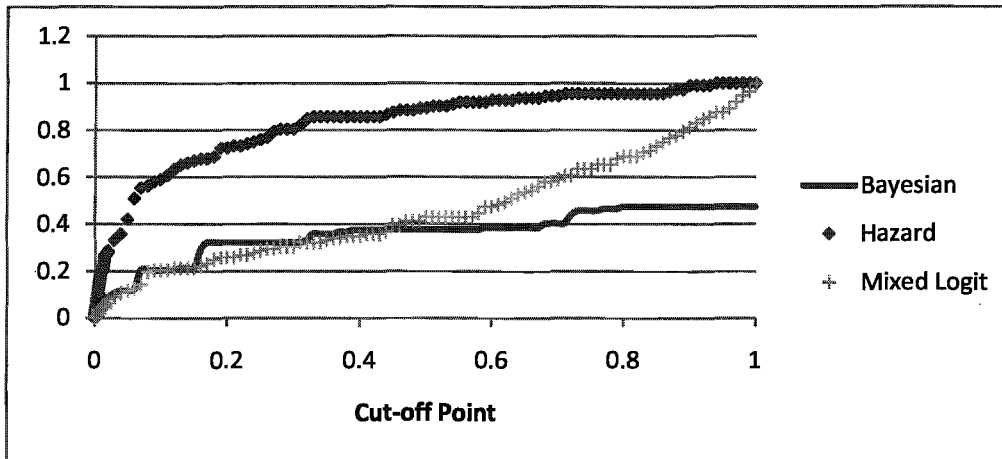
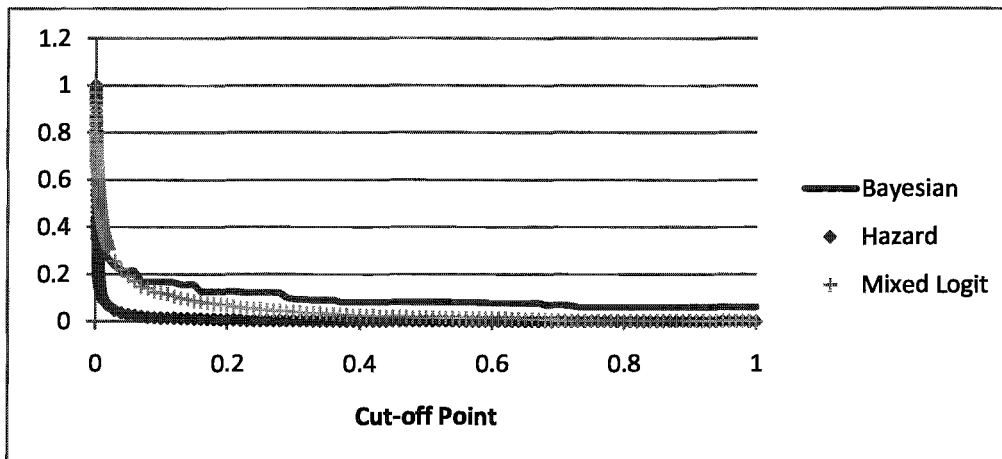


Figure 3. Rate of Type II Error



If the classification error (represented by rate of two types of errors) or the classification accuracy (calculated by one minus classification error) is the validation method applied to compare bankruptcy models, figures 2 and 3 indicate that one can have different conclusions by choosing different cut-off points.

There are three main methods of choosing cut-off points. Johnson et al. (2007) suggest using a cut-off point of 50%, because prior probability and costs are difficult to incorporate. Grice et al. (2001), Zmijewski (1984) and Ohlson (1980) use 50% as a cut-off point in their studies. Martin (1977) recommends the use of the industry failure rate (0.41%) as a cut-off point, assuming that a prior probability for group membership is

equal to the sample probability. Begley et al. (1996) used 3.8%, a rate obtained from Ohlson (1980), that minimizes the sum of type I and type II errors.

Table 3. Predictive Results with Different Cut-off point

Cut-off point	Bayesian			Hazard			Mixed Logit		
	Type I	Type II	Type I+II	Type I	Type II	Type I+II	Type I	Type II	Type I+II
50%	0.375	0.079688	0.454688	0.892857	0.000781	0.893638	0.428571	0.014844	0.443415
0.41%	0.026786	0.535938	0.562723	0.080357	0.254688	0.335045	0.008929	0.579688	0.588616
3.80%	0.116071	0.216406	0.332478	0.357143	0.035938	0.39308	0.107143	0.21875	0.325893

Table 3 reports the results based on different cut-off points used in previous studies. Within each model, Type I (II) is the rate of type I (II) error; Type I+II is the sum of type I errors and type II errors. If the cut-off point is 50%, when the type I error is the main concern for validation, the Bayesian model outperforms the others. However, when the sum of the type I errors and type II errors is the main concern, the Hazard and Mixed Logit models will dominate when using the cut-off points of 0.41% and 3.8% respectively. Obviously, such results are arbitrary and don't reflect the true quality of the models. The above cut-off point choices have two limitations. First, an inappropriate assumption of equal costs of misclassification is made, if the sum of two errors is used as the validation method. Altman et al. (1977) estimated that cost of a type I error is 35 times greater than that of a type II error. This is due to the fact that, when type I errors occur, creditors lose their total principle, in contrast to the opportunity cost resulting from the occurrence of type II error occurs. Secondly, the differences of numbers of firms in each group are generally ignored. The number of bankrupt firms is far fewer than that of non-bankrupt firms. According to Zmijewski (1984), the frequency of financial failure has not exceeded .75% since 1934.

5.2 Optimal Cut-off Point

An appropriate cut-off point should minimize the cost of misclassification, as given

by equation (16).

$$\min_{\alpha} C = \sum_i^n [C_{I,i} \cdot I(\hat{P}_i \leq \alpha)]^{y_i} \cdot [C_{II,i} \cdot I(\hat{P}_i > \alpha)]^{(1-y_i)} \quad (16)$$

Where

C is the total cost of misclassification

$C_{I,i}, C_{II,i}$ are the cost of type I and type II error for firm I respectively

α is the cut-off point

\hat{P}_i is the fitted probability of bankruptcy for firm i

y_i is the observed status of firm i

$I()$ is the indicator function that returns 1 if logical function is satisfied, otherwise returns 0

Weiss and Capkun (2004) provide a methodology to estimate the 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 (17) provides the minimization of the simplified total cost.

$$\min_{\alpha} C = C_I \cdot P_I(\alpha) \cdot Q_I + C_{II} \cdot P_{II}(\alpha) \cdot Q_{II} \quad (17)$$

P_I, P_{II} are rates of type I, type II error, a function of cut-off point α

Q_I, Q_{II} are rates of bankrupt, non-bankrupt firms in the sample

Equation (17) cannot be solved in a closed form. Thus the optimal cut-off point is not known in advanced for a hold-out sample. A proxy of the optimal cut-off point can be found using various methods. One way will be using the cut-off points discussed in the literature. Another way is to estimate the cut-off point using the training sample. To determine if an empirical cut-off point, estimated by using the training sample, is the optimal proxy for the one in the holdout sample, we run random selection sample 30 times. The training sample is used to evaluate prior 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, 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 reported in Table 4.

Table 4 illustrates results using various cut-off points. The first column contains the ratios of the cost of type I errors 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 a robust check. The results suggest that the empirical optimal cut-off point calculated with the training sample is the best proxy for the true optimal cut-off point for the holdout sample since it dominates most of the specifications of type I costs over type II costs by producing the least cost of misclassification. Such results are consistent for all three models. Following the estimated optimal cut-off point, the industry bankruptcy rate (0.41%) is considered as the second best cut-off point. It is not appropriate to use it when the ratio of type I cost and type II cost is low, since most of the companies will be classified as bankrupt, reducing the cost of type I error, while simultaneously increasing the 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.

Table 4. Random Selection and Test Results

$\frac{C_I}{C_{II}}$	α				α				α			
	0.41%	3.8%	50%	opti	0.41%	3.8%	50%	opti	0.41%	3.8%	50%	opti
	<i>Bayesian Model</i>				<i>Hazard Model</i>				<i>Mixed Logit Model</i>			
1	0.5274	0.2332	0.0913	0.0776*	0.4157	0.0828	0.0713	0.0660*	0.2621	0.0676*	0.0744	0.0668*
10	0.5438	0.3228	0.4595	0.2977*	0.4450	0.3337	0.6982	0.2672*	0.3090	0.3982	0.7263	0.2578*
20	0.5620	0.4224*	0.8687	0.4345*	0.4776	0.6125	1.3948	0.3591*	0.3612*	0.7655	1.4506	0.3633*
35	0.5893*	0.5718*	1.4824	0.6289	0.5264	1.0306	2.4396	0.4340*	0.4395*	1.3165	2.5371	0.4504*
50	0.6166	0.7212	2.0962	0.5771*	0.5752	1.4488	3.4845	0.5049*	0.5177*	1.8675	3.6236	0.5320*
75	0.6621*	0.9702	3.1191	0.6535*	0.6565*	2.1457	5.2259	0.6332*	0.6481*	2.7858	5.4345	0.6896*
100	0.7076*	1.2192	4.1420	0.7169*	0.7379*	2.8427	6.9674	0.7381*	0.7785*	3.7040	7.2453	0.7677*

1. "opti" is the estimated optimal cut-off point evaluated with training sample.

2. * indicates that the cost of misclassification is statistically least in one C_I/C_{II} ratio under 5% significant level. For certain 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 opti are not significantly different while they both are significantly less than the other two costs under cut-off point 0.41% and 50%.

5.3 Randomly Selected Samples

As shown in section 5.1, different choices of cut-off points can lead to different conclusions about the comparative power of prediction of various models. As a result, tests should be conducted under conditions that are free of the distortion posed by an arbitrary cut-off point while maintaining the true characteristics of the models. The results in section 5.2, suggest that the cut-off points estimated from the training sample are more preferable than the other specifications. Therefore we use this empirical cut-off point to compare the three new models. The 30 random samples in section 5.2 are used again. Pair-sample t-tests are performed to generate the total costs of misclassification for each pair of models. The results are shown in Table 5 which is in fact the re-production of Table 4 using the optimal cut-off point.

Table 5. Costs of Misclassification

C_I/C_{II}	Bayesian	Hazard	Mixed Logit
1	0.0776	0.0660*	0.0668*
10	0.2977	0.2672*	0.2578*
20	0.4345	0.3591*	0.3633*
35	0.6289	0.4340*	0.4504*
50	0.5771	0.5049*	0.5320*
75	0.6535*	0.6332*	0.6896*
100	0.7169*	0.7381*	0.7677*

“*” indicates the least cost at 5% significant level (two or three “*” in one row indicates that they are not statistically different)

Hazard and Mixed Logit models are effectively no different in terms of predictive power for all specification of type I and type II cost. When the ratio of the cost of type I errors to cost of type II errors reaches a very high level, the Bayesian model is comparable to the other two.

A shortcoming of randomly selected samples can result in situations where ex post results are used to explain and predict prior events, thus creating a noisy sample.

5.4 Samples in Different Business Cycles and Sub Cycles.

Since random selection produces results that may be tarnished with the noisy sample, we conduct our analysis around the 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 the preceding cycle and the cost of misclassification is calculated with the subsequent cycle. Since the first business cycle is short and sample size is fairly small, we combine the first two cycles. Table 6 reports the results on business cycle. Table 7 reports the results on sub cycles using the same methodology.

Table 6. Results Estimated across Business Cycles

	Bayesian		Hazard		Mixed Logit	
PANEL A						
	<u>1st & 2nd Cycle → 3rd Cycle</u>					
C_I/C_{II}	opti	cost	opti	cost	opti	cost
1	5.452403	0.143556	0.029408	0.089796	0.075185	0.117455
10	0.011548	0.379282	0.000432	0.36898	0.009964	0.279772
20	0.011548	0.403752	0.000432	0.409796	0.004163	0.355628
35	0.011548	0.440457	0.000432	0.47102	0.003113	0.458401
50	0.011548	0.477162	0.000432	0.532245	0.003113	0.519576
75	0.011548	0.538336	0.000136	0.761633	0.003113	0.621533
100	9.67e-05	0.723491	0.000136	0.863673	0.003113	0.723491
PANEL B						
	<u>3rd Cycle → 4th Cycle</u>					
C_I/C_{II}	opti	cost	opti	cost	opti	cost
1	1.501003	0.071558	0.274224	0.059675	0.422968	0.059783
10	0.189025	0.279891	0.012941	0.203436	0.080687	0.354167
20	0.028605	0.508152	0.008952	0.274864	0.080687	0.499094
35	0.023976	0.557971	0.007701	0.37613	0.061825	0.571558
50	0.023976	0.598732	0.006077	0.468354	0.029012	0.541667
75	0.023976	0.666667	0.006077	0.581374	0.029012	0.609601
100	0.023976	0.734601	0.006077	0.694394	0.029012	0.677536

Numbers in bold-face type are the lowest cost within three models for each C_I/C_{II}

Table 7. Results Estimated across Sub Cycles

		Bayesian		Hazard		Mixed Logit	
PANEL A							
<u>6th sub cycle → 7th sub cycle</u>							
C_I/C_{II}	opti	cost	opti	cost	opti	cost	
1	0.403524	0.246377	0.263798	0.267606	0.370651	0.173913	
10	0.196552	1.173913	0.015411	0.535211	0.113745	0.898551	
20	0.082942	1.797101	0.008594	0.690141	0.062985	0.347826	
35	0.019332	0.594203	0.007341	1.126761	0.062985	0.565217	
50	0.019332	0.811594	0.005905	1.549296	0.028562	0.869565	
75	0.019332	1.173913	0.005905	2.253521	0.028562	1.231884	
100	0.019332	1.536232	0.005905	2.957746	0.028562	1.594203	
PANEL B							
<u>7th sub cycle → 8th sub cycle</u>							
C_I/C_{II}	opti	cost	opti	cost	opti	cost	
1	0.767288	0.115942	0.000482	0.100483	0.033753	0.24058	
10	0.038165	0.677295	0.000303	0.206763	0.000329	0.392271	
20	0.038165	0.686957	0.000303	0.274396	0.000329	0.44058	
35	0.038165	0.701449	0.000303	0.375845	0.000329	0.513043	
50	0.038165	0.715942	0.000303	0.477295	0.000329	0.585507	
75	0.038165	0.740097	0.000303	0.646377	0.000329	0.70628	
100	0.038165	0.764251	0.000303	0.815459	0.000329	0.827053	

The two tables indicate that the optimal cut-off point is decreasing as cost of type I errors increases. However, the decrease is mitigated by the rise in type II error. From the tables, it can be seen that as the ratio of C_I/C_{II} is high, this indicates the cut-off point decrease to a very small number.

With respect to total cost, the results are in contradiction to those reported in randomly selected samples. Panel A of Table 6 shows that Bayesian model outperforms the others except when the ratio of C_I/C_{II} is lower than 35. The Hazard model generates less misclassification cost overall as shown in Panel B of Table 6. In general, the Mixed logit model underperforms the others. Such results are consistent for tests on sub cycles as shown in Table 7. The Hazard model dominates most of the specification of the ratio of C_I/C_{II} , however the Bayesian model has better performance when the ratio of C_I/C_{II} is high.

Based the results obtained from randomly selected samples and cross-business-cycle

samples, the Hazard model has the highest predictive power. The predictive power of the Bayesian model in higher range of type I costs is, in general, higher than the other two. This conclusion is consistent with random sample and across business-cycle.

Figures 4 and 5, which utilize randomly selected sample, provide the estimated fitted probability for firms with bankrupt or non-bankrupt status. They demonstrate the outperformance of the Bayesian model when type I costs are high. In both figures, the horizontal axis represents fitted probability and the vertical axis represents the number of firms.

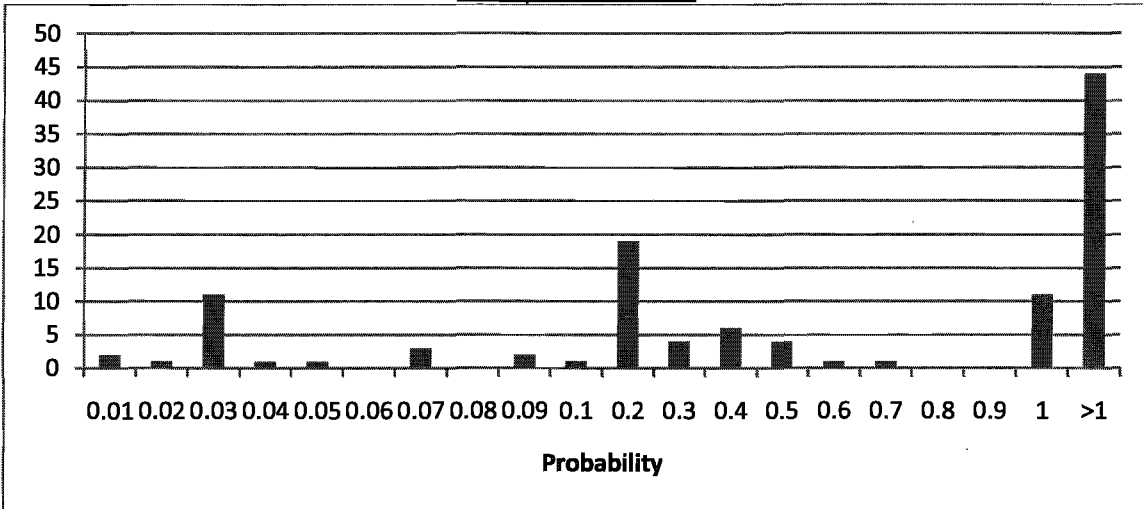
As shown in Figure 4, the patterns are quite similar between the Hazard and Mixed Logit models, whose fitted probability mass is between 0 to 0.1. However, for the Bayesian model, a majority of firms have fitted probabilities higher than 1.

Figure 5 shows that there are more than 1000 non-bankrupt firms have fitted probability between 0 to 0.01 for the Hazard and Mixed Logit models. For the Bayesian model, only 660 firms have fitted probabilities in the lowest range.

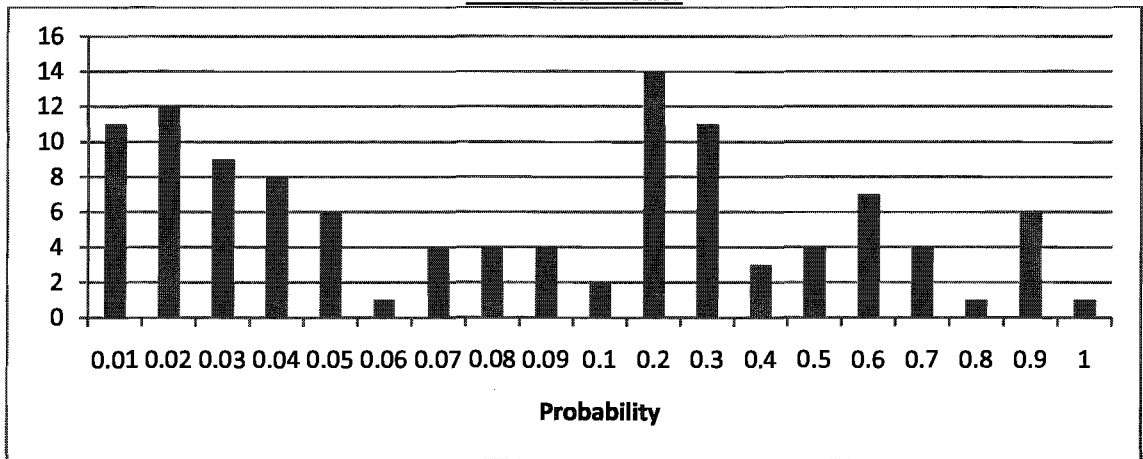
Figures 4 and 5 show two main differences between the Bayesian model and the Hazard and Mixed Logit models: 1. In general, the fitted probabilities generated from the Hazard and Mixed Logit models are lower than the ones from the Bayesian model; 2. The fitted probabilities generated from the Hazard and Mixed Logit models are more concentrated in lowest range. When the type I costs are high, the empirical cut-off points will be forced to be a very small number. If the cut-off points are small, for the Hazard and Mixed Logit models, more firms need to be classified, because most of the firms' fitted probabilities are concentrated within the lowest range. This fact will result in either higher total costs of type I errors or total costs of type II errors.

Figure 4. Bankrupt firms' fitted Probabilities

a. Bayesian Model



b. Hazard Model



c. Mixed Logit Model

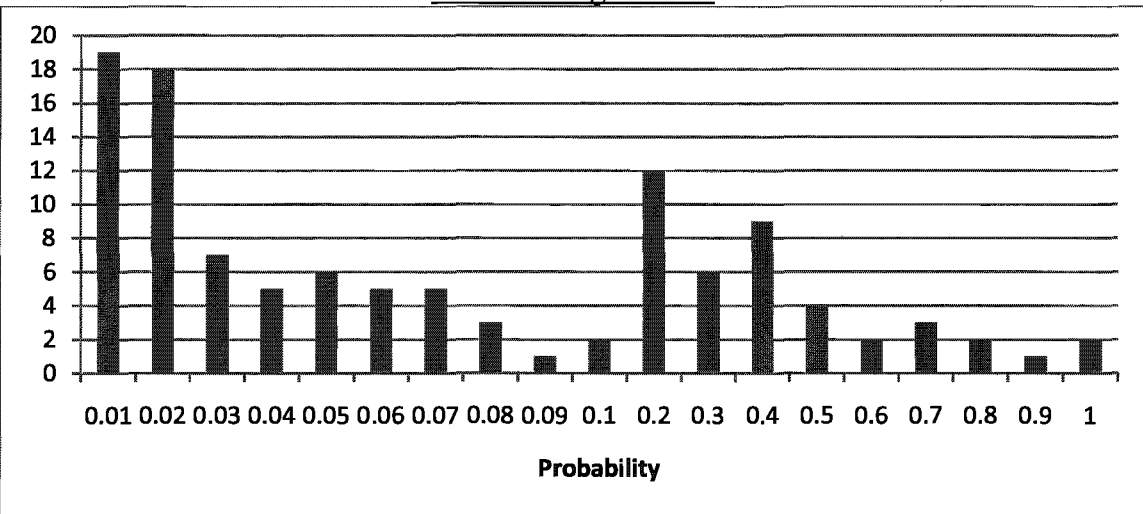
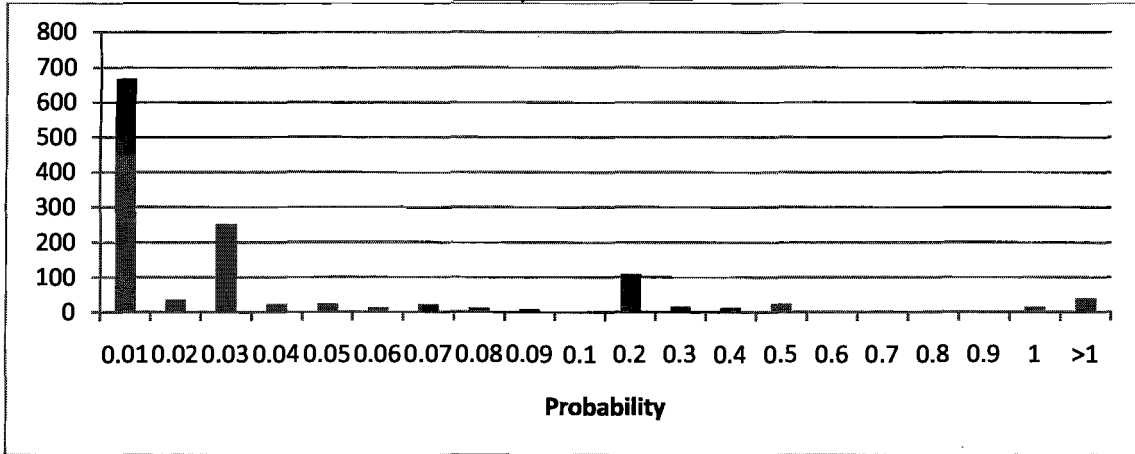
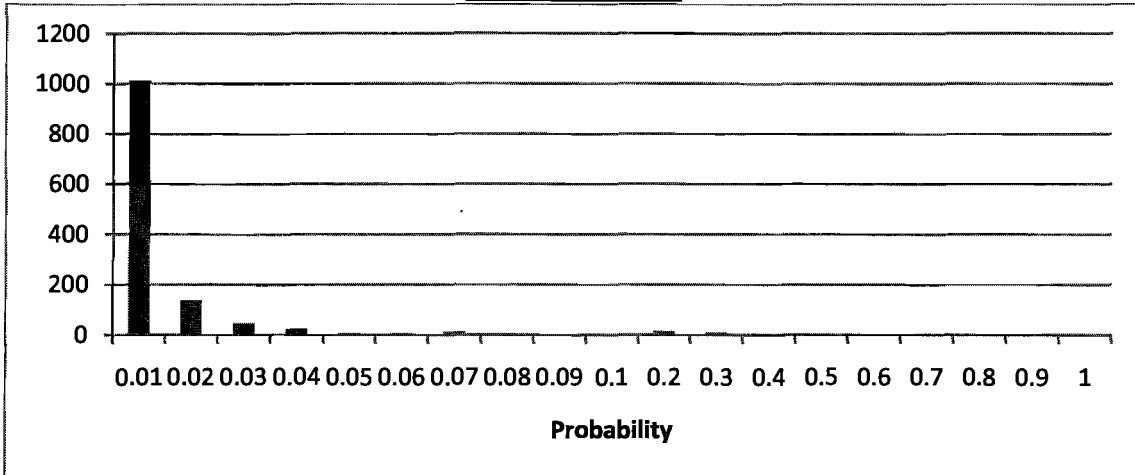


Figure 5. Non-Bankrupt firms' fitted Probabilities

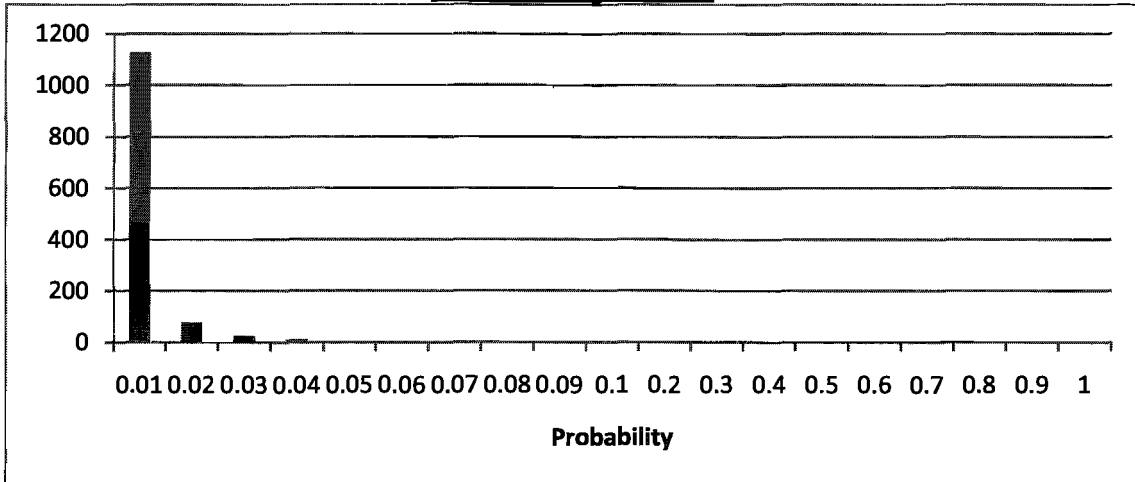
a. Bayesian Model



b. Hazard Model



c. Mixed Logit Model



6. Conclusion

Three models- Bayesian, Hazard and Mixed Logit- are currently applied to bankruptcy prediction. All three models are theoretically attractive. This study compared their empirical performance on predicting corporate bankruptcy. To make fair comparisons, factors that may unreasonably diminish or inflate predictive accuracy should be controlled or adjusted. In addition to “backcasting”, choice-based sample bias, change of financial ratio etc as identified by previous studies, it was found that the selection of cut-off points and sampling procedure influences predictive results.

To reduce the level of arbitrariness in selecting a cut-off point, we use the optimal cut-off point estimated by using the training sample as a proxy for the true optimal cut-off point in the holdout sample. We show that this proxy results in least cost of misclassification when compared to the cut-off points used in previous studies. Furthermore using this empirical optimal cut-off point, the results indicate that the Hazard and Mixed Logit models are statistically superior to the Bayesian model based on randomly selected samples when the ratio of the cost of the type I errors and the cost of the type II errors is low.

When the ratio of the cost of a type I error and the cost of a type II error is high, the Bayesian model appears to be equivalent to the other two models in randomly selected sample. Furthermore, with the cross-business-cycle sample and the cross-sub-cycle sample, the results indicate that the Bayesian model dominates the other two models when the ratio of Type I and II costs is high. Taking into account for results generated from both sampling methods, the Hazard model performs relatively better in general. However, the higher predictive power of the Bayesian model, when the ratio of the cost of Type I errors to the cost of Type II errors is high, is consistent across all sampling methods.

Such an advantage of the Bayesian model may make it more attractive in the current

economic environment where many large firms have come under financial distress, including such cases as Enron, Lehman and GM. These companies were financed with a 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 the credit spread of such large firms 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 study leaves several issues in question; they will be the subject of future studies. Unlike this thesis, where we took a methodological approach which focuses on the sensitivity of the performance of the Bayesian, Hazard and Mixed Logit models, one can use an empirical approach to further analyze which model is sensitive to what factors or what factors may influence which model the most e.g. (Beaver et al 2005; Grice and Dugan 2001). In addition, rather than using a wide specification of cost, research such as Weiss and Capkun (2004), develop model that more accurately estimates the cost of both errors on an individual level helping to better compare bankruptcy prediction models.

Canada already has, and the United States will probably soon adopt the International Financial Reporting Standards (IFRS). Thus, using the Bayesian, Hazard and Mixed Logit models, to study how the adoption of these standards will affect accounting numbers to predict bankruptcy and to judge the capacity of the IFRS to produce more relevant accounting numbers.

Finally, two major differences characterize US and Canadian markets: the level of litigation risk and the standard setting approach. An interesting study will be to take an international perspective and determine the differential impact of financial ratios and from a qualitative perspective, corporate governance quality in predicting bankruptcies.

7. Reference

- Agarwal, V., & Taffler, R. (2008). Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models. *Journal of Banking and Finance* , 32, 1541-1551.
- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* , 23, 589-609.
- Altman, E. I. (2007). Revisiting Credit Scoring Models in a Basel 2 Environment. Lecture at National Taiwan University. Available from Internet: <www.fin.ntu.edu.tw/~hwangdar/94emba19.ppt> .
- Altman, E. I., & McGough, T. P. (1974). Evaluation of A Company as A Going Concern. *The Journal of Accountancy* , (Dec.), 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, 29-54.
- 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, 505-529.
- 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.
- Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research* , 4, 71-111.
- 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, 93-122.
- 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, 267-284.
- Bellovary, J. L., Giacomino, D., & Akers, M. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education* , Winter 33.
- Chava, S., & R. Jarrow. (2004). Bankruptcy prediction with industry effect, market

- versus accounting variables, and reduced form of credit risk models. *Review of Finance* 8 (4): 537-569.
- 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, 151-166.
- Hensher, D. A., & S. Jones. (2007). Forecasting corporate bankruptcy: optimizing the performance of the mixed logit model. *Abacus*. 43 (3): 241-264.
- Hillegeist, S., D. Cram, E. Keating, & K. Lundstedt. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*. 9 (1):5-34.
- 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, 409-431.
- Hsieh, S. J. (1993). A Note on The Optimal Cut-off point Point in Bankruptcy Prediction Models. *Journal of Business Finance and Accounting*, 20 (3).
- Hull, J., & White, A. (2000). Valuing Credit Default Swaps I: No Counterparty Default Risk. *Journal of Derivatives*, 8, 29-40.
- Johnson, R. A., & D. W. Wichern, *Applied Multivariate Statistical Analysis (sixth edition* (pp. 575-649). Pearson Education Inc.
- 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.
- Katz, S., Lilien, S., & Nelson, B. (1985). Stock Market Behavior around Bankruptcy Model Distress and Recovery Predictions. *Financial Analysts Journal*, January-February, 70-73.
- Martin, D. (1977). Early Warning of Bank Failure. *Journal of Banking and Finance*, 1, 249-276.
- McKee, T. E. (2003). Rough Sets Bankruptcy Prediction Models versus Auditor Signaling Rates. *Journal of Forecasting*, 22, 569-586.
- 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.

- Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance* , 29, 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., & C.M. Molinero. (2004). Predicting Corporate Failure in the UK: A Multidimensional Scaling Approach. *Journal of Business Finance and Accounting*. 31 (5-6): 677-710.
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* , 18 (1), 109-131.
- Platt, H. D., & Platt, M. B. (Summer 2002). Predicting Corporate Financial Distress: Reflections on Choice-Based Sample Bias. *Journal of Economics and Finance* , 26 (2), 184-199.
- Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques-A review. *European Journal of Operational Research* , 180, 1-28.
- 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.
- Robertson, J., & Mills, R. (1991). The Uses and Abuses of Corporate Prediction Models. *Management Accounting* , 20-22.
- Sarkar, S., & Sriram, R. S. (2001). Bayesian Models for Early Warning of Bank Failures. *Management Science* , 47 (11), 1457-1475.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business* , 74 (1).
- Sun, L., & Shenoy, P. P. (2007). Using Bayesian Networks for Bankruptcy Prediction: Some Methodological Issues. *European Journal of Operational Research* , 180 (2), 738-753.
- Tam, K. Y. (1991). Neural Network Models and the Prediction of Bank Bankruptcy. *OMEGA* , 19 (5), 429-445.
- Thomson, J. B. (1991). Predicting Bank failures in the 1980s. *Economic Review, Federal Reserve Bank of Cleveland* , 27, 9-20.

- Verikas, A., Kalsyte, Z., Bacauskiene, M., & Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques. *Soft Computing* , 14, 995–1010.
- Webster, E., and D. B. Thornton. 2005. Earnings quality under rules-versus principles-based accounting standards: A test of the skinner hypothesis. *Canadian Accounting Perspectives*. 4 (2): 167-192.
- Weiss, L. A., & Capkun, V. (2004, Working Paper). The Impact of Incorporating the Cost of Errors into Bankruptcy Prediction Models.
- West, R. C. (1985). A Factor-Analytic Approach to Bank Condition. *Journal of Banking and Finance* , 9, 253-266.
- Whalen, G., & Thomson, J. B. (1988). Using Financial Data to Identify Changes in Bank Condition. *Economic Review, Federal Reserve Bank of Cleveland* , 24, 17-26.
- Wiginton, J. C. (Sep. 1980). A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior. *Journal of Financial and Quantitative Analysis* , 15 (3).
- Wilson, R. L., & Sharda, R. (1994). Bankruptcy Prediction Using Neural Networks. *Decision Support Systems* , 11, 545-557.
- Zavgren, C. (1983). The Prediction of Corporate failure: The State of The art. *Journal of Accounting Literature* , 2, 1-37.
- Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* , 22, 59-82.