國 立 交 通 大 學

管理科學系

博 士 論 文

No. 023

預測公司破產事件之研究

On Bankruptcy Prediction

研 究 生：黃瑞卿

指導教授：李昭勝 教授

中 華 民 國 九 十 五 年 十 月

國 立 交 通 大 學

管理科學系

博 士 論 文

No. 023

預測公司破產事件之研究

On Bankruptcy Prediction

研 究 生：黃瑞卿

研究指導委員會：王耀德 教授

許和鈞 教授

謝國文 教授

指導教授：李昭勝 教授

中 華 民 國 九 十 五 年 十 月

預測公司破產事件之研究

On Bankruptcy Prediction

研 究 生：黃瑞卿　　Student：Ruey-Ching Hwang

指導教授：李昭勝　　Advisor：Jack C. Lee

國 立 交 通 大 學

管理科學系

博 士 論 文

A Thesis

Submitted to Department of Management Science

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Management

October 2006

Hsin-Chu, Taiwan, Republic of China

中華民國九十五年十月

國立交通大學
研究所博士班

論文口試委員會審定書

本校　管 理 科 學 系　博 士 班　黃 瑞 卿　君

所提論文　預測公司破產事件之研究

合於博士資格水準、業經本委員會評審認可。

口試委員：

指導教授：

系主任：　　教授

中華民國　九十五　年　十　月　二十　日

On Bankruptcy Prediction

A DISSERTATION OF Ruey-Ching Hwang
WAS ACCEPTED AS PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Committee

Dissertation advisor

Director
Department of Management Science

National Chiao Tung University

October 20, 2006

預測公司破產事件之研究

學生：黃瑞卿　　　　　　　　　　　　　　　　指導教授：李昭勝博士

國立交通大學管理科學系博士班

摘　　　要

本文使用半母數羅吉特模型（semiparametric logit model）建立一個公司破產事件的預測方法，並將之應用在追蹤性(prospective)或稱簡單隨機(simple random)資料，以及個案控制（case-control）或稱選擇性（choice-based）資料。我們使用區域概似方法（local likelihood approach）估計半母數羅吉特模型中未知參數，且研究這些估計式的漸近偏差量與變異數（asymptotic bias and variance)。我們證明當應用這個半母數羅吉特模型至前述兩種不同類型資料上，其所對應的破產預測方法是相同的。因此我們的預測方法可以直接應用到這兩種重要類型的資料。實證研究結果顯示，我們的預測方法較 Altman (1968)的區別分析模型（discriminant analysis model)、Ohlson（1980）的線性羅吉特模型（linear logit model)、以及 Merton (1974)與 Bharath and Shumway (2004) 的 KMV-Merton 模型等所建立的預測方法，能夠產生較小的樣本外誤差率（out-of-sample error rate)。

另外，本文使用離散型倖存模型（discrete-time survival model； Allison, 1982)，預測公司發生財務危機的機率。我們以最大概似法（maximum likelihood method)估計該模型的參數值，導出參數估計式的漸近常態分配(asymptotic normal distribution)，進而估計公司發生財務危機的機率。藉由此機率估計值，我們可建立財務危機預警模型，並用以分析及預測台灣股票上市公司發生財務危機的機率。實證研究結果顯示，本文所介紹的離散型倖存模型對公司財務危機的預測，比線性羅吉特模型，有更好的樣本外預測能力。

關鍵詞：個案控制資料、離散型倖存模型、區別分析模型、KMV-Merton 模型、線性羅吉特模型、追蹤性資料、半母數羅吉特模型、型 I 誤差率、型 II 誤差率。

On Bankruptcy Prediction

Student: Ruey-Ching Hwang Advisor: Dr. Jack C. Lee

Department of Management Science

National Chiao Tung University

ABSTRACT

Bankruptcy prediction methods based on a semiparametric logit model are proposed for prospective (simple random) and case-control (choice-based) data. The unknown quantities in the model are estimated by the local likelihood approach, and the resulting estimators are analyzed through their asymptotic biases and variances. Our semiparametric bankruptcy prediction methods using these two types of data are shown to be essentially equivalent. Thus our proposed prediction model can be directly applied to data sampled from the two important designs. Empirical studies demonstrate that our prediction method is more powerful than alternatives based on the discriminant analysis model (Altman 1968), the linear logit model (Ohlson 1980), and the KMV-Merton model (Merton 1974; Bharath and Shumway 2004), in the sense of yielding smaller out-of-sample error rates.

The discrete-time survival model (Allison 1982) is applied to predict the probability of financial distress. The maximum likelihood method is employed to estimate the values of parameters in the model. The resulting estimates are analyzed by their asymptotic normal distributions, and are used to estimate the probability of financial distress for each firm under study. Using such estimated probability, a strategy is developed to identify failing firms, and is applied to study the probability of financial distress for firms listed in Taiwan Stock Exchange. Empirical studies demonstrate that our strategy developed from the discrete-time survival model can yield more accurate out-of-sample forecasts than the alternative method based on the linear logit model in Ohlson (1980).

Keywords: case-control data, discrete-time survival model, discriminant analysis model, KMV-Merton model, linear logit model, prospective data, semiparametric logit model, type I error rate, type II error rate.

ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. Jack C. Lee, for his constant support, enthusiastic guidance, and helpful criticism in all parts of the research presented in this dissertation. I aslo wish to thank the other members of my committee, Dr. Chuang-Chang Chang, Dr. K.F. Cheng, Dr. Huimin Chung, Dr. Tsung-I Lin, Dr. Her-Jiun Sheu, Dr. G. Shieh, and Dr. Yau-De Wang for their valuable suggestions which substantially improve the research results.

I am grateful to the Department of Management Science and Graduate Institute of Finance for providing a good environment for study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION TO BANKRUPTCY PREDICTION METHODS

1.1 Introduction

Academics, practitioners and regulators have routinely used models to predict the bankruptcy of companies. For example, the discriminant analysis model (DAM) has been a popular technique for studying the financial health of a corporate; see Altman (1968). Other frequently referred models include the models by Ohlson (1980) and Zmijewski (1984). The former bankruptcy prediction method is based on a linear logit model (LLM). The latter, on the other hand, is based on a probit model. Grice and Dugan (2001) recently cautioned the routine application of these two probabilistic models of bankruptcy. Their study showed that using the prediction models to time periods and industries other than those used to develop the models may result in significant decline in prediction accuracies.

Bankruptcy prediction methods using other models or concepts include, for example, the recursive partition model (Frydman, Altman, and Kao 1985), expert systems (Messier and Hansen 1988), chaos theory (Lindsay and Campbell 1996), neural networks (Koh and Tan 1999), survival analysis (Lane, Looney, and Wansley 1986; Shumway 2001; Chava and Jarrow 2004), rough set theory (McKee 2003), KMV-Merton model (KMV; Merton 1974; Bharath and Shumway 2004; Vassalou and Xing 2004), and support vector machines (Härdle et al., 2006). Basically, these methods are more complicated in computation and interpretation than the above probabilistic models.

The bankruptcy prediction model in Ohlson (1980) postulates that the logit function of bankruptcy probability is a linear function of the predictors. Nine predictors were selected for developing his model because they appeared to be the ones most frequently mentioned in the literature. The main reason of using the LLM is due to its simplicity in computation and interpretation. There are many software packages having logistic regression capabilities, for example, BMDP, EGRET, GAUSS, GLIM, and SAS, etc.

Thus LLM can be easily updated or revised as long as there are new observations of the same predictors or new predictive variables available for analysis. For a detailed introduction of the LLM, see the monograph by Hosmer and Lemeshow (1989).

When appropriate, the LLM has definite advantages. For example, the corresponding inferential methods usually have nice efficiency properties. Also, the parameters generally have some physical meaning which makes them interpretable and of interest in their own right. If the assumed linear logit function is grossly in error, then the advantages of the LLM will not be realized. Thus, there are few benefits from using a poorly specified LLM. See the discussion and Figure 2 of Härdle, Moro, and Schäfer (2006). Their results show that the relation between the bankruptcy probability and predictors, such as net income change and company size, may not be monotonic. The LLM is most appropriate when theory, past experience, or other sources are available that provide detailed knowledge about the data under study. Sometimes, based on previous experience, there are reasons for modelling the logit function of bankruptcy probability as a particular parametric function of predictors, which may not be linear. However, a general drawback of such parametric modelling is that if one chooses a parametric family that is not of appropriate form, at least approximately, then there is still a danger of reaching erroneous inference.

The first focus of this dissertation is to consider a robust method, against misspecification of the parametric logit model relation, by introducing a semiparametric logit model (SLM; Zhao, Kristal, and White 1996) for predicting bankruptcy. This model is basically very similar to the LLM, except that some unspecified function replaces the linear function to model the relation between the predictors and the logit function of bankruptcy probability. Thus, clearly, the SLM is much more general and flexible in predicting the bankruptcy of a firm. Since the SLM is developed without assuming a parametric form for the logit function, there is some loss in the interpretability and efficiency of estimators obtained in this fashion. In contrast to physics or engineering, it may not be often appropriate to give a specific functional relationship between the probability of bankruptcy and the predictors in finance fields. Härdle, Moro, and Schäfer (2006) also propose a flexible but fully nonparametric approach for predicting bankruptcy. They use support vector machines to generate nonlinear score function of predictors, and then employ nonparametric technique to map scores into bankruptcy probabilities. Their work presents a new trend in bankruptcy analysis.

On the other hand, there is another potential pitfall of the LLM. It is static in nature, since it uses only one set of predictor values collected at a specific time point for each firm. The static model is generally not appropriate for predicting bankruptcy because it ignores both facts that the characteristics of firms change through time as well as bankruptcy does not often occur. For more discussions of the drawback to the static model, see Shumway (2001).

To avoid the drawback to the static model, Shumway (2001) applies the idea of discrete-time survival analysis (Cox and Oakes 1984) to develop the so-called discrete-time hazard model. The model has the advantage of using all available historical information to determine each firm's bankruptcy risk at each point in time, hence it is a dynamic forecasting model. It is also called the discrete-time survival model (DSM) in Allison (1982). The values of parameters in Shumway's dynamic prediction model are estimated by using the same approach as those in the multiperiod logit model (Pagano, Panetta, and Zingales 1998). However, theoretically, the multiperiod logit model assumes the predictor values collected for each firm at all time points are independent. Clearly, such predictor values are dependent, and the assumption does not hold in practice. Thus, asymptotic properties of the resulting estimates of parameters in Shumway's dynamic prediction model can not be obtained from the multiperiod logit model.

The second focus of this dissertation is to employ directly the DSM to predict bankruptcy, and ignore the estimation procedure of the multiperiod logit model. The values of parameters in the DSM are estimated by the maximum likelihood method. The advantages of direct employment of the DSM include, for example, using all available historical information to determine each firm's bankruptcy risk at each point in time, assuming the predictor values of each firm are dependent. Hence, it is more general and flexible for the DSM to predict bankruptcy. The DSM has been successfully applied in many fields including, for example, social science (Allison 1982), econometrics (Lancaster 1990), education (Singer and Willett 1993), and biostatistics (Klein and Moeschberger 1997).

The rest of this chapter is organized as follows. In Section 1.2, three important sampling schemes including the prospective (simple random), the case-control (choice-based), and the discrete-time survival data for bankruptcy prediction study are described. The data of the first two types are for static forecasting models, the LLM, the SLM, the KMV, and the DAM, and the data of the third type are for the dynamic forecasting model, the DSM. In Sections 1.3-1.7, five bankruptcy prediction models, the LLM, the SLM, the KMV, the DAM, and the DSM, are introduced respectively. In Section 1.8, bankruptcy prediction devices based on the five bankruptcy prediction models are presented. Finally, Section 1.9 contains a brief summary of the results obtained.

1.2 Three Sampling Schemes

In this section, the formulation of the data used in this dissertation for bankruptcy prediction study will be given. Three types of data including the prospective, the case-control, and the discrete-time survival data will be described in sequence.

Most bankruptcy prediction methods were developed on training samples. Usually, the training sample consists of the data of n companies collected for some time period by a simple random sampling scheme from the distribution of (X, Z). For the i-th company, $i = 1, \cdots, n$, we observe values (Y_i, x_i, z_i), where $Y_i = 1$ indicating that the i-th company is in the state of bankruptcy and 0, otherwise, and $x_i = (x_{i1}, \cdots, x_{id})^T$ and $z_i = (z_{i1}, \cdots, z_{iq})^T$ are values of the vectors of explanatory variables (X, Z) used to forecast failure. Here X and Z are the d-dimensional continuous and the q-dimensional discrete variables, respectively, and the upper index T stands for the transpose of a matrix. Hence we have the prospective sample

$$(Y_i,\ x_i,\ z_i), \quad i = 1,\ \cdots,\ n.$$

For example, in Ohlson (1980), there were 9 financial variables being used for developing his bankruptcy prediction model. Among these explanatory variables, 7 $(= d)$ of them are continuous variables and 2 $(= q)$ are binary variables.

On the other hand, the case-control data for bankruptcy prediction are composed of two simple random samples. One is selected from the population of bankrupt companies, and called the case sample. The other is selected from the population of nonbankrupt companies, and called the control sample. An important special case of the case-control study is the stratified (matched) case-control study. In the latter study, the number of cases (bankrupt companies) and controls (nonbankrupt companies) need not to be constant across strata, but most matched designs include one case and one to five controls per stratum and are thus referred to as 1-M matched designs. For a detailed introduction of the (matched) case-control data, see the monograph by Hosmer and Lemeshow (1989).

According to the case-control sampling, the case-control data are composed of a random sample of nonbankrupt companies of n_0 observations (controls), say $(x_1,\ z_1)$, $\cdots$, $(x_{n_0},\ z_{n_0})$, from the conditional distribution of $(X,\ Z)$ given $Y = 0$, and an independent random sample of bankrupt companies of n_1 observations (cases), say $(x_{n_0+1},\ z_{n_0+1})$, $\cdots$, $(x_n,\ z_n)$, where $n = n_0 + n_1$, from the conditional distribution of $(X,\ Z)$ given $Y = 1$. Here $Y_i = 1$ indicating that the i-th company is in the state of bankruptcy and 0, otherwise. Hence we have the case-control sample

$$(Y_i,\ x_i,\ z_i), \quad i = 1, \cdots, n, \quad \text{where} \quad Y_i = 0 \quad \text{for} \quad i \le n_0, \quad \text{and} \quad 1 \quad \text{for} \quad i > n_0.$$

We now close this section by giving the formulation of the discrete-time survival data. The data are generated by three steps. Firstly, both the sampling period and the sampling criteria are selected. For example, the sampling period may be taken as the one starting from January of the year 1981 to December of the year 1999, and the sampling criteria may be defined as those firms starting to be listed in Taiwan Stock Exchange during the sampling period. Secondly, n companies satisfying the sampling criteria are selected. Finally, all available historical information occurred at the discrete time points in the sampling period of n selected companies are collected. Hence we have the discrete-time survival data

$$(t_i,\ Y_i,\ x_{i,1},\ \cdots,\ x_{i,t_i},\ z_{i,1},\ \cdots,\ z_{i,t_i}),\ \text{ for }\ i = 1,\ \cdots,\ n.$$

Here $t_i \in \{1, 2, \cdots, m\}$ denotes the duration time of the i-th company in the sampling period, and m is a positive integer standing for the length of the sampling period. Also, at the duration time t_i, $Y_i = 0$ indicates that the i-th company is nonbankrupt, and 1 the i-th company is bankrupt. Further, $x_{i,j}$ and $z_{i,j}$ are values of the d-dimensional continuous and q-dimensional discrete explanatory variables X and Z collected at the duration time j, respectively in each case, for each $j = 1, \cdots, t_i$ and for the i-th company.

1.3 The LLM

In this section, the formulation of the LLM using the prospective sample as well as that using the case-control sample for predicting bankruptcy will be introduced.

Given the prospective sample $(Y_i,\ x_i,\ z_i)$, $i = 1, \cdots, n$, the LLM is defined by assuming the bankruptcy probability for the company with the predictor values $(X, Z) = (x, z)$ to be

$$p(Y = 1 \mid X = x,\ Z = z) = \frac{\exp(\alpha + \beta\ x + \theta\ z)}{1 + \exp(\alpha + \beta\ x + \theta\ z)}, \tag{1}$$

or written in the form of the logit function of bankruptcy probability

$$\text{logit}\{p(Y = 1 \mid X = x, Z = z)\} = \log\left\{\frac{p(Y = 1 \mid X = x, Z = z)}{1 - p(Y = 1 \mid X = x, Z = z)}\right\} = \alpha + \beta x + \theta z.$$

Here α, β, and θ are 1×1, $1 \times d$, and $1 \times q$ vectors of logistic parameters, respectively. For the company with predictor values (x_0, z_0), its predicted bankruptcy probability

$$\hat{p}(Y = 1 \mid X = x_0,\ Z = z_0) = \frac{\exp(\hat{\alpha} + \hat{\beta}\ x_0 + \hat{\theta}\ z_0)}{1 + \exp(\hat{\alpha} + \hat{\beta}\ x_0 + \hat{\theta}\ z_0)} \tag{2}$$

is the logistic distribution evaluated at the predicted score $\hat{\alpha} + \hat{\beta}\ x_0 + \hat{\theta}\ z_0$. Here $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ are maximum likelihood estimates for α, β, and θ, respectively, based on the prospective sample from the LLM (1).

The maximum likelihood approach for producing $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ in (2) is now described. For this, set $\eta = (\alpha,\ \beta,\ \theta)^T$. Using the prospective sample from the LLM (1), the log-likelihood function of η is

$$\ell_{LLM}(\eta) = \sum_{i=1}^{n} \left[Y_i\ (\alpha + \beta\ x_i + \theta\ z_i) - \log\{1 + \exp(\alpha + \beta\ x_i + \theta\ z_i)\} \right].$$

Then $(\hat{\alpha},\ \hat{\beta},\ \hat{\theta})^T$ may be taken as the solution of the normal equations

$$\frac{\partial \ell_{LLM}(\eta)}{\partial \eta} = \sum_{i=1}^{n} \left[Y_i - \frac{\exp(\alpha + \beta\ x_i + \theta\ z_i)}{1 + \exp(\alpha + \beta\ x_i + \theta\ z_i)} \right] \begin{bmatrix} 1 \\ x_i \\ z_i \end{bmatrix} = 0. \tag{3}$$

Due to the consistency of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ (Hosmer and Lemeshow, 1989), the predicted bankruptcy probability $\frac{\exp(\hat{\alpha}+\hat{\beta}\ x_0+\hat{\theta}\ z_0)}{1+\exp(\hat{\alpha}+\hat{\beta}\ x_0+\hat{\theta}\ z_0)}$ in (2) does converge to the true bankruptcy probability $\frac{\exp(\alpha+\beta\ x_0+\theta\ z_0)}{1+\exp(\alpha+\beta\ x_0+\theta\ z_0)}$ in (1) for the company with predictor values $(x_0,\ z_0)$. By this fact, it will be used in Section 1.8 to construct a bankruptcy prediction device for prospective data from the LLM (1).

On the other hand, using the case-control sample from the LLM (1) and treating the sample as if it was a prospective sample from the LLM (1), the maximum likelihood estimates for logistic parameters α, β, and θ are now given. Applying the case-control sample $(Y_i = 0,\ x_i,\ z_i)$ for $i \leq n_0$ and $(Y_i = 1,\ x_i,\ z_i)$ for $i > n_0$ from the LLM (1) to the normal equations (3), Prentice and Pyke (1979) show that the resulting maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\theta}$ of logistic parameters α, β, and θ, respectively, converge to their true values, except the intercept estimate $\hat{\alpha}$, as both sample sizes of control and case data become large. This intercept estimate $\hat{\alpha}$ approaches the quantity $\alpha + c^*$, where

$$c^* = \log\{p(Y=0) \;/\; p(Y=1)\} \;+\; \log(n_1 \;/\; n_0).$$

Using the case-control sample, inferences about the constant c^* are not possible since such data generally provide no information about the population frequency of bankrupt companies. Unfortunately, due to the inconsistency of the intercept estimate $\hat{\alpha}$ and the fact that the unknown quantity c^* is generally not equal to 0, the resulting predicted bankruptcy probability $\frac{\exp(\hat{\alpha}+\hat{\beta}\; x_0+\hat{\theta}\; z_0)}{1+\exp(\hat{\alpha}+\hat{\beta}\; x_0+\hat{\theta}\; z_0)}$, obtained by plugging all these estimates of coefficients into (2), does not converge to the true bankruptcy probability $\frac{\exp(\alpha+\beta\; x_0+\theta\; z_0)}{1+\exp(\alpha+\beta\; x_0+\theta\; z_0)}$ in (1), but approaches $\frac{\exp(\alpha+c^*+\beta\; x_0+\theta\; z_0)}{1+\exp(\alpha+c^*+\beta\; x_0+\theta\; z_0)}$, for the company with predictor values $(x_0,\; z_0)$. This is the major difference between applying the LLM to the prospective sample and to the case-control sample. Although the predicted bankruptcy probability derived by the case-control sample from the LLM (1) does not estimate the true bankruptcy probability, we will discuss in Section 1.8 that it still can be used to develop a bankruptcy prediction device for case-control data from the LLM (1).

The main advantage of the LLM lies in its simplicity of computation and interpretation, but the model may not be efficient for the purpose of prediction. Sometimes, based on previous experience, there are reasons for modelling the logit function of bankruptcy probability as a particular parametric function of $(X,\; Z)$, which may not be linear. However, a general drawback of such parametric modelling is that if one chooses a parametric family that is not of appropriate form, at least approximately, then there is a danger of reaching erroneous prediction. The limitation of LLM can be overcome by removing the restriction that the logit function of $p(Y=1 \mid X=x,\; Z=z)$ belongs to a parametric family. In Section 1.4, we shall use some unspecified function $H(\cdot)$ to replace the linear function to model the relation between the continuous predictors and the logit function of bankruptcy probability.

1.4 The SLM

In this section, the formulation of the SLM using the prospective sample and that using the case-control sample will be given. The SLM is defined similarly to the LLM (1) by replacing the linear relationship $\alpha + \beta\ x$ of the continuous predictor X in the logit function of the LLM with an unknown function $H(x)$.

Given the prospective sample $(Y_i,\ x_i,\ z_i)$, $i = 1, \cdots, n$, the SLM is defined by assuming the bankruptcy probability for the company with the predictor values $(X, Z) = (x, z)$ to be

$$p(Y = 1 \mid X = x,\ Z = z) = \frac{\exp\{H(x) + \theta\ z\}}{1 + \exp\{H(x) + \theta\ z\}}, \tag{4}$$

or written in the form of the logit function of bankruptcy probability

$$\text{logit}\{p(Y = 1 \mid X = x, Z = z)\} = \log\left\{\frac{p(Y = 1 \mid X = x, Z = z)}{1 - p(Y = 1 \mid X = x, Z = z)}\right\} = H(x) + \theta z.$$

Here, we only assume $H(x)$ to be a smooth function of the value x of the continuous predictor X, otherwise, it is not specified. Also, θ is a $1 \times q$ vectors of logistic parameters, as it does in the LLM (1). Clearly, this is a very flexible prediction model. For the company with predictor values $(x_0,\ z_0)$, its predicted probability of bankruptcy is thus defined as

$$\hat{p}(Y = 1 \mid X = x_0,\ Z = z_0) = \frac{\exp\{\hat{H}(x_0) + \hat{\theta}\ z_0\}}{1 + \exp\{\hat{H}(x_0) + \hat{\theta}\ z_0\}}, \tag{5}$$

the logistic distribution evaluated at the predictive score $\hat{H}(x_0) + \hat{\theta}\ z_0$. Here $\hat{H}(x_0)$ and $\hat{\theta}$ are estimates derived by applying the local likelihood method to the prospective sample from the SLM (4).

The local likelihood approach for producing $\hat{H}(x_0)$ and $\hat{\theta}$ in (5) is now introduced. This approach is composed of three steps. In the first step, an initial local likelihood estimate $\hat{H}_1(x_0)$ of $H(x_0)$ is generated. There exists many methods for estimating $H(x_0)$. One of these methods with simple idea is the local likelihood method; see

Tibshirani and Hastie (1987). This method is to first choose a positive scalar constant b_θ, also called the bandwidth, and define a neighborhood of x_0 as

$$N(x_0; b_\theta) = \{t = (t_1, \cdots, t_d)^T : |t_j - x_{0j}| \leq b_\theta, \ \text{ for } \ j = 1, \ \cdots, \ d\},$$

where $x_0 = (x_{01}, \cdots, x_{0d})^T$. Then the idea of the local likelihood method is to apply both concepts of the weighted likelihood method using partial sample

$$S(x_0; b_\theta) = \{(Y_i, x_i, z_i) : x_i \in N(x_0; b_\theta), \ \text{ for } \ i = 1, \ \cdots, \ n\},$$

and the first order Taylor approximation

$$H(x_i) \approx H(x_0) + H^{(1)}(x_0)^T \ (x_i - x_0) \equiv \alpha + \beta \ (x_i - x_0),$$

for each $x_i \in N(x_0; \ b_\theta)$. Here the larger the value of b_θ, the larger the number of data points contained in $S(x_0; \ b_\theta)$. Also, the parameters α and β are 1×1 and $1 \times d$ vectors of parameters, respectively, as they are in the LLM. But, they now stand for the unknown quantities $H(x_0)$ and $H^{(1)}(x_0)^T$, respectively, and $H^{(1)}(x_0)$ is the $d \times 1$ vector of partial derivatives of $H(x_0)$.

Specifically, to produce $\hat{H}_1(x_0)$, a bankruptcy probability model developed by the above arguments

$$p(Y = 1 \mid X = x, \ Z = z) = \frac{\exp\{\alpha + \beta \ (x - x_0) + \theta \ z\}}{1 + \exp\{\alpha + \beta \ (x - x_0) + \theta \ z\}} \tag{6}$$

is imposed to the prospective sample $(Y_i, \ x_i, \ z_i)$, $i = 1, \cdots, n$, from the SLM (4) with $x_i \in N(x_0; \ b_\theta)$. Given the value of b_θ and the resulting bankruptcy probability model (6) for the prospective sample from the SLM (4) with $x_i \in N(x_0; \ b_\theta)$, the local log-likelihood function of $\eta = (\alpha,\, \beta,\, \theta)^T$ is defined by

$$\begin{aligned}\ell_{SLM}(\eta;\; x_0) \;\; &= \;\; \sum_{i=1}^{n} Y_i\, \{\alpha + \beta\, (x_i - x_0) + \theta\, z_i\}\, K_{b_\theta,i} \;\; - \\ &\quad\;\; \sum_{i=1}^{n} \log[1 + \exp\{\alpha + \beta\, (x_i - x_0) + \theta\, z_i\}]\, K_{b_\theta,i},\end{aligned}$$

where $K_{b_\theta,i} = K_{b_\theta}(x_i - x_0) = \prod_{j=1}^{d} K\{(x_{i,j} - x_{0,j})/b_\theta\}$. Here $K(\cdot)$ is called the kernel function, and is used to compute the weight assigned to the data. It is usually taken as a symmetric and unimodal probability density function over $[-1,\, 1]$. Hence it gives positive weight to the data inside the neighborhood sample $S(x_0; b_\theta)$ and weight 0 outside. The larger weights are given to data points with X values closer to x_0 and smaller weights to those with X values far from x_0. However, the results from the literature show that the choice of the density function $K(\cdot)$ is not very important in the local fitting. A popular choice of $K(\cdot)$ is the Epanechnikov kernel defined as

$$K(u) = (3/4)\,(1 - u^2)\, I(|u| \le 1);$$

see Wand and Jones (1995), due to its computational convenience and optimal performance (for example it minimizes mean square error among all nonnegative kernel functions).

Set the first element $\hat{\alpha}$ of the solution $\hat{\eta} = (\hat{\alpha},\, \hat{\beta},\, \hat{\theta})$ of the normal equations

$$\frac{\partial \ell_{SLM}(\eta; x_0)}{\partial \eta} = \sum_{i=1}^{n} \left[Y_i - \frac{\exp(\alpha + \beta\, (x_i - x_0) + \theta\, z_i)}{1 + \exp(\alpha + \beta\, (x_i - x_0) + \theta\, z_i)} \right] \begin{bmatrix} 1 \\ x_i - x_0 \\ z_i \end{bmatrix} K_{b_\theta,i} = 0 \quad (7)$$

as the initial local likelihood estimate $\hat{H}_1(x_0)$ of $H(x_0)$. By the same arguments for the consistency of the maximum likelihood estimate $\hat{\alpha}$ derived by (3) for the LLM (1), $\hat{H}_1(x_0)$ is a consistent estimate of $H(x_0)$. For this fact, see also Fan, Heckman, and Wand (1995).

Note that the concept of local inference is well established in regression analysis; see also Wand and Jones (1995). There are two major strategies considered in the local likelihood approach: using linear approximation (the first order Taylor approximation) for each $H(x_i)$ with $x_i \in N(x_0;\, b_\theta)$, and using the partial (local) sample $S(x_0;\, b_\theta)$ to derive the maximum local likelihood estimates. This method is directly analogous to the LLM, except that here we have used the concept of local fitting.

In the second step, the estimate $\hat{\theta}$ required in (5) is generated by applying the simple logistic regression analysis. To estimate the value of θ, we shall replace the unknown quantity $H(x_i)$ in the SLM (4) with its initial local likelihood estimate $\hat{H}_1(x_i)$, for each $i = 1, \cdots, n$, fit the bankruptcy probability by the resulting model

$$p(Y = 1 \mid X = x_i,\ Z = z_i) = \frac{\exp\{\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i\}}{1 + \exp\{\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i\}}, \qquad (8)$$

and use the prospective sample from the SLM (4) to maximize the corresponding pseudo profile log-likelihood function with respect to $\phi = (\alpha_0,\, \theta)^T$. Here α_0 is a normalizing constant which makes the bankruptcy probability function (8) be integrated to 1.

Specifically, using the bankruptcy probability model (8) and the prospective sample from the SLM (4), the pseudo profile log-likelihood function of $\phi = (\alpha_0,\, \theta)^T$ is

$$\hat{\ell}_{SLM}(\phi) = \sum_{i=1}^{n} \left[Y_i\ \{\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i\} - \log[1 + \exp\{\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i\}] \right].$$

Set $(\hat{\alpha}_0,\, \hat{\theta})^T$ as the solution of the normal equations

$$\frac{\partial \hat{\ell}_{SLM}(\phi)}{\partial \phi} = \sum_{i=1}^{n} \left[Y_i - \frac{\exp(\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i)}{1 + \exp(\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i)} \right] \begin{bmatrix} 1 \\ z_i \end{bmatrix} = 0. \qquad (9)$$

Hence the required estimate $\hat{\theta}$ of θ in (5) is obtained. By the results in Hosmer and Lemeshow (1989), the consistency of $\hat{\theta}$ for θ can be seen.

Finally, in the third step, the local likelihood estimate $\hat{H}(x_0)$ required in (5) is produced. To produce the value of $\hat{H}(x_0)$, follow the same arguments in the first step, replace the unknown quantity θ with $\hat{\theta}$ obtained in the second step, use the value of bandwidth b_H, fit the bankruptcy probability by the resulting model

$$p(Y = 1 \mid X = x_i,\ Z = z_i) = \frac{\exp\{\alpha^* + \beta\ (x_i - x_0) + \hat{\theta}\ z_i\}}{1 + \exp\{\alpha^* + \beta\ (x_i - x_0) + \hat{\theta}\ z_i\}} \tag{10}$$

for the prospective sample from the SLM (4) with $x_i \in N(x_0;\ b_H)$, and maximize the corresponding pseudo profile local log-likelihood function with respect to $\xi = (\alpha^*,\ \beta)^T$. Here α^* and β stand for $H(x_0) + \alpha_1$ and $H^{(1)}(x_0)^T$, respectively, where α_1 is a normalizing constant which makes the bankruptcy probability function (10) be integrated to 1.

Specifically, using the value of bandwidth b_H, the resulting bankruptcy probability model (10), and the prospective sample from the SLM (4), the pseudo profile local log-likelihood function of $\xi = (\alpha^*,\ \beta)^T$ is

$$\begin{aligned}\hat{\ell}_{SLM}(\xi;\ x_0) \quad = \quad & \sum_{i=1}^{n} Y_i\ \{\alpha^* + \beta\ (x_i - x_0) + \hat{\theta}\ z_i\}\ K_{b_H,i}\ - \\ & \sum_{i=1}^{n}\ \log[1 + \exp\{\alpha^* + \beta\ (x_i - x_0) + \hat{\theta}\ z_i\}]\ K_{b_H,i}.\end{aligned}$$

Set $(\hat{\alpha}^*,\ \hat{\beta})$ as the solution of the normal equations

$$\frac{\partial \hat{\ell}_{SLM}(\xi; x_0)}{\partial \xi} = \sum_{i=1}^{n} \left[Y_i - \frac{\exp(\alpha^* + \beta(x_i - x_0) + \hat{\theta} z_i)}{1 + \exp(\alpha^* + \beta(x_i - x_0) + \hat{\theta} z_i)} \right] \begin{bmatrix} 1 \\ x_i - x_0 \end{bmatrix} K_{b_H,i} = 0. \tag{11}$$

Combining the consistency of $\hat{\theta}$ and the consistency of the maximum likelihood estimates $(\hat{\alpha}^*,\ \hat{\beta})$ for $(\alpha^*,\ \beta)$, we see that the value of α_1 converges to 0, as the sample size of prospective data become large. Hence $\hat{\alpha}^*$ is a consistent estimate of $H(x_0)$, and the required estimate $\hat{H}(x_0)$ of $H(x_0)$ in (5) may be taken as $\hat{H}(x_0) = \hat{\alpha}^*$. For this fact, see also Fan, Heckman, and Wand (1995).

By the consistency of $\hat{H}(x_0)$ and $\hat{\theta}$, the corresponding predicted bankruptcy probability $\frac{\exp\{\hat{H}(x_0)+\hat{\theta}\ z_0\}}{1+\exp\{\hat{H}(x_0)+\hat{\theta}\ z_0\}}$ in (5) approaches the true bankruptcy probability $\frac{\exp\{H(x_0)+\theta\ z_0\}}{1+\exp\{H(x_0)+\theta\ z_0\}}$ in (4) for the company with predictor values (x_0, z_0), as the sample size of prospective data become large. By this fact, it will be used in Section 1.8 to construct a bankruptcy prediction device for prospective data from the SLM (4).

On the other hand, using the case-control sample from the SLM (4) and treating the sample as if it was a prospective sample from the SLM (4), the local likelihood estimates for $H(x_0)$ and θ are now given. Applying the case-control sample ($Y_i = 0$, x_i, z_i) for $i \leq n_0$ and ($Y_i = 1$, x_i, z_i) for $i > n_0$ from the SLM (4) to the normal equations (7), (9), and (11), the local likelihood estimates $\hat{H}_1(x_0)$ and $\hat{H}(x_0)$ for $H(x_0)$ and $\hat{\theta}$ for θ can be produced. The consistency of both $\hat{H}_1(x_0)$ and $\hat{H}(x_0)$ for $H(x_0) + c^*$, and $\hat{\theta}$ for θ will be shown in Chapter II. Here

$$c^* = \log\{p(Y = 0) \ / \ p(Y = 1)\} \ + \ \log(n_1 \ / \ n_0)$$

has been defined in Section 1.3.

Unfortunately, due to the consistency of $\hat{H}(x_0)$ for $H(x_0) + c^*$ and the fact that the unknown quantity c^* is generally not equal to 0, the resulting predicted bankruptcy probability $\frac{\exp\{\hat{H}(x_0)+\hat{\theta}\ z_0\}}{1+\exp\{\hat{H}(x_0)+\hat{\theta}\ z_0\}}$, obtained by plugging these $\hat{H}(x_0)$ and $\hat{\theta}$ into (5), does not converge to the true bankruptcy probability $\frac{\exp\{H(x_0)+\theta\ z_0\}}{1+\exp\{H(x_0)+\theta\ z_0\}}$ in (4), but approaches $\frac{\exp\{c^*+H(x_0)+\theta\ z_0\}}{1+\exp\{c^*+H(x_0)+\theta\ z_0\}}$, for the company with predictor values (x_0, z_0). This is the major difference between applying the SLM to the prospective sample and to the case-control sample. Although the predicted bankruptcy probability (5) derived by the case-control sample from the SLM (4) does not estimate the true bankruptcy probability, we will discuss in Section 1.8 that it still can be used to develop a bankruptcy prediction device for case-control data from the SLM (4). The same conclusions have also been reached for the LLM in Section 1.3.

1.5 The KMV

In this section, the KMV producing a probability of default for each firm under study will be introduced. The detailed computational procedure of the default probability can be referred to Bharath and Shumway (2004).

The KMV has two particularly important assumptions. The first one is that the total value of a firm is assumed to follow geometric Brownian motion:

$$\frac{dV}{V} = \mu \ dt + \sigma_V \ dZ.$$

Here V is the total value of a firm, μ is the expected continuously compounded return on V, σ_V is the volatility of firm value, and Z is a standard Wiener process. The second assumption of the KMV is that the firm has issued just one discount bond maturing in T periods. Under these two assumptions, the equity of the firm is a call option on the underlying value of the firm with a strike price equal to the face value of the firm's debt with a time-to-maturity of T. By the Black-Scholes call option pricing model, the equity value of a firm satisfies

$$E = V \ N(d_1) - e^{-rT} \ B \ N(d_2). \tag{12}$$

Here E is the market value of the firm's equity, B is the face value of the firm's debt, r is the risk-free interest rate, $N(\cdot)$ is the cumulative standard normal distribution function, and

$$d_1 = \frac{\ln(V \ / \ B) \ + \ (r \ + \ \sigma_V^2 \ / \ 2) \ T}{\sigma_V \ \sqrt{T}}, \qquad d_2 = d_1 - \sigma_V \ \sqrt{T}.$$

This formula (12) is called Black-Scholes-Merton option valuation equation. Hence, the value of equity is a function of the value of the firm and time, so it follows directly from Ito's lemma that $\sigma_E = (V \ / \ E) \ (\partial E \ / \ \partial V) \ \sigma_V$. By Black-Scholes-Merton option valuation equation, it can be shown that $\partial E/\partial V = N(d_1)$, so that under the assumptions of the KMV , the volatilities of the firm value and its equity can be expressed by

$$\sigma_E = (V \ / \ E) \ N(d_1) \ \sigma_V. \tag{13}$$

In order to implement the KMV, first it needs to compute the market value of the firm's equity E by multiplying the firm's shares outstanding by its current stock price. Second, it needs to estimate the volatility of equity from either historical stock returns data or from option implied volatility data. Third, it needs to choose a forecasting horizon T and a measure of the face value of the firm's debt B. For example, it is common to assume $T = 1$, and take the book value of the firm's total liabilities to be the face value of the firm's debt. Fourth, it needs to collect values of the risk-free interest rate. After performing these steps, we have values for each of the variables in equations (12) and (13) except for the total value of the firm V and the volatility of firm value σ_V. Finally, it needs to simultaneously solve equations (12) and (13) numerically for values of V and σ_V. Once this numerical solution is obtained, by the assumptions of the KMV, the probability of default can be calculated as

$$\pi_{KMV} = N(-DD),$$

where

$$DD = \frac{\ln(V \ / \ B) + (\mu - \sigma_V^2 \ / \ 2) \ T}{\sigma_V \ \sqrt{T}}.$$

Simultaneously solving equations (12) and (13) is reasonably straightforward. However, the probability of default is not obtained by simply solving these two equations numerically. Crosbie and Bohn (2001) explain that "In practice the market leverage moves around far too much for [equation (13)] to provide reasonable results." To resolve this problem, we compute the probability of default by implementing a complicated iterative procedure suggested by Bharath and Shumway (2004). The iterative procedure includes the following steps. First, we propose an initial value of $\sigma_V = \sigma_E$ $\{E/(E + B)\}$ and we use this value of σ_V and equation (12) to infer the market value of each firm's assets at the end of every month for the previous year. Recall that E

is the market value of each firm's equity and is calculated from CRSP database as the product of share price at the end of the month and the number of shares outstanding, B is the sum of the debt in current liabilities and one half of long term debt, σ_E is the annualized percent standard deviation of returns and is estimated from the prior year stock return data for each month, r is the risk free interest rate, and $T = 1$. Here r is taken as 1-Year Treasury Constant Maturity Rate obtained from the Board of Governors of the Federal Reserve system, and it is available from the website at http://research.stlouisfed.org/fred/data/irates/gs1. We then calculate the implied log return on assets each month and use that returns series to generate new estimates of σ_V and μ. We iterate on σ_V in this manner until it converges (so the absolute difference in adjacent σ_V is less than 10^{-3}).

1.6 The DAM

In this section, the formulation of the DAM proposed by Altman (1968) for predicting bankruptcy will be introduced.

Given the sample (Y_i, x_i), $i = 1, \cdots, n$, the DAM for predicting bankruptcy for the company with the predictor value $X = x$ is to compute the discriminant function value:

$$DFV = (\overline{x}_1 - \overline{x}_0)^T \; S_{pooled}^{-1} \; x,$$

where

$$\overline{x}_0 = \left\{ \sum_{i=1}^{n} x_i \; I(Y_i = 0) \right\} \; / \; \left\{ \sum_{i=1}^{n} I(Y_i = 0) \right\},$$

$$\overline{x}_1 = \left\{ \sum_{i=1}^{n} x_i \; I(Y_i = 1) \right\} \; / \; \left\{ \sum_{i=1}^{n} I(Y_i = 1) \right\},$$

$$S_{pooled} = \left\{ \sum_{i=1}^{n} (x_i - \overline{x}_1)(x_i - \overline{x}_1)^T I(Y_i = 1) + \sum_{i=1}^{n} (x_i - \overline{x}_0)(x_i - \overline{x}_0)^T I(Y_i = 0) \right\} /(n-2).$$

The larger the value of DFV, the larger the possibility that the company having the predictor value $X = x$ bankrupts. Given a cutoff value v, if $DFV > v$, then the company is classified to be in the status of bankruptcy, otherwise it is classified as a healthy company. The method for computing the optimal cutoff value v^* for v will be introduced in Section 1.8. For a detailed introduction of the DAM, see Johnson and Wichern (2002).

1.7 The DSM

In this section, the formulation of the DSM using the discrete-time survival data for predicting bankruptcy will be introduced.

The DSM is expressed by the likelihood function of the discrete-time survival data. Recall the discrete-time survival data given in Section 1.2:

$$(t_i,\ Y_i,\ x_{i,1},\ \cdots,\ x_{i,t_i},\ z_{i,1},\ \cdots,\ z_{i,t_i}),\ \text{ for }\ i = 1,\ \cdots,\ n.$$

Here $t_i \in \{1,\ 2,\ \cdots,\ m\}$ denotes the duration time of the i-th company in the sampling period, and m is a positive integer standing for the length of the sampling period. Also, at the duration time t_i, $Y_i = 0$ indicates that the i-th company is nonbankrupt, and 1 the i-th company is bankrupt. Further, $x_{i,j}$ and $z_{i,j}$ are values of the d-dimensional continuous and q-dimensional discrete explanatory variables X and Z collected at the duration time j, respectively in each case, for each $j = 1,\ \cdots,\ t_i$ and for the i-th company.

To give the likelihood function of the discrete-time survival data, set $f(t, x, z;\ \psi)$ as the conditional frequency function of T given $(X,\ Z) = (x,\ z)$. Here T is a discrete random variable, $T \in N = \{1,\ 2,\ \cdots\}$, standing for the duration time of a given company, $(x,\ z)$ are the values of explanatory variables $(X,\ Z)$ observed at the duration time $T = t$, and ψ is a vector of parameters. Also, set the survivor function of the given company as

$$S(t, x, z;\ \psi) = 1 - \sum_{j<t} f(j, x, z;\ \psi) = p(T \geq t \mid x, z;\ \psi). \tag{14}$$

The survivor function (14) gives the probability of nonbankruptcy before the duration time t for the given company. Further, set the hazard function of the given company as

$$h(t, x, z;\ \psi) = \frac{f(t, x, z;\ \psi)}{S(t, x, z;\ \psi)} = p(T = t \mid T \geq t, x, z;\ \psi). \tag{15}$$

The hazard function (15) gives the probability of bankruptcy happened instantly at the duration time t for the given company which is nonbankrupt before the duration time t.

We now give the likelihood function of the discrete-time survival data (t_i, Y_i, $x_{i,1}$, $\cdots$, x_{i,t_i}, $z_{i,1}$, $\cdots$, z_{i,t_i}), for $i = 1, \cdots, n$. Using the conditional frequency function $f(t, x, z; \psi)$, it can be written as

$$\begin{aligned} L(\psi) &= \prod_{i=1}^{n} p(T_i = t_i \mid x_{i,t_i}, z_{i,t_i};\ \psi)^{Y_i}\ p(T_i > t_i \mid x_{i,t_i}, z_{i,t_i};\ \psi)^{1-Y_i} \\ &= \prod_{i=1}^{n} f(t_i, x_{i,t_i}, z_{i,t_i};\ \psi)^{Y_i}\ p(T_i > t_i \mid x_{i,t_i}, z_{i,t_i};\ \psi)^{1-Y_i}. \end{aligned}$$

Using elementary properties of conditional probabilities, each of the two probabilities $f(t_i, x_{i,t_i}, z_{i,t_i};\ \psi)$ and $p(T_i > t_i \mid x_{i,t_i}, z_{i,t_i};\ \psi)$ in the likelihood function $L(\psi)$ can be expressed as a function of the hazard function (15). For this, using (15) and replacing $S(t, x, z; \psi)$ with $f(t, x, z; \psi) + p(T > t \mid x, z; \psi)$, we have

$$f(t, x, z;\ \psi) = h(t, x, z;\ \psi)\ \{f(t, x, z;\ \psi) + p(T > t \mid x, z;\ \psi)\}.$$

On both sides of the equation, first subtracting $f(t, x, z;\ \psi)\ h(t, x, z;\ \psi)$, and then dividing by $h(t, x, z; \psi)$, we have

$$\{1 - h(t, x, z;\ \psi)\}\ S(t, x, z;\ \psi) = p(T > t \mid x, z;\ \psi).$$

Replacing $S(t, x, z; \psi)$ with $p(T > t - 1 \mid x, z; \psi)$, and applying the result iteratively to $p(T > j \mid x, z; \psi)$ for each $j = 1, \cdots, t$, we have

$$\prod_{j=1}^{t}\{1 - h(j, x, z;\ \psi)\} = p(T > t \mid x, z;\ \psi). \tag{16}$$

On the other hand, using (15) and (16), we have

$$f(t, x, z;\ \psi) = h(t, x, z;\ \psi)\ \prod_{j=1}^{t-1}\{1 - h(j, x, z;\ \psi)\}. \tag{17}$$

Substituting (16) and (17) into the above likelihood function $L(\psi)$ for the discrete-time survival data, it can be expressed as

$$L(\psi) = \prod_{i=1}^{n}\ \left\{\frac{h(t_i, x_{i,t_i}, z_{i,t_i};\ \psi)}{1 - h(t_i, x_{i,t_i}, z_{i,t_i};\ \psi)}\right\}^{Y_i}\ \prod_{j=1}^{t_i}\ \{1 - h(j, x_{i,j}, z_{i,j};\ \psi)\}. \tag{18}$$

Note that the hazard function (15) can be of any functional form whose values are all in the interval (0, 1). In this dissertation, for simplicity of presentation, it is taken as a logistic function

$$h(t, x, z;\ \psi) = \frac{\exp\{\alpha_0 + \alpha_1\ g(t) + \beta\ x + \theta\ z\}}{1 + \exp\{\alpha_0 + \alpha_1\ g(t) + \beta\ x + \theta\ z\}},$$

where $\psi = (\alpha_0, \alpha_1, \beta, \theta)^T$. Here α_0, α_1, β, and θ are 1×1, 1×1, $1\times d$, and $1\times q$ vectors of logistic parameters, respectively. In this dissertation, the function $g(t)$ is taken as $g(t) = \log(t)$. Substituting the logistic hazard function $h(t, x, z;\ \psi) = \frac{\exp\{\alpha_0+\alpha_1\ g(t)+\beta\ x+\theta\ z\}}{1+\exp\{\alpha_0+\alpha_1\ g(t)+\beta\ x+\theta\ z\}}$ and the natural logarithm function $g(t) = \log(t)$ into the likelihood function $L(\psi)$ in (18), the DSM for the discrete-time survival data is expressed as

$$L(\psi) = \prod_{i=1}^{n}\ [\exp\{\alpha_0 + \alpha_1\ \log(t_i) + \beta\ x_{i,t_i} + \theta\ z_{i,t_i}\}^{Y_i}\ \times$$

$$\prod_{j=1}^{t_i}\ [1 + \exp\{\alpha_0 + \alpha_1\ \log(j) + \beta\ x_{i,j} + \theta\ z_{i,j}\}]^{-1}. \tag{19}$$

If the function $g(t)$ of the duration time t is taken as the natural logarithm function, then the resulting DSM (19) is an accelerated failure-time model; see Lancaster (1990). Such logistic hazard function with $g(t) = \log(t)$ is also considered by Shumway (2001).

We now give the maximum likelihood estimates $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}$, and $\hat{\theta}$ of α_0, α_1, β, and θ, respectively. Using the DSM (19) and the discrete-time survival data, the log-likelihood function of $\psi = (\alpha_0,\, \alpha_1,\, \beta,\, \theta)^T$ is

$$\begin{aligned}\ell_{DSM}(\psi) &= \sum_{i=1}^{n} Y_i \,\{\alpha_0 + \alpha_1 \;\log(t_i) + \beta\; x_{i,t_i} + \theta\; z_{i,t_i}\} \;- \\ &\quad \sum_{i=1}^{n}\sum_{j=1}^{t_i} \;\log[1 + \exp\{\alpha_0 + \alpha_1\; \log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}].\end{aligned}$$

Then $\hat{\psi} = (\hat{\alpha}_0,\, \hat{\alpha}_1,\, \hat{\beta},\, \hat{\theta})^T$ may be taken as the solution of the normal equations

$$\begin{aligned}0 &= \frac{\partial \ell_{DSM}(\psi)}{\partial \psi} \\ &= \sum_{i=1}^{n} Y_i \begin{bmatrix} 1 \\ \log(t_i) \\ x_{i,t_i} \\ z_{i,t_i} \end{bmatrix} - \sum_{i=1}^{n}\sum_{j=1}^{t_i} \frac{\exp\{\alpha_0 + \alpha_1\; \log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}}{1 + \exp\{\alpha_0 + \alpha_1\; \log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}} \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix}.\end{aligned}$$

Using the maximum likelihood estimates $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}$, and $\hat{\theta}$, for the company with predictor values $(X,\, Z) = (x_0,\, z_0)$ at the duration time t, its predicted probability of instant bankruptcy

$$h(t, x_0, z_0;\; \hat{\psi}) = \frac{\exp\{\hat{\alpha}_0 + \hat{\alpha}_1\; \log(t) + \hat{\beta}\; x_0 + \hat{\theta}\; z_0\}}{1 + \exp\{\hat{\alpha}_0 + \hat{\alpha}_1\; \log(t) + \hat{\beta}\; x_0 + \hat{\theta}\; z_0\}} \tag{20}$$

is the logistic hazard function evaluated at the predicted score $\hat{\alpha}_0 + \hat{\alpha}_1 \log(t) + \hat{\beta}\; x_0 + \hat{\theta}$ z_0. Due to the consistency of maximum likelihood estimates $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}$, and $\hat{\theta}$ (Section 3.3 of Cox and Oakes 1984), the predicted probability of instant bankruptcy $\exp\{\hat{\alpha}_0 + \hat{\alpha}_1$ $\log(t) + \hat{\beta}\; x_0 + \hat{\theta}\; z_0\} \;/\; [1 + \exp\{\hat{\alpha}_0 + \hat{\alpha}_1\; \log(t) + \hat{\beta}\; x_0 + \hat{\theta}\; z_0\}]$ in (20) converges to the true probability of instant bankruptcy $\frac{\exp\{\alpha_0 + \alpha_1\; \log(t) + \beta\; x_0 + \theta\; z_0\}}{1 + \exp\{\alpha_0 + \alpha_1\; \log(t) + \beta\; x_0 + \theta\; z_0\}}$, the logistic hazard function evaluated at the true score $\alpha_0 + \alpha_1 \log(t) + \beta\; x_0 + \theta\; z_0$, for the company with predictor values $(x_0,\, z_0)$ at the duration time t. By this fact, it will be used in Section 1.8 to construct a bankruptcy prediction device for the DSM with the discrete-time survival data.

1.8 Bankruptcy Prediction Devices

In this section, we shall develop bankruptcy prediction methods for the five models, LLM, SLM, KMV, DAM, and DSM, in sequence.

Using the prospective training sample from the LLM (1), we determine a $p^* \in (0, 1)$ value to make bankruptcy prediction for the company with predictor values $(x_0,\, z_0)$. By the consistency of its predicted bankruptcy probability $\hat{p}(Y = 1 \mid X = x_0,\, Z = z_0)$ derived from (2), if it satisfies

$$\hat{p}(Y = 1 \mid X = x_0,\ Z = z_0) > p^*,$$

then the company is classified to be in the status of bankruptcy, otherwise it is classified as a healthy company. To decide a proper cut-off point p^*, usually one would use the training sample to evaluate the performance of the classification scheme. In doing so, there are two types of "in-sample" error rate occurred in this evaluation based on the training sample:

$$\text{type I error rate} \qquad \alpha_{in}(p) = \frac{\sum_{i=1}^{n} Y_i\, I\{\hat{p}(Y = 1 \mid X = x_i,\ Z = z_i) \le p\}}{\sum_{i=1}^{n} Y_i},$$

and

$$\text{type II error rate} \qquad \beta_{in}(p) = \frac{\sum_{i=1}^{n} (1 - Y_i)\, I\{\hat{p}(Y = 1 \mid X = x_i,\ Z = z_i) > p\}}{\sum_{i=1}^{n} (1 - Y_i)}.$$

Here $I(\cdot)$ stands for the indicator function. Using the training sample and the cut-off point p, $\alpha_{in}(p)$ is the rate of misclassifying bankrupt company to healthy company, and $\beta_{in}(p)$ is the rate of misclassifying healthy company to bankrupt company. To keep these error rates to be as small as possible, we determine a proper cut-off point p^* such that

$$\tau_{in}(p^*) = \alpha_{in}(p^*) + \beta_{in}(p^*) = \min_{p\in[0,1],\ \alpha_{in}(p)\le u} \{\alpha_{in}(p) + \beta_{in}(p)\}.$$

That is to control the in-sample type I error rate $\alpha_{in}(p)$ to be at most u, so that the sum of the two in-sample error rates $\tau_{in}(p) = \alpha_{in}(p) + \beta_{in}(p)$ is minimal. This is essential if the type I error would cause much more severe losses to the investors. On the other hand, if classifying healthy firms as being bankrupt would cause more severe losses to the investors, we might control the in-sample type II error rate $\beta_{in}(p)$ to be at most u. In practice, the value of $u \in [0, 1]$ is determined by the investor. If $u = 1$, then there is no restriction on the magnitude of in-sample type I and II error rates (Altman, 1968; Ohlson, 1980; Begley, Ming, and Watts, 1996). Since the value of p^* depends on that of u, it is also denoted by $p^*(u)$.

On the other hand, using the case-control training sample from the LLM (1) and treating the sample as if it was a prospective sample from the LLM (1), by the results in Section 1.3, the corresponding predicted bankruptcy probability $\frac{\exp(\hat\alpha+\hat\beta\ x_0+\hat\theta\ z_0)}{1+\exp(\hat\alpha+\hat\beta\ x_0+\hat\theta\ z_0)}$ obtained from (2) does not converge to the true bankruptcy probability $\frac{\exp(\alpha+\beta\ x_0+\theta\ z_0)}{1+\exp(\alpha+\beta\ x_0+\theta\ z_0)}$ in (1), but approaches $\frac{\exp(\alpha+c^*+\beta\ x_0+\theta\ z_0)}{1+\exp(\alpha+c^*+\beta\ x_0+\theta\ z_0)}$. Here $c^* = \log\{p(Y = 0)/p(Y = 1)\} + \log(n_1/n_0)$. This drawback is caused by the fact that the resulting maximum likelihood estimates $\hat\beta$ and $\hat\theta$ of logistic parameters β and θ in the LLM (1) converge to their true values, respectively, but $\hat\alpha$ approaches the quantity $\alpha + c^*$, as both sample sizes of control and case data become large. This is the major difference between applying the LLM to case-control data and to prospective data.

Fortunately, we still can use the predicted bankruptcy probability $\frac{\exp(\hat\alpha+\hat\beta\ x_0+\hat\theta\ z_0)}{1+\exp(\hat\alpha+\hat\beta\ x_0+\hat\theta\ z_0)}$ obtained by the case-control sample from the LLM (1) to develop a bankruptcy prediction device by applying the following simple equivalent inequalities:

$$\frac{\exp\{\alpha+\beta\ x+\theta\ z\}}{1+\exp\{\alpha+\beta\ x+\theta\ z\}} > p,$$

if and only if

$$\frac{\exp\{\alpha + c^* + \beta\ x + \theta\ z\}}{1 + \exp\{\alpha + c^* + \beta\ x + \theta\ z\}} > \frac{p\ \exp(c^*)}{(1-p) + p\ \exp(c^*)} \equiv p_{c^*}.$$

This result is to say that using the probability $\frac{\exp(\alpha+\beta\ x+\theta\ z)}{1+\exp(\alpha+\beta\ x+\theta\ z)}$ to define classification device with cut-off point p is equivalent to using the probability $\frac{\exp(\alpha+c^*+\beta\ x+\theta\ z)}{1+\exp(\alpha+c^*+\beta\ x+\theta\ z)}$ to define classification device with cut-off point p_{c^*}. Hence we may pretend the predicted bankruptcy probability $\frac{\exp(\hat{\alpha}+\hat{\beta}\ x_0+\hat{\theta}\ z_0)}{1+\exp(\hat{\alpha}+\hat{\beta}\ x_0+\hat{\theta}\ z_0)}$ obtained by the case-control sample from the LLM (1) to be the estimate of the true bankruptcy probability and use it to determine the corresponding proper cut-off point $p^*(u)$. Then the bankruptcy prediction device for the case-control sample from the LLM (1) can be obtained.

Note that above bankruptcy prediction methods built for the LLM (1) using the two types of data are essentially equivalent. Based on the same arguments, similar bankruptcy prediction devices can be developed directly for the SLM (4) using the two types of data by replacing respectively (2) and $\hat{\alpha} + \hat{\beta}\ x_0$ with (5) and $\hat{H}(x_0)$. Hence the bankruptcy prediction methods constructed for the SLM (4) using the two types of data are also essentially equivalent.

Note also that the above method for computing the optimal cutoff value for the LLM (1) can be similarly applied for both the KMV and the DAM by replacing $\hat{p}(Y = 1 \mid X = x_0, Z = z_0)$ with π_{KMV} and DFV to derive their optimal cutoff values π^*_{KMV} and v^*, respectively in each case. Given the optimal cutoff value π^*_{KMV}, if $\pi_{KMV} > \pi^*_{KMV}$, then the company with the probability of default π_{KMV} is classified to be in the status of bankruptcy, otherwise it is classified as a healthy company. Similarly, given the optimal cutoff value v^*, if $DFV > v^*$, then the company with the discriminant function value DFV is classified to be in the status of bankruptcy, otherwise it is classified as a healthy company.

We now give the bankruptcy prediction method for the DSM. Using the discrete-time survival data and following the same arguments of the bankruptcy method based on the LLM with prospective data, we determine a $p^* \in (0, 1)$ value to make bankruptcy prediction for the company with predictor values $(x_0,\ z_0)$ at the duration time t. By the consistency of its predicted probability of instant bankruptcy $h(t, x_0, z_0;\ \hat{\psi})$, if it satisfies

$$h(t, x_0, z_0;\ \hat{\psi}) > p^*,$$

then, at the duration time t, the company is classified to be in the status of bankruptcy, otherwise it is classified as a healthy company. To decide a proper cut-off point p^*, we use the data $(t_i,\ Y_i,\ x_{i,ti},\ z_{i,t_i})$, for $i = 1, \cdots, n$, to evaluate the performance of the classification scheme. In doing so, there are two types of "in-sample" error rate occurred in this evaluation based on the data $(t_i,\ Y_i,\ x_{i,ti},\ z_{i,t_i})$, for $i = 1, \cdots, n$:

$$\text{type I error rate} \qquad \alpha_{in}(p) = \frac{\sum_{i=1}^{n}\ Y_i\ I\{h(t_i, x_{i,t_i}, z_{i,t_i};\ \hat{\psi}) \le p\}}{\sum_{i=1}^{n}\ Y_i},$$

and

$$\text{type II error rate} \qquad \beta_{in}(p) = \frac{\sum_{i=1}^{n}\ (1 - Y_i)\ I\{h(t_i, x_{i,t_i}, z_{i,t_i};\ \hat{\psi}) > p\}}{\sum_{i=1}^{n}\ (1 - Y_i)}.$$

Here $I(\cdot)$ stands for the indicator function. To keep these two error rates to be as small as possible, we determine a proper cut-off point $p^* = p^*(u)$ for the bankruptcy prediction method based on the DSM such that

$$\tau_{in}\{p^*(u)\} = \alpha_{in}\{p^*(u)\} + \beta_{in}\{p^*(u)\} = \min_{p\in[0,1],\ \alpha_{in}(p)\le u} \{\alpha_{in}(p) + \beta_{in}(p)\},$$

for each $u \in [0,\ 1]$.

1.9 Summary of Results

In Chapter II, the SLM (4) with case-control sampling is applied to estimate bankruptcy probabilities for firms collected from Compustat North America (COMPUSTAT) and Center for Research in Security Prices (CRSP) databases. The unknown quantities in the model are estimated by the local likelihood approach, and the resulting estimators are analyzed through their asymptotic biases and variances. Both a real data example and a simulation study demonstrate that, given case-control data, our semiparametric prediction method based on the SLM (4) is more powerful than the prediction method based on the LLM (1), the KMV, and the DAM, in the sense of yielding smaller out-of-sample error rate.

In Chapter III, the DSM (19) with discrete-time survival data is applied to estimate the probabilities of financial distress for firms listed in Taiwan Stock Exchange. Since there are only few bankrupt firms in Taiwan, it is difficult to predict bankruptcy well. In this case, to provide more failure firms to proceed our research, we replaced our target on bankruptcy prediction with financial distress prediction. According to the definition of financial distress given by Taiwan Stock Exchange, financial distress companies are those whose stocks were delisted, stopped trading, or traded by cash. The maximum likelihood method is employed to estimate the values of parameters in the DSM, and the resulting estimators are analyzed by their asymptotic normal distributions. Empirical studies demonstrate that the financial distress prediction method based on the dynamic DSM (19) using discrete-time survival data can yield more accurate forecasts than the alternative method based on the static LLM (1), in the sense of yielding smaller out-of-sample error rate.

In Chapter IV, some concluding remarks and future research topics for bankruptcy prediction methods based on each of the static SLM (4) and the dynamic DSM (19) are presented.

CHAPTER II

SEMIPARAMETRIC BANKRUPTCY PREDICTION METHODS

2.1 Introduction

As introduced in Sections 1.3 and 1.4, the SLM (4) is a robust method against misspecification of the parametric logit model relation for predicting bankruptcy. This model is basically very similar to the LLM (1), except that some unspecified function $H(x)$ replaces the linear function $\alpha + \beta\, x$ to model the relation between the continuous predictors and the logit function of bankruptcy probability. Thus, clearly, the SLM is much more general and flexible in predicting the bankruptcy of a firm.

In this section, we shall first study the asymptotic properties of estimators $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$ given in Section 1.4 for the SLM with case-control data. Then the finite sample performance of the bankruptcy prediction method based on the SLM with case-control data is studied through a real data example and a simulation study. For these, the composition of the case-control sample and the formulations of these estimators are recalled.

According to the case-control sampling, we draw a random sample of nonbankrupt companies of n_0 observations (controls), say $(x_1, z_1), \cdots, (x_{n_0}, z_{n_0})$, from the conditional distribution of predictors (X, Z) given $Y = 0$, and an independent random sample of bankrupt companies of n_1 observations (cases), say $(x_{n_0+1}, z_{n_0+1}), \cdots, (x_n, z_n)$, where $n = n_0 + n_1$, from the conditional distribution of predictors (X, Z) given $Y = 1$. Here $Y_i = 1$ indicating that the i-th company is in the state of bankruptcy and 0, otherwise. Hence we have the case-control sample (Y_i, x_i, z_i), $i = 1, \cdots, n$, where $Y_i = 0$ for $i \leq n_0$, and 1 for $i > n_0$. Set $f(x, z)$ as the frequency function of predictors (X, Z), and $f_0(x, z) = f(x, z \mid Y = 0)$ and $f_1(x, z) = f(x, z \mid Y = 1)$ as the conditional frequency functions of (X, Z) given $Y = 0$ and 1, respectively. Then from Bayes theorem and the SLM (4), these two conditional frequency functions can be related by

$$f_1(x,\ z) = f_0(x,\ z)\ \exp\{H^*(x) + \theta\ z\}, \tag{21}$$

where

$$H^*(x) = H(x) \;+\; \log\{p(Y=0) \;/\; p(Y=1)\}.$$

Given the case-control sample and the bandwidth parameters b_θ and b_H, by the development of the logistic regression in the case-control setting given in both Section 1.4 of this dissertation and Section 6.3 of Hosmer and Lemeshow (1989), the log-likelihood functions for producing $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$ in the SLM (4) with kernel function K can be expressed respectively as

$$\ell_1(\alpha, \beta, \theta;\ x) = (-1) \sum_{i=1}^{n} \log[1 + \exp\{\alpha + \beta\ (x_i - x) + \theta\ z_i)\}]\ K_{b_\theta}(x_i - x) + \sum_{i=n_0+1}^{n} \{\alpha + \beta\ (x_i - x) + \theta\ z_i\}\ K_{b_\theta}(x_i - x), \tag{22}$$

$$\ell_2(\alpha_0,\ \theta) = (-1) \sum_{i=1}^{n} \log[1 + \exp\{\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i)\}] + \sum_{i=n_0+1}^{n} \{\alpha_0 + \hat{H}_1(x_i) + \theta\ z_i\}, \tag{23}$$

$$\ell_3(\alpha^*, \beta;\ x) = (-1) \sum_{i=1}^{n} \log[1 + \exp\{\alpha^* + \beta\ (x_i - x) + \hat{\theta}\ z_i)\}]\ K_{b_H}(x_i - x) + \sum_{i=n_0+1}^{n} \{\alpha^* + \beta\ (x_i - x) + \hat{\theta}\ z_i\}\ K_{b_H}(x_i - x). \tag{24}$$

Note that the parameter β in (22) and (24) represents the unknown quantity $H^{(1)}(x)^T$, as it did in $\ell_{SLM}(\eta;\ x_0)$ and $\hat{\ell}_{SLM}(\xi;\ x_0)$ in Section 1.4 with x_0 replaced by x. But α and α^* in (22) and (24) stands for $H(x) + c^*$, not as they did in $\ell_{SLM}(\eta;\ x_0)$ and $\hat{\ell}_{SLM}(\xi;\ x_0)$ in Section 1.4 for $H(x)$ with x_0 replaced by x, respectively in each case. The maximum likelihood estimates of α and α^* produced from (22) and (24) would be estimates for $H(x)+c^*$, not for $H(x)$. This fact causes the major difference between the applications of the logit model to the case-control sample and the prospective sample (see discussions in Section 1.4).

This chapter is organized as follows. Section 2.2 presents asymptotic properties of estimators $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$ introduced in Section 1.4 for the SLM (4) with case-control data. To illustrate the bankruptcy prediction method based on the SLM with case-control data, a real data set is analyzed in Section 2.3. Simulation results which give additional insight of the bankruptcy prediction method are contained in Section 2.4. Section 2.5 gives concluding remarks and future research topics. Finally, sketches of the proofs are given in Section 2.6.

2.2 Theoretical Results

In this section, we shall study the asymptotic properties of $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$. For this, we need the following conditions:

(C1) Kernel function $K(u)$ is a symmetric and Lipschitz continuous probability density function supported on $[-1, 1]$, and is bounded above zero on $[-1/2, 1/2]$.

(C2) $n_0/n \to \zeta \in (0, 1)$, as $n \to \infty$.

(C3) Bandwidth parameters b_θ, $b_H \in [\delta\, n^{-1+\delta}, \delta^{-1} n^{-\delta}]$, for some δ satisfying $0 < \delta < 1/2$. They also satisfy $n\, b_\theta^{d+2} >> 1 >> b_\theta$ and $n\, b_H^d >> 1 >> b_H >> b_\theta$. The notation $a_n >> b_n$ means that $b_n/a_n \to 0$, as $n \to \infty$.

(C4) The d-variate function $H(x)$ is defined on $[0, 1]^d$, and each of its second order partial derivatives is Lipschitz continuous on $[0, 1]^d$.

(C5) Under control and case populations, their respective marginal densities $f_0(x)$ and $f_1(x)$ of X are Lipschitz continuous and bounded above zero on $[0, 1]^d$. Also, their corresponding conditional probabilities $f_0(z \mid x)$ and $f_1(z \mid x)$ of Z given $X = x$ can not be zero or one for each given x, and are Lipschitz continuous with respect to x.

Conditions (C1)-(C4) are regular for the usual nonparametric regression analysis. The support $[0, 1]^d$ in (C4) and (C5) of the d-dimensional variable X is given for simplicity of presentation and for studying the asymptotic behavior of $\hat{H}_1(x)$ and $\hat{H}(x)$ on the boundary region of the support of $f_0(x)$ and $f_1(x)$. It can be replaced with any bounded region $\Omega \subset R^d$, and the asymptotic properties for the resulting $\hat{H}_1(x)$, $\hat{\theta}$, and

$\hat{H}(x)$ remain unchanged. The first part of condition (C5) guarantees that the design points X, under control and case populations, have no holes on $[0,\ 1]^d$. The second part of (C5) makes sure that the Hessian matrix for each of $\ell_1(\alpha, \beta, \theta;\ x)$ and $\ell_3(\alpha, \beta;\ x)$ is invertible.

In order to give concise expressions for the asymptotic properties of $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$, we need more notations. Let

$$x = (x_1,\ \cdots,\ x_d)^T, \qquad t = (t_1,\ \cdots,\ t_d)^T, \qquad H_{ij}(x) = \partial^2/(\partial x_i\ \partial x_j)\ H(x),$$

$$m_i = \max\{-1,\ (x_i - 1)/b\}, \qquad k_i = \min\{1,\ x_i/b\}, \qquad K^{\#}(t) = \prod_{j=1}^{d} K(t_j),$$

$$\lambda_0 = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} K^{\#}(t)\ dt, \qquad \tau_0 = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} K^{\#}(t)^2\ dt,$$

$$\lambda_{i,k} = \int_{m_i}^{k_i} u^k\ K(u)\ du, \qquad c_{ij} = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} t_i\ t_j\ K^*(t)\ dt,$$

Q be the collection of all values of the discrete q-dimensional variable Z, and I_1 be the $(1+q) \times (1+q)$ identity matrix with the first column vector of the identity matrix deleted, for $i,\ j = 1, \cdots, d$ and $k \geq 0$. Here $K^*(t)$ is the d-variate Lejeune-Sarda kernel function of order two (Lejeune and Sarda 1992). In particular, given the point $x \in [0,\ 1]^d$, the kernel function K, and the bandwidth b, $K^*(t)$ can be expressed as

$$K^*(t) = \left\{\prod_{i=1}^{d} \lambda_{i,0}^{-1}\ K(t_i)\right\} \left\{1 - \sum_{i=1}^{d} (t_i\ \lambda_{i,0} - \lambda_{i,1})\ \lambda_{i,1}\ (\lambda_{i,0}\ \lambda_{i,2} - \lambda_{i,1}^2)^{-1}\right\},$$

and its corresponding values c_{ij} become

$$c_{ij} = \begin{cases} (\lambda_{i,2}^2 - \lambda_{i,1}\ \lambda_{i,3})\ (\lambda_{i,0}\ \lambda_{i,2} - \lambda_{i,1}^2)^{-1}, & \text{for } i = j, \\ (-1)\ \lambda_{i,1}\ \lambda_{i,0}^{-1}\ \lambda_{j,1}\ \lambda_{j,0}^{-1}, & \text{for } i \neq j. \end{cases}$$

Define

$$r(x,\ z) = \frac{f_0(x,\ z)\ \exp\{H(x) + c^* + \theta\ z\}}{\zeta + (1-\zeta)\ \exp\{H(x) + c^* + \theta\ z\}},$$

$$D_0(x) = \sum_{z \in Q} r(x, z), \qquad D_1(x) = \sum_{z \in Q} z\ r(x, z), \qquad D_2(x) = \sum_{z \in Q} z\ z^T\ r(x, z),$$

$$D_j = \int_0^1 \cdots \int_0^1 D_j(x)\ dx, \quad \text{for} \quad j = 0, 1, 2, \qquad D = \begin{pmatrix} D_0, & D_1^T \\ D_1, & D_2 \end{pmatrix}.$$

Also, define quantities related to the asymptotic biases and variances of $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$ in the following:

$$c_H(x;\ b) = (1/2) \sum_{i=1}^{d} \sum_{j=1}^{d} c_{ij}\ H_{ij}(x),$$

$$v_{H,1}(x;\ b) = D_0(x)^{-1} \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} K^*(t)^2\ dt,$$

$$v_{H,2}(x;\ b) = \lambda_0^{-2}\ \tau_0\ D_0(x)^{-1}\ D_1^T(x)\ \left\{D_0(x)\ D_2(x) - D_1(x)\ D_1^T(x)\right\}^{-1}\ D_1(x),$$

$$c_\theta(b) = (-1)\ \left[\int_0^1 \cdots \int_0^1 \{D_0(x), D_1^T(x)\}\ c_H(x; b)\ dx\right]\ D^{-1}\ I_1,$$

$$v_\theta = D_2^{-1}\ D_1\ (D_0 - D_1^T\ D_2^{-1}\ D_1)^{-1}\ D_1^T\ D_2^{-1}\ +\ D_2^{-1}.$$

If x is in the interior region $[b,\ 1-b]^d$ of $[0,\ 1]^d$, then it can be seen that the values of $c_H(x;\ b)$ and $v_{H,1}(x;\ b)$ become

$$c_H(x;\ b) = (1/2)\ \left\{\int_{-1}^{1} u^2\ K(u)\ du\right\}\ \left\{\sum_{i=1}^{d} H_{ii}(x)\right\},$$

$$v_{H,1}(x;\ b) = D_0(x)^{-1}\ \left\{\int_{-1}^{1} K(u)^2\ du\right\}^d.$$

The following Theorem 2.1 states the asymptotic bias and variance for $\hat{H}_1(x)$, and those for $\hat{\theta}$ and $\hat{H}(x)$. The proofs will be given in Section 2.6.

Theorem 2.1. Under the SLM and the case-control sample, suppose that conditions (C1)-(C5) are satisfied. For each $x \in [0,\ 1]^d$ and as $n \to \infty$,

$$\text{Bias}\{\hat{H}_1(x)\} = \text{E}\{\hat{H}_1(x)\} - H(x) - c^* = b_\theta^2\ c_H(x;\ b_\theta) + O(b_\theta^3 + n^{-1} b_\theta^{-d}), \qquad (25)$$

$$\text{Var}\{\hat{H}_1(x)\} = n^{-1} b_\theta^{-d}\, \zeta^{-1}(1-\zeta)^{-1}\, \{v_{H,1}(x;\ b_\theta) + v_{H,2}(x;\ b_\theta)\} + O(n^{-1} b_\theta^{-d+1}), \quad (26)$$

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = b_\theta^2\, c_\theta(b_\theta) + O(b_\theta^3 + n^{-1} b_\theta^{-d}), \quad (27)$$

$$\text{Var}(\hat{\theta}) = n^{-1}\, \zeta^{-1}(1-\zeta)^{-1}\, v_\theta + O(n^{-1} b_\theta), \quad (28)$$

$$\text{Bias}\{\hat{H}(x)\} = \text{E}\{\hat{H}(x)\} - H(x) - c^* = b_H^2\, c_H(x;\ b_H) + O(b_H^3 + b_\theta^2 + n^{-1} b_H^{-d}), \quad (29)$$

$$\text{Var}\{\hat{H}(x)\} = n^{-1} b_H^{-d}\, \zeta^{-1}(1-\zeta)^{-1}\, v_{H,1}(x;\ b_H) + O(n^{-1} b_H^{-d+1}). \quad (30)$$

Remark 2.1. {The optimal kernel function K and the magnitudes of the optimal bandwidth parameters b_θ^* and b_H^* for constructing $\hat{\theta}$ and $\hat{H}(x)$} By Theorem 8 of Fan, Gasser, Gijbels, Brockmann, and Engel (1993) and our Theorem 2.1, the optimal K satisfying the conditions in (C1) for constructing $\hat{H}(x)$ is the Epanechnikov kernel $K(u) = (3/4)\ (1-u^2)\ I(|u| \leq 1)$, for each $x \in [0,\ 1]^d$, in the sense of having smaller asymptotic mean square error. On the other hand, by (29) and (30), the optimal choice of b_H, in terms of having smallest mean square error of $\hat{H}(x)$, is $b_H^* = c_H^*\, n^{-1/(d+4)}$, where c_H^* is a constant depending on the unknown factors $H(\cdot)$, θ, and $f_0(x,\ z)$. Similarly, by (25)-(28) and (C3), the optimal value b_θ^* of b_θ, in terms of having smallest mean square error of $\hat{\theta}$, satisfies the condition $n^{-1/(d+4)} >> b_\theta^* >> n^{-1/(d+2)}$. Hence we conclude that the value of b_H^* is of larger order than that of b_θ^*, and that the mean square error of $\hat{H}(x)$ using the optimal bandwidth parameter b_H^* is of smaller order in magnitude than that of $\hat{H}(x)$ using b_θ^*.

Remark 2.2. {Selection of values of b_θ and b_H for practically constructing $\hat{\theta}$ and $\hat{H}(x)$, respectively} The practical implementation of SLM requires the specification of each value of b_θ and b_H. It determines how many data points should be included in the SLM. The optimal values b_θ^* and b_H^* of b_θ and b_H, respectively, can be determined by minimizing the mean square errors of the resulting $\hat{\theta}$ and $\hat{H}(x)$. Theoretical results in Remark 2.1 show that b_H^* is of larger order in magnitude than b_θ^*. Although such theoretical results give some indication on how to select bandwidth parameters b_θ and b_H, they are not available in practice, since they depend on the unknown $H(\cdot)$, θ, and density function of the predictors. Thus, in real applications, we would suggest to consider the in-sample type I and II error rates defined in Section 1.8 as functions of the cut-off point p and values of b_θ and b_H, denoted as $\alpha_{in}(p, b_\theta, b_H)$ and $\beta_{in}(p, b_\theta, b_H)$, respectively. The cut-off point and the bandwidth parameters are then simultaneously determined so that the sum of the two in-sample error rates $\tau_{in}(p, b_\theta, b_H) = \alpha_{in}(p, b_\theta, b_H) + \beta_{in}(p, b_\theta, b_H)$ is minimal, subject to the constrains: $p \in [0, 1]$, $\alpha_{in}(p, b_\theta, b_H) \leq u$, and $b_H \geq b_\theta$, for each given value of $u \in [0, 1]$. Let $\hat{p}(u)$, $\hat{b}_\theta(u)$, and $\hat{b}_H(u)$ denote such selected values for $p^*(u)$, b^*_θ, and b^*_H, respectively.

2.3 A Real Data Example

In this section, a real case-control data set is analyzed using our method SLM and prediction rules DAM, LLM and KMV. McKee (2003) pointed out that company asset size and industry are significant factors affecting bankruptcy status. Thus an ideal approach is to stratify companies according to industry and asset size and determine prediction model for each stratum. Unfortunately, we did not have enough data from COMPUSTAT and CRSP databases for doing so. Thus, to illustrate our method, we simply used two controls to match with one case so that they had the same standard industrial classification (SIC) code and similar company asset size from the same year. By doing this, it is clear that the company asset size has no more power in discriminating the bankruptcy status of the company and thus will not be included in the analysis of our example.

We now introduce the case-control data set. The data set contains 79 companies that were delisted and declared bankruptcy (cases) during the period 1994 to 2002 by COMPUSTAT as meeting the Chapter 11 Bankruptcy or Chapter 7 Liquidation. After identifying these companies filing for bankruptcy, both COMPUSTAT and CRSP databases were searched to locate the latest annual financial data prior to the delisting date. Thus the annual financial data for the identified bankrupt companies were from the period 1993 to 2001. Among the 79 selected bankrupt companies, each was matched with two nonbankrupt companies, except 2 companies only matched with one

Table 1: The SIC codes of companies in our case-control sample.

SIC category	number of bankrupt companies	number of nonbankrupt companies
1000 – 1999	4	8
2000 – 2999	11	22
3000 – 3999	21	40
4000 – 4999	5	10
5000 – 5999	18	36
6000 – 6999	3	6
7000 – 7999	13	26
8000 – 8999	4	8
Total companies:	79	156

nonbankrupt company each, due to the incompleteness of the two databases. Hence our data set also contains 156 nonbankrupt companies (controls). The total number of companies in this research was $n = 235$. The financial institutions were eliminated from the sample due to the unique capital requirements and regulatory structure in that industry group.

We note that COMPUSTAT provides 233 companies whose common stocks were traded in New York Stock Exchange, American Stock Exchange or NASDAQ, and which were declared bankrupt during the period 1994 to 2002. But since COMPUSTAT and CRSP databases contain many missing values for the predictors studied in our example, we only found 79 bankrupt companies with complete predictor values. There is no additional criteria imposed to the bankrupt companies in our case-control sample. The problem of missing data is not unusual in applications, especially when there are many predictive variables used in the model. As long as the missingness occurs "at random" then it will not introduce systematic biases in our analyses (Little and Rubin, 2002). We have no reason not to believe that the missingness occurred in COMPUSTAT and CRSP databases is "missing at random".

The information about industry and that about company asset size of the selected companies are given in Tables 1 and 2, respectively. The two-sample median test was performed to test the null hypothesis of equal magnitude of the asset size for nonbankrupt company and that for bankrupt company. The p-value given in Table

Table 2: Summary statistics of company asset sizes (in million US dollars) from our case-control sample.

	79 bankrupt companies	156 nonbankrupt companies	median-stat (p-value)
mean	105.103	150.508	0.092 (0.927)
median	32.211	33.599	
std	290.254	808.741	
min	1.447	1.636	
max	2345.800	9794.400	

2 shows that there is no significant difference between both company asset sizes at significance level 0.05. This result indicates that our matching process has successfully created similar asset sizes for bankrupt and nonbankrupt companies in our case-control sample.

For predicting bankruptcy, the values of the 9 variables used by Ohlson (1980) and the 2 variables suggested by Shumway (2001) were collected for our selected companies from COMPUSTAT and CRSP databases. The 11 predictive variables are as follows:

1. TLTA = Total liabilities divided by total assets.

2. WCTA = Working capital divided by total assets.

3. CLCA = Current liabilities divided by current assets.

4. NITA = Net income divided by total assets.

5. FUTL = Funds provided by operations divided by total liabilities.

6. CHIN = $(NI_t - NI_{t-1})$ / $(|NI_t| + |NI_{t-1}|)$, where NI_t is net income for the most recent period.

7. INTWO = One if net income was negative for the last two years, zero otherwise.

8. OENEG = One if total liabilities exceed total assets, zero otherwise.

9. SIZE = Logarithm of total asset divided by GNP price-level index. The index assumes a base value of 100 for 1991.

10. Relative Size = Logarithm of each firm's market equity value divided by the total NYSE / AMEX / NASDAQ market equity value.

11. Excess Return = Monthly return on the firm minus the value-weighted CRSP NYSE / AMEX / NASDAQ index return cumulated to obtain the yearly return.

Table 3: Summary statistics of variables in our case-control sample.

variable	mean	median	std	min	max	median-stat (p-value)
			79 bankrupt companies			
TLTA	0.801	0.747	0.435	0.020	2.450	−5.432 (0.000)
WCTA	0.040	0.075	0.387	−1.192	0.980	4.511 (0.000)
CLCA	1.711	0.857	3.545	0.020	23.214	−4.603 (0.000)
NITA	−0.423	−0.161	0.649	−2.833	0.182	5.891 (0.000)
FUTL	−0.335	−0.051	0.921	−4.953	1.279	5.339 (0.000)
CHIN	−0.251	−0.363	0.655	−1.000	1.000	3.130 (0.002)
INTWO	0.570	1	0.498	0	1	−4.612 (0.000)
OENEG	0.190	0	0.395	0	1	−3.844 (0.000)
Excess Return	−0.254	−0.634	1.258	−1.320	6.617	3.682 (0.000)
Relative Size	−5.803	−5.830	0.675	−7.379	−4.577	4.234 (0.000)
π_{KMV}	0.413	0.331	0.383	0.000	1.000	−6.537 (0.000)
			156 nonbankrupt companies			
TLTA	0.486	0.478	0.273	0.029	1.926	
WCTA	0.276	0.291	0.258	−0.592	0.921	
CLCA	0.707	0.509	0.796	0.055	6.904	
NITA	−0.079	0.024	0.386	−3.800	0.249	
FUTL	−0.030	0.110	0.715	−3.387	2.544	
CHIN	−0.015	0.052	0.573	−1.000	1.000	
INTWO	0.263	0	0.442	0	1	
OENEG	0.038	0	0.193	0	1	
Excess Return	−0.131	−0.289	0.631	−1.246	2.503	
Relative Size	−5.284	−5.320	0.659	−6.838	−2.821	
π_{KMV}	0.114	0.001	0.241	0.000	0.989	

Note that Ohlson (1980) suggested using the first 9 variables as predictive variables. But in this dissertation we only used the first 8 variables as the predictive variables in our case-control data analysis. The $9th$ variable, SIZE, was not used as a predictive variable because the total asset had already been used as the matching factor in the process of selecting the case-control sample for study. The last 2 variables are the market-driven variables used in Shumway (2001).

Pairwise scatter diagrams of our case-control sample on the continuous variables are presented in Figure 1. From the figure, it is clear that the distributions of these variables are fat-tailed and skewed, and it is very difficult to perform bankruptcy prediction visually, since most data points are clustered together.

The summary statistics of the 10 predictive variables considered in our case-control data analysis are presented in Table 3. For each of these 10 variables, the two-sample median test was performed to test the null hypothesis of equal magnitude for nonbankrupt company and for bankrupt company. The p-value in Table 3 shows that the null hypothesis of equal magnitude for cases and controls is significant at 0.05 level for each predictive variable. This result indicates that each of these variables should be an effective predictive variable. On the other hand, the summary statistics and the frequency distribution of the values of π_{KMV} for the selected companies in our case-control data analysis are shown respectively in Table 3 and Figure 2. The results also indicate that π_{KMV} has good predictive power.

Given our case-control sample, the bankruptcy prediction rules associated with DAM, LLM, KMV and SLM were estimated. Their performance was measured by the out-of-sample error rate. The out-of-sample error rate was computed on each of the 100 testing samples randomly selected from the given case-control sample. Each testing sample was composed of 50% of bankrupt companies and their matched nonbankrupt companies. The data not included in the testing sample were taken as the training sample, and were used to develop the prediction rule.

Under SLM, kernel function K was taken as the Epanechnikov kernel $K(u) = (3/4)(1-u^2)\, I(|u| \leq 1)$. To compute the out-of-sample error rate for the prediction rule based on SLM on each testing sample, the procedure given in Remark 2.2 for computing the in-sample total error rate $\tau_{in}(p,\, b_\theta,\, b_H) = \alpha_{in}(p,\, b_\theta,\, b_H) + \beta_{in}(p,\, b_\theta,\, b_H)$ on the training sample was applied to choose the values of $(p,\, b_\theta,\, b_H)$. We computed $\tau_{in}(p,\, b_\theta,\, b_H)$ on the equally spaced logarithmic grid of $1001 \times 501 \times 501$ values of $(p,\, b_\theta,\, b_H)$ in $[0, 1] \times [1/10, 15] \times [1/10, 15]$. Given each value of $u \in [0, 1]$, the global minimizer $\{\hat{p}(u), \hat{b}_\theta(u), \hat{b}_H(u)\}$ of $\tau_{in}(p,\, b_\theta,\, b_H)$ on the grid points with the restrictions $\alpha_{in}(p,\, b_\theta,\, b_H) \leq u$ and $b_H > b_\theta$ was taken as the selected values for $(p,\, b_\theta,\, b_H)$.

Using the selected values of $\{\hat{p}(u), \hat{b}_\theta(u), \hat{b}_H(u)\}$ and the training sample, the values $\hat{H}(x_j)$ and $\hat{\theta}$ were computed for each data point $(x_j,\, z_j)$ in the testing sample. The company with the predictor values $(x_j,\, z_j)$ in the testing sample was classified as a

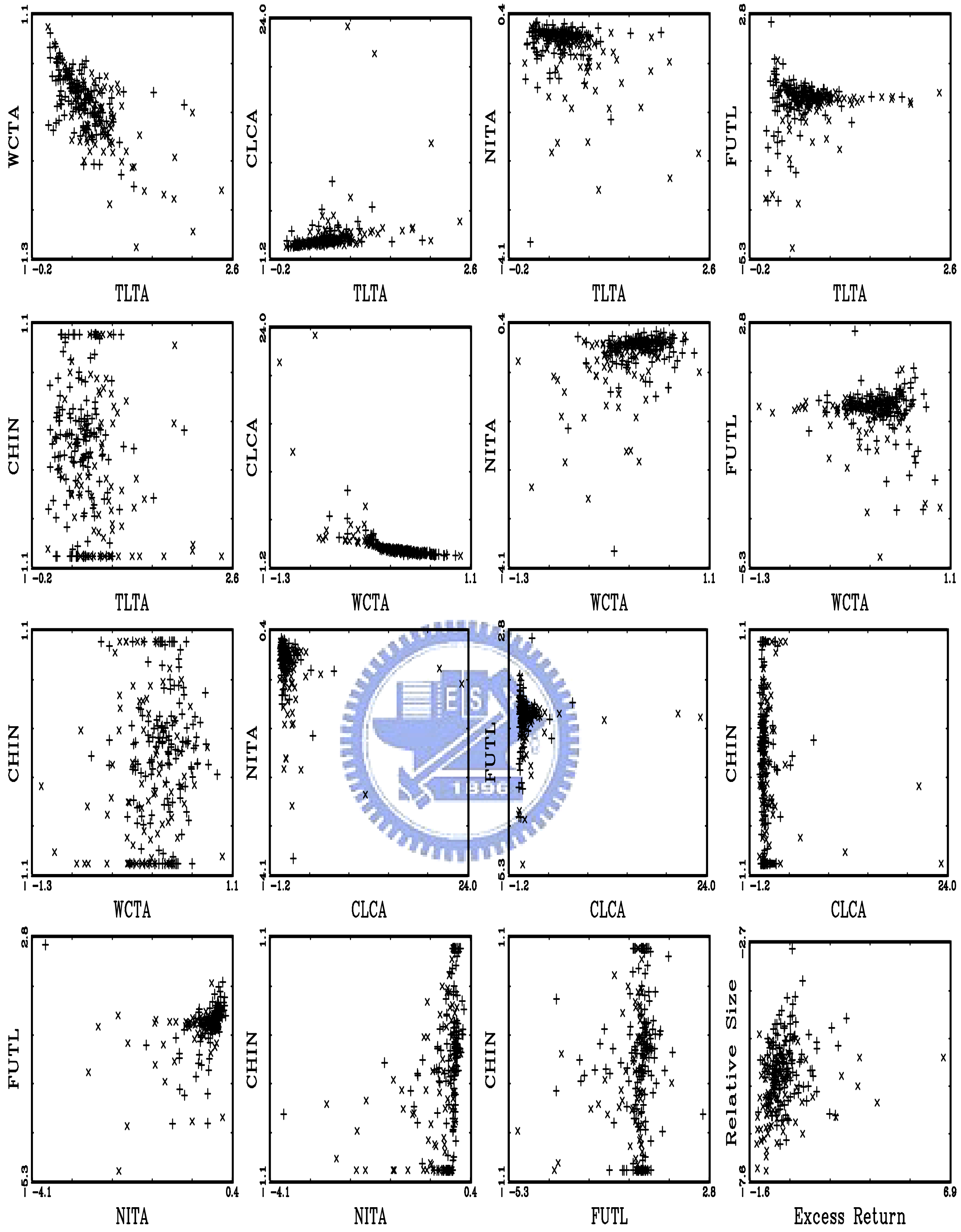

Figure 1: Pairwise scatter diagrams of our case-control sample. Given the values of Shumway's 2 market-driven variables, Excess Return and Relative Size, and Ohlson's 6 continuous variables in our case-control sample, their pairwise scatter diagrams are presented. Each panel plots 156 nonbankrupt companies (pluses) and 79 bankrupt companies (stars) selected from COMPUSTAT and CRSP databases.

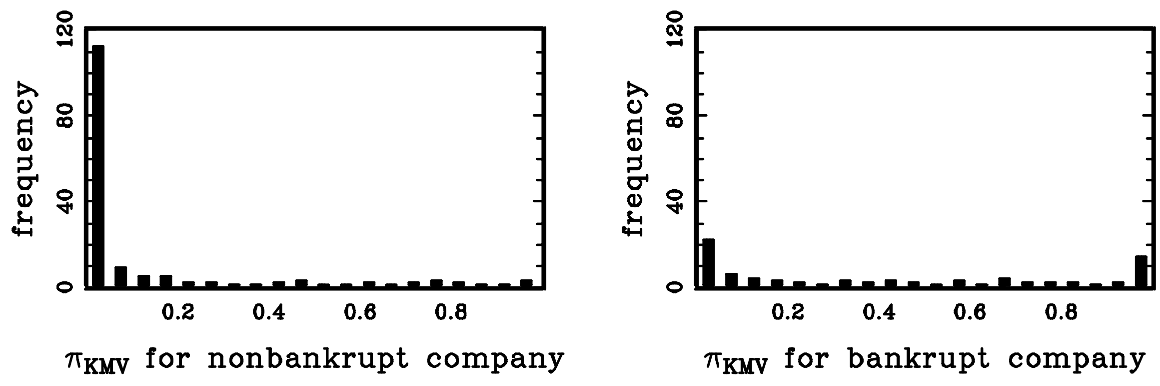

Figure 2: The frequency histogram of the values of π_{KMV} in our case-control sample. The frequency histogram of the values of π_{KMV} for the 156 nonbankrupt companies, and that for the 79 bankrupt companies in our case-control sample are plotted in the left and the right panels, respectively.

bankrupt company if

$$\hat{\psi}_j = \frac{\exp\{\hat{H}(x_j) + \hat{\theta}\ z_j\}}{1 + \exp\{\hat{H}(x_j) + \hat{\theta}\ z_j\}} > \hat{p}(u),$$

otherwise a healthy company. After the classification procedure was completed for each company in the testing sample, the out-of-sample error rates

$$\alpha_{SLM}(u) = \frac{\sum_{j:(x_j,z_j) \text{ in testing sample}} Y_j\ I\{\hat{\psi}_j \leq \hat{p}(u)\}}{\sum_{j:(x_j,z_j) \text{ in testing sample}} Y_j},$$

$$\beta_{SLM}(u) = \frac{\sum_{j:(x_j,z_j) \text{ in testing sample}} (1 - Y_j)\ I\{\hat{\psi}_j > \hat{p}(u)\}}{\sum_{j:(x_j,z_j) \text{ in testing sample}} (1 - Y_j)},$$

$$\tau_{SLM}(u) = \alpha_{SLM}(u) + \beta_{SLM}(u),$$

of the bankruptcy prediction rule based on SLM were computed, for each given value of u. For the given value of u, $\alpha_{SLM}(u)$ is the out-of-sample type I error rate of classifying the bankrupt companies to healthy ones, and $\beta_{SLM}(u)$ is the out-of-sample type II error rate of classifying the healthy companies to bankrupt ones from the testing sample. After the computational procedure was completed for each testing sample, the average of each out-of-sample error rate over the 100 testing samples was computed.

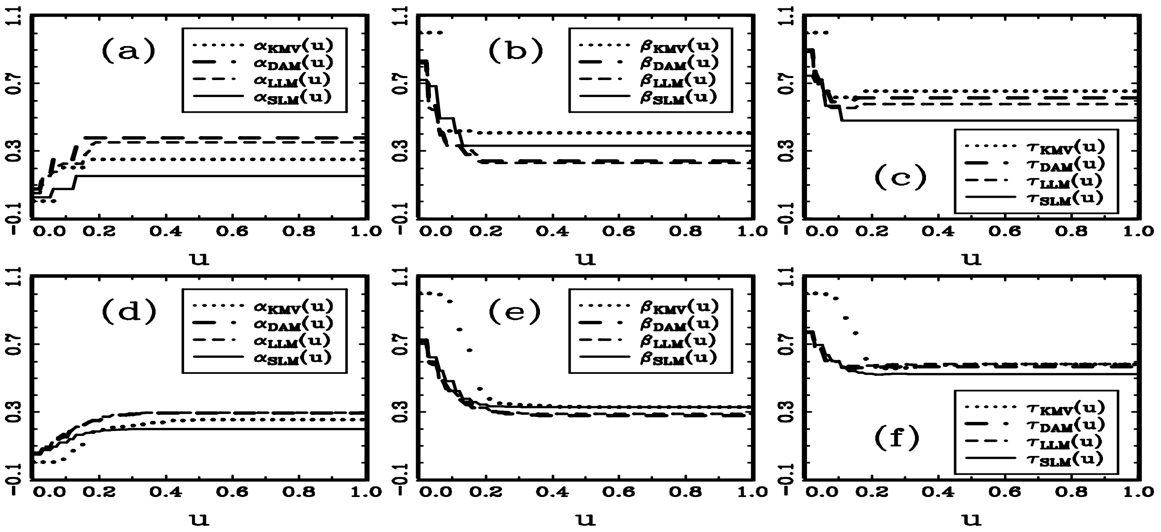

Figure 3: The out-of-sample error rates obtained by applying KMV, DAM, LLM, and SLM to our case-control sample. Panels (a)-(c) show three out-of-sample error rates of the prediction methods derived from one testing sample. Panels (d)-(f) show sample averages of the three out-of-sample error rates over the 100 testing samples. Each testing sample was composed of 50% of bankrupt companies and their matched nonbankrupt companies in our case-control sample.

The same computational procedures were applied to the prediction rules based on DAM, LLM and KMV. Let $\{\alpha_{DAM}(u),\ \beta_{DAM}(u),\ \tau_{DAM}(u)\}$, $\{\alpha_{LLM}(u),\ \beta_{LLM}(u),\ \tau_{LLM}(u)\}$ and $\{\alpha_{KMV}(u),\ \beta_{KMV}(u),\ \tau_{KMV}(u)\}$ be similarly defined as the out-of-sample error rates for DAM, LLM and KMV. The prediction results obtained by applying the four discussed bankruptcy prediction rules to our case-control data are shown in Figure 3 and Table 4.

Figure 3 presents the three (averaged) out-of-sample error rates for the four prediction models under one (one hundred) testing sample(s). These error rates were derived under the constraint that the type I error rate was at most u. If no such constraint is required, we simply take $u = 1$ and the related out-of-sample error rates are given in Table 4. For the case of $u = 1$, both SLM and KMV give smaller out-of-sample type I error rates than DAM and LLM. Nevertheless, KMV has the largest out-of-sample type II error rate among the four competing prediction rules. DAM and LLM show rather similar behavior in the sense of having almost the same averaged out-of-sample types I and II error rates. In terms of the total error rate, however, Table 4 confirms that SLM

Table 4: Numerical results of the out-of-sample error rates obtained by applying KMV, DAM, LLM, and SLM to our case-control sample. Given the value of $u = 1$, the values of the three out-of-sample error rates shown in (a)-(c) of Figure 3 are presented, and those shown in (d)-(f) of Figure 3 are given in parentheses.

	KMV	DAM	LLM	SLM
type I error rate	0.250 (0.253)	0.375 (0.290)	0.350 (0.296)	0.200 (0.202)
type II error rate	0.405 (0.328)	0.241 (0.278)	0.228 (0.287)	0.291 (0.321)
total error rate	0.655 (0.581)	0.616 (0.568)	0.578 (0.583)	0.491 (0.523)

has the best overall performance. Thus it is fair to say that by a reasonable margin, the most accurate model listed in Table 4 is the SLM.

From Figure 3, we find out that the similar conclusions as those shown in Table 4 can also be reached. For $u \leq 0.2$, KMV has the smallest averaged out-of-sample type I error rate. However, it also has the largest averaged type II error rate in this range. For $u > 0.2$, KMV has similar averaged type II error rate as SLM but larger type I error rate than SLM. For $u \in [0, 1]$, DAM and LLM show very similar performance. However, comparing the four prediction rules based on averaged out-of-sample total error rate, Figure 3 shows that SLM has the best overall performance.

2.4 A Simulation Study

In this section, a simulation study was performed to compare the performance of the prediction rules based on DAM, LLM and SLM. We first introduce the simulation settings. The dimension of the continuous predictor X was $d = 2$, and that of the discrete predictor Z was $q = 1$. Two skewed and fat-tailed distributions for the simulated $X = (X_1, X_2)$ were considered. In the first case, the skewed Student t distribution (Fernandez and Steel, 1998) with degrees of freedom k and scale parameter s was considered. The simulated control (nonbankruptcy) X_i values were taken from the skewed Student t distribution with $(k, s) = (3, 2)$ for $i = 1$, and $(7, 4)$ for $i = 2$, and those for case (bankruptcy) values $(k, s) = (5, -3)$ for $i = 1$, and $(5, 2)$ for $i = 2$ were used. In the second case, the Pareto distribution (Siegrist, 2005) with shape parameter a and scale parameter s was considered. Similarly, the values of (a, s) of both Pareto random variables X_1 and X_2 for controls were $(3, 2)$ and $(7, 4)$, and those for cases were $(5, -3)$

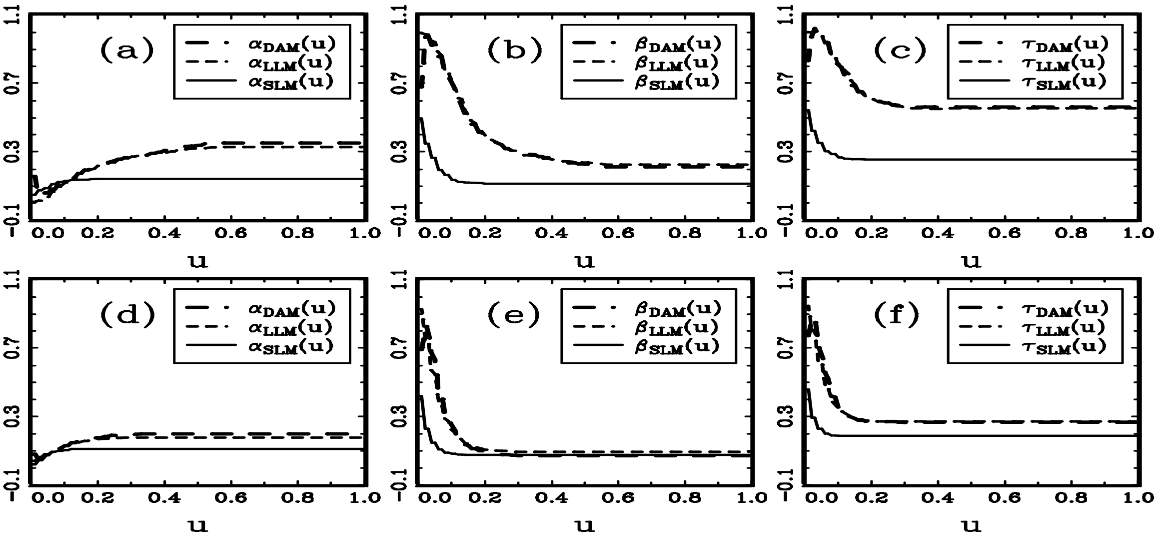

Figure 4: The out-of-sample error rates obtained by applying DAM, LLM, and SLM to our simulated case-control data with the skewed Student t distribution for X. Given the case of $(\mu_0, \mu_1) = (-0.1, 0.1)$, panels (a)-(c) show sample averages of the three out-of-sample error rates of the prediction methods over the 100 simulated case-control data sets. For each simulated case-control data set, one testing sample was randomly selected, and was composed of 50% of cases and their matched controls. The corresponding results for the case of $(\mu_0, \mu_1) = (-0.3, 0.3)$ are shown in panels (d)-(f).

and (5, 2), respectively.

Given marginal distributions of X, the simulated control X values with size 200 were generated using mean vector $(\mu_0, 0)$ and covariance matrix $\begin{pmatrix} 1/25, & -1/250 \\ -1/250, & 1/25 \end{pmatrix}$, and their associated Z values were independently generated from a binary random variable with the probability $p(Z = 1) = 1/3$. The simulated case X values with size 100 were similarly generated with mean vector $(\mu_1, 0)$ and covariance matrix $\begin{pmatrix} 1/4, & 1/8 \\ 1/8, & 1/4 \end{pmatrix}$, and their associated Z values were independently generated from a binary random variable with the probability $p(Z = 1) = 2/3$. Two sets of the values $(\mu_0, \mu_1) = (-0.1, 0.1)$ and $(-0.3, 0.3)$ were considered. For each distribution of X and each set of the values (μ_0, μ_1), one hundred independent sets of the case-control data were generated. Given each case-control data set, one testing sample was randomly selected, and was composed of 50% of cases and their matched controls.

Three bankruptcy prediction methods based on DAM, LLM and SLM were con-

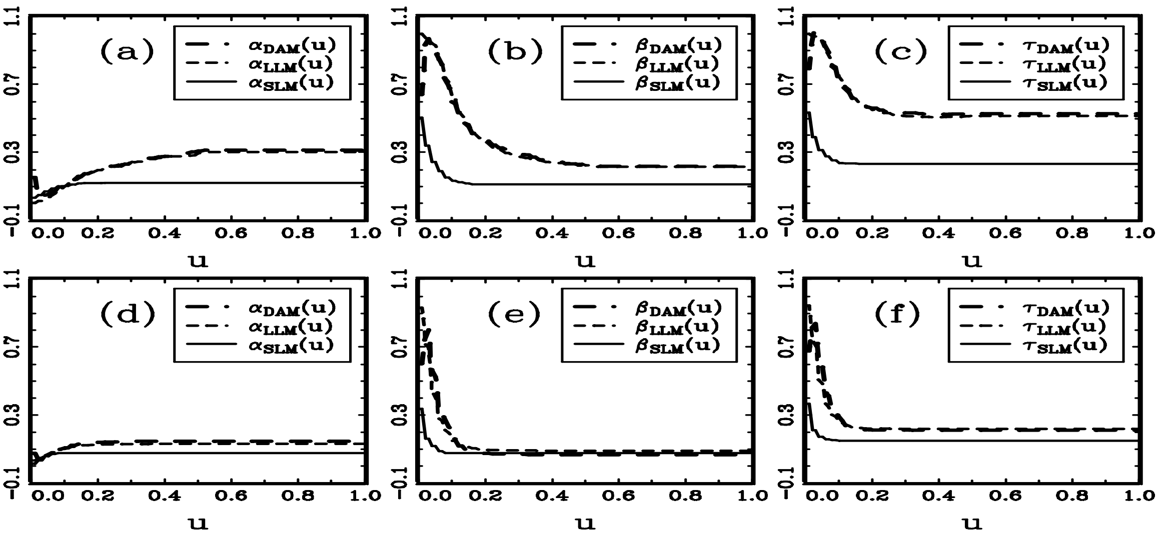

Figure 5: The out-of-sample error rates obtained by applying DAM, LLM, and SLM to our simulated case-control data with the Pareto distribution for X. The captions for (a)-(f) of Figure 5 are the same as those of Figure 4 with the skewed Student t distribution of X replaced by the Pareto distribution for X.

sidered in this simulation study. The computational procedures and the measures of performance presented in Section 2.3 were applied to the three prediction methods. For SLM, the equally spaced logarithmic grid of 201×201 values of $(b_\theta,\ b_H)$ in $[1/10,\ 2] \times [1/10,\ 2]$ were employed for selecting values of $(b_\theta,\ b_H)$, and the Epanechnikov kernel $K(u) = (3/4)\ (1-u^2)\ I(|u| \leq 1)$ was used. The simulation results are presented in Figures 4 and 5.

Figure 4 presents averages of out-of-sample error rates over the 100 simulated data sets for the three bankruptcy prediction methods under the case of the skewed Student t distribution for X. From the figure, our SLM performs better than DAM and LLM, since for most values of u, our prediction method has smaller average of out-of-sample error rate of any type. Further, the smaller the difference $|\mu_0 - \mu_1|$ is, the larger advantage of SLM will achieve over DAM and LLM. The forecasting performance of the three prediction methods under the case of the Pareto distribution for X is shown in Figure 5. The results from Figure 5 also confirm that SLM has the best overall performance.

2.5 Discussion

In this chapter, bankruptcy prediction methods based on the SLM are proposed for the prospective and the case-control data. Our SLM is developed by replacing the linear logit function of the LLM with an unknown but smooth logit function. Hence it is more flexible and robust than the LLM. The estimators for the unknown quantities in the SLM are computed by the local likelihood method, and their large sample properties are studied through their asymptotic biases and variances. We point out that, under the case-control data, the estimated bankruptcy probability does not estimate the true bankruptcy probability, unless the quantity $c^* = \log\{p(D = 0)/p(D = 1)\} + \log(n_1/n_0)$ is 0. In contrast, for the prospective data, our estimated probability does estimate the true bankruptcy probability. This is the major difference between the applications of the logit model to the case-control sample and the prospective sample. However, using the fact that the logistic distribution is strictly monotonically increasing, we discover that such estimated probability can still be used to develop a bankruptcy prediction device. To decide the optimal prediction rule, we propose to control the in-sample type I (II) error rate to be at most u, so that the sum of in-sample type I and II error rates is minimal. This is sometimes essential since the type I (II) error would cause much more severe losses to the investors. The value of $u \in [0, 1]$ is determined by the investor. If $u = 1$, then there is no restriction on the magnitude of in-sample type I and II error rates. Our results from one real data example based on eight predictor variables of Ohlson (1980) and two market-driven variables of Shumway (2001) and simulations confirm that the SLM performs better than the DAM, LLM and KMV, in the sense of having smaller out-of-sample total error rate.

In applications of the SLM to the bankruptcy prediction problem, we need to decide proper values of the bandwidth parameters b_θ and b_H. In this chapter, we suggest to estimate the bandwidth parameters so that the sum of the corresponding in-sample type I and II errors is minimized subject to some restrictions. This approach may suffer from the heavy computational burden. One possible remedy for this drawback is to use the plug-in method to estimate these bandwidth parameters. For example, we may determine the bandwidth parameter to minimize the estimated mean square error of each estimator $\hat{H}(x)$ and $\hat{\theta}$. For more discussion of the plug-in method, see Jones, Marron, and Sheather (1996).

Two possible extensions of SLM are outlined below. Firstly, in the SLM, we assume that $H(x)$ is an unknown but smooth function, and a local likelihood method has been developed to estimate $H(x)$. However, the resulting estimator $\hat{H}(x)$ suffers from the curse of dimensionality, that is, as the dimension of the continuous predictor X increases, the performance of the resulting $\hat{H}(x)$ deteriorates. For example, from Remark 1 of Appendix A, the minimum mean square error of $\hat{H}(x)$ with respect to b_H is of order $n^{-4/(d+4)}$ in magnitude. Such mean square error increases as the value of d increases. To avoid such drawback, one possible remedy is to consider an additive model for $H(x)$:

$$H(x) = H_1(x_1) + \cdots + H_d(x_d),$$

as described by Hastie and Tibshirani (1990). Here $x = (x_1, \cdots, x_d)^T$ and $H_i(x_i)$ is any unknown but smooth function of x_i, the i-th component of x, for each $i = 1, \cdots, d$.

Secondly, the logit function of our SLM is basically an additive model with $H(X)$ and $\theta\, Z$. This assumption will be violated if X and Z are interactive. A possible solution to this problem is to introduce a nonparametric interaction such as $G(X)\, Z$, where $G(\cdot)$ is a q-dimensional row vector of unknown but smooth functions, in the model. It will be interesting to study the estimates of functions $H(x)$ and $G(x)$ simultaneously.

2.6 Sketches of the Proofs

In this section, sketches of the proofs for Theorem 2.1 will be given. The following notations will be used. Let $\ell_j^{(1)}$ and $\ell_j^{(2)}$ be the gradient vector and the Hessian matrix of ℓ_j, for each j = 1,2,3, given in (22)-(24), respectively. Also, $H^{(2)}(x)$ is the Hessian matrix of $H(x)$. Define P_0 as the event that the number of control data points falling into the neighborhood $N(x;\, b/2)$ of x is less than $\rho_0 n_0 \int_{N(x;\, b/2)} f_0(t)\, dt$, where ρ_0 is a positive constant satisfying $\rho_0 \leq 1/4$, and Q_0 the event that the number of control data points falling into the neighborhood $N(x;\, b)$ of x is greater than $\varphi_0 n_0 \int_{N(x;\, b)} f_0(t)\; dt$, where φ_0 is a positive constant satisfying $\varphi_0 \geq \exp(1)$. The definition of neighborhood $N(x;\, b)$ of the given point x has been given in Section 1.4. The events P_1 and Q_1 are similarly defined for case data points with n_0, f_0, ρ_0, and φ_0 replaced respectively by n_1, f_1, ρ_1, and φ_1.

The proofs of the asymptotic bias and variance for each of the estimators $\hat{H}_1(x)$, $\hat{\theta}$, and $\hat{H}(x)$ are given below in sequence.

Proof of the asymptotic bias and variance for $\hat{H}_1(x)$. Set $\eta = (\alpha,\, \beta,\, \theta)^T$ and $\hat{\eta} = (\hat{\alpha},\, \hat{\beta},\, \hat{\theta})^T$, the maximizer of $\ell_1(\alpha, \beta, \theta;\, x)$ in (22). By the first order Taylor expansion, we have

$$0 = \ell_1^{(1)}(\hat{\eta};\; x) = \ell_1^{(1)}(\eta;\; x) + \ell_1^{(2)}(\eta^*;\; x)\; (\hat{\eta} - \eta), \tag{31}$$

for each $x \in [0,\, 1]^d$, where η^* lies in the line segment connecting η and $\hat{\eta}$.

Using conditions (C1)-(C5), (21), and the large deviation theorem in Section 10.3.1 of Serfling (1980), a straightforward calculation leads to the following asymptotic results: as $n \to \infty$,

$$P(P_0 \cup Q_0 \cup P_1 \cup Q_1) = O\{\exp(-n\; b_\theta)\}, \tag{32}$$

$$E\{\ell_1^{(1)}(\eta;\; x)\} = (1/2)\; n\; b_\theta^2\; \zeta(1-\zeta)\; A_1 + O(n\; b_\theta^3 + b_\theta^{-d}), \tag{33}$$

$$E\{\ell_1^{(2)}(\eta;\; x)\} = (-1)\; n\; \zeta(1-\zeta)\; B_1 + O(n\; b_\theta), \tag{34}$$

$$Var\{\ell_1^{(1)}(\eta;\; x)\} = n\; b_\theta^{-d}\; \zeta(1-\zeta)\; C_1 + O(n\; b_\theta^{-d+1}), \tag{35}$$

for each η. Here

$$A_1 = \{u_0(x)\; D_0(x),\; u_1^T(x)\; D_0(x),\; u_0(x)\; D_1^T(x)\}^T,$$

$$B_1 = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} \Lambda_1(x,t)\; K^{\#}(t)\; dt,$$

$$C_1 = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} \Lambda_1(x,t)\; K^{\#}(t)^2\; dt,$$

where

$$u_0(x) = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} \{t^T\ H^{(2)}(x)\ t\}\ K^{\#}(t)\ dt,$$

$$u_1(x) = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} \{t^T\ H^{(2)}(x)\ t\}\ t\ K^{\#}(t)\ dt,$$

$$\Lambda_1(x, t) = \begin{pmatrix} D_0(x), & t^T\ D_0(x), & D_1^T(x) \\ t\ D_0(x), & t\ t^T\ D_0(x), & t\ D_1^T(x) \\ D_1(x), & D_1(x)\ t^T, & D_2(x) \end{pmatrix}.$$

Using condition (C3) and the results of (32)-(35) and comparing the magnitudes of $\ell_1^{(1)}(\eta;\ x) = O_p(n\ b_\theta^2 + n^{1/2}\ b_\theta^{-d/2})$ and $\ell_1^{(2)}(\eta^*;\ x) = O_p(n)$ in (31), we have

$$\hat{\eta} - \eta = o_p(1). \tag{36}$$

Using (31)-(36) and approximations to the standard errors of functions of random variables in Section 10.5 of Stuart and Ord (1987), the results of the asymptotic bias and variance of $\hat{H}_1(x)$ in (25) and (26) follow, respectively.

Proof of the asymptotic bias and variance for $\hat{\theta}$. Set $\phi = (\alpha_0,\ \theta)$ and $\hat{\phi} = (\hat{\alpha}_0,\ \hat{\theta})$, the maximizer of $\ell_2(\alpha_0,\ \theta)$ in (23). Using the fact that α_0 is a normalizing constant for $f_1(x,\ z) = f_0(x,\ z)\ \exp\{\alpha_0 + \hat{H}_1(x) + \theta\ z\}$, the results of the asymptotic bias and variance of $\hat{H}_1(x)$ in (25) and (26), (C1)-(C5), (21), and approximations to the standard errors of functions of random variables, through a straightforward calculation, we have

$$\exp(\alpha_0) = (1 + b_\theta^2\ c_1 + c_2)^{-1}\ \{1 + o_p(1)\},$$

where

$$c_1 = \int_0^1 \cdots \int_0^1\ f_1(x)\ c_H(x;\ b_\theta)\ dx, \qquad c_2 = \int_0^1 \cdots \int_0^1\ f_1(x)\ [\hat{H}_1(x) - E\{\hat{H}_1(x)\}]\ dx.$$

Next, using this result and (C1)-(C5), through a straightforward calculation, we have, as $n \to \infty$,

$$E\{\ell_2^{(1)}(\phi)\} = (-1)\ n\ b_\theta^2\ \zeta(1-\zeta)\ \times$$

$$[\int_0^1 \cdots \int_0^1 \{D_0(x),\ D_1^T(x)\}\ c_H(x; b_\theta)\ dx\ -\ (D_0,\ D_1^T)\ c_1] + O(n\ b_\theta^3), \tag{37}$$

$$E\{\ell_2^{(2)}(\phi)\} = (-1)\ n\ \zeta(1-\zeta)\ D + O(n\ b_\theta), \tag{38}$$

$$Var\{\ell_2^{(1)}(\phi)\} = n\ \zeta(1-\zeta)\ \{D\ -\ (D_0,\ D_1^T)^T\ (D_0,\ D_1^T)\} + O(n\ b_\theta), \tag{39}$$

for each ϕ.

Following the same arguments as those of (31) and (36) and using (37)-(39), we have $\hat{\phi} - \phi = o_p(1)$. Combining this result and using approximations to the standard errors of functions of random variables in Section 10.5 of Stuart and Ord (1987), the results of the asymptotic bias and variance of $\hat{\theta}$ in (27) and (28) follow, respectively.

Proof of the asymptotic bias and variance for $\hat{H}(x)$. Set $\xi = (\alpha^*, \beta)^T$ and $\hat{\xi} = (\hat{\alpha}^*, \hat{\beta})^T$, the maximizer of $\ell_3(\alpha^*, \beta; x)$ in (24), where $\alpha^* = H^*(x) + \alpha_1$ and α_1 is a normalizing constant. Using (C1)-(C5), (21), the asymptotic bias and variance of $\hat{\theta}$ in (27) and (28), and approximations to the standard errors of functions of random variables, we have, as $n \to \infty$,

$$E\{\ell_3^{(1)}(\xi;\ x)\} = (1/2)\ n\ b_H^2\ \zeta(1-\zeta)\ A_3 + O(n\ b_H^3), \tag{40}$$

$$E\{\ell_3^{(2)}(\xi;\ x)\} = (-1)\ n\ \zeta(1-\zeta)\ B_3 + O(n\ b_H), \tag{41}$$

$$Var\{\ell_3^{(1)}(\xi;\ x)\} = n\ b_H^{-d}\ \zeta(1-\zeta)\ C_3 + O(n\ b_H^{-d+1}), \tag{42}$$

for each ξ, where

$$A_3 = \{u_0(x),\ u_1^T(x)\}^T\ D_0(x),$$

$$B_3 = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} \Lambda_3(x,t)\ K^{\#}(t)\ dt,$$

$$C_3 = \int_{m_1}^{k_1} \cdots \int_{m_d}^{k_d} \Lambda_3(x,t)\ K^{\#}(t)^2\ dt,$$

$$\Lambda_3(x,t) = \begin{pmatrix} 1, & t^T \\ t, & t\,t^T \end{pmatrix} D_0(x).$$

Following the same arguments as those of (31) and (36) and using (32) and (40)-(42), we have $\hat{\xi} - \xi = o_p(1)$. Combining this result with (40)-(42) and using approximations to the standard errors of functions of random variables, the results of the asymptotic bias and variance of $\hat{H}(x)$ in (29) and (30) follow, respectively. Hence the proof of Theorem 2.1 is completed.

CHAPTER III

DYNAMIC PREDICTION METHODS FOR BANKRUPTCY AND FINANCIAL DISTRESS

3.1 Introduction

As introduced in Sections 1.1 and 1.7, the DSM (19) is a dynamic forecasting method against the static forecasting method LLM (1). The DSM (19) has the advantage of using all available historical information to determine each firm's bankruptcy risk at each point in time; but the LLM (1) uses only one set of predictor values collected at a specific time point for each firm.

In this chapter, we shall first study asymptotic properties of maximum likelihood estimators $\hat{\psi} = (\hat{\alpha}_0,\ \hat{\alpha}_1,\ \hat{\beta},\ \hat{\theta})^T$ given in Section 1.5 for the parameters $\psi = (\alpha_0,\ \alpha_1,\ \beta,\ \theta)^T$ in DSM (19) with discrete-time survival data. Then the practical performance of the bankruptcy prediction method based on the DSM (19) with discrete-time survival data is studied through a real data example. For these, the composition of the discrete-time survival data and the formulation of the DSM (19) are recalled.

According to the discrete-time survival sampling introduced in Section 1.3, the discrete-time survival data are expressed as

$$(t_i,\ Y_i,\ x_{i,1},\ \cdots,\ x_{i,t_i},\ z_{i,1},\ \cdots,\ z_{i,t_i}),\quad \text{for}\quad i = 1,\ \cdots,\ n.$$

Here $t_i \in \{1,\ 2,\ \cdots,\ m\}$ denotes the duration time of the i-th company in the sampling period, and m is a positive integer standing for the length of the sampling period. Also, at the duration time t_i, $Y_i = 0$ indicates that the i-th company is nonbankrupt, and 1 the i-th company is bankrupt. Further, $x_{i,j}$ and $z_{i,j}$ are values of the d-dimensional continuous and q-dimensional discrete explanatory variables X and Z collected at the duration time j, respectively in each case, for each $j = 1,\ \cdots,\ t_i$ and for the i-th company.

Combining the discrete-time survival data, the logistic hazard function, and the natural logarithm function of the duration time, the DSM (19) is expressed by the log-likelihood function for the discrete-time survival data as

$$\begin{aligned}\ell_{DSM}(\psi) \;=\; & \sum_{i=1}^{n} Y_i \,\{\alpha_0 + \alpha_1 \;\log(t_i) + \beta\; x_{i,t_i} + \theta\; z_{i,t_i}\} \;- \\ & \sum_{i=1}^{n} \sum_{j=1}^{t_i} \log[1 + \exp\{\alpha_0 + \alpha_1 \;\log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}].\end{aligned}$$

Using the log-likelihood function $\ell_{DSM}(\psi)$, maximum likelihood estimators $\hat{\psi} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}, \hat{\theta})^T$ for $\psi = (\alpha_0, \alpha_1, \beta, \theta)^T$ may be taken as the solution of the normal equations

$$\begin{aligned}0 \;=\; & \frac{\partial \ell_{DSM}(\psi)}{\partial \psi} \\ =\; & \sum_{i=1}^{n} Y_i \begin{bmatrix} 1 \\ \log(t_i) \\ x_{i,t_i} \\ z_{i,t_i} \end{bmatrix} - \sum_{i=1}^{n} \sum_{j=1}^{t_i} \frac{\exp\{\alpha_0 + \alpha_1 \;\log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}}{1 + \exp\{\alpha_0 + \alpha_1 \;\log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}} \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix}.\end{aligned}$$

By the functional form of the normal equations, we can not derive a closed-form solution for $\psi = (\alpha_0, \alpha_1, \beta, \theta)^T$ from the normal equations. But, practically, there are many software packages including, for example, S-plus, Gauss, and SAS providing available procedures to solve the normal equations. In this dissertation, we use the Gauss software to process the computational work.

This chapter is organized as follows. Section 3.2 presents asymptotic properties of maximum likelihood estimators $\hat{\psi} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}, \hat{\theta})^T$ for the DSM (19) with the discrete-time survival data. To illustrate the bankruptcy prediction method based on the DSM (19) with the discrete-time survival data, a real data set is analyzed in Section 3.3. Finally, the concluding remarks and future research topics are given in Section 3.4.

3.2 Theoretical Results

In this section, we shall study the asymptotic properties of maximum likelihood estimators $\hat{\psi} = (\hat{\alpha}_0,\, \hat{\alpha}_1,\, \hat{\beta},\, \hat{\theta})^T$ for the parameters $\psi = (\alpha_0,\, \alpha_1,\, \beta,\, \theta)^T$ in the DSM (19).

By the properties of maximum likelihood estimators given in Section 3.3 of Cox and Oakes (1984), $\hat{\psi} = (\hat{\alpha}_0,\, \hat{\alpha}_1,\, \hat{\beta},\, \hat{\theta})^T$ are asymptotically normally distributed as

$$\hat{\psi} \;\approx\; N\,\{\psi,\; I_n(\psi)^{-1}\},$$

where

$$\begin{aligned} I_n(\psi) &= (-1)\; E\left\{\frac{\partial^2\, \ell_{DSM}(\psi)}{\partial\psi\; \partial\psi^T}\right\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{t_i} \frac{\exp\{\alpha_0 + \alpha_1\; \log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}}{[1+\exp\{\alpha_0 + \alpha_1\; \log(j) + \beta\; x_{i,j} + \theta\; z_{i,j}\}]^2} \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix} \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix}^T . \end{aligned}$$

In practice, we can use

$$\begin{aligned} I_n(\hat{\psi}) &= (-1)\; \frac{\partial^2\, \ell_{DSM}(\psi)}{\partial\psi\; \partial\psi^T}\bigg|_{\psi=\hat{\psi}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{t_i} \frac{\exp\{\hat{\alpha}_0 + \hat{\alpha}_1\; \log(j) + \hat{\beta}\; x_{i,j} + \hat{\theta}\; z_{i,j}\}}{[1+\exp\{\hat{\alpha}_0 + \hat{\alpha}_1\; \log(j) + \hat{\beta}\; x_{i,j} + \hat{\theta}\; z_{i,j}\}]^2} \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix} \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix}^T \end{aligned}$$

to estimate $I_n(\psi)$. Replacing the quantity $I_n(\psi)^{-1}$ in the above asymptotic normal distribution of $\hat{\psi}$ with its estimate $I_n(\hat{\psi})^{-1}$, the resulting distribution can be used to derive the confidence interval estimate and test the values of ψ.

3.3 A Real Data Example

In this section, the DSM (19) with the discrete-time survival data was applied to the data occurred in Taiwan. Our discrete-time survival data were drawn by three steps. In the first step, the sampling period was taken as the one during January of the year 1981 to December of the year 1999, and the sampling criterion was defined as those firms starting to be listed in Taiwan Stock Exchange during the sampling period. By characteristics of industries, the financial institutions (with industry code M2800: Banking and Insurance) were eliminated from the sample due to the unique capital requirements and regulatory structure in that industry group. Also, electronic companies (with industry code M2300: Electron) were not considered because such companies provide much less historical data and have much smaller financial failure rate, compared to traditional companies. Further, the companies providing incomplete values of explanatory variables were excluded. Hence, our discrete-time survival data were selected from the traditional companies providing complete values of explanatory variables.

In the second step, 249 companies satisfying the above sampling considerations were selected, and called the in-sample companies. Finally, in the third step, the historical data of the 249 selected in-sample companies were drawn from the financial database provided by Taiwan Economic Journal Co. Ltd.

Note that since there are only few bankrupt companies among the 249 selected ones, it is difficult to predict bankruptcy well. In this case, to provide more failure companies to proceed our research, we replaced our target on bankruptcy prediction with financial distress prediction. According to the definition of financial distress given by Taiwan Stock Exchange, financial distress companies are those whose stocks were delisted, stopped trading, or traded by cash.

For predicting financial distress, two different sets of predictors were considered in the DSM (19). The first set of predictors was that used in Altman (1968), and the second was that employed in Zmijewski (1984). The two sets of predictors are given as follows.

Altman's predictors:

1. WC/TA = Working capital / Total assets.

2. RE/TA = Retained earnings / Total assets.

Table 5: The information about industry and financial status of in-sample companies.

Industry	number of distress companies	number of healthy companies
M1100 Cement	0	4
M1200 Food	3	17
M1300 Plastics	2	14
M1400 Textiles	3	35
M1500 Electric, Machinery	1	16
M1600 Appliance, Cable	0	6
M1700 Chemica	0	15
M1800 Glass, Ceramics	1	5
M1900 Paper, Pulp	1	1
M2000 Steel, Iron	6	21
M2100 Rubber	0	6
M2200 Automobile	3	3
M2500 Construction	4	28
M2600 Transportation	1	14
M2700 Tourism	0	4
M2900 Department Stores	0	7
M9900 Other	4	24
Total companies	29	220

3. EBIT/TA = Earnings before interest and taxes / Total assets.

4. ME/TL = Market value of equity / Book value of total debts.

5. S/TA = Sales / Total assets.

Zmijewski's predictors:

1. NI/TA = Net income / Total assets.

2. TL/TA = Total debts / Total assets.

3. CA/CL = Current assets / Current liabilities.

Since there is no discrete predictor considered by Altman and Zmijewski, our discrete-time survival data only contain continuous predictors. To perform the financial distress prediction, the duration time t required by the DSM (19) was taken as the firm's trading age (Shumway 2001).

The values of Altman's and Zmijewski's predictors were collected for our selected 249 in-sample companies from the Taiwan Economic Journal database. For each se-

Table 6: Summary statistics of variables in our discrete-time survival data.

	mean	median	std	min	max
$\log(t)$	1.3867	1.6094	0.7535	0.0000	2.8332
WC/TA	0.1486	0.1395	0.1803	−1.0713	1.1979
RE/TA	0.0533	0.0593	0.1232	−1.6818	0.6752
EBIT/TA	0.0592	0.0602	0.0854	−1.0061	0.6526
ME/TL	5.4051	3.3237	6.4633	0.0355	78.7321
S/TA	0.7228	0.6318	0.4568	−0.1665	4.1400
NI/TA	0.0347	0.0407	0.0984	−1.6825	0.6407
TL/TA	0.4000	0.3942	0.1690	0.0485	1.5139
CA/CL	1.9803	1.4929	1.6762	0.0129	21.1094

lected company, the annual values of Altman's and Zmijewski's predictors were collected during the sampling period. The information about industry and financial status of our selected 249 in-sample companies are given in Table 5. The summary statistics of $\log(t)$ and the predictors considered by Altman and Zmijewski in our selected discrete-time survival data are given in Table 6.

In this dissertation, in order to predict financial distress, the data used in the LLM (1) and DSM (19) were standardized so that the prediction results are unaffected by range, outliers, and other factors. The maximum likelihood estimates for parameters in the DSM (19) using the discrete-time survival data with Altman's predictors, and those in the LLM (1) using the last annual data of the discrete-time survival data with Altman's predictors are presented in Table 7. By characteristics of Altman's predictors, the larger the values of Altman's predictors, the smaller the probability of financial distress. Combining the result and the fact that the logistic function is strictly increasing, the coefficient estimates of Altman's predictors in both models LLM (1) and DSM (19) should be negative. From Table 7, the coefficient estimates of the three predictors, WC/TA, RE/TA, and S/TA, are negative in both the DSM (19) and LLM (1). But those of the two predictors, EBIT/TA and ME/TL, are positive in both models. The tests of the latter two coefficient estimates are not significant at 0.05 level in the DSM (19). Therefore, the resulting coefficient estimates of Altman's predictors in the DSM (19) are reasonable. On the other hand, the tests of the latter two coefficient

Table 7: The estimated values of parameters in each of the DSM and the LLM using our discrete-time survival data with Altman's predictors. A z statistic was given to test the significance of the value of each parameter. The value given in the parenthesis stands for the p-value of the corresponding z test. Each value marked by * denotes that the corresponding coefficient is significant at 0.05 level.

prediction model	$\log(t)$	WC/TA	RE/TA	EBIT/TA	ME/TL	S/TA
DSM	−0.0151	−0.0291	−0.3219	0.0879	0.0486	−0.2000
	(0.747)	(0.548)	(0.000)*	(0.078)	(0.300)	(0.670)
LLM	−0.1383	−0.1561	−3.9517	1.8612	0.4580	−0.2876
	(0.326)	(0.405)	(0.000)*	(0.000)*	(0.001)*	(0.080)

estimates are significant at 0.05 level in the LLM (1). Hence, it is unreasonable in this situation. The coefficient test of the firm's trading age is not significant in each of the DSM (19) and LLM (1).

Table 8 shows the maximum likelihood estimates of the parameters in the DSM (19) and the LLM (1), similar to Table 7, but with Zmijewski's predictors instead of Altman's predictors. By characteristics of Zmijewski's predictors, the larger the values of NI/TA and CA/CL, and the smaller the value of TL/TA, the smaller the probability of financial distress. Combining the result and the fact that the logistic function is strictly increasing, the coefficient estimates of the two predictors, NI/TA and CA/CL, should be negative and that of TL/TA should be positive. From Table 8, the coefficient estimates of the predictors, NI/TA and TL/TA , are negative and positive, respectively, in the DSM (19). The same remark is also made for the LLM. Those of the predictors CA/CL are positive in both the DSM (19) and LLM (1). The test of coefficient estimate of the predictor CA/CL is not significant at 0.05 level in the DSM (19). Therefore, the resulting coefficient estimates of Zmijewski's predictors in the DSM (19) are reasonable. On the other hand, the test of coefficient estimate of the predictor CA/CL is significant at 0.05 level in the LLM (1). Hence, it is unreasonable in this situation. The coefficient test of the firm's trading age is not significant in each of the DSM (19) and LLM (1).

In this illustration, the selected optimal cutoff value $\hat{p}^*$ was taken as the minimizer of the sum of in-sample type I and type II error rates without restriction on the magnitude

Table 8: The estimated values of parameters in each of the DSM and the LLM using our discrete-time survival data with Zmijewski's predictors. A z statistic was given to test the significance of the value of each parameter. The value given in the parenthesis stands for the p-value of the corresponding z test. Each value marked by * denotes that the corresponding coefficient is significant at 0.05 level.

prediction model	$\log(t)$	NI/TA	TL/TA	CA/CL
DSM	0.0009	-0.2421	0.0704	0.0522
	(0.985)	$(0.000)^*$	(0.146)	(0.267)
LLM	-0.0223	-1.5164	0.6676	0.3827
	(0.874)	$(0.000)^*$	$(0.000)^*$	$(0.006)^*$

Table 9: The selected optimal cutoff values $\hat{p}^*$ obtained by applying each of the DSM and the LLM to our discrete-time survival data with Altman's predictors, and those with Zmijewski's predictors.

predictors	DSM	LLM
Altman	0.5281	0.4808
Zmijewski	0.5476	0.5621

of in-sample type I error rate. The type I and II error rates have been introduced in Section 1.8. For each of the DSM (19) and LLM (1), Table 9 shows the selected optimal cutoff values $\hat{p}^*$ using the discrete-time survival data with each set of Altman's and Zmijewski's predictors.

In order to compare the financial distress prediction performance of LLM (1) and DSM (19), the 220 healthy companies in the sampling period were used to predict their financial status in the out-of-sample period. These companies were called the out-of-sample companies. The out-of-sample period was taken as the one during January of the year 2000 to December of the year 2002. Table 10 presents their industry and financial status in the out-of-sample period.

The last annual values of predictive variables in the out-of-sample period were collected for the out-of-sample companies. These data was called the out-of-sample data.

Table 10: The information about industry and financial status of out-of-sample companies.

Industry	number of distress companies	number of healthy companies
M1100 Cement	0	4
M1200 Food	4	13
M1300 Plastics	1	13
M1400 Textiles	3	32
M1500 Electric, Machinery	1	15
M1600 Appliance, Cable	1	5
M1700 Chemica	0	15
M1800 Glass, Ceramics	1	4
M1900 Paper,Pulp	0	1
M2000 Steel, Iron	5	16
M2100 Rubber	0	6
M2200 Automobile	0	3
M2500 Construction	9	19
M2600 Transportation	0	14
M2700 Tourism	0	4
M2900 Department Stores	1	6
M9900 Other	3	21
Total companies	29	191

The out-of-sample data is expressed as

$$(t_k,\ \tilde{Y}_k,\ x_{k,t_k}),\quad \text{for}\quad k = 1,\ ...,\ n_0.$$

Here $n_0 = \sum_{i=1}^{n}(1 - Y_i)$ is the number of the out-of-sample companies, n is the number of the in-sample companies, and t_k denotes the duration time of the k-th out-of-sample company in the out-of-sample period. Also, at the duration time t_k, $\tilde{Y}_k = 0$ indicates that the k-th out-of-sample company is healthy, and 1 the k-th out-of-sample company is of financial distress. Further, x_{k,t_k} denotes the values of Altman's or Zmijewski's predictors collected at the duration time t_k.

Given each set of Altman's and Zmijewski's predictors, the financial distress prediction performance of the LLM (1) and that of the DSM (19) were compared on their out-of-sample error rates. Given each set of Altman's and Zmijewski's predictors, the out-of-sample error rates of the DSM were calculated as follows. First, use the maximum likelihood estimates $\hat{\psi}$ of parameters ψ in Tables 7 and 8 for the DSM to calculate the predicted probability of financial distress for the company with predictor values (t_k, x_{k,t_k}), for each $k = 1, ..., n_0$. Second, use the resulting predicted probability of financial distress to compare with the selected optimal cutoff value $\hat{p}^*$ in Table 9 for the DSM. The k-th company with predictor values (t_k, x_{k,t_k}) is classified as a financial distress company if

$$\hat{h}_k = h(t_k, x_{i,t_k};\ \hat{\psi}) = \frac{\exp\{\hat{\alpha}_0 + \hat{\alpha}_1\ \log(t_k) + \hat{\beta}\ x_{i,t_k}\}}{1 + \exp\{\hat{\alpha}_0 + \hat{\alpha}_1\ \log(t_k) + \hat{\beta}\ x_{i,t_k}\}} > \hat{p}^*,$$

otherwise a healthy company. Finally, the three out-of-sample error rates of the DSM corresponding to each set of Altman's and Zmijewski's predictors defined by

$$\alpha_{DSM} = \left\{\sum_{k=1}^{n_0} \tilde{Y}_k\ I(\hat{h}_k \le \hat{p}^*)\right\} \Big/ \left\{\sum_{k=1}^{n_0} \tilde{Y}_k\right\},$$

$$\beta_{DSM} = \left\{\sum_{k=1}^{n_0} (1 - \tilde{Y}_k)\ I(\hat{h}_k > \hat{p}^*)\right\} \Big/ \left\{\sum_{k=1}^{n_0} (1 - \tilde{Y}_k)\right\},$$

$$\tau_{DSM} = \alpha_{DSM} + \beta_{DSM},$$

were computed. Here α_{DSM} is the error rate of misclassifying the financial distress companies to healthy ones, and β_{DSM} is the error rate of misclassifying the healthy companies to financial distress ones for the prediction rule DSM.

The same computational procedures were also applied to the prediction rule based on the LLM (1). See Section 1.3 for a detailed introduction of the computational procedures for the LLM (1). Let α_{LLM}, β_{LLM}, and τ_{LLM} be similarly defined as the out-of-sample error rates for the LLM (1) corresponding to each set of Altman's and Zmijewski's predictors.

Given each set of Altman's and Zmijewski's predictors, Table 11 presents the financial distress prediction performance of the DSM (19) and that of the LLM (1). Using Zmijewski's predictors, Table 11 shows that the performance of the DSM (19) is nearly

Table 11: The out-of-sample error rates obtained by applying each of the DSM and the LLM to our discrete-time survival data with Altman's predictors, and those with Zmijewski's predictors.

predictors	α_{DSM}	β_{DSM}	τ_{DSM}
Altman	0.0345	0.2513	0.2858
Zmijewski	0.2413	0.1675	0.4088
predictors	α_{LLM}	β_{LLM}	τ_{LLM}
Altman	0.1724	0.2251	0.3975
Zmijewski	0.2413	0.1518	0.3931

equal to that of the LLM (1). But, using Altman's predictors, Table 11 shows that the performance of the DSM (19) is much better than that of the LLM (1). Hence, from Table 11, to predict the financial distress for the traditional companies listed in Taiwan Stock Exchange, we suggest using the DSM (19) with Altman's predictors.

3.4 Discussion

In this chapter, the prediction of financial distress based on the DSM (19) are proposed for the discrete-time survival data collected in Taiwan. The DSM (19) has the advantage of using all available historical information to determine each firm's bankruptcy risk at each point in time, and it is a dynamic forecasting model.

The maximum likelihood method is employed to estimate the values of parameters ψ of the DSM (19), and the resulting estimators $\hat{\psi}$ are asymptotically normally distributed as $\hat{\psi} \approx N\{\psi,\, I_n(\psi)^{-1}\}$, where $I_n(\psi) = (-1)E\left\{\frac{\partial^2\, \ell_{DSM}(\psi)}{\partial\psi\, \partial\psi^T}\right\}$. In practice, we use $I_n(\hat{\psi}) = (-1)\, \frac{\partial^2\, \ell_{DSM}(\psi)}{\partial\psi\, \partial\psi^T}\Big|_{\psi=\hat{\psi}}$ to estimate $I_n(\psi)$. Replacing the quantity $I_n(\psi)^{-1}$ in the asymptotic normal distribution of $\hat{\psi}$ with its estimate $I_n(\hat{\psi})^{-1}$, the resulting distribution can be used to derive the confidence interval estimate and test the values of ψ.

To decide the optimal prediction rule, we propose to define the optimal cutoff value $\hat{p}^*$ as the minimizer of the sum of in-sample type I and type II error rates without restriction on the magnitude of the in-sample type I error rate. Based on our real data example, our DSM (19) performs better than LLM (1), in the sense of having much smaller and having almost equal out-of-sample total error rates using Altman's and Zmijewski's predictors, respectively. In conclusion, it is better for the financial distress prediction of the traditional companies listed in Taiwan Stock Exchange to use the DSM (19) with Altman's predictors.

Although the DSM (19) has the above advantages, it has some practical drawbacks. Firstly, it needs to collect each firm's time-series data. However, such time-series data may be incomplete and some ad hoc imputation methods are frequently employed. For example, Shumway (2001) suggested substituting variable values from past years for the missing values in the cases when the explanatory variables are not completely observed. Secondly, there will be a problem encountered in economical structure change. Grice and Dugan (2001) showed that using the prediction models to time periods other than those used to develop the models may result in significant decline in prediction accuracies. Finally, the hazard function employed by the DSM (19) is a linear logistic function, which is not robust with respect to the misspecification of the linear relation. This is the same problem happened to the LLM (1). Hence, to avoid such drawback to the DSM (19), one remedy is to consider the hazard function as a semiparametric logistic function

$$h(t,\ x,\ z) = \frac{\exp\{H(t,\ x) + \theta\ z\}}{1 + \exp\{H(t,\ x) + \theta\ z\}},$$

as discussed in Section 1.4 and Chapter 2. Here $H(t,\ x)$ is an unknown, but smooth function of $(t,\ x)$. It will be interesting to study the financial distress prediction performance for the model combining both the SLM and DSM.

CHAPTER IV
CONCLUSIONS

Ohlson (1980) proposed the LLM to predict bankruptcy. The LLM postulates that the logit function of bankruptcy probability is a linear function of the predictors. The main reason of using the LLM is due to its simplicity in computation and interpretation. When appropriate, the LLM has definite advantages. For example, the corresponding inferential methods usually have nice efficiency properties. Also, the parameters generally have some physical meaning which makes them interpretable and of interest in their own right. However, if the assumed linear logit function is grossly in error, then the advantages of the LLM will not be realized. Thus, there are few benefits from using a poorly specified LLM. Härdle, Moro, and Schäfer (2006) shows that the relation between the bankruptcy probability and predictors, such as net income change and company size, may not be monotonic.

The LLM is most appropriate when theory, past experience, or other sources are available that provide detailed knowledge about the data under study. Sometimes, based on previous experience, there are reasons for modelling the logit function of bankruptcy probability as a particular parametric function of predictors, which may not be linear. However, a general drawback of such parametric modelling is that if one chooses a parametric family that is not of appropriate form, at least approximately, then there is still a danger of reaching erroneous inference.

Under the circumstance, the first focus of this dissertation is to consider a robust method, against misspecification of the parametric logit model relation, by introducing the SLM for predicting bankruptcy. The SLM is basically very similar to the LLM, except that some unknown but smooth function replaces the linear function to model the relation between the predictors and the logit function of bankruptcy probability. Thus, clearly, the SLM is much more general and flexible in predicting the bankruptcy of a firm.

In this dissertation, bankruptcy prediction methods based on the SLM are proposed for both the prospective and the case-control data. The unknown parameters and prediction probabilities in the model are estimated by the local likelihood approach, and the resulting estimators are analyzed through their asymptotic biases and variances. The bankruptcy prediction methods using these two types of data are shown to be essentially equivalent. Thus our SLM can be directly applied to data sampled from the two important designs. One real data example and simulations confirm that our prediction method is more powerful than those based on the DAM, LLM, and KMV, in the sense of yielding smaller out-of-sample error rates.

There are two possible extensions for the SLM. Firstly, in the SLM, we assume the logit function of bankruptcy probability as $H(x) + \theta\ z$, where $H(x)$ is an unknown but smooth function. A local likelihood method has been developed to estimate $H(x)$. However, the resulting estimator $\hat{H}(x)$ suffers from the curse of dimensionality, that is, as the dimension of the continuous predictor X increases, the performance of the resulting $\hat{H}(x)$ deteriorates. To avoid such drawback, one possible remedy is to consider an additive model for $H(x)$:

$$H(x) = H_1(x_1) + \ \cdots \ + H_d(x_d),$$

as described by Hastie and Tibshirani (1990). Here $x = (x_1, \cdots, x_d)^T$ and $H_i(x_i)$ is any unknown but smooth function of x_i, the i-th component of x, for each $i = 1, \cdots, d$.

Secondly, the logit function of bankruptcy probability in our SLM is basically an additive model with $H(X)$ and $\theta\ Z$. This assumption will be violated if X and Z are interactive. A possible solution to this problem is to introduce a nonparametric interaction such as $G(X)\ Z$, where $G(\cdot)$ is a q-dimensional row vector of unknown but smooth functions, in the model. It will be interesting to study the estimates of functions $H(x)$ and $G(x)$ simultaneously.

On the other hand, there is another potential pitfall for the LLM. It is static in nature, since it uses only one set of predictor values collected at a specific time point for each firm. The static model is generally not appropriate for predicting bankruptcy because it ignores both facts that the characteristics of firms change through time as well as bankruptcy does not often occur. The same drawback to the LLM also happens to our proposed SLM. To avoid the drawback, the DSM (Allison 1982; Shumway 2001) is suggested using the idea of discrete-time survival analysis. It has the advantage of using all available historical information to determine each firm's bankruptcy risk at each point in time, hence it is a dynamic forecasting model.

However, the values of parameters in Shumway's dynamic prediction model are estimated by using the same approach as those in the multiperiod logit model (Pagano, Panetta, and Zingales 1998). Theoretically, the multiperiod logit model assumes the predictor values collected for each firm at all time points are independent. Clearly, such predictor values are dependent, and the assumption does not hold in practice. Thus, asymptotic properties of the resulting estimates of parameters in Shumway's dynamic prediction model can not be obtained from the multiperiod logit model.

By the considerations, the second focus of this dissertation is to employ directly the DSM to predict bankruptcy. In this dissertation, the predictor values of each firm in the DSM are assumed to be dependent, and the values of their parameters are estimated by the maximum likelihood method. Asymptotic normalities of the resulting estimators of parameters are obtained. Thus, the inferences about the parameters are available.

In practice, the DSM was applied to predict financial distress using the discrete-time survival data collected in Taiwan. Since there are only few bankrupt companies collected, it is difficult to predict bankruptcy well. In this case, to provide more failure companies to proceed our research, we replaced our target on bankruptcy prediction with financial distress prediction. According to the definition of financial distress given by Taiwan Stock Exchange, financial distress companies are those whose stocks were delisted, stopped trading, or traded by cash. Based on our real data example, our DSM performs better than the LLM, in the sense of having much smaller and having almost equal out-of-sample total error rates using Altman's and Zmijewski's predictors, respectively. The empirical result shows that it is better for the financial distress prediction of the traditional companies listed in Taiwan Stock Exchange to use the DSM with Altman's predictors.

There is one possible extension for the DSM. The hazard function employed by the DSM is defined by

$$h(t, x, z;\ \psi) = \frac{\exp\{\alpha_0 + \alpha_1\ \log(t) + \beta\ x + \theta\ z\}}{1 + \exp\{\alpha_0 + \alpha_1\ \log(t) + \beta\ x + \theta\ z\}},$$

where $\psi = (\alpha_0,\ \alpha_1,\ \beta,\ \theta)^T$. It is a linear logistic hazard function. Thus, it is not robust with respect to the misspecification of the linear relation. This is the same problem happened to the LLM. Hence, to avoid such drawback to the DSM, the same remedy improving the drawback to the LLM can be applied to the DSM by considering the hazard function as a semiparametric logistic function

$$h(t,\ x,\ z) = \frac{\exp\{H(t,\ x) + \theta\ z\}}{1 + \exp\{H(t,\ x) + \theta\ z\}}.$$

Here $H(t,\ x)$ is an unknown but smooth function of $(t,\ x)$. It will be interesting to study the performance of prediction methods based on the resulting semiparametric dynamic model.

BIBLIOGRAPHY

Allison, P. D., "Discrete-time methods for the analysis of event histories," Sociological Methodology, 13, pp. 61-98, 1982.

Altman, E. I., "Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy," Journal of Finance, 23, pp. 589-609, 1968.

Begley, J., Ming, J., and Watts, S., "Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models," Review of Accounting Studies, 1, pp. 267-284, 1996.

Bharath, S., and Shumway, T., "Forecasting default with the KMV-Merton model," manuscript, University of Michigan, 2004.

Chava, S. and Jarrow, R. A., "Bankruptcy prediction with industry effects," Review of Finance, 8, pp. 537-569, 2004.

Cox, D. R. and Oakes, D., Analysis of Survival Data, Chapman, New York, 1984.

Crosbie, P. J. and Bohn, J. R., "Modeling Default Risk," KMV Corporation, http://www.kmv.com, 2001.

Fan, J., Gasser, T., Gijbels, I., Brookmann, M., and Engel, M., "Local polynomial fitting: A standard for nonparametric regression," Discussion paper 9315, Institut de Statistique, Universite Catholique de Louvain, Belgium, 1993.

Fan, J., Heckman, N. E., and Wand, M. P., "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions," Journal of the American Statistical Association 90, pp. 141-150, 1995.

Fernandez, C. and Steel, M. F. J., "On Bayesian modeling of fat tails and skewness," Journal of the American Statistical Association, 93, pp. 359-371, 1998.

Frydman, H., Altman, E. I., and Kao, D. L., "Introducing recursive partitioning for financial classification: the case of financial distress," Journal of Finance, 40, pp. 269-291, 1985.

Grice, J. S. and Dugan, M. T., "The limitations of bankruptcy prediction models: Some cautions for the research," Review of Quantitative Finance and Accounting, 17, pp. 151-166, 2001.

Härdle W., Moro R. A, and Schäfer D., "Graphical data representation in bankruptcy analysis," in Handbook of Computational Statistics, Härdle, W. (ed), Springer, Berlin, 2006.

Hastie, T. and Tibshirani, R., Generalized Additive Models, Chapman and Hall, London, 1990, .

Hosmer, D. J. and Lemeshow, S., Applied Logistic Regression, Wiley, New York, 1989.

Johnson, R. A. and Wichern, D. W., Applied Multivariate Statistical Analysis, Prentice Hall, New York, 2002.

Jones, M. C., Marron, J. S., and Sheather, S. J., "A brief survey of bandwidth selection for density estimation," Journal of the American Statistical Association, 91, pp. 401-407, 1996.

Klein, J. P. and Moeschberger M. L., Survival Analysis: Techniques for Censored and Truncated Data, Springer, New York, 1997.

Koh, H. C. and Tan, S. S., "A neural network approach to the prediction of going concern status," Accounting and Business Research, 29, pp. 211-216, 1999.

Lancaster, T., The Econometric Analysis of Transition Data, Cambridge University Pr- ess, New York, 1990.

Lane, W., Looney, S., and Wansley, J., "An application of the cox proportional hazards model to bank failure," Journal of Banking and Finance, 10, pp. 511-531, 1986.

Lejeune, M. and Sarda, P., "Smooth estimators of distribution and density functions," Computational Statistics and Data Analysis, 14, pp. 457-471, 1992.

Lindsay, D. H., and Campbell, A., "A chaos approach to bankruptcy prediction," Journal of Applied Business Research, 12, pp. 1-9, 1996.

Little, R. J. A. and Rubin, D. B. Statistical Analysis with Missing Data, Wiley, New York, 2002.

Mckee, T. E., "Rough sets bankruptcy prediction models versus auditor signaling rates," Journal of Forecasting, 22, pp. 569-586, 2003.

Messier, J. W. and Hansen, J. V., "Inducing rules for expert system development: an example using default and bankruptcy data," Management Science, 34, pp. 1403-1415, 1988.

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

Ohlson, J., "Financial ratios and the probabilistic prediction of bankruptcy," Journal of Accounting Research, 18, pp. 109-131, 1980.

Pagano, M., Panetta, F., and Zingales, L., "Why do companies go public? An empirical analysis," Journal of Finance, 53, pp. 27-64, 1998.

Prentice, R. L. and Pyke, R., "Logistic disease incidence models and case-control studies," Biometrika, 66, pp. 403-411, 1979.

Serfling, R. J., Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.

Shumway, T., "Forecasting bankruptcy more accurately: a simple hazard model," Journal of Business, 74, pp. 101-124, 2001.

Siegrist, K., "The Pareto distribution," http://www.ds.unifi.it/VL/VL_EN/special/special12.html [3 February 2006], 2005.

Singer, J. D. and Willett, J. B., “It’s about time: using discrete-time survival analysis to study duration and the timing of events,” Journal of Educational Statistics, 18, pp. 155-195, 1993.

Stuart, A. and Ord, J. K., Kendall’s Advanced Theory of Statistics, Volume 1, Oxford University Press, New York, 1987.

Tibshirani, R., and Hastie, T., “Local likelihood estimation,” Journal of the American Statistical Association, 82, pp. 559-568, 1987.

Vassalou, M. and Xing, Y., “Default risk in equity returns,” Journal of Finance, 59, pp. 831-868, 2004.

Wand, M. P., and Jones, M. C., Kernel Smoothing, Chapman, London, 1995.

Zhao, L. P., Kristal, A. R., and White, E., “Estimating relative risk functions in case-control studies using a nonparametric logistic regression,” American Journal of Epidemiology, 144, pp. 598-609, 1996.

Zmijewski, M. E., “Methodological issues related to the estimation of financial distress prediction models,” Journal of Accounting Research, 22, pp. 59-82, 1984.