

## A Semiparametric Method for Predicting Bankruptcy

RUEY-CHING HWANG,<sup>1,2,\*</sup> K. F. CHENG<sup>3,4</sup> AND JACK C. LEE<sup>5</sup>

<sup>1</sup> *Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan*

<sup>2</sup> *Department of Finance and Banking, Dahan Institute of Technology, Hualien, Taiwan*

<sup>3</sup> *Institute of Statistics, National Central University, Jhongli, Taiwan*

<sup>4</sup> *Biostatistical Center, China Medical University, Taichung, Taiwan*

<sup>5</sup> *Graduate Institute of Finance, National Chiao Tung University, Hsinchu, Taiwan*

### ABSTRACT

Bankruptcy prediction methods based on a semiparametric logit model are proposed for simple random (prospective) and case–control (choice-based; retrospective) 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 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. One real data example and simulations confirm that our prediction method is more powerful than alternatives, in the sense of yielding smaller out-of-sample error rates. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS linear logit model; out-of-sample error rate; semiparametric logit model

### 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 company; see Altman (1968). Other frequently referred models include those 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

---

\* Correspondence to: Ruey-Ching Hwang, Department of Finance and Banking, Dahan Institute of Technology, 1 Su-Ren Street, Hualien 97145, Taiwan.  
E-mail: sunny@ms01.dahan.edu.tw

of bankruptcy. Their study showed that using 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 *et al.*, 1985), expert systems (Messier and Hansen, 1988), chaos theory (Lindsay and Campbell, 1996), neural networks (Koh and Tan, 1999), survival analysis (Lane *et al.*, 1986; Shumway, 2001; Chava and Jarrow, 2004), rough set theory (McKee, 2003), KMV–Merton model (KMV; Bharath and Shumway, 2004; Vassalou and Xing, 2004) and support vector machines (Härdle *et al.*, 2006).

Ohlson (1980) stated that the main reason of using the LLM is its simplicity in computation and interpretation. There are many software packages having logistic regression capabilities, for example, BMDP, EGRET, GLIM and SAS. 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. Another interesting reason, not stated in Ohlson (1980), for using the LLM is that the approach gives us a great flexibility in using different sampling schemes for collecting observations for building prediction models. Farewell (1979) and Prentice and Pyke (1979) justified that under the LLM one can analyze case–control (choice-based; retrospective) observations as if they were sampled from a simple random sampling scheme (prospective observations). The observations in Ohlson's empirical study were case–control observations. His prediction model had been estimated as if the observations were prospective observations (but without any justification).

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 numbers of cases (bankrupt companies) and controls (nonbankrupt companies) are matched according to some stratifying variables. Usually, in most matched case–control designs, each case is matched with one to five controls per stratum. For a detailed introduction of the LLM and the (matched) case–control data, see, for example, the monograph by Hosmer and Lemeshow (1989).

Applying case–control data to the LLM, Prentice and Pyke (1979) showed that the resulting estimators of the coefficients converge to their true values, except the intercept estimator, as both sample sizes of control and case data become large. Due to the incorrect estimation of the intercept, the 'estimated' bankruptcy probability, obtained by plugging all these estimates of coefficients into the LLM, does not converge to the true bankruptcy probability. However, since the logistic distribution is a strictly monotone increasing function, such estimated bankruptcy probability still can be employed to develop a prediction device as the LLM does under the prospective data. For further discussion of these facts, see the next section.

Ohlson's bankruptcy prediction model 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. One potential pitfall of this model is that it assumes a linear relation between the predictors and the logit function of bankruptcy probability. This approach in general is not robust with respect to the misspecification of the linear relation. See the discussion and Figure 2 of Härdle *et al.* (2006). Their results show that the relation between bankruptcy probability and predictors, such as net income change and company size, may not be monotonic.

The main focus of this paper is to consider a robust method, against misspecification of the relation between the predictors and the logit of bankruptcy probability, by introducing a semiparametric logit model (SLM; Zhao *et al.*, 1996) for predicting bankruptcy. This model is basically very

similar to the LLM, except that some unspecified function  $H(\cdot)$  replaces the linear function to model the relation between the predictors and the logit of bankruptcy probability. Thus, clearly, the SLM is much more general and flexible in predicting the bankruptcy of a firm. Härdle *et al.* (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 a nonparametric technique to map scores into bankruptcy probabilities. Their work presents a new trend in bankruptcy analysis.

In LLM, there are a finite number of coefficients and thus the usual likelihood method can be applied to estimate the coefficients. In contrast, conceptually, there are infinitely many parameters in the unspecified function  $H(\cdot)$  in SLM. We shall use the local likelihood method (Tibshirani and Hastie, 1987; Wand and Jones, 1995) to estimate the unknown function  $H(\cdot)$  and hence develop a bankruptcy prediction method. Specifically, given the SLM and any particular predictor value  $x_0$ , the quantity  $H(x_0)$  can be estimated using weighted linear logistic regression with weights determined by the local neighborhood of  $x_0$ . Thus the application of the SLM is rather straightforward and efficient.

We may consider SLM as an extension of LLM. This is a useful extension, since in many practical applications it is not easy to confirm the relation between the logit and a particular set of predictors. Usually, linear function only serves as an approximation and it may not be good enough for the true model. It is possible, for a particular set of predictors, to add polynomial terms in the linear function to achieve the same superior fit that the SLM provides. However, this polynomial model may not be suitable for other sets of predictive variables. Further, we require a large dataset and powerful lack-of-fit tests to confirm the polynomial model to be appropriate. In contrast, our SLM provides a useful alternative approach. Using this approach, we do not need strategies for building the relation between the logit and the predictors, and it is quite simple to apply.

The remainder of the paper is organized as follows. In next section, our methodology for predicting bankruptcy based on SLM is developed using concepts similar to those based on LLM. The third section describes one real case-control dataset and provides some summary statistics. The summary statistics show that the predictors under consideration have reasonable power in discriminating the bankruptcy status of the company. The real dataset is analyzed using SLM and the alternative methods such as DAM, LLM and KMV. The prediction ability of each method is described by the out-of-sample error rate. The results about out-of-sample error rates are summarized in the third section. Furthermore, the out-of-sample error rate results from some Monte Carlo simulations are given in the fourth section. From comparisons of these results, we can confirm that SLM has the best performance. The fifth section contains concluding remarks and future research topics. Our theoretical results are presented in Appendix A. Finally, sketches of the proofs are given in Appendix B.

## METHODOLOGY

In this section we describe the formulation of SLM, develop the methodology for estimating its unknown quantities, and build a bankruptcy prediction method based on SLM for both simple random and case-control data.

### Linear logit model versus semiparametric logit model

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. For the  $i$ th company,  $i = 1, \dots, n$ , we observe values  $(D_i, x_i, z_i)$  where  $D_i = 1$

indicates that the  $i$ th company is in the state of bankruptcy and 0, otherwise, and  $x_i = (x_{i1}, \dots, x_{id})^T$  and  $z_i = (z_{i1}, \dots, z_{iq})^T$  are values of the vectors of explanatory variables used to forecast failure. Here we have used  $X$  and  $Z$ , respectively, to represent the continuous and discrete variables, and employed the upper index  $T$  to stand for the transpose of a vector. For example, in Ohlson (1980), there were nine financial variables being used for developing his bankruptcy prediction model. Among these explanatory variables, seven ( $= d$ ) of them are continuous variables and two ( $= q$ ) are binary variables.

Given the prospective training sample and the values  $(x, z)$  of the predictors  $(X, Z)$ , the LLM is defined by assuming the bankruptcy probability to be

$$p(D=1 | X=x, Z=z) = \frac{\exp(\eta + \zeta x + \theta z)}{1 + \exp(\eta + \zeta x + \theta z)},$$

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

$$\text{logit}\{p(D=1 | X=x, Z=z)\} = \log\left\{\frac{p(D=1 | X=x, Z=z)}{1 - p(D=1 | X=x, Z=z)}\right\} = \eta + \zeta x + \theta z.$$

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

$$\hat{p}(D=1 | X=x_0, Z=z_0) = \frac{\exp(\hat{\eta} + \hat{\zeta}x_0 + \hat{\theta}z_0)}{1 + \exp(\hat{\eta} + \hat{\zeta}x_0 + \hat{\theta}z_0)}$$

is the logistic distribution evaluated at the predicted score  $\hat{\eta} + \hat{\zeta}x_0 + \hat{\theta}z_0$ , where  $\hat{\eta}$ ,  $\hat{\zeta}$  and  $\hat{\theta}$  are the maximum likelihood estimates based on the prospective training sample  $\{(D_i, x_i, z_i), i = 1, \dots, n\}$ .

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 modeling the logit function of bankruptcy probability as a particular function of  $(X, Z)$ , which may not be linear. However, there is a general drawback to such parametric modeling. If one chooses a parametric family that is not of appropriate form, at least approximately, then the resulting model-based bankruptcy probability prediction might not correctly estimate the true bankruptcy probability, and there is a danger of reaching erroneous prediction.

The limitation of LLM can be overcome by removing the restriction that the logit function of  $p(D=1 | X=x, Z=z)$  belongs to a parametric family. This approach may lead to the following SLM:

$$p(D=1 | X=x, Z=z) = \frac{\exp\{H(x) + \theta z\}}{1 + \exp\{H(x) + \theta z\}}. \quad (1)$$

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. 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

$$\hat{p}(D=1 | X=x_0, Z=z_0) = \frac{\exp\{\hat{H}(x_0) + \hat{\theta}z_0\}}{1 + \exp\{\hat{H}(x_0) + \hat{\theta}z_0\}}, \quad (2)$$

the logistic distribution evaluated at the predicted score  $\hat{H}(x_0) + \hat{\theta}z_0$ . Here  $\hat{H}(x_0)$  and  $\hat{\theta}$  are estimates deriving from the prospective training sample and local likelihood method.

There exists many methods for estimating  $H(x_0)$ , for  $x_0 = (x_{01}, \dots, x_{0d})^T$ . One of these methods with a simple idea is the local likelihood method (see Tibshirani and Hastie, 1987). This approach is to first choose a positive scalar constant  $b$  and define a neighborhood of  $x_0$  as  $N(x_0; b) = \{t = (t_1, \dots, t_d)^T : |t_j - x_{0j}| \leq b, \text{ for each } j = 1, \dots, d\}$ . Then the idea of the local likelihood method is to apply both concepts of the likelihood method using partial sample  $S(x_0; b) = \{(D_i, x_i, z_i) : x_i \in N(x_0; b)\}$ , 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  $x_i \in N(x_0; b)$ . Here the parameters  $\alpha$  and  $\beta$  stand for the unknown quantities  $H(x_0)$  and  $H^{(1)}(x_0)^T$ , respectively, and  $H^{(1)}(x_0)$  is the column vector of partial derivatives of  $H(x_0)$ . Specifically, we find the maximizer  $(\hat{\alpha}, \hat{\beta}, \hat{\theta})$  of the 'local' log-likelihood function

$$\ell = (-1) \sum_{i: x_i \in N(x_0; b)} \log[1 + \exp\{\alpha + \beta(x_i - x_0) + \theta z_0\}] + \sum_{i: x_i \in N(x_0; b)} D_i \{\alpha + \beta(x_i - x_0) + \theta z_0\}, \quad (3)$$

use  $\hat{H}(x_0) = \hat{\alpha}$  to estimate  $H(x_0)$ , and combine estimates  $\hat{H}(x_0)$  and  $\hat{\theta}$  to develop a prediction of bankruptcy probability (2).

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  $x_i \in N(x_0; b)$ , and using the partial (local) sample  $S(x_0; b)$  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.

We now determine a  $p^* \in [0, 1]$  value to make bankruptcy prediction for the company with predictor values  $(x_0, z_0)$ . That is, if its predicted bankruptcy probability  $\hat{p}(D = 1 | X = x_0, Z = z_0)$  derived from (2) and (3) satisfies

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

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

$$\begin{aligned} \text{type I error rate } \alpha_{in}(p) &= \frac{\sum_{i=1}^n D_i I\{\hat{p}(D = 1 | X = x_i, Z = z_i) \leq p\}}{\sum_{i=1}^n D_i}, \\ \text{type II error rate } \beta_{in}(p) &= \frac{\sum_{i=1}^n (1 - D_i) I\{\hat{p}(D = 1 | X = x_i, Z = z_i) > p\}}{\sum_{i=1}^n (1 - D_i)}. \end{aligned}$$

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 a bankrupt company as a healthy company, and  $\beta_{in}(p)$  is the rate of misclassifying a healthy company as a bankrupt company. To keep these error rates as small as possible, we might determine a proper cut-off point  $p^*$  such that

$$\tau_m(p^*) = \alpha_m(p^*) + \beta_m(p^*) = \min_{p \in [0,1], \alpha_m(p) \leq u} \{\alpha_m(p) + \beta_m(p)\}.$$

That is to control the in-sample type I error rate  $\alpha_m(p)$  to be at most  $u$ , so that the sum of the two in-sample error rates  $\tau_m(p) = \alpha_m(p) + \beta_m(p)$  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  $\beta_m(p)$  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 *et al.*, 1996). Since the value of  $p^*$  depends on that of  $u$ , it is also denoted by  $p^*(u)$ .

### More general estimates of bankruptcy probability

Suppose we define  $K(u)$  to be the uniform probability density function over  $[-1, 1]$ , then the local log-likelihood function (3) can be expressed in a different form:

$$\begin{aligned} \ell = & (-1) \sum_{i=1}^n \log[1 + \exp\{\alpha + \beta(x_i - x_0) + \theta z_0\}] K_b(x_i - x_0) \\ & + \sum_{i=1}^n D_i \{\alpha + \beta(x_i - x_0) + \theta z_0\} K_b(x_i - x_0), \end{aligned} \quad (4)$$

where  $K_b(x_i - x_0) = \prod_{j=1}^d K\{(x_{ij} - x_{0j})/b\}$ . From this expression we see that the local log-likelihood function (4) can be considered as a weighted log-likelihood function which gives weight 1 to the data inside the neighborhood sample  $S(x_0; b)$  and weight 0 outside. Conceptually, a different weighting scheme can also be employed for defining a different weighted log-likelihood function, so that larger weights are given to data points with  $X$  values closer to  $x_0$  and smaller weights to those with  $X$  values far from  $x_0$ . This can be achieved by introducing a unimodal probability density function that is symmetric about 0 to replace the uniform density for  $K(\cdot)$  in (4). However, the results from the literature show that the choice of the density function  $K(\cdot)$ , also called the kernel function, is not very important in the local fitting. A popular choice of  $K(\cdot)$  is the Epanechnikov kernel defined as  $K(u) = (3/4)(1 - u^2)I(|u| \leq 1)$  (see Wand and Jones, 1995), due to its computational convenience and optimal performance (for example, it minimizes mean square error among all non-negative kernel functions).

### Refined estimates of bankruptcy probability

It was pointed out in Wand and Jones (1995) that the estimator of bankruptcy probability produced from (2) and (4) is consistent under very general conditions. However, our theoretical results in Appendix A show that more refined estimators of  $H(x_0)$  and  $\theta$ , in terms of smaller asymptotic mean square error, can be achieved. The computation procedure of the refined estimates includes the following three steps.

*Step 1:* For each  $x_i$ ,  $i = 1, \dots, n$ , use the local log-likelihood function (4) with the Epanechnikov kernel to compute the estimate  $\hat{H}(x_i)$ .

*Step 2:* Replace the unknown quantities  $H(x_i)$  in model (1) with their estimates  $\hat{H}(x_i)$ , for  $i = 1, \dots, n$ , fit the bankruptcy probability by the resulting model



$$p(D=1 | X = x_i, Z = z_i) = \frac{\exp\{\theta_0 + \hat{H}(x_i) + \theta z_i\}}{1 + \exp\{\theta_0 + \hat{H}(x_i) + \theta z_i\}}$$

and use the training sample to maximize the corresponding pseudo log-likelihood function with respect to  $\theta_0$  and  $\theta$ . Here  $\theta_0$  is a normalizing constant which makes the bankruptcy probability function be integrated to 1. Let the maximum likelihood estimate of  $\theta$  be denoted as  $\hat{\theta}_R$ , the refined estimate of  $\theta$ .

Step 3: Replace  $\theta$  in the local log-likelihood function (4) by  $\hat{\theta}_R$  and use the training sample to maximize the resulting pseudo local log-likelihood function with respect to  $\alpha$  and  $\beta$ . Let  $(\hat{\alpha}_R, \hat{\beta}_R)$  be the maximizer, then the refined estimate of  $H(x_0)$  is  $\hat{H}_R(x_0) = \hat{\alpha}_R$ .

The refined predicted bankruptcy probability

$$\hat{p}_R(D=1 | X = x_0, Z = z_0) = \frac{\exp\{\hat{H}_R(x_0) + \hat{\theta}_R z_0\}}{1 + \exp\{\hat{H}_R(x_0) + \hat{\theta}_R z_0\}}$$

is the logistic distribution evaluated at the predicted score  $\hat{H}_R(x_0) + \hat{\theta}_R z_0$ . Note that this new procedure is no more complicated than the procedure we discussed earlier. Since the computation of  $\hat{\theta}_R$  depends only on the training sample, we consider  $\hat{\theta}_R$  as a constant in the prediction system. For any new values  $(x_0, z_0)$  of the predictors  $(X, Z)$ , we only need to apply Step 3 to compute  $\hat{H}_R(x_0)$  and then the predicted bankruptcy probability.

**The selection of constant b**

The practical implementation of SLM requires the specification of the constant  $b$ , also called the bandwidth. It determines how many data points should be included in the SLM. From the last two subsections, we see that the same bandwidth has been used for computing the refined estimates  $\hat{H}_R(x_0)$  and  $\hat{\theta}_R$ . Alternatively, we may consider using bandwidth  $b_\theta$  to compute  $\hat{\theta}_R$  (Steps 1 and 2), and employing bandwidth  $b_H$  to compute  $\hat{H}_R(x_0)$  (Step 3). The optimal values  $b_\theta^*$  and  $b_H^*$  of  $b_\theta$  and  $b_H$ , respectively, can be determined by minimizing the mean square errors of the resulting  $\hat{\theta}_R$  and  $\hat{H}_R$ . Theoretical results in Appendix A show that  $b_H^*$  is of larger order in magnitude than  $b_\theta^*$ . Although such theoretical results give some indication on how to select bandwidth parameters  $b_\theta$  and  $b_H$ , they are not available in practice, since they depend on the unknown  $H(\cdot)$ ,  $\theta$  and density function of the predictors. Thus, in real applications, we would suggest considering the in-sample type I and II error rates defined in the first subsection as functions of  $p, b_\theta$  and  $b_H$ , denoted as  $\alpha_{in}(p, b_\theta, b_H)$  and  $\beta_{in}(p, b_\theta, b_H)$ , respectively. The bandwidth parameters and the cut-off point 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 constraints  $p \in [0, 1]$ ,  $\alpha_{in}(p, b_\theta, b_H) \leq u$  and  $b_H \geq b_\theta$ . Note that such selected values  $\hat{p}(u)$ ,  $\hat{b}_\theta(u)$  and  $\hat{b}_H(u)$  for  $p^*(u)$ ,  $b_\theta^*$  and  $b_H^*$ , respectively, also depend on the training sample. Hence they are also considered as constants in the prediction system.

**Applications to case-control data**

The purpose of this paper is to develop a semiparametric bankruptcy prediction model. Previous results show that a fairly simple SLM can be developed using the training sample, which consists of simple random observations. However, as we stated earlier, many financial data are sampled using

case-control designs. For example, in Ohlson, the bankrupt companies were generated from a list of failed companies (cases) satisfying certain inclusion criteria, and a sample of nonbankrupt companies was obtained from COMPUSTAT (controls). Further, in the analysis given by Grice and Dugan (2001), the bankrupt companies were collected from those reported by COMPUSTAT meeting certain conditions (cases), and the nonbankrupt companies were collected from those that did not receive poor S&P ratings (controls). So the basic question is: can we use the case-control data to develop SLM as the LLM did? The answer is yes, but it needs justification.

Applying case-control data to the SLM as if they were simple random observations is permissible, since our theoretical results in Appendix A show that the value of  $\theta$  can be consistently estimated, and the unknown function  $H(x)$  can be correctly estimated, up to the unknown additive constant

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

Using the case-control data, inferences about the constant  $c^*$  are not possible since such data generally provide no information about the population frequency of bankrupt companies. Thus treating the case-control data as if they were simple random observations and applying the procedures outlined in the third subsection, we can only estimate  $H(x_0) + c^*$  and  $\theta$ . Specifically, let  $\hat{H}_{CC}(x_0)$  and  $\hat{\theta}_{CC}$  be denoted as the estimators derived from the third subsection, then  $\hat{\theta}_{CC}$  consistently estimates  $\theta$ , but for any  $x_0$   $\hat{H}_{CC}(x_0)$  estimates  $H(x_0) + c^*$ . Combining the inconsistency of  $\hat{H}_{CC}(x_0)$  and the fact that the unknown quantity  $c^*$  is generally not equal to 0, the resulting predicted score  $\hat{H}_{CC}(x_0) + \hat{\theta}_{CC} z_0$  does not estimate the true score  $H(x_0) + \theta z_0$ , and thus the predicted bankruptcy probability

$$\hat{p}_{CC}(D=1 | X = x_0, Z = z_0) = \frac{\exp\{\hat{H}_{CC}(x_0) + \hat{\theta}_{CC} z_0\}}{1 + \exp\{\hat{H}_{CC}(x_0) + \hat{\theta}_{CC} z_0\}}$$

does not estimate the true bankruptcy probability (1) with predictor values  $(x_0, z_0)$ . This is the major difference between applying the SLM to case-control data and to prospective data.

Fortunately, we still can use  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$  to develop a bankruptcy prediction device by applying the following simple equivalent inequalities:

$$\frac{\exp\{H(x) + \theta z\}}{1 + \exp\{H(x) + \theta z\}} > p$$

if and only if

$$\frac{\exp\{H(x) + c^* + \theta z\}}{1 + \exp\{H(x) + c^* + \theta z\}} > p^* = \frac{p \exp(c^*)}{(1-p) + p \exp(c^*)}.$$

This result is to say that using the probability  $\frac{\exp\{H(x) + \theta z\}}{1 + \exp\{H(x) + \theta z\}}$  to define a classification device

with cut-off point  $p$  is equivalent to using the probability  $\frac{\exp\{H(x) + c^* + \theta z\}}{1 + \exp\{H(x) + c^* + \theta z\}}$  to define a classification device with cut-off point  $p^*$ . Hence we may pretend  $\hat{p}_{CC}(D = 1 | X = x_0, Z = z_0)$  to be



the estimate of the true bankruptcy probability and as before use it to determine the associated cut-off point  $p^*(u)$  and bandwidth parameters  $b_{\theta}^*$  and  $b_H^*$  (see the fourth subsection), and then the semi-parametric bankruptcy prediction device.

Before closing this section, we remark that, based on the same argument for bankruptcy prediction, the methods of LLM using prospective data and case-control data are considered to be essentially equivalent.

### A REAL DATA EXAMPLE

In this section, a real case-control dataset is analyzed using our SLM method 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 industry 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 dataset. The dataset contains 79 companies that were delisted and declared bankruptcy (cases) during the period 1994–2002 by COMPUSTAT as meeting 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 1993–2001. Among the 79 selected bankrupt companies, each was matched with two non-bankrupt companies, except for two companies only matched with one nonbankrupt company each, due to the incompleteness of the two databases. Hence our dataset 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 on the New York Stock Exchange, American Stock Exchange or NASDAQ, and that were declared bankrupt during the period 1994–2002. 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 are no additional criteria imposed on 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 occurring in COMPUSTAT and CRSP databases is 'missing at random'.

Information about industry and company asset size of the selected companies is given in Tables I and II, respectively. The two-sample median test was performed to test the null hypothesis of equal magnitude of the asset size for a nonbankrupt company and that for a bankrupt company. The  $p$ -value given in Table II 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.

Table I. The SIC codes of 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

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

	79 bankrupt companies	156 nonbankrupt companies	Median-stat ( <i>p</i> -value)
Mean	105.103	150.508	0.092 (0.927)
Median	32.211	33.599	
SD	290.254	808.741	
Min.	1.447	1.636	
Max.	2345.800	9794.400	

For predicting bankruptcy, the values of the nine variables used by Ohlson (1980) and the two 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.

Note that Ohlson (1980) suggested using the first nine variables as predictive variables. But in this paper we only used the first eight variables as the predictive variables in our case-control data analysis. The ninth 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 two variables are the market-driven variables used in Shumway (2001).

For predicting bankruptcy, we further computed the KMV–Merton default probabilities  $\pi_{KMV}$  for the selected companies in our case–control data analysis. The detailed computation procedure of  $\pi_{KMV}$  can be referred to Bharath and Shumway (2004).

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 III. For each of these 10 variables, the two-sample median test was performed to test the null hypothesis of equal magnitude for a nonbankrupt company and for a bankrupt company. The  $p$ -value in Table III shows that the null hypothesis of equal magnitude for cases and controls is significant at the 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  $\pi_{KMV}$  for the selected companies in our case–control data analysis are shown respectively in Table III and Figure 2. The results also indicate that  $\pi_{KMV}$  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, which 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. To compute the out-of-sample error rate for the prediction rule based on SLM on each testing sample, the procedure given in the second section 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  $\hat{H}_{CC}(x_i)$  and  $\hat{\theta}_{CC}$  were computed for each data point  $(x_i, z_i)$  in the testing sample. The company with the predictor values  $(x_i, z_i)$  in the testing sample was classified as a bankrupt company if

$$\hat{\psi}_i = \frac{\exp\{\hat{H}_{CC}(x_i) + \hat{\theta}_{CC}z_i\}}{1 + \exp\{\hat{H}_{CC}(x_i) + \hat{\theta}_{CC}z_i\}} > \hat{p}(u),$$

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

$$\alpha_{SLM}(u) = \frac{\sum_{i:(x_i, z_i) \text{ in testing sample}} D_i I\{\hat{\psi}_i \leq \hat{p}(u)\}}{\sum_{i:(x_i, z_i) \text{ in testing sample}} D_i},$$

$$\beta_{SLM}(u) = \frac{\sum_{i:(x_i, z_i) \text{ in testing sample}} (1 - D_i) I\{\hat{\psi}_i > \hat{p}(u)\}}{\sum_{i:(x_i, z_i) \text{ in testing sample}} (1 - D_i)},$$

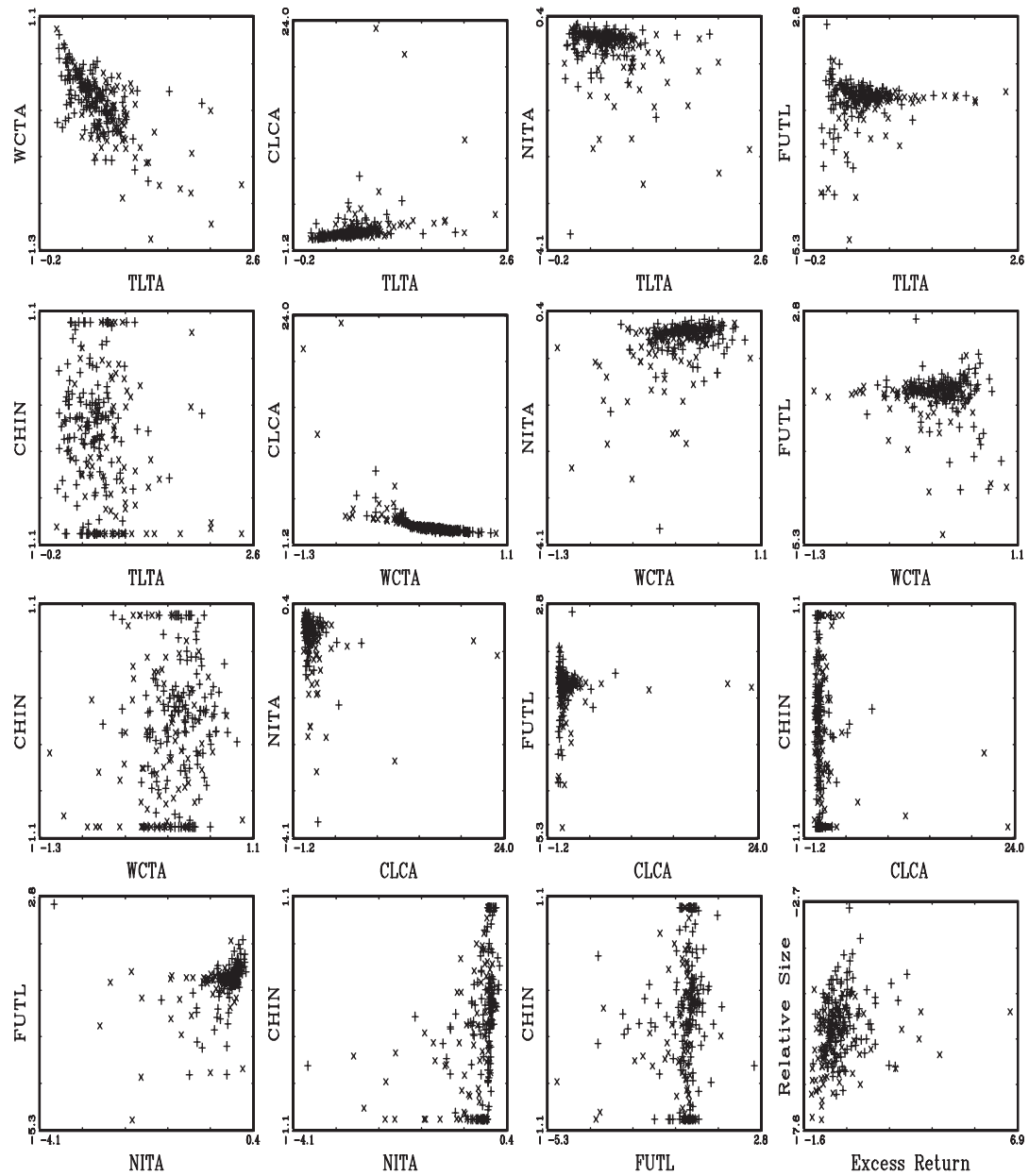


Figure 1. Pairwise scatter diagrams of our case-control sample on Shumway's two market-driven variables, Excess Return and Relative Size, and Ohlson's six continuous variables. Each graph plots 156 nonbankrupt companies (+) and 79 bankrupt companies (x) selected from COMPUSTAT and CRSP databases

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

Variable	Mean	Median	SD	Min.	Max.	Median-stat ( <i>p</i> -value)
<i>79 bankrupt companies</i>						
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)
$\pi_{KMV}$	0.413	0.331	0.383	0.000	1.000	-6.537 (0.000)
<i>156 nonbankrupt companies</i>						
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	
$\pi_{KMV}$	0.114	0.001	0.241	0.000	0.989	

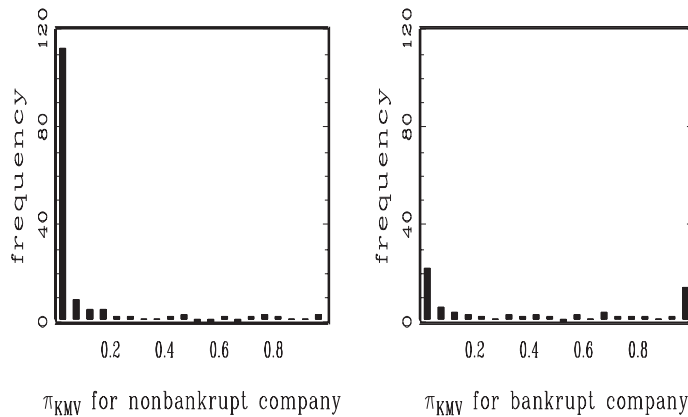


Figure 2. The frequency histogram of the values of  $\pi_{KMV}$  for the 156 nonbankrupt companies and that for the 79 bankrupt companies in our case-control sample

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

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 as healthy ones, and  $\beta_{SLM}(u)$  is the out-of-sample type II error rate of classifying the healthy companies as bankrupt ones from the testing sample. After the computation procedure was completed for each testing sample, the average of each out-of-sample error rate over the 100 testing samples was computed.

The same computation procedures were applied to the prediction rules based on DAM, LLM and KMV. The bankruptcy prediction method based on KMV was performed by taking the predicted bankruptcy probability  $\hat{p}(D = 1 \mid X = x_i, Z = z_i)$  as the value of KMV–Merton default probability  $\pi_{KMV}$  associated with the  $i$ th company. 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 results for applying the four discussed bankruptcy prediction rules to our case–control data are shown in Figure 3 and Table IV.

Figure 3 presents the three (averaged) out-of-sample error rates for the four prediction models under one (100) 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 IV. 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 IV confirms that SLM has the best overall performance. Thus it is fair to say that by a reasonable margin the most accurate model listed in Table IV is the SLM.

From Figure 3, we find that similar conclusions to those shown in Table IV can also be drawn. 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 a similar averaged type II error rate to SLM but a 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.

## SIMULATION STUDIES

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's  $t$  distribution (Fernandez and Steel, 1998) with degrees of freedom  $k$  and scale parameter  $s$  was considered. The simulated control (nonbankruptcy)  $X_j$  values were taken from the skewed Student's  $t$  distribution with  $(k, s) = (3, 2)$  for  $j = 1$ , and  $(7, 4)$  for  $j = 2$ , and those for case (bankruptcy) values  $(k, s) = (5, -3)$  for  $j = 1$ , and  $(5, 2)$  for  $j = 2$  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  $(5, -3)$  and  $(5, 2)$ , respectively.



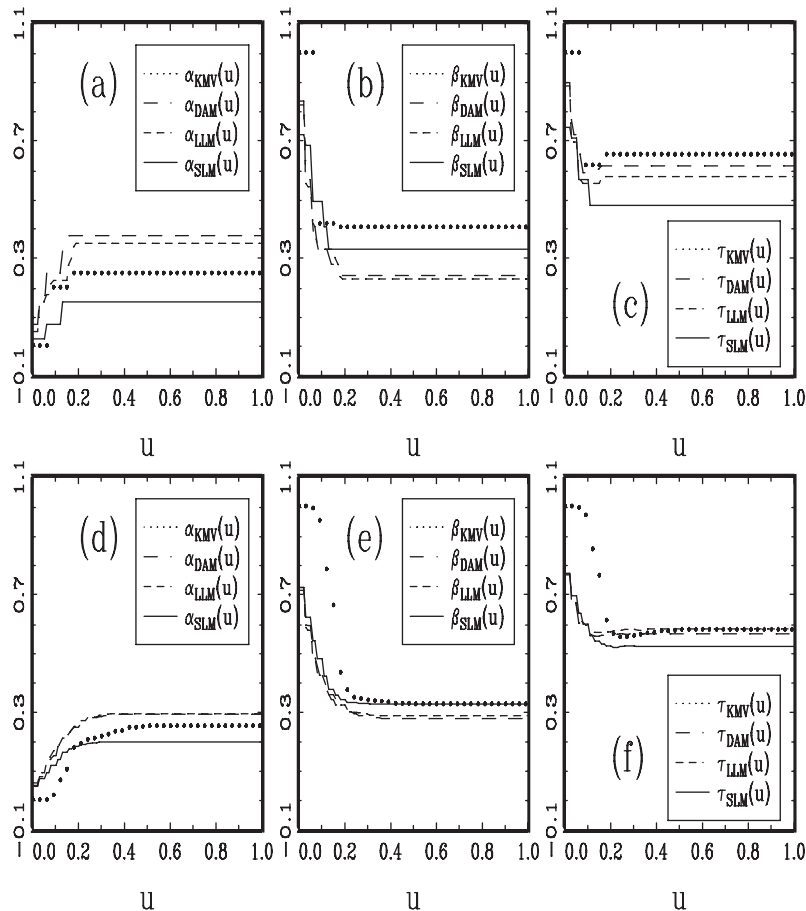


Figure 3. Parts (a)–(c) show respectively three out-of-sample error rates of the prediction methods associated with KMV, DAM, LLM, and SLM estimated with one testing sample. Parts (d)–(f) show respectively the averages of the three out-of-sample error rates over the 100 testing samples. Each testing sample was composed of 50% of bankrupt companies and their matched nonbankrupt companies in our case–control sample

Table IV. Given the value of  $u = 1$ , the values of the three out-of-sample error rates shown in (a)–(c) of Figure 3, and those shown in (d)–(f) of Figure 3 (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)

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 values  $(\mu_0, \mu_1)$ , 100 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 considered in this simulation study. The computation procedures and the measures of performance presented in the third section were applied to the three prediction methods. For SLM, the equally spaced logarithmic grid of  $201 \times 201$  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 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 datasets for the three bankruptcy prediction methods under the case of the skewed Student's  $t$  distribution for  $X$ . From the figure, our SLM performs better than DAM and LLM, since for most values of  $u$  our prediction method has a smaller average of out-of-sample error rate of any type. Further, the smaller the absolute difference  $|\mu_0 - \mu_1|$ , the more significant the advantage of SLM over DAM and LLM. The forecasting performance of the three prediction methods in 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.

## CONCLUDING REMARKS AND FUTURE RESEARCH TOPICS

In this paper, bankruptcy prediction methods based on 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 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 monotone increasing, we discover that such estimated probability still can 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

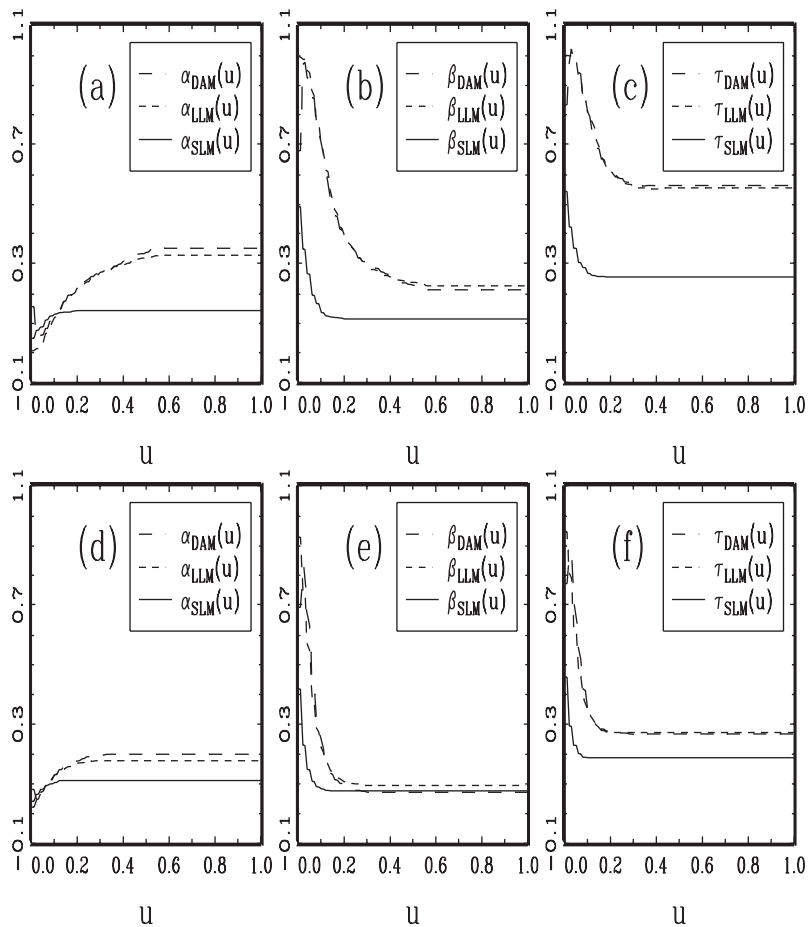


Figure 4. Given  $(\mu_0, \mu_1) = (-0.1, 0.1)$  in the case of the skewed Student's  $t$  distribution for  $X$ , parts (a)–(c) show respectively the sample averages of the three out-of-sample error rates of the prediction methods based on DAM, LLM and SLM over the 100 simulated case–control datasets. For each simulated case–control dataset, 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 parts (d)–(f)

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 with case–control data, the relation between the  $d$ -dimensional predictive variable  $x$  and the logit of bankruptcy probability can be obtained from the plot of the estimated function  $\hat{H}_{CC}(x)$ . For example, the relation between the  $j$ th component of  $x$  and the logit of bankruptcy probability can be seen from the plot of  $\hat{H}_{CC}(x)$  with the fixed values of other components of  $x$  with respect to the  $j$ th component of  $x$ , for each  $j = 1, \dots, d$ . The same remark applies to the plot of  $\hat{H}_R(x)$  with prospective data.

In order to estimate the SLM, we need to decide proper values of the bandwidth parameters  $b_\theta$  and  $b_H$ . In this paper, we suggest estimating the bandwidth parameters so that the sum of the

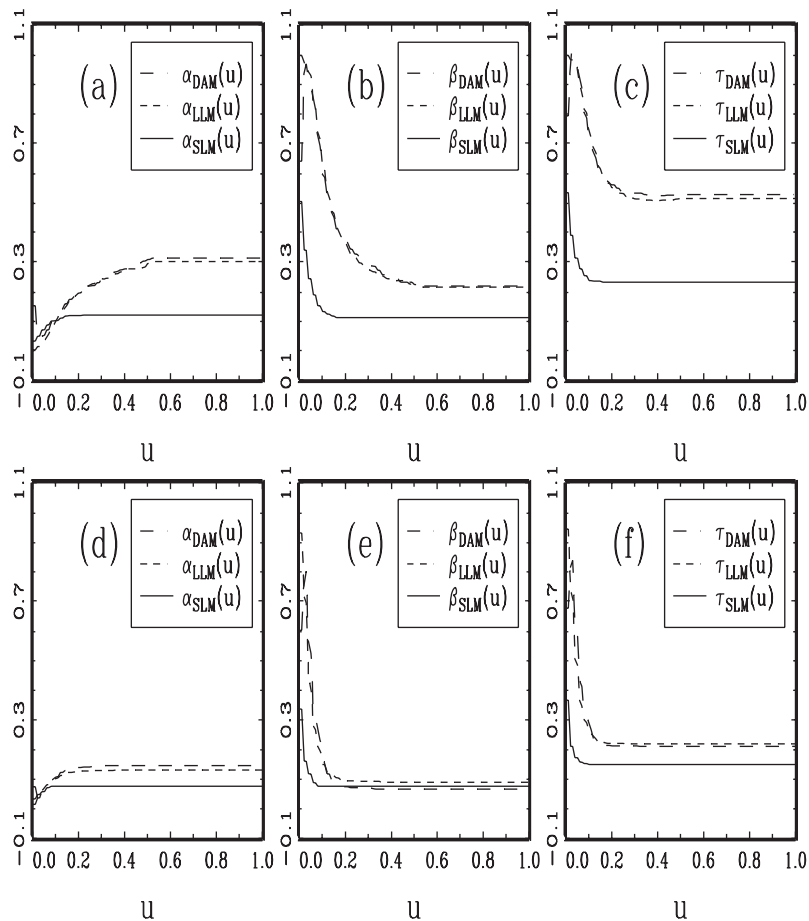


Figure 5. As Figure 4, with the skewed Student's  $t$  distribution for  $X$  replaced by the Pareto distribution for  $X$

corresponding in-sample type I and II error rates 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, given case-control data, we may determine the bandwidth parameter to minimize the estimated mean square error of each estimator  $\hat{H}_{CC}(x)$  and  $\hat{\theta}_{CC}$ . For more discussion of the plug-in method, see, for example, Härdle *et al.* (1992) and Jones *et al.* (1996).

In the case of massive data, a remedy for reducing the computational burden for SLM is to bin the data. See Fan and Marron (1994) and the monograph by Härdle (1991) for a detailed introduction of the binning method. On the other hand, in the case that there is a large number of predictors available, we may perform variable selection for SLM. See, for example, Härdle *et al.* (2006) for a variable selection method based on forward selection. These two topics are important in practice. They are our future research topics.

Three extensions of SLM are outlined below. Firstly, Shumway (2001) criticized that the bankruptcy model using single period data is static in nature, since it ignores the fact that the character-

istics of firms change through time. To avoid the drawback to the static model, Shumway (2001) and Chava and Jarrow (2004) apply the idea of survival analysis (Cox and Oakes, 1984) to develop the discrete-time hazard model (DHM) for bankruptcy prediction. Their DHM has the advantage of using all available history information to determine each firm’s bankruptcy risk at each point in time; hence it is a dynamic forecasting model. However, their DHM is also based on the assumption of simple linear logit relation for the hazard function. This assumption may not be valid as in the LLM. We point out that the idea of SLM can also be directly applied to the DHM using panel data.

Secondly, 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  $\hat{H}_{CC}(x)$  suffers from the curse of dimensionality; that is, as the dimension of the continuous predictor  $X$  increases, the performance of the resulting  $\hat{H}_{CC}(x)$  deteriorates. For example, from Remark 1 of Appendix A, the minimum mean square error of  $\hat{H}_{CC}(x)$  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 a drawback, one possible remedy is to consider an additive model for  $H(x)$ :

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

as described by Hastie and Tibshirani (1990). Here  $x = (x_1, \dots, x_d)^T$  and  $H_j(x_j)$  is any unknown but smooth function of  $x_j$ , the  $j$ th component of  $x$ , for each  $j = 1, \dots, d$ . This is basically equivalent to mapping the original predictors to the transformed variables having the desired linear relation in the LLM.

Finally, 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.

### APPENDIX A: THEORETICAL RESULTS

In Appendix A, we shall study the asymptotic properties of estimators  $\hat{H}(x)$ ,  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$  given in the second section with case–control data. 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  $(x_1, z_1), \dots, (x_{n_0}, z_{n_0})$ , from the conditional distribution of  $(X, Z)$  given  $D = 0$ , and an independent random sample of bankrupt companies of  $n_1$  observations (cases), say  $(x_{n_0+1}, z_{n_0+1}), \dots, (x_n, z_n)$ , where  $n = n_0 + n_1$ , from the conditional distribution of  $(X, Z)$  given  $D = 1$ . Hence we have the case–control sample  $(D_i, x_i, z_i)$ ,  $i = 1, \dots, n$ , where  $D_i = 0$  for  $i \leq n_0$ , and 1 for  $i > n_0$ . Thus if  $f_0(x, z) = f(x, z | D = 0)$  and  $f_1(x, z) = f(x, z | D = 1)$  are the conditional frequency functions of  $(X, Z)$  given  $D = 0$  and 1, respectively, then from Bayes theorem and (1), these two frequency functions can be related by

$$f_1(x, z) = f_0(x, z) \exp\{H^*(x) + \theta z\}, \tag{5}$$

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

Given the case–control sample and the bandwidth parameters  $b_\theta$  and  $b_H$ , by the development of the logistic regression in the case–control setting given in Section 6.3 of Hosmer and Lemeshow

(1989), the log-likelihood functions corresponding to  $\hat{H}(x)$ ,  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$  given in the second section with general kernel function  $K$  can be expressed respectively as

$$\begin{aligned} \ell_1(\alpha, \beta, \theta) = & (-1) \sum_{i=1}^n \log[1 + \exp\{\alpha + \beta(x_i - x) + \theta z_i\}] K_{b_\theta}(x_i - x) \\ & + \sum_{i=n_0+1}^n \{\alpha + \beta(x_i - x) + \theta z_i\} K_{b_\theta}(x_i - x), \end{aligned} \quad (6)$$

$$\ell_2(\theta_0, \theta) = (-1) \sum_{i=1}^n \log[1 + \exp\{\theta_0 + \hat{H}(x_i) + \theta z_i\}] + \sum_{i=n_0+1}^n \{\theta_0 + \hat{H}(x_i) + \theta z_i\}, \quad (7)$$

$$\begin{aligned} \ell_3(\alpha, \beta) = & (-1) \sum_{i=1}^n \log[1 + \exp\{\alpha + \beta(x_i - x) + \hat{\theta}_{CC} z_i\}] K_{b_H}(x_i - x) \\ & + \sum_{i=n_0+1}^n \{\alpha + \beta(x_i - x) + \hat{\theta}_{CC} z_i\} K_{b_H}(x_i - x). \end{aligned} \quad (8)$$

Note that the parameter  $\beta$  in (6) and (8) represents the unknown quantity  $H^{(1)}(x)^T$ , as it did in (3) and (4), with  $x$  replaced by  $x_0$ . But  $\alpha$  in (6) and (8) stands for  $H(x) + c^*$ , not as it did in (3) and (4) for  $H(x)$ , with  $x$  replaced by  $x_0$ . Here  $c^* = \log\{p(D=0)/p(D=1)\} + \log(n_1/n_0)$ , and it has been defined in the second section. Hence the maximum likelihood estimates of  $\alpha$  produced from (6) and (8) 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.

To study the asymptotic properties of  $\hat{H}(x)$ ,  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$ , 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.  $n_0/n \rightarrow \zeta \in (0, 1)$ , as  $n \rightarrow \infty$ .
- C3. Bandwidth parameters  $b_\theta, b_H \in [\delta n^{-1+\delta}, \delta^{-1} n^{-\delta}]$ , for some  $\delta$  satisfying  $0 < \delta < 1/2$ . They also satisfy  $nb_\theta^{d+2} \gg 1 \gg b_\theta$  and  $nb_H^d \gg 1 \gg b_H \gg b_\theta$ . The notation  $a_n \gg b_n$  means that  $b_n/a_n \rightarrow 0$ , as  $n \rightarrow \infty$ .
- C4. The  $d$ -variate function  $H(x)$  is defined on  $[0, 1]^d$ , and each of its second-order partial derivative 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  $f_0(z|x)$  and  $f_1(z|x)$  of  $Z$  given  $X = x$  cannot 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. It can be replaced with any bounded region  $\Omega \subset R^d$ , and the asymptotic properties for the resulting  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$  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  $\ell_1(\alpha, \beta, \theta)$  and  $\ell_3(\alpha, \beta)$  is invertible.

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



$$x = (x_1, \dots, x_d)^T, \quad t = (t_1, \dots, t_d)^T, \quad H_{ij}(x) = \partial^2 / (\partial x_i \partial x_j) H(x), \quad K^\#(t) = \prod_{j=1}^d K(t_j),$$

$$m_i = \max\{-1, (x_i - 1)/b\}, \quad k_i = \min\{1, x_i/b\}, \quad c_{ij} = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} t_i t_j K^*(t) dt,$$

$$\lambda_0 = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} K^\#(t) dt, \quad \tau_0 = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} K^\#(t)^2 dt, \quad \lambda_{i,k} = \int_{m_i}^{k_i} u^k K(u) du,$$

$Q$  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  $i, j = 1, \dots, d$  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

$$r(x, z) = \frac{f_0(x, z) \exp\{H(x) + c^* + \theta z\}}{\zeta + (1 - \zeta) \exp\{H(x) + c^* + \theta z\}}, \quad D_0(x) = \sum_{z \in Q} r(x, z), \quad D_1(x) = \sum_{z \in Q} z r(x, z),$$

$$D_2(x) = \sum_{z \in Q} z z^T r(x, z), \quad D_j = \int_0^1 \dots \int_0^1 D_j(x) dx, \quad \text{for } j = 0, 1, 2, \quad D = \begin{pmatrix} D_0, & D_1^T \\ D_1, & D_2 \end{pmatrix}.$$

Define quantities related to the asymptotic biases and variances of  $\hat{H}(x)$ ,  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$  in the following:

$$c_H(x; b) = (1/2) \sum_{i=1}^d \sum_{j=1}^d c_{ij} H_{ij}(x), \quad v_{H,1}(x; b) = D_0(x)^{-1} \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} K^*(t)^2 dt,$$

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

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

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

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

$$c_H(x; b) = (1/2) \left\{ \int_{-1}^1 u^2 K(u) du \right\} \left\{ \sum_{i=1}^d H_{ii}(x) \right\}, \quad v_{H,1}(x; b) = D_0(x)^{-1} \left\{ \int_{-1}^1 K(u)^2 du \right\}^d.$$

The following Theorem 1 states the asymptotic bias and variance for  $\hat{H}(x)$ , and those for  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$ . Its proof will be given in Appendix B.

**Theorem 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  $n \rightarrow \infty$ :

$$\text{Bias}\{\hat{H}(x)\} = E\{\hat{H}(x)\} - H(x) - c^* = b_\theta^2 c_H(x; b_\theta) + O(b_\theta^3 + n^{-1} b_\theta^{-d}), \tag{9}$$

$$\text{Var}\{\hat{H}(x)\} = n^{-1} b_\theta^{-d} \zeta^{-1} (1 - \zeta)^{-1} \{v_{H,1}(x; b_\theta) + v_{H,2}(x; b_\theta)\} + O(n^{-1} b_\theta^{-d+1}), \tag{10}$$

$$\text{Bias}\{\hat{\theta}_{CC}\} = E\{\hat{\theta}_{CC}\} - \theta = b_\theta^2 c_\theta(b_\theta) + O(b_\theta^3 + n^{-1} b_\theta^{-d}), \tag{11}$$

$$\text{Var}\{\hat{\theta}_{CC}\} = n^{-1} \zeta^{-1} (1 - \zeta)^{-1} v_\theta + O(n^{-1} b_\theta), \tag{12}$$

$$\text{Bias}\{\hat{H}_{CC}(x)\} = E\{\hat{H}_{CC}(x)\} - H(x) - c^* = b_H^2 c_H(x; b_H) + O(b_H^3 + b_\theta^2 + n^{-1} b_H^{-d}), \tag{13}$$

$$\text{Var}\{\hat{H}_{CC}(x)\} = n^{-1} b_H^{-d} \zeta^{-1} (1 - \zeta)^{-1} v_{H,1}(x; b_H) + O(n^{-1} b_H^{-d+1}). \tag{14}$$

**Remark 1:** The optimal kernel function  $K$  and the magnitudes of the optimal bandwidth parameters  $b_\theta^*$  and  $b_H^*$  for constructing  $\hat{\theta}_{CC}$  and  $\hat{H}_{CC}(x)$ . By Theorem 8 of Fan *et al.* (1993) and our Theorem 1, the optimal  $K$  satisfying the conditions in C1 for constructing  $\hat{H}_{CC}(x)$  is the Epanechnikov kernel, for each  $x \in [0, 1]^d$ , in the sense of having smaller asymptotic mean square error. On the other hand, by (13) and (14), the optimal choice of  $b_H$ , in terms of having smallest mean square error of  $\hat{H}_{CC}(x)$ , is  $b_H^* = c_H^* n^{-1/(d+4)}$ , where  $c_H^*$  is a constant depending on the unknown factors  $H(\cdot)$ ,  $\theta$  and  $f_0(x, z)$ . Similarly, by (9)–(12) and C3, the optimal value  $b_\theta^*$  of  $b_\theta$ , in terms of having smallest mean square error of  $\hat{\theta}_{CC}$ , satisfies the condition  $n^{-1/(d+4)} \gg b_\theta^* \gg n^{-1/(d+2)}$ . Hence we conclude that the value of  $b_H^*$  is of larger order than that of  $b_\theta^*$ , and that the mean square error of  $\hat{H}_{CC}(x)$  using the optimal bandwidth parameter  $b_H^*$  is of smaller order in magnitude than that of  $\hat{H}(x)$  using  $b_\theta^*$ .

### APPENDIX B: SKETCHES OF THE PROOFS

In Appendix B, sketches of the proof for Theorem 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  $\ell_j$ , for each  $j = 1, 2, 3$ , given in (6)–(8), 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  $\rho_0$  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  $\varphi_0$  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 (3). The events  $P_1$  and  $Q_1$  are similarly defined for case data points with  $n_0, f_0, \rho_0$  and  $\varphi_0$  replaced respectively by  $n_1, f_1, \rho_1$  and  $\varphi_1$ .

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

**Proof of the asymptotic bias and variance for  $\hat{H}(x)$**

Set  $w = (\alpha, \beta, \theta)^T$  and  $\hat{w} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})^T$ , the maximizer of  $\ell_1(\alpha, \beta, \theta)$  in (6). By the first-order Taylor expansion, we have

$$0 = \ell_1^{(1)}(\hat{w}) = \ell_1^{(1)}(w) + \ell_1^{(2)}(w^*)(\hat{w} - w), \tag{15}$$

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

Using conditions C1–C5, (5), and the large deviation theorem in Section 10.3.1 of Serfling (1980), a straightforward calculation leads to the following asymptotic results: as  $n \rightarrow \infty$ ,

$$P(P_0 \cup Q_0 \cup P_1 \cup Q_1) = O\{\exp(-nb_\theta)\}, \tag{16}$$

$$E\{\ell_1^{(1)}(w)\} = (1/2)nb_\theta^2\zeta(1-\zeta)A_1 + O(nb_\theta^3 + b_\theta^{-d}), \tag{17}$$

$$E\{\ell_1^{(2)}(w)\} = (-1)n\zeta(1-\zeta)B_1 + O(nb_\theta), \tag{18}$$

$$\text{var}\{\ell_1^{(1)}(w)\} = nb_\theta^{-d}\zeta(1-\zeta)C_1 + O(nb_\theta^{-d+1}), \tag{19}$$

for each  $w$ . Here

$$A_1 = \{u_0(x)D_0(x), u_1^T(x)D_0(x), u_0(x)D_1^T(x)\}^T, \\ B_1 = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} \Lambda_1(x, t)K^\#(t)dt, \quad C_1 = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} \Lambda_1(x, t)K^\#(t)^2 dt,$$

where

$$u_0(x) = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} \{t^T H^{(2)}(x)t\}K^\#(t)dt, \quad u_1(x) = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} \{t^T H^{(2)}(x)t\}tK^\#(t)dt, \\ \Lambda_1(x, t) = \begin{pmatrix} D_0(x), & t^T D_0(x), & D_1^T(x) \\ tD_0(x), & tt^T D_0(x), & tD_1^T(x) \\ D_1(x), & D_1(x)t^T, & D_2(x) \end{pmatrix}.$$

Using condition C3 and the results of (16)–(19) and comparing the magnitudes of  $\ell_1^{(1)}(w) = O_p(nb_\theta^2 + n^{1/2}b_\theta^{-d/2})$  and  $\ell_1^{(2)}(w^*) = O_p(n)$  in (15), we have

$$\hat{w} - w = o_p(1). \tag{20}$$

Using (15)–(20) 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  $\hat{H}(x)$  in (9) and (10) follow, respectively.

**Proof of the asymptotic bias and variance for  $\hat{\theta}_{CC}$**

Set  $\gamma = (\theta_0, \theta)$  and  $\hat{\gamma} = (\hat{\theta}_0, \hat{\theta}_{CC})$ , the maximizer of  $\ell_2(\theta_0, \theta)$  in (7). Using the fact that  $\theta_0$  is a normalizing constant for  $f_1(x, z) = f_0(x, z)\exp\{\hat{H}(x) + \theta_0 + \theta_z\}$ , the results of the asymptotic bias and variance of  $\hat{H}(x)$  in (9) and (10), C1–C5, (5), and approximations to the standard errors of functions

of random variables, through a straightforward calculation, we have  $\exp(\theta_0) = (1 + b_{\theta}^2 c_1 + c_2)^{-1} \{1 + o_p(1)\}$ . Here  $c_1 = \int_0^1 \dots \int_0^1 f_i(x) c_H(x; b_{\theta}) dx$ ,  $c_2 = \int_0^1 \dots \int_0^1 f_i(x) [\hat{H}(x) - E\{\hat{H}(x)\}] dx$ . Next, using this result and C1–C5, through a straightforward calculation, we have, as  $n \rightarrow \infty$ ,

$$E\{\ell_2^{(1)}(\gamma)\} = (-1)nb_{\theta}^2\zeta(1-\zeta)\left[\int_0^1 \dots \int_0^1 \{D_0(x), D_1^T(x)\}c_H(x; b_{\theta})dx - (D_0, D_1^T)c_1\right] + O(nb_{\theta}^3), \tag{21}$$

$$E\{\ell_2^{(2)}(\gamma)\} = (-1)n\zeta(1-\zeta)D + O(nb_{\theta}), \tag{22}$$

$$\text{var}\{\ell_2^{(1)}(\gamma)\} = n\zeta(1-\zeta)\left\{D - (D_0, D_1^T)^T(D_0, D_1^T)\right\} + O(nb_{\theta}), \tag{23}$$

for each  $\gamma$ .

Following the same arguments as those of (15) and (20) and using (21)–(23), we have  $\hat{\gamma} - \gamma = o_p(1)$ . 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  $\hat{\theta}_{CC}$  in (11) and (12) follow, respectively.

**Proof of the asymptotic bias and variance for  $\hat{H}_{CC}(x)$**

Set  $\eta = (\alpha, \beta)^T$  and  $\hat{\eta} = (\hat{\alpha}, \hat{\beta})^T$ , the maximizer of  $\ell_3(\alpha, \beta)$  in (8). Using C1–C5, (5), the asymptotic bias and variance of  $\hat{\theta}_{CC}$  in (11) and (12), and approximations to the standard errors of functions of random variables, we have, as  $n \rightarrow \infty$ ,

$$E\{\ell_3^{(1)}(\eta)\} = (1/2)nb_H^2\zeta(1-\zeta)A_3 + O(nb_H^3), \tag{24}$$

$$E\{\ell_3^{(2)}(\eta)\} = (-1)n\zeta(1-\zeta)B_3 + O(nb_H), \tag{25}$$

$$\text{var}\{\ell_3^{(1)}(\eta)\} = nb_H^{-d}\zeta(1-\zeta)C_3 + O(nb_H^{-d+1}), \tag{26}$$

for each  $\eta$ , where

$$A_3 = \{u_0(x), u_1^T(x)\}^T D_0(x), \quad B_3 = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} \Lambda_3(x, t)K^{\#}(t)dt,$$

$$c_3 = \int_{m_1}^{k_1} \dots \int_{m_d}^{k_d} \Lambda_3(x, t)K^{\#}(t)^2 dt, \quad \Lambda_3(x, t) = \begin{pmatrix} 1, & t^T \\ t, & tt^T \end{pmatrix} D_0(x).$$

Following the same arguments as those of (15) and (20) and using (16) and (24)–(26), we have  $\hat{\eta} - \eta = o_p(1)$ . Combining this result with (24)–(26) and using approximations to the standard errors of functions of random variables, the results of the asymptotic bias and variance of  $\hat{H}_{CC}(x)$  in (13) and (14) follow, respectively. Hence the proof of Theorem 1 is completed.

ACKNOWLEDGEMENTS

The authors thank the Editor and two referees for their kind suggestions, which greatly improved the presentation of this paper. The research of the second and the third authors was supported by the National Science Council, Taiwan, Republic of China.

## REFERENCES

- Altman EI. 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *Journal of Finance* **23**: 589–609.
- Begley J, Ming J, Watts S. 1996. Bankruptcy classification errors in the 1980s: an empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies* **1**: 267–284.
- Bharath ST, Shumway T. 2004. Forecasting default with the KMV–Merton model. AFA 2006 Boston Meetings paper. <http://ssrn.com/abstract=637342> [3 February 2006].
- Chava S, Jarrow RA. 2004. Bankruptcy prediction with industry effects. *Review of Finance* **8**: 537–569.
- Cox DR, Oakes D. 1984. *Analysis of Survival Data*. Chapman & Hall: London.
- Fan J, Gasser T, Gijbels I, Brookmann M, Engel J. 1993. Local polynomial fitting: A standard for nonparametric regression. Discussion paper 9315, Institut de Statistique, Universite Catholique de Louvain, Belgium.
- Fan J, Marron JS. 1994. Fast implementations of nonparametric curve estimators. *Journal of Computational and Graphical Statistics* **3**: 35–56.
- Farewell V. 1979. Some results on the estimation of logistic models based on retrospective data. *Biometrika* **66**: 27–32.
- Fernandez C, Steel MFJ. 1998. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**: 359–371.
- Frydman H, Altman EI, Kao DL. 1985. Introducing recursive partitioning for financial classification: the case of financial distress. *Journal of Finance* **40**: 269–291.
- Grice JS, Dugan MT. 2001. The limitations of bankruptcy prediction models: Some cautions for the research. *Review of Quantitative Finance and Accounting* **17**: 151–166.
- Härdle W. 1991. *Smoothing Techniques: With Implementation in S*. Springer: Berlin.
- Härdle W, Hart J, Marron JS, Tsybakov AB. 1992. Bandwidth selection for average derivative estimation. *Journal of the American Statistical Association* **87**: 218–226.
- Härdle W, Moro RA, Schäfer D. 2006. Graphical data representation in bankruptcy analysis. In *Handbook of Computational Statistics, Data Visualization*, Chen CH, Härdle W, Urwin A (eds.). Springer: Heidelberg.
- Hastie T, Tibshirani R. 1990. *Generalized Additive Models*. Chapman & Hall: London.
- Hosmer DJ, Lemeshow S. 1989. *Applied Logistic Regression*. Wiley: New York.
- Jones MC, Marron JS, Sheather SJ. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* **91**: 401–407.
- Koh HC, Tan SS. 1999. A neural network approach to the prediction of going concern status. *Accounting and Business Research* **29**: 211–216.
- Lane W, Looney S, Wansley J. 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance* **10**: 511–531.
- Lejeune M, Sarda P. 1992. Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis* **14**: 457–471.
- Lindsay DH, Campbell A. 1996. A chaos approach to bankruptcy prediction. *Journal of Applied Business Research* **12**: 1–9.
- Little RJA, Rubin DB. 2002. *Statistical Analysis with Missing Data*. Wiley: New York.
- McKee TE. 2003. Rough sets bankruptcy prediction models versus auditor signaling rates. *Journal of Forecasting* **22**: 569–586.
- Messier JW, Hansen JV. 1988. Inducing rules for expert system development: an example using default and bankruptcy data. *Management Science* **34**: 1403–1415.
- Ohlson J. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* **18**: 109–131.
- Prentice RL, Pyke R. 1979. Logistic disease incidence models and case–control studies. *Biometrika* **66**: 403–411.
- Serfling RJ. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley: New York.
- Shumway T. 2001. Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business* **74**: 101–124.
- Siegrist K. 2005. The Pareto distribution. [http://www.ds.unifi.it/VL/VL\\_EN/special/special12.html](http://www.ds.unifi.it/VL/VL_EN/special/special12.html) [3 February 2006].
- Stuart A, Ord JK. 1987. *Kendall's Advanced Theory of Statistics*, Vol. 1. Oxford University Press: New York.
- Tibshirani R, Hastie T. 1987. Local likelihood estimation. *Journal of the American Statistical Association* **82**: 559–568.

- Vassalou M, Xing Y. 2004. Default risk in equity returns. *Journal of Finance* **59**: 831–868.
- Wand MP, Jones MC. 1995. *Kernel Smoothing*. Chapman & Hall: London.
- Zhao LP, Kristal AR, White E. 1996. Estimating relative risk functions in case-control studies using a nonparametric logistic regression. *American Journal of Epidemiology* **144**: 598–609.
- Zmijewski ME. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* **22**: 59–82.

*Authors' biographies:*

**Ruey-Ching Hwang** is Associate Professor of Finance at Dahan Institute of Technology. Her research interests include forecasting, kernel smoothing, quantitative finance, and risk management.

**K. F. Cheng** is Professor of Statistics at National Central University, and Professor of Biostatistics at China Medical University. His research interests include biostatistics and quantitative finance.

**Jack C. Lee**, an ASA Fellow, is University Chair Professor of Finance at National Chiao Tung University. His research interests include Bayesian inference, financial statistics, forecasting, growth curve, and time series analysis.

*Authors' addresses:*

**Ruey-Ching Hwang**, Department of Finance and Banking, Dahan Institute of Technology, 1 Su-Ren Street, Hualien, Taiwan 97145, Taiwan.

**K. F. Cheng**, Institute of Statistics, National Central University, 300 Zhongda Road, Zhongli, Taiwan 32001, Taiwan.

**Jack C. Lee**, Graduate Institute of Finance, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu, Taiwan 30050, Taiwan.