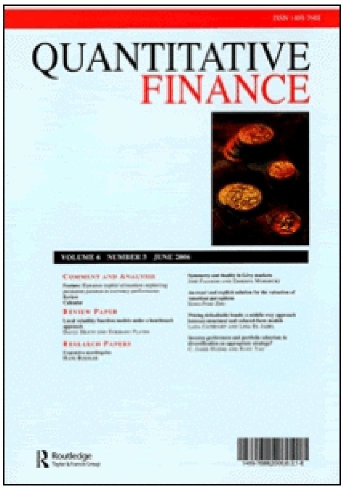
This article was downloaded by: *[National Dong Hwa University]*
On: *21 October 2010*
Access details: *Access Details: [subscription number 918619350]*
Publisher *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Quantitative Finance
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713665537

Predicting bankruptcy using the discrete-time semiparametric hazard model

K. F. Cheng[a]; C. K. Chu[b]; Ruey-Ching Hwang[c]
[a] Biostatistics Center, China Medical University, Taichung, Taiwan, and Institute of Statistics, National Central University, Jhongli, Taiwan [b] Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan [c] Department of Finance, National Dong Hwa University, Hualien, Taiwan

First published on: 23 July 2009

To cite this Article Cheng, K. F. , Chu, C. K. and Hwang, Ruey-Ching(2010) 'Predicting bankruptcy using the discrete-time semiparametric hazard model', Quantitative Finance, 10: 9, 1055 — 1066, First published on: 23 July 2009 (iFirst)
To link to this Article: DOI: 10.1080/14697680902814274
URL: http://dx.doi.org/10.1080/14697680902814274

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Quantitative Finance, Vol. 10, No. 9, November 2010, 1055–1066

Routledge
Taylor & Francis Group

Predicting bankruptcy using the discrete-time semiparametric hazard model

K. F. CHENG†, C. K. CHU‡ and RUEY-CHING HWANG*§

†Biostatistics Center, China Medical University, Taichung, Taiwan, and Institute of Statistics, National Central University, Jhongli, Taiwan
‡Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan
§Department of Finance, National Dong Hwa University, Hualien, Taiwan

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

(*Received 25 October 2007; in final form 6 February 2009*)

The usual bankruptcy prediction models are based on single-period data from firms. These models ignore the fact that the characteristics of firms change through time, and thus they may suffer from a loss of predictive power. In recent years, a discrete-time parametric hazard model has been proposed for bankruptcy prediction using panel data from firms. This model has been demonstrated by many examples to be more powerful than the traditional models. In this paper, we propose an extension of this approach allowing for a more flexible choice of hazard function. The new method does not require the assumption of a parametric model for the hazard function. In addition, it also provides a tool for checking the adequacy of the parametric model, if necessary. We use real panel datasets to illustrate the proposed method. The empirical results confirm that the new model compares favorably with the well-known discrete-time parametric hazard model.

Keywords: Discrete-time hazard model; Local likelihood; Out-of-sample error rate; Panel data; Semiparametric model

1. Introduction

Bankruptcy prediction has been routinely applied by academics, practitioners, and regulators. The well-known prediction models include the discriminant analysis model (Altman 1968), the KMV-Merton model (Merton 1974, Vassalou and Xing 2004), the linear logit model (Ohlson 1980), and the probit model (Zmijewski 1984), to name only a few. The common principle of these approaches is that the models are developed using only single-period data from the studied firms. Shumway (2001) criticized that such prediction processes are static in nature, since they ignore the changing characteristics of firms through time. In order to avoid the possible loss of predictive power due to the use of static models, Shumway (2001) and Chava and Jarrow (2004) suggested that a discrete-time hazard model (DHM) could be used for bankruptcy prediction. Their analyses are based on applying the idea of survival analysis (Cox and Oakes 1984). This novel model has the advantage of using all available information of firms to build up a prediction system so that each firm's bankruptcy risk at each time point can be determined. Thus the model is a dynamic forecasting model. Other bankruptcy forecasting models based on multiple-period data include, for example, Hillegeist *et al.* (2004), Bharath and Shumway (2008), and Chava *et al.* (2008) using the same idea of survival analysis, and Duffie (2005) and Duffie *et al.* (2007) making use of different ideas in point processes. Approaches based on neural networks (Atiya 2001), support vector machines (Härdle *et al.* 2008), and Bayesian networks (Sun and Shenoy 2007), etc., have also been introduced in the literature for bankruptcy prediction.

According to Shumway (2001) and Chava and Jarrow (2004), the important parameters in DHM are determined by maximizing a log-likelihood function. However, their approach is not flexible enough for modeling the hazard function. One major assumption needed in their DHM is that the hazard function has to be a parametric function such as a simple linear logistic function. Unfortunately, the parametric model assumption is not always true in all applications. Härdle *et al.* (2008) also pointed out that

*Corresponding author. Email: rchwang@mail.ndhu.edu.tw

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2010 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/14697680902814274

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

many well-known parametric models for bankruptcy prediction are not proper. In general, it is difficult to ensure that conclusions based on the parametric model are meaningful unless we have a large dataset and powerful lack-of-fit tests to confirm that the parametric model is most appropriate in bankruptcy prediction. The latter often seems to be impossible, particularly when one does not have a sufficient number of sample observations for analysis. To avoid this potential pitfall, we show in this paper that the idea of the semiparametric logit model (Hwang *et al.* 2007) can be directly extended to the DHM using panel data. Specifically, we shall propose a discrete-time semiparametric hazard model (DSHM) for bankruptcy prediction. This model is built on the work of DHM but needs not assume any parametric form for the hazard function. If necessary, the result of the proposed modeling strategy can also guide us as to how to determine the most appropriate parametric form for the hazard function.

In the literature, two types of hazard function have been considered. The first type of hazard function, after taking logit transformation, is a linear function of predictors; see Shumway (2001) and Chava and Jarrow (2004). The second type uses the discrete-time proportional hazard function; see Allison (1982). Our semiparametric approach can be applied on these two types of hazard function. However, our development of DSHM is mainly based on the logistic hazard function, since this function is often used for predicting bankruptcy.

The remainder of this paper is organized as follows. In section 2, we first outline the basic idea of DHM. We then point out that the rationale of DSHM is similar to that of the semiparametric logit model of Hwang *et al.* (2007). The method is developed under the concept of local likelihood, and it turns out that the important estimators needed in DSHM can be derived from solving a simple system of weighted normal equations. Thus, the required computation is as simple as that in DHM. In section 3, we illustrate our method using four panel datasets based on the predictors suggested by Ohlson (1980) and Shumway (2001). Each panel dataset was analysed using DSHM, DHM, and modified DHM. The modified DHM is an improved parametric version of DHM using results developed from DSHM. The predictive power of each method was measured by out-of-sample error rates. Based on the error rates summarized in section 3, we conclude that DSHM has better performance in all cases. Sometimes, the improvement of both DSHM and modified DHM over DHM is very significant, depending on the predictors selected in the study. This shows that DSHM has potential as a powerful bankruptcy prediction model. Finally, concluding remarks appear in section 4.

2. Methodology

In this section, we describe the basic idea of DSHM, develop estimating equations for unknown quantities and introduce a tool for visually checking the adequacy of the linear logistic hazard function in DHM. Before doing this, we first briefly review the basic steps for deriving DHM.

2.1. *DHM*

The DHM can be formally defined from the log-likelihood function of the panel data. The model has the advantage of using all available information to predict each firm's bankruptcy risk at each point in time. In the following, we describe the structure of the panel data used in the prediction model.

The panel data are determined by two factors. They are the sampling period and sampling criteria. In this paper, the panel datasets analysed in section 3 were sampled from January 1984 to December 2000, and all firms starting their listing on the New York Stock Exchange, American Stock Exchange, or NASDAQ during the sampling period were recruited in the sample. All information at the discrete time points during the sampling period were collected from both COMPUSTAT and CRSP databases. Assume that there are n selected companies under the particular sampling scheme. We denote the panel data by

$$\{(Y_{i,j}, x_{i,j}, z_{i,j}),\ j = 1, \ldots, t_i,\ i = 1, \ldots, n\}.$$

Here, for the ith firm in the dataset, we denote $t_i \in \{1, \ldots, \xi\}$ to be the length of the firm's duration during the sampling period, and ξ is a positive integer indicating the total length of the sampling period. At the last observation time t_i, $Y_{i,t_i} = 1$ indicates that the ith company is bankrupt, and $Y_{i,t_i} = 0$ otherwise. At the observation time $j < t_i$, we always have $Y_{i,j} = 0$. Finally, we let $x_{i,j}$ and $z_{i,j}$ be values of the $d \times 1$ continuous and $q \times 1$ discrete explanatory variables X and Z collected at time j, respectively.

The log-likelihood function of the panel data has been given in (21) of Allison (1982). It is expressed as

$$\ell_{\mathrm{DHM}} = \sum_{i=1}^{n} Y_{i,t_i} \log\left\{\frac{h\left(t_i, x_{i,t_i}, z_{i,t_i}\right)}{1 - h\left(t_i, x_{i,t_i}, z_{i,t_i}\right)}\right\} + \sum_{i=1}^{n} \sum_{j=1}^{t_i} \log\{1 - h(j, x_{i,j}, z_{i,j})\}.$$

Here, $h(j, x_{i,j}, z_{i,j})$ is the value of the hazard function indicating the probability of bankruptcy instantly occurring at time j for the ith company which is non-bankrupt before time j, for each $j = 1, \ldots, t_i$ and $i = 1, \ldots, n$.

Note that the hazard function $h(t, x, z)$ in ℓ_{DHM} can be of any functional form with values in the interval $(0, 1)$. Shumway (2001) considered a linear logistic function for the hazard function:

$$h(t, x, z) = \frac{\exp\{\alpha_1 + \beta_1 \log(t) + \gamma_1 x + \theta_1 z\}}{1 + \exp\{\alpha_1 + \beta_1 \log(t) + \gamma_1 x + \theta_1 z\}},$$

where α_1, β_1, γ_1, and θ_1 are 1×1, 1×1, $1 \times d$, and $1 \times q$ vectors of parameters, respectively. Given the linear

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

logistic hazard function, the resulting log-likelihood of the panel data becomes

$$\ell = \sum_{i=1}^{n} Y_{i,t_i}\left\{\alpha_1 + \beta_1 \log(t_i) + \gamma_1 x_{i,t_i} + \theta_1 z_{i,t_i}\right\} - \sum_{i=1}^{n}\sum_{j=1}^{t_i} \log[1 + \exp\{\alpha_1 + \beta_1 \log(j) + \gamma_1 x_{i,j} + \theta_1 z_{i,j}\}].$$

The maximum likelihood estimates of parameters α_1, β_1, γ_1, and θ_1 can be simply obtained by solving the normal equations:

$$0 = \sum_{i=1}^{n} Y_{i,t_i}\begin{bmatrix} 1 \\ \log(t_i) \\ x_{i,t_i} \\ z_{i,t_i} \end{bmatrix} - \sum_{i=1}^{n}\sum_{j=1}^{t_i} \frac{\exp\{\alpha_1 + \beta_1 \log(j) + \gamma_1 x_{i,j} + \theta_1 z_{i,j}\}}{1 + \exp\{\alpha_1 + \beta_1 \log(j) + \gamma_1 x_{i,j} + \theta_1 z_{i,j}\}} \times \begin{bmatrix} 1 \\ \log(j) \\ x_{i,j} \\ z_{i,j} \end{bmatrix}.$$

Based on the maximum likelihood estimates $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\gamma}_1$, and $\hat{\theta}_1$, if a firm has predictor values (x_0, z_0) at time t_0, then its predicted instant bankruptcy probability can be given by

$$\hat{h}(t_0, x_0, z_0) = \frac{\exp\{\hat{\alpha}_1 + \hat{\beta}_1 \log(t_0) + \hat{\gamma}_1 x_0 + \hat{\theta}_1 z_0\}}{1 + \exp\{\hat{\alpha}_1 + \hat{\beta}_1 \log(t_0) + \hat{\gamma}_1 x_0 + \hat{\theta}_1 z_0\}}.$$

Cox and Oakes (1984) showed that the maximum likelihood estimates $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\gamma}_1$, and $\hat{\theta}_1$ are consistent for α_1, β_1, γ_1, and θ_1, respectively. Thus, the resulting predicted instant bankruptcy probability $\hat{h}(t_0, x_0, z_0)$ converges to the true instant bankruptcy probability $h(t_0, x_0, z_0)$. This result shows that DHM should be an efficient bankruptcy prediction model if the hazard function is correctly specified.

2.2. DSHM

The main advantage of DHM lies in its simplicity of computation and interpretation, but the linear logistic function for modeling the hazard function may not be proper. If one chooses a parametric hazard function that is not appropriate, then the resulting model-based instant bankruptcy probability prediction might not correctly estimate the true probability, and there is a danger of coming to an erroneous prediction.

The limitation of DHM can be improved by removing the restriction that the hazard function belongs to a particular parametric family. In this paper, we suggest a DSHM, which is more flexible in modeling the hazard function. The DSHM is constructed by replacing the parametric hazard function in DHM with a semiparametric hazard function. That is, we assume the hazard function belongs to the family

$$h^*(t, x, z) = \frac{\exp\{\beta \log(t) + m(x) + \theta z\}}{1 + \exp\{\beta \log(t) + m(x) + \theta z\}}.$$

Here, β and θ are unknown parameters, and $m(x)$ is an unknown but smooth function of the value x of the d-dimensional continuous predictor X. Following the same development of ℓ, the corresponding log-likelihood function of the panel data based on our DSHM is expressed by

$$\ell^* = \sum_{i=1}^{n} Y_{i,t_i}\left\{\beta \log(t_i) + m\left(x_{i,t_i}\right) + \theta z_{i,t_i}\right\} - \sum_{i=1}^{n}\sum_{j=1}^{t_i} \log[1 + \exp\{\beta \log(j) + m(x_{i,j}) + \theta z_{i,j}\}].$$

For a company with predictor values (x_0, z_0) at time t_0, if β, $m(x_0)$, and θ can be efficiently estimated by $\hat{\beta}$, $\hat{m}(x_0)$, and $\hat{\theta}$, respectively, then the firm's instant bankruptcy probability can be predicted by

$$\hat{h}^*(t_0, x_0, z_0) = \frac{\exp\left\{\hat{\beta} \log(t_0) + \hat{m}(x_0) + \hat{\theta} z_0\right\}}{1 + \exp\left\{\hat{\beta} \log(t_0) + \hat{m}(x_0) + \hat{\theta} z_0\right\}}.$$

In sections 2.3 and 2.4, we show how to estimate parameters β, $m(x_0)$, and θ using a local likelihood method. The advantage of this approach will be seen from empirical studies given in section 3.

2.3. A local likelihood method

There exist many well-known methods for estimating β, $m(x_0)$, and θ, where x_0 is any given value of the d-dimensional continuous predictor X. One of these methods with a simple idea is the local likelihood method; see, for example, Tibshirani and Hastie (1987), Staniswallis (1989), Fan *et al.* (1995), and Hwang *et al.* (2007). The basic rational of the local likelihood method is to center the data around x_0 and weight the likelihood in such a way that it places more emphasis on those observations nearest to x_0.

The idea of the local likelihood method can be simply explained by first introducing a neighborhood $S(x_0) = \{x = (x_1, \ldots, x_d)^T : \|x - x_0\| \le b\}$ of x_0. Here b is some positive constant to be determined later by the sampled data, and called the bandwidth. The notation $\|x\|$ denotes the Euclidean distance of the given vector x. If the value of b is small enough and $x_{i,j}$ belongs to $S(x_0)$, then Taylor's first-order expansion states that

$$m(x_{i,j}) \approx m(x_0) + m^{(1)}(x_0)^T (x_{i,j} - x_0),$$

and such $m(x_{i,j})$ in the likelihood can be written as $\mu_0 + \mu_1(x_{i,j} - x_0)$, where we denote $m(x_0)$ by μ_0 and $m^{(1)}(x_0)^T$ by μ_1. Note that μ_0 is a scalar parameter and μ_1 is a $1 \times d$ vector of parameters.

To make an inference about $\omega=(\mu_0,\beta,\mu_1,\theta)$, we suggest modifying the likelihood function ℓ^* to the following 'local (weighted)' log-likelihood function:

$$\ell_0^*(\omega;x_0)=\sum_{i=1}^n Y_{i,t_i}\{\mu_0+\beta\log(t_i)+\mu_1(x_{i,t_i}-x_0)+\theta z_{i,t_i}\}\times W(x_{i,t_i})-\sum_{i=1}^n\sum_{j=1}^{t_i}\log[1+\exp\{\mu_0+\beta\log(j)+\mu_1(x_{i,j}-x_0)+\theta z_{i,j}\}]W(x_{i,j}).$$

Here $W(x)$ is called the weight function and the simplest weight $W(x_{i,j})$ assigned to the observation $(Y_{i,j},x_{i,j},z_{i,j})$ is the indicator value $I\{x_{i,j}\in S(x_0)\}$. However, a more general weighting scheme can be used for defining the local likelihood. This can be achieved, for example, by introducing a symmetric and unimodal probability density function $K_b(x)$, and defining $W(x_{i,j})=K_b(x_{i,j}-x_0)$. In the paper, we suggest that $K_b(x)$ be taken as the joint probability density function of d independent normal random variables $N(0,b^2)$. Given such $K_b(x)$ we point out that if the value of $b^{-1}\|x_{i,j}-x_0\|$ becomes larger because of choosing a smaller value of b, then the effect of the observation $(Y_{i,j},x_{i,j},z_{i,j})$ on estimating the important parameters in DSHM will tend to be smaller or even non-existent. This indicates that the value of b can be used to control sample observations for inclusion in the analysis. The results in the literature show that the choice of bandwidth b plays an important role in the analysis. Some discussions of the above weighting method can be found in the monographs of Eubank (1988), Müller (1988), Härdle (1990, 1991), Scott (1992), Wand and Jones (1995), Fan and Gijbels (1996), and Simonoff (1996), etc. In this paper, we select $W(x_{i,j})=K_b(x_{i,j}-x_0)$ in all analyses.

Set $\tilde\omega=(\tilde\mu_0,\tilde\beta,\tilde\mu_1,\tilde\theta)$ as the maximizer of $\ell_0^*(\omega;x_0)$. The maximum local likelihood estimate $\tilde\omega$ can also be equivalently obtained by solving a system of weighted normal equations

$$0=\sum_{i=1}^n Y_{i,t_i}\begin{bmatrix}1\\ \log(t_i)\\ x_{i,t_i}-x_0\\ z_{i,t_i}\end{bmatrix}K_b(x_{i,t_i}-x_0)-\sum_{i=1}^n\sum_{j=1}^{t_i}\frac{\exp\{\mu_0+\beta\log(j)+\mu_1(x_{i,j}-x_0)+\theta z_{i,j}\}}{1+\exp\{\mu_0+\beta\log(j)+\mu_1(x_{i,j}-x_0)+\theta z_{i,j}\}}\times\begin{bmatrix}1\\ \log(j)\\ x_{i,j}-x_0\\ z_{i,j}\end{bmatrix}K_b(x_{i,j}-x_0).$$

We define $\tilde m(x_0)=\tilde\mu_0$ to indicate that it is an estimate of $m(x_0)$. We also point out that β and θ are global parameters and their corresponding estimates produced from $\tilde\omega$ may not be efficient, since such estimates are derived by maximizing a local log-likelihood depending on x_0. In section 2.4, we show how more efficient estimates of β, $m(x_0)$, and θ can be achieved.

2.4. More powerful estimates of parameters in DSHM

More powerful estimates of β, $m(x_0)$, and θ can be derived using the following two-step procedure. We first note that, for each value $x_{i,j}$, an initial estimate $\tilde m(x_{i,j})$ of $m(x_{i,j})$ can be obtained by the method outlined in section 2.3. The two-step procedure includes the following.

Step 1: β and θ are estimated by maximizing the pseudo log-likelihood

$$\ell_1^*(\beta,\theta)=\sum_{i=1}^n Y_{i,t_i}\{\beta\log(t_i)+\tilde m(x_{i,t_i})+\theta z_{i,t_i}\}-\sum_{i=1}^n\sum_{j=1}^{t_i}\log[1+\exp\{\beta\log(j)+\tilde m(x_{i,j})+\theta z_{i,j}\}],$$

or, equivalently, solving equations

$$0=\sum_{i=1}^n Y_{i,t_i}\begin{bmatrix}\log(t_i)\\ z_{i,t_i}\end{bmatrix}-\sum_{i=1}^n\sum_{j=1}^{t_i}\frac{\exp\{\beta\log(j)+\tilde m(x_{i,j})+\theta z_{i,j}\}}{1+\exp\{\beta\log(j)+\tilde m(x_{i,j})+\theta z_{i,j}\}}\begin{bmatrix}\log(j)\\ z_{i,j}\end{bmatrix}.$$

Let the estimates of (β,θ) be $(\hat\beta,\hat\theta)$, the maximizer of $\ell_1^*(\beta,\theta)$. Here $\ell_1^*(\beta,\theta)$ is obtained by replacing each $m(x_{i,j})$ in ℓ^* with its initial estimate $\tilde m(x_{i,j})$.

Step 2: $m(x_0)$ is estimated by maximizing the pseudo local log-likelihood

$$\ell_2^*(\mu_0,\mu_1;x_0)=\sum_{i=1}^n Y_{i,t_i}\left\{\mu_0+\hat\beta\log(t_i)+\mu_1(x_{i,t_i}-x_0)+\hat\theta z_{i,t_i}\right\}\times K_g(x_{i,t_i}-x_0)-\sum_{i=1}^n\sum_{j=1}^{t_i}\log[1+\exp\{\mu_0+\hat\beta\log(j)+\mu_1(x_{i,j}-x_0)+\hat\theta z_{i,j}\}]K_g(x_{i,j}-x_0),$$

or, equivalently, solving equations

$$0=\sum_{i=1}^n Y_{i,t_i}\begin{bmatrix}1\\ x_{i,t_i}-x_0\end{bmatrix}K_g(x_{i,t_i}-x_0)-\sum_{i=1}^n\sum_{j=1}^{t_i}\frac{\exp\{\mu_0+\hat\beta\log(j)+\mu_1(x_{i,j}-x_0)+\hat\theta z_{i,j}\}}{1+\exp\{\mu_0+\hat\beta\log(j)+\mu_1(x_{i,j}-x_0)+\hat\theta z_{i,j}\}}\times\begin{bmatrix}1\\ x_{i,j}-x_0\end{bmatrix}K_g(x_{i,j}-x_0).$$

Set $(\hat\mu_0,\hat\mu_1)$ as the maximizer of $\ell_2^*(\mu_0,\mu_1;x_0)$. The estimate of $m(x_0)$ is given by $\hat m(x_0)=\hat\mu_0$. Here $\ell_2^*(\mu_0,\mu_1;x_0)$ is obtained by replacing β and θ in $\ell_0^*(\omega;x_0)$ with their estimates produced in step 1.

We note that in step 2 we have used a different bandwidth g in the local likelihood method. We allow b and g to be different in the analysis but emphasize that both values will be determined by the sampled data (see our proposal given in section 2.6). We suggest that the final estimates of β, $m(x_0)$, and θ be defined by $\hat\beta$, $\hat m(x_0)$, and $\hat\theta$. Also, at time t_0, the predicted instant bankruptcy

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

probability of a firm with predictor values (x_0, z_0) is suggested to be defined by

$$\hat{h}^*(t_0, x_0, z_0) = \frac{\exp\left\{\hat{\beta}\log(t_0) + \hat{m}(x_0) + \hat{\theta}z_0\right\}}{1 + \exp\left\{\hat{\beta}\log(t_0) + \hat{m}(x_0) + \hat{\theta}z_0\right\}}.$$

2.5. Selecting parametric hazard function using $\hat{m}(x)$

The estimated function $\hat{m}(x)$ of $m(x)$ can be used to determine the functional form of the logit of hazard function. We recall in the usual DHM that a linear logistic function is assumed for the hazard function. That is, the logit transformation of the hazard function is a linear function of the d-dimensional continuous predictors. We note that, for the jth predictor value x_j in x, the relation between the logit-transformed hazard function and x_j can be determined visually by plotting $\{x_j, \hat{m}(x)\}$, for each $j = 1, \ldots, d$. Here x in the plot of $\{x_j, \hat{m}(x)\}$ has the jth component as x_j, but all other components are fixed at their sample median levels, since the distribution of the explanatory variable in the financial field is usually fat-tailed and skewed. Using the plots, a proper functional form of the logit of the hazard function can be determined. For example, if the plot of $\{x_j, \hat{m}(x)\}$, for some j, presents a cubic relation, then the relation between the logit-transformed hazard function and x_j should be an order-three polynomial. In the empirical examples discussed in section 3, we apply this strategy to propose a new parametric hazard function. We denote the parametric hazard function derived from using the plots of $\{x_j, \hat{m}(x)\}$, for $j = 1, \ldots, d$, by $h^{\#}(t, x, z)$. The DHM based on such a data-based parametric hazard function $h^{\#}(t, x, z)$ is denoted DHM$^{\#}$ in the analysis.

2.6. Bankruptcy prediction

Theoretical argument shows $\hat{h}^*(t_0, x_0, z_0)$ to be a consistent estimator of the instant bankruptcy probability. This means that a reliable bankruptcy prediction system can be established based on using estimate $\hat{h}^*(t_0, x_0, z_0)$. In this paper, we suggest that if a firm has predictor values x_0 and z_0 at time t_0 and the calculated probability $\hat{h}^*(t_0, x_0, z_0)$ is no more than a given cut-off value p, then this firm is classified to be in a healthy status. Otherwise, it is classified to be in a bankruptcy status.

To decide a proper cut-off value p, usually one would use all of the panel data to evaluate the performance of the classification scheme. For simplicity of computation, we suggest only using the dataset $\{(Y_{i,t_i}, x_{i,t_i}, z_{i,t_i}), i = 1, \ldots, n\}$, collected at the last observation time of each company in the sampling period. There are two types of 'in-sample' error rates occurring in this evaluation:

$$\text{Type I error rate} \quad \alpha_{\text{in}}(p) = \frac{\left[\sum_{i=1}^{n} Y_{i,t_i} I\left\{\hat{h}^*(t_i, x_{i,t_i}, z_{i,t_i}) \le p\right\}\right]}{\left[\sum_{i=1}^{n} Y_{i,t_i}\right]},$$

and

$$\text{Type II error rate} \quad \beta_{\text{in}}(p) = \frac{\left[\sum_{i=1}^{n} (1 - Y_{i,t_i}) I\left\{\hat{h}^*(t_i, x_{i,t_i}, z_{i,t_i}) > p\right\}\right]}{\left[\sum_{i=1}^{n} (1 - Y_{i,t_i})\right]},$$

where $p \in [0, 1]$ and $I(\cdot)$ stands for the indicator function. Using the cut-off value p, $\alpha_{\text{in}}(p)$ is the rate of misclassifying a bankrupt company as a healthy company, and $\beta_{\text{in}}(p)$ is the rate of misclassifying a healthy company as a bankrupt company.

To keep these two error rates as small as possible, we determine a proper cut-off value p^* for the bankruptcy prediction method based on DSHM such that

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

for each $u \in [0, 1]$. That is to control the in-sample type I error rate $\alpha_{\text{in}}(p)$ to be at most u, so that the sum of the two in-sample error rates is minimal. Controlling the magnitude of $\alpha_{\text{in}}(p)$ 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 investor, we might control the in-sample type II error rate $\beta_{\text{in}}(p)$ instead. In practice, the value of u is determined by the investor. If there is no restriction on the magnitude of $\alpha_{\text{in}}(p)$ and $\beta_{\text{in}}(p)$, then we simply take $u = 1$ (Altman 1968, Ohlson 1980, Begley *et al.* 1996).

Recall that the DSHM also depends on the bandwidths b and g. Thus we need to generalize the previous method for defining p^*. We suggest considering the in-sample type I and II error rates as functions of p, b, and g, denoted, respectively, as $\alpha_{\text{in}}(p, b, g)$ and $\beta_{\text{in}}(p, b, g)$. For each given $u \in [0, 1]$, the proper cut-off value p^* and bandwidths b and g are then determined simultaneously by minimizing

$$\tau_{\text{in}}(p, b, g) = \alpha_{\text{in}}(p, b, g) + \beta_{\text{in}}(p, b, g)$$

with respect to (p, b, g) under the constraints: $p \in [0, 1]$, $b > 0$, $g > 0$, and $\alpha_{\text{in}}(p, b, g) \le u$. Such values for p^*, b, and g are denoted, respectively, as $\hat{p}(u)$, $\hat{b}(u)$, and $\hat{g}(u)$.

2.7. Measuring prediction performance

The performance of the bankruptcy prediction rule based on DSHM is measured by the 'out-of-sample' error rates. To compute these error rates, the out-of-sample data are selected. In contrast, the panel data used to build the bankruptcy prediction rule are considered as the 'in-sample' data. The out-of-sample data are generated similarly to the panel data for building prediction models. The out-of-sample period is from January 2001 to December 2004. The out-of-sample companies include all healthy firms in the panel data and the new firms beginning their listing on the New York Stock Exchange, American Stock Exchange, or NASDAQ during the out-of-sample period. Assume that there are n_0 out-of-sample companies. All predictor values occurring at the last observation time of the n_0 out-of-sample companies in the out-of-sample period were also collected from both

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

COMPUSTAT and CRSP databases. The out-of-sample data are denoted by

$$\{(\tilde{Y}_{k,\tilde{t}_k}, \tilde{x}_{k,\tilde{t}_k}, \tilde{z}_{k,\tilde{t}_k}),\ k = 1, \ldots, n_0\}.$$

Here, for the kth out-of-sample company, $\tilde{t}_k \in \{1, \ldots, \xi + \xi_0\}$ denotes the length of duration, where ξ_0 is a positive integer indicating the length of the out-of-sample period. At the last observation time $\tilde{t}_k$, $\tilde{Y}_{k,\tilde{t}_k} = 1$ indicates that the kth company is bankrupt, and $\tilde{Y}_{k,\tilde{t}_k} = 0$ otherwise. Further, $\tilde{x}_{k,\tilde{t}_k}$ and $\tilde{z}_{k,\tilde{t}_k}$ are values of explanatory variables X and Z collected at time $\tilde{t}_k$, respectively.

Given each value of $u \in [0, 1]$, the out-of-sample error rates for the bankruptcy prediction rule based on DSHM are defined by

Type I error rate

$$\alpha_{\text{out}}(u) = \frac{\left[\sum_{k=1}^{n_0} \tilde{Y}_{k,\tilde{t}_k} I\left\{\hat{h}^*(\tilde{t}_k, \tilde{x}_{k,\tilde{t}_k}, \tilde{z}_{k,\tilde{t}_k}) \le \hat{p}(u)\right\}\right]}{\left[\sum_{k=1}^{n_0} \tilde{Y}_{k,\tilde{t}_k}\right]},$$

Type II error rate

$$\beta_{\text{out}}(u) = \frac{\left[\sum_{k=1}^{n_0} (1 - \tilde{Y}_{k,\tilde{t}_k}) I\left\{\hat{h}^*(\tilde{t}_k, \tilde{x}_{k,\tilde{t}_k}, \tilde{z}_{k,\tilde{t}_k}) > \hat{p}(u)\right\}\right]}{\left[\sum_{k=1}^{n_0} (1 - \tilde{Y}_{k,\tilde{t}_k})\right]},$$

and the total error rate is $\tau_{\text{out}}(u) = \alpha_{\text{out}}(u) + \beta_{\text{out}}(u)$.

Given the out-of-sample data, the out-of-sample error rates can be similarly defined for the bankruptcy prediction rules based on DHM and DHM$^\#$

3. Empirical studies

In this section, empirical studies are conducted to compare the performance of the prediction rules based on DHM, DHM$^\#$ and DSHM.

3.1. *The data*

Four panel datasets were considered for empirical studies. The predictors considered were the accounting variables and market-driven variables suggested by Ohlson (1980) and Shumway (2001). Ohlson (1980) suggested using nine accounting variables:

- WCTA = Working capital divided by total assets,
- TLTA = Total liabilities divided by total assets,
- NITA = Net income divided by total assets,
- CLCA = Current liabilities divided by current assets,
- FUTL = Funds provided by operations divided by total liabilities,
- CHIN = $(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$, where NI_t is net income for the most recent period,
- SIZE = Logarithm of total assets divided by GNP price-level index, where the index assumes a base value of 100 for 1984,
- INTWO = One if net income was negative for the last two years, zero otherwise,
- OENEG = One if total liabilities exceed total assets, zero otherwise.

Shumway (2001) suggested using only two accounting variables, TLTA and NITA, in the model. Besides accounting variables, Shumway (2001) and Chava and Jarrow (2004) further suggested using market-driven variables such as

- RSIZE = Logarithm of each firm's market equity value divided by the total NYSE/AMEX/NASDAQ market equity value,
- EXRET = Monthly return on the firm minus the value-weighted CRSP NYSE/AMEX/NASDAQ index return cumulated to obtain the yearly return,

as well as the variable

LNAGE = Logarithm of firm age

for prediction. Here the firm age is defined as the number of calendar years it has been traded during the sampling period on the New York Stock Exchange, American Stock Exchange, or NASDAQ (Shumway 2001).

Based on these predictors, we studied the performance of the prediction rules using combinations of accounting and market-driven variables. We considered two studies, with and without market-driven variables, for each set of accounting variables. The variable LNAGE was always included in the prediction models, since the models considered in this paper depend on the hazard function (see definitions of DHM and DSHM). Later, we shall report the empirical results of the prediction rules using the four different sets of panel data.

The sampling period of each of the four panel datasets (for building prediction model) was taken from January 1984 to December 2000. The out-of-sample period (for measuring prediction performance) was from January 2001 to December 2004. All firms starting their listing on the New York Stock Exchange, American Stock Exchange, or NASDAQ during both sampling periods are included in the studies, except that the financial institutions were eliminated from the sample due to the unique capital requirements and regulatory structure in that industry group. All panel and out-of-sample datasets were selected from both COMPUSTAT and CRSP databases. Companies that were delisted and declared bankruptcy by CRSP as meeting the delisting codes 400–490, 572, and 574 were considered bankrupt, otherwise healthy.

Note that COMPUSTAT and CRSP databases contain many missing values for the predictors in each study. However, in the analysis we only considered those companies in the dataset with complete predictor values. The problem of missing data is not unusual in applications, especially when there are many predictive variables used in the model. But as long as the missingness occurs at random, the complete-data analysis will not introduce systematic biases (Allison 2001, Little and Rubin 2002). Here we have no reason not to believe that the missingness occurring in the COMPUSTAT and CRSP databases is missing at random.

In each study, the DHM with linear logistic hazard function $h(t, x, z)$ and the DSHM with semiparametric logistic hazard function $h^*(t, x, z)$ were considered. A modified DHM, denoted by DHM$^{\#}$, using the data-based parametric hazard function $h^{\#}(t, x, z)$ suggested from the result of DSHM, is also included in the analysis so that a comparison can be made.

3.2. Computational procedures

In computing the DSHM, the values of the continuous predictors were first divided by their respective sample standard deviations so that all variables have the same scale. This is important, since it can avoid the influence of a predictor with very large range in estimating the optimal values of (p, b, g) and in reading the plots of $\{x_j, \hat{m}(x)\}$, for $j = 1, \ldots, d$.

A grid-search approach was used in computing the optimal values of (p, b, g). First, the values of $\tau_{\text{in}}(p, b, g)$ on an equally spaced logarithmic grid of $1001 \times 51 \times 51$ values of (p, b, g) in $[10^{-5}, 1] \times [0.5, 5] \times [0.5, 5]$ were computed. See Marron and Wand (1992) for a discussion that an equally spaced grid of parameters is typically not a very efficient design for this type of grid search. Given each value of $u \in [0, 1]$, the global minimizer $\{\hat{p}(u), \hat{b}(u), \hat{g}(u)\}$ of $\tau_{\text{in}}(p, b, g)$ on the grid points with restriction $\alpha_{\text{in}}(p, b, g) \leq u$ was taken as the optimal values of (p, b, g). Based on these optimal values, the out-of-sample error rates, functions of u, can then be computed according to our previous definitions.

In addition, using $\{\hat{p}(1), \hat{b}(1), \hat{g}(1)\}$, the plot of $\{x_j, \hat{m}(x)\}$ can be produced for each continuous predictor. We note that, in the plot of $\{x_j, \hat{m}(x)\}$, we have taken the left and the right boundary points of its horizontal axes as the 0.5 and 99.5 percentiles of the values of the jth component of the continuous predictor X, for each j. These plots are used to visually check the adequacy of the order-one polynomial function assumed for each continuous predictor in the linear logistic hazard function of DHM. The empirical results given below show that, sometimes, the order-one polynomial functions should be replaced by order-two or -three polynomials in order to yield better predictive power.

3.3. Results based on using Ohlson's accounting variables with and without market-driven variables included

Given the two datasets with and without the market-driven variables included, table 1 reports the summary statistics and the estimated coefficients of DHM, and figure 1 presents the plot of $\{x_j, \hat{m}(x)\}$ for each continuous predictor. Table 1 shows that the values of the estimated coefficients for variables TLTA, NITA, CLCA, and SIZE in panel A, and those for variables TLTA, FUTL, and SIZE in panel B do not agree with their expected signs. This result indicates that the linear logit of the hazard function of DHM for each of the two datasets might not be suitable. The slope of each curve in figure 1 agrees with

Table 1. Summary statistics of the panel dataset and the estimated coefficients of DHM using Ohlson's accounting variables with and without market-driven variables.

Variable	Mean	Median	Standard deviation	Minimum	Maximum	Estimated coefficient of DHM (p-value)
Panel A: Without market-driven variables 78 bankrupt companies, 2275 healthy companies, and 14,066 firm years						
Intercept						−5.397 (0.001)
WCTA	0.299	0.299	1.733	−202	0.995	−1.272 (0.002)
TLTA	0.471	0.431	1.736	0.001	203	−0.681 (0.091)
NITA	−0.182	0.036	14.501	−1719	1.421	0.031 (0.779)
CLCA	0.650	0.444	3.326	0.002	215.667	−0.001 (0.908)
FUTL	−0.136	0.116	4.285	−38.061	464.448	−0.031 (0.585)
CHIN	0.072	0.097	0.644	−1	1	−0.931 (0.001)
SIZE	−0.170	−0.238	1.224	−7.462	4.742	0.055 (0.586)
INTWO	0.254	0	0.435	0	1	0.774 (0.006)
OENEG	0.028	0	0.164	0	1	1.942 (0.001)
LNAGE	1.283	1.386	0.776	0	2.708	0.171 (0.287)
Panel B: With market-driven variables 77 bankrupt companies, 2192 healthy companies, and 13,400 firm years						
Intercept						−12.557 (0.001)
WCTA	0.304	0.302	1.771	−202	0.987	−1.164 (0.007)
TLTA	0.464	0.430	1.774	0.001	203	−0.554 (0.178)
NITA	−0.168	0.038	14.854	−1719	1.421	−0.134 (0.226)
CLCA	0.607	0.441	2.707	0.002	203	0.018 (0.077)
FUTL	−0.083	0.124	4.327	−38.061	464.448	0.001 (0.975)
CHIN	0.077	0.101	0.643	−1	1	−0.823 (0.001)
SIZE	−0.114	−0.188	1.205	−7.462	4.742	0.443 (0.001)
INTWO	0.239	0	0.426	0	1	0.764 (0.006)
OENEG	0.022	0	0.146	0	1	2.047 (0.001)
RSIZE	−4.793	−4.803	0.755	−8.584	−1.451	−1.428 (0.001)
EXRET	0.900	−0.358	16.761	−10.893	867.761	−0.117 (0.211)
LNAGE	1.275	1.386	0.775	0	2.708	−0.051 (0.820)

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

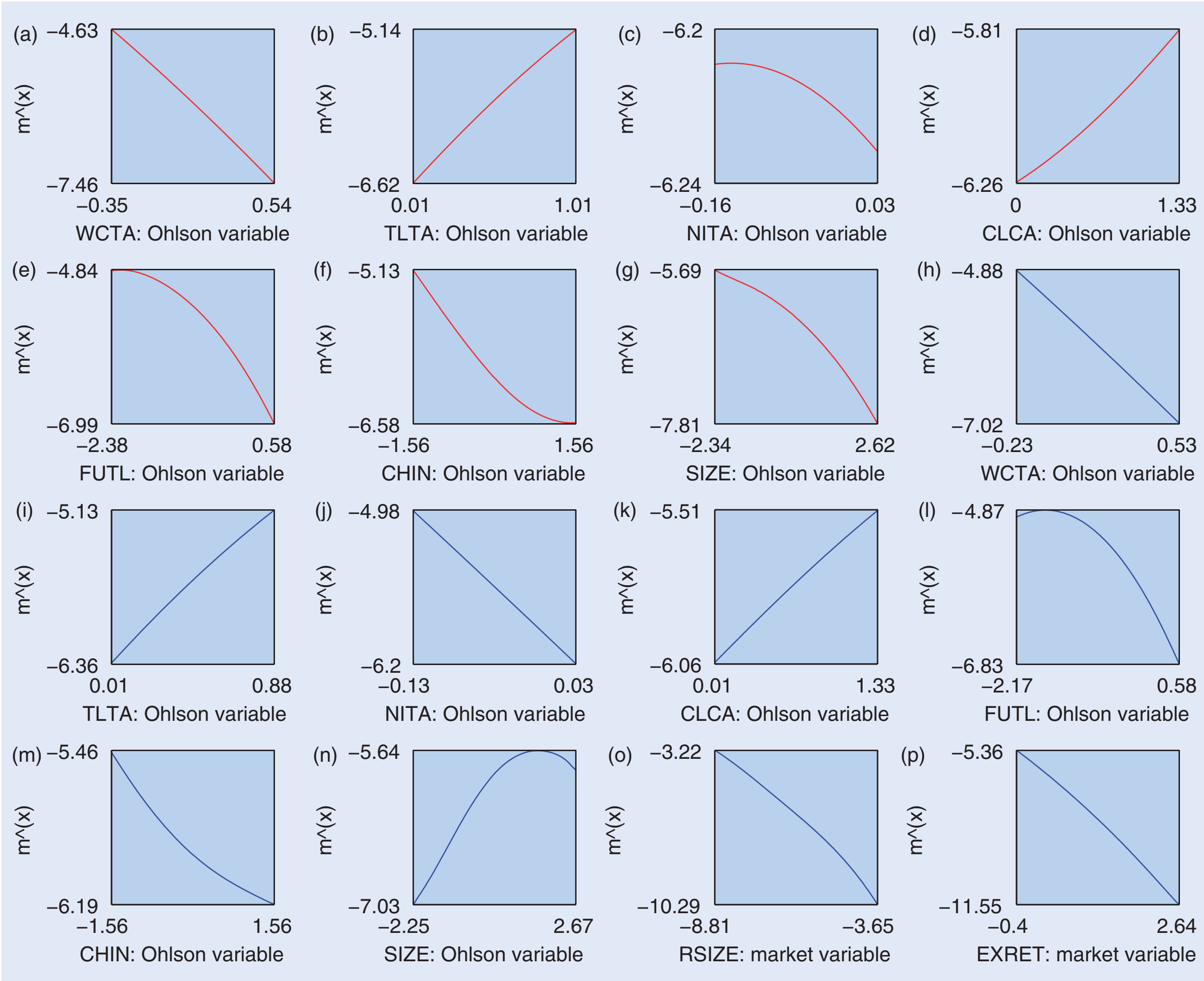


Figure 1. Plots of marginal relations between the logit-transformed hazard function and predictors. Panels (a)–(g) show the plots of $\{x_j, \hat{m}(x)\}$ resulting from DSHM solely using Ohlson's accounting variables. Panels (h)–(p) show the plots of $\{x_j, \hat{m}(x)\}$ resulting from DSHM using Ohlson's accounting variables and market-driven variables. The value of x in the plot of $\{x_j, \hat{m}(x)\}$ has the jth component as x_j, but all other components fixed at their sample median level.

the expected direction of the corresponding variable effect, except variable SIZE in panel (n). Panels (e) and (g) of figure 1 indicate that the order-one polynomials for the two variables FUTL and SIZE in DHM should be replaced by order-two polynomials if only Ohlson's accounting variables were considered in the study. Panels (l) and (n) of figure 1 indicate that the order-two polynomials should be applied on the variables FUTL and SIZE in the DHM if the market-driven variables are also included in the analysis. Figure 2 shows the out-of-sample error rates of the three prediction rules based on DHM, DHM$^{\#}$, and DSHM for the two given datasets. We first see that the prediction models based on using parametric hazard functions are in general conservative in the sense of having smaller type I error rates than the expected upper bound u. In contrast, the type I error rates of the DSHM are close to the designed upper bounds in the cases of $u \leq 0.20$. On the other hand, the type II and the total error rates of the DSHM are much smaller than those of the parametric models when $u \leq 0.20$. One can see that, in the case of solely using Ohlson's accounting variables for analysis, the largest percentage decrease of the total error rate of the DSHM over the DHM is 55%. In the case of including market-driven variables in the analysis, the largest percentage decrease becomes 63%. We also point out that, in the two cases considered in figure 2, the improvement of DHM with the data-based parametric hazard function $h^{\#}(t, x, z)$ over that with the linear logistic hazard function $h(t, x, z)$ is limited when $u \leq 0.20$. This result suggests that the DHM with the data-based hazard function may not always improve the performance of DHM with a simple linear logistic hazard function.

3.4. Results based on using Shumway's accounting variables with and without market-driven variables included

We next report the results of the prediction models DHM, DHM$^{\#}$, and DSHM using Shumway's accounting variables with and without market-driven variables included.

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

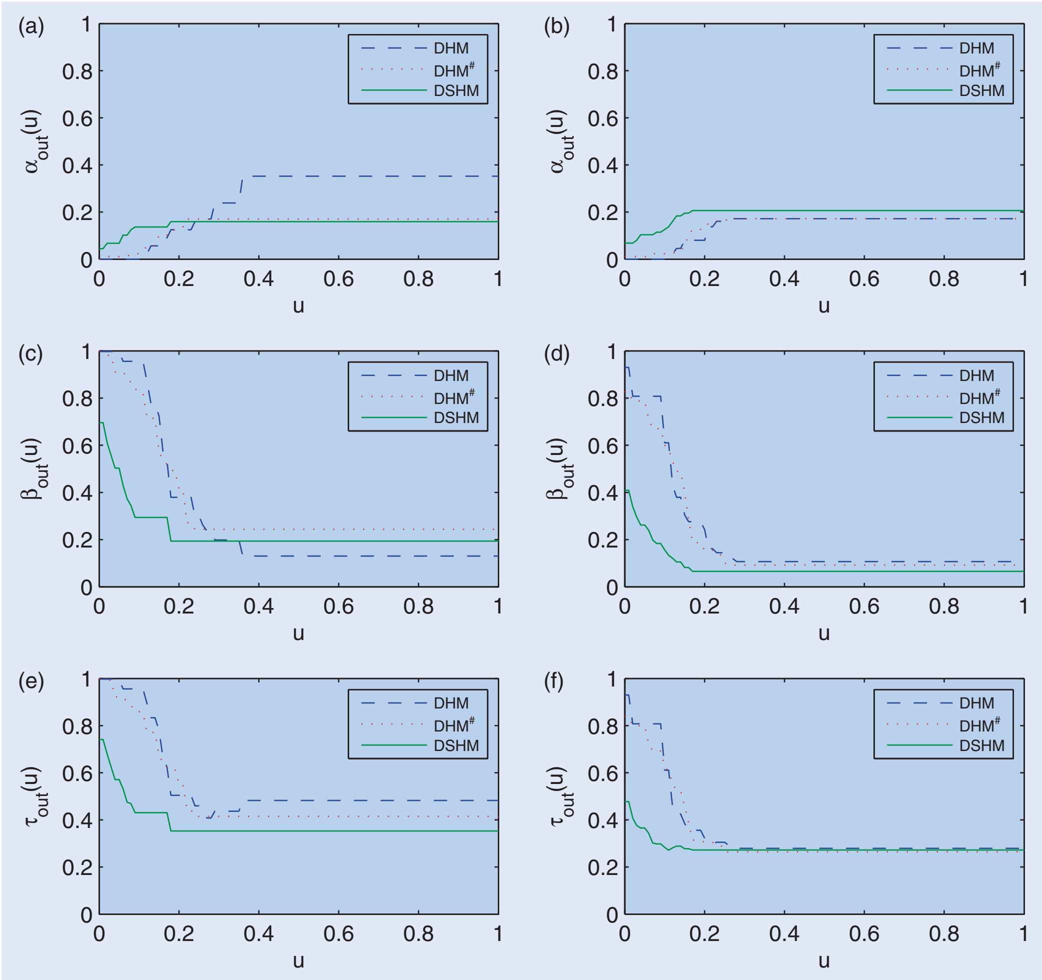


Figure 2. The performance of the three prediction rules based on DHM (dashed curve), DHM$^{\#}$ (dotted curve), and DSHM (solid curve) using Ohlson's accounting variables with and without market-driven variables. Panels (a), (c), and (e) show, respectively, the out-of-sample type I, type II, and total error rates of the prediction methods solely using Ohlson's accounting variables. Panels (b), (d), and (f) show the three out-of-sample error rates using Ohlson's accounting variables and market-driven variables. In each panel, the data-based parametric hazard function $h^{\#}(t, x, z)$ of DHM$^{\#}$ was identical to the linear logistic hazard function $h(t, x, z)$ of DHM except that the order-one polynomials for the two variables FUTL and SIZE were replaced by order-two polynomials.

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

Table 2 reports the summary statistics and the estimated coefficients of DHM. It shows that the values of the estimated coefficients of Shumway's accounting and the two market-driven variables all agree with their expected signs in the study. Figure 3 presents a plot of $\{x_j, \hat{m}(x)\}$ for each continuous predictor, and shows that the slope of the curve in each panel agrees with the expected direction of the variable effect. Panels (a) and (b) of figure 3 show that the order-one polynomials for the two accounting variables in DHM are proper in the study without market-driven variables. However, simply for comparison, we naively used order-two polynomials for variables TLTA and NITA to define DHM$^{\#}$. On the other hand, from panels (c)–(f) of figure 3, we see that the order-one polynomial for the variable RSIZE in DHM should be replaced by an order-three polynomial. Figure 4 shows the out-of-sample error rates of the prediction rules based on DHM, DHM$^{\#}$, and DSHM. Inspecting the results given in figure 4, we see that, in the range $u \leq 0.20$, the type I error rates of DHM and DSHM are basically very similar. However, the total error rate of the DSHM is in general smaller than that of the DHM, for all $u \in [0, 1]$. The largest percentage decrease of the total error rate by the DSHM over the DHM is 17% when the market-driven variables are not included in the analysis, and 21% when the market-driven variables are included. The figure also shows that the improvement of DHM$^{\#}$ over DHM is minimal in the case without including the market-driven variables. This result is reasonable, since the corresponding order-one polynomials for the accounting variables are proper for modeling the hazard function. However, the figure shows that the improvement of DHM$^{\#}$ over DHM is significant when the market-driven variables are included in the model.

4. Concluding remarks

In this paper, a bankruptcy prediction method based on DSHM is proposed. This is an extension of the DHM

Table 2. Summary statistics of the panel dataset and the estimated coefficients of DHM using Shumway's accounting variables with and without market-driven variables.

Variable	Mean	Median	Standard deviation	Minimum	Maximum	Estimated coefficient of DHM (p-value)
Panel A: Without market-driven variables						
92 bankrupt companies, 2368 healthy companies, and 14,846 firm years						
Intercept						−5.712 (0.001)
TLTA	0.478	0.441	1.692	0.001	203	0.959 (0.001)
NITA	−0.174	0.035	14.115	−1719	1.421	−0.096 (0.267)
LNAGE	1.290	1.386	0.777	0	2.708	0.055 (0.696)
Panel B: With market-driven variables						
91 bankrupt companies, 2281 healthy companies, and 14,140 firm years						
Intercept						−12.197 (0.001)
TLTA	0.471	0.440	1.728	0.001	203	1.342 (0.001)
NITA	−0.161	0.037	14.460	−1719	1.421	−0.080 (0.448)
RSIZE	−4.797	−4.808	0.753	−8.584	−1.451	−1.258 (0.001)
EXRET	0.818	−0.368	16.328	−10.893	867.761	−0.186 (0.026)
LNAGE	1.282	1.386	0.777	0	2.708	−0.258 (0.180)

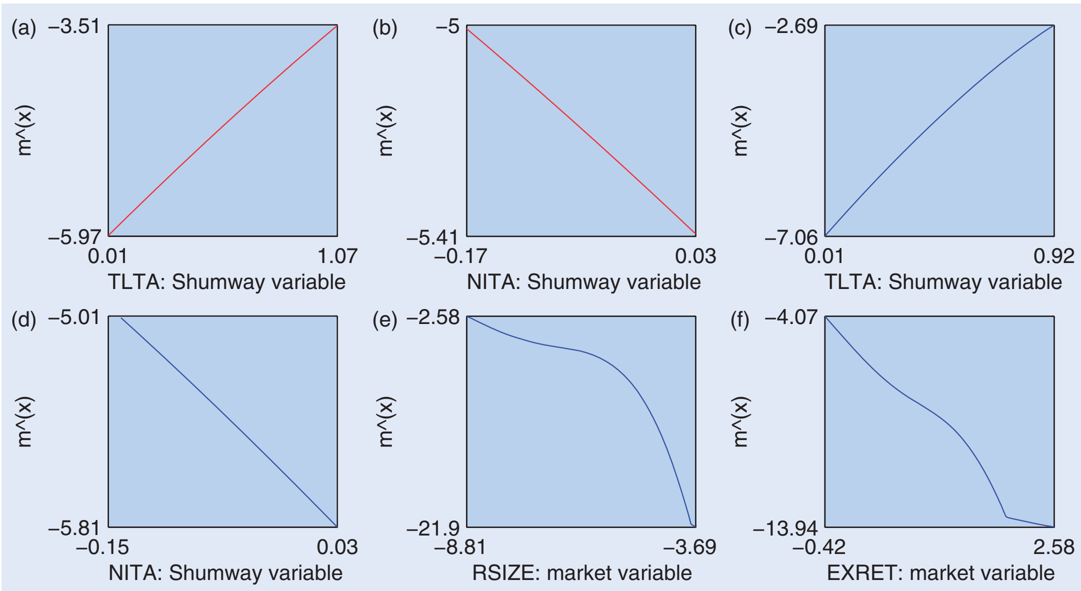


Figure 3. Plots of marginal relations between the logit-transformed hazard function and predictors. Panels (a) and (b) show the plots of $\{x_j, \hat{m}(x)\}$ resulting from DSHM solely using Shumway's accounting variables. Panels (c)–(f) show the plots of $\{x_j, \hat{m}(x)\}$ resulting from DSHM using Shumway's accounting variables and market-driven variables. The value of x in the plot of $\{x_j, \hat{m}(x)\}$ has the jth component as x_j, but all other components fixed at their sample median level.

proposed by Shumway (2001) and Chava and Jarrow (2004). The DHM assumes that the logit transformation of the hazard function is a linear function of the predictors. In contrast, the DSHM only assumes that the transformed function is a smooth function of the continuous predictors. This gives the prediction model more freedom in modeling the underlying hazard function. We point out that the estimates in the DSHM are derived from using the local likelihood method. It can be shown that, under very general conditions, the computed instant bankruptcy probability using DSHM consistently estimates the true instant bankruptcy probability. Thus the DSHM is a reliable prediction rule.

One additional advantage of using DSHM is that by plotting $\{x_j, \hat{m}(x)\}$ for each continuous predictor, one can visually check the adequacy of the parametric DHM. If the parametric model is not proper, the results from DSHM can also guide us on how to make a better selection of parametric model. Sometimes, using parametric modeling is important, particularly when one has too many predictor variables to be considered simultaneously and does not have enough sample data to estimate them non-parametrically.

We have considered four studies to investigate the finite sample performance of the DSHM. The four studies were based on the accounting and market-driven variables proposed by Ohlson (1980) and Shumway (2001). The results of the four studies demonstrate that the DSHM improves the performance of DHM in the prediction of bankruptcy. The DSHM generally has smaller out-of-sample total error rates in all studies. Such an advantage of the DSHM over the DHM in the case of solely using accounting variables is more significant than that in the case of employing both accounting and

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

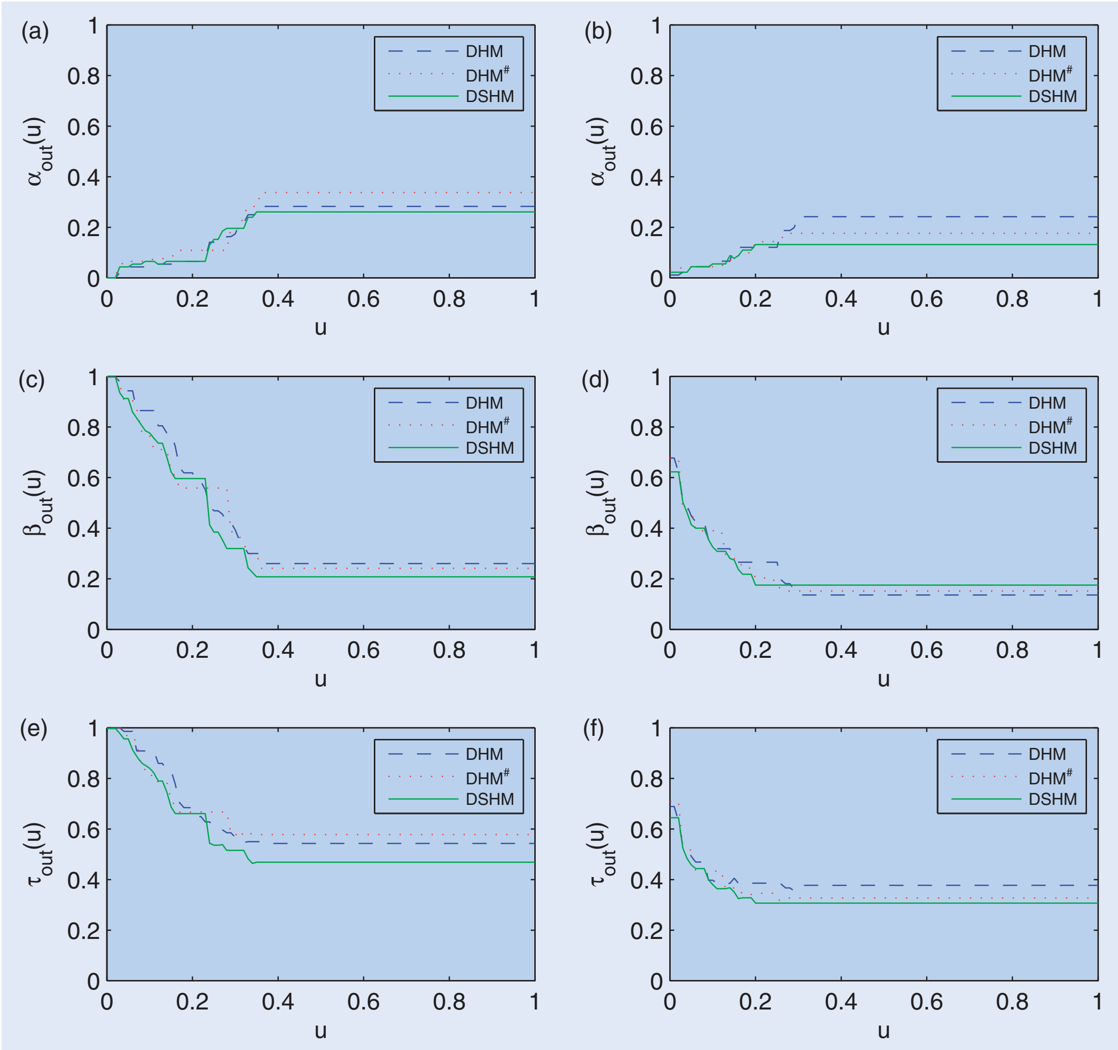


Figure 4. The performance of the three prediction rules based on DHM (dashed curve), DHM$^{\#}$ (dotted curve), and DSHM (solid curve) using Shumway's accounting variables with and without market-driven variables. Panels (a), (c), and (e) show, respectively, the out-of-sample type I, type II, and total error rates of the prediction methods solely using Shumway's accounting variables. The data-based parametric hazard function $h^{\#}(t, x, z)$ of DHM$^{\#}$ in each of panels (a), (c), and (e) used order-two polynomials for variables TLTA and NITA (this model is over-parameterized, since DHM is approximately correct, but included here simply for comparison). Panels (b), (d), and (f) show the three out-of-sample error rates using Shumway's accounting variables and market-driven variables. The data-based parametric hazard function $h^{\#}(t, x, z)$ of DHM$^{\#}$ in each of panels (b), (d), and (f) was identical to the linear logistic hazard function $h(t, x, z)$ in DHM except that the order-one polynomial for the variable RSIZE was replaced by an order-three polynomial.

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

market-driven variables. This result is particularly useful when applying DSHM to those companies not listed in stock exchanges. Shumway (2001) pointed out that, in general, the prediction performance of DHM using both accounting and market-driven variables is better than that solely employing accounting variables. Our empirical results confirm that such an advantage of market-driven variables can also be applied to DSHM.

Note that, in our development of DSHM, we have used a logistic hazard function $h(t, x, z)$ as a basis. We remark that Allison (1982) considered a discrete-time proportional hazard function defined by

$$\eta(t, x, z) = 1 - \exp[-\exp\{\alpha_1 + \beta_1 \log(t) + \gamma_1 x + \theta_1 z\}]$$

in the analysis. The discrete-time proportional hazard function was derived from the well-known proportional hazard function of Cox (1972). Using the same rationale as given in section 2, we can also modify the discrete-time proportional hazard function as the discrete-time semi-parametric proportional hazard function

$$\eta^{*}(t, x, z) = 1 - \exp[-\exp\{\beta \log(t) + m(x) + \theta z\}]$$

for bankruptcy prediction. Our unreported empirical results from the four panel datasets studied in this paper show that the performance of DHM using $\eta(t, x, z)$ is similar to that employing $h(t, x, z)$. The same remark also applies to DSHM when replacing $h^{*}(t, x, z)$ by $\eta^{*}(t, x, z)$.

More investigation of DSHM is necessary. Firstly, in applications, it is not clear how long a sampling period should be used so that a powerful prediction model can be developed. This is important, since if it is long, then there will be many missing data. Secondly, in some practical applications, such as credit rating, we are interested in predicting the rating of a particular company. Thus it is useful to study how to extend the prediction methods to such a situation. Thirdly, in this paper, the performance of DSHM was only studied using firm-specific variables including accounting and market-driven variables. Other important firm-specific variables, such as the KMV-Merton default probability, and industry effects and macroeconomic variables have been considered by Chava and Jarrow (2004), Hillegeist *et al.* (2004), Duffie *et al.* (2007), Bharath and Shumway (2008), and Chava *et al.* (2008). It would be of interest to study the effects of these variables on our semiparametric approach in the future. Further, to account for the heterogeneity, a latent variable method can also be considered; see Duffie *et al.* (2009) and Chava *et al.* (2008). Finally, we remark that the DSHM depends on the logistic hazard function. The robustness of the use of this particular hazard function is still not clear. If it is not robust, then the local quasi-likelihood approach (Fan *et al.* 1995) or the local semilikelihood approach (Claeskens and Aerts 2000, Claeskens and Keilegom 2003) can be considered.

Downloaded By: [National Dong Hwa University] At: 02:28 21 October 2010

Acknowledgements

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

References

Allison, P.D., Discrete-time methods for the analysis of event histories. *Sociol. Methodol.*, 1982, **13**, 61–98.

Allison, P.D., *Missing Data*, 2001 (SAGE Publications: London).

Altman, E.I., Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *J. Finan.*, 1968, **23**, 589–609.

Atiya, A.F., Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Trans. Neural Ntwks*, 2001, **12**, 929–935.

Begley, J., Ming, J. and Watts, S., Bankruptcy classification errors in the 1980s: an empirical analysis of Altman's and Ohlson's models. *Rev. Account. Stud.*, 1996, **1**, 267–284.

Bharath, S.T. and Shumway, T., Forecasting default with the Merton distance to default model. *Rev. Finan. Stud.*, 2008, **21**, 1339–1369.

Chava, S. and Jarrow, R.A., Bankruptcy prediction with industry effects. *Rev. Finan.*, 2004, **8**, 537–569.

Chava, S., Stefanescu, C. and Turnbull, S., Modeling the loss distribution, 2008. Available online at: http://faculty.london.edu/cstefanescu/research.html (accessed 21 April 2008).

Claeskens, G. and Aerts, M., Bootstrapping local polynomial estimators in likelihood-based models. *J. Statist. Plann. Infer.*, 2000, **86**, 63–80.

Claeskens, G. and Keilegom, I.V., Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, 2003, **31**, 1852–1884.

Cox, D.R., Regression models and life-tables (with discussion). *J. R. Statist. Soc., Ser. B*, 1972, **34**, 187–220.

Cox, D.R. and Oakes, D., *Analysis of Survival Data*, 1984 (Chapman and Hall: London).

Duffie, D., Credit risk modeling with affine process. *J. Bank. Finan.*, 2005, **29**, 2751–2802.

Duffie, D., Eckner, A., Horel, G. and Saita, L., Frailty correlated default, 2009. Available online at: http://www.afajof.org/afa/forthcoming/5279.pdf (accessed 12 January 2009).

Duffie, D., Saita, L. and Wang, K., Multi-period corporate default prediction with stochastic covariates. *J. Finan. Econ.*, 2007, **83**, 635–665.

Eubank, R.L., *Spline Smoothing and Nonparametric Regression*, 1988 (Marcel Dekker: New York).

Fan, J. and Gijbels, I., *Local Polynomial Modeling and its Application—Theory and Methodologies*, 1996 (Chapman and Hall: New York).

Fan, J., Heckman, N.E. and Wand, M.P., Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Am. Statist. Assoc.*, 1995, **90**, 141–150.

Härdle, W., *Applied Nonparametric Regression*, 1990 (Cambridge University Press: Cambridge).

Härdle, W., *Smoothing Techniques: with Implementation in S*, 1991 (Springer: Berlin).

Härdle, W., Moro, R.A. and Schäfer, D., Graphical data representation in bankruptcy analysis. In *Handbook of Data Visualization*, edited by C.H. Chen, W. Härdle and A. Unwin, pp. 853–872, 2008 (Springer: Berlin).

Hillegeist, S.A., Keating, E.K., Cram, D.P. and Lundstedt, K.G., Assessing the probability of bankruptcy. *Rev. Account. Stud.*, 2004, **9**, 5–34.

Hwang, R.C., Cheng, K.F. and Lee, J.C., A semiparametric method for predicting bankruptcy. *J. Forecast.*, 2007, **26**, 317–342.

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

Marron, J.S. and Wand, M.P., Exact mean integrated square error. *Ann Statist.*, 1992, **20**, 712–736.

Merton, R.C., On the pricing of corporate debt: the risk structure of interest rates. *J. Finan.*, 1974, **29**, 449–470.

Müller, H.G., *Nonparametric Regression Analysis of Longitudinal Data*, 1988 (Springer: Berlin).

Ohlson, J., Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.*, 1980, **18**, 109–131.

Scott, D.W., *Multivariate Density Estimation: Theory, Practice, and Visualization*, 1992 (Wiley: New York).

Shumway, T., Forecasting bankruptcy more accurately: a simple hazard model. *J. Bus.*, 2001, **74**, 101–124.

Simonoff, J.S., *Smoothing Methods in Statistics*, 1996 (Springer: New York).

Staniswallis, J.G., The kernel estimate of a regression function in likelihood-based models. *J. Am Statist. Assoc.*, 1989, **84**, 276–283.

Sun, L. and Shenoy, P.P., Using Bayesian networks for bankruptcy prediction: some methodological issues. *Eur. J. Oper. Res.*, 2007, **180**, 738–753.

Tibshirani, R. and Hastie, T., Local likelihood estimation. *J. Am. Statist. Assoc.*, 1987, **82**, 559–568.

Vassalou, M. and Xing, Y., Default risk in equity returns. *J. Finan.*, 2004, **59**, 831–868.

Wand, M.P. and Jones, M.C., *Kernel Smoothing*, 1995 (Chapman and Hall: London).

Zmijewski, M.E., Methodological issues related to the estimation of financial distress prediction models. *J. Account. Res.*, 1984, **22**, 59–82.