Journal of Empirical Finance 17 (2010) 120–137

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Predicting issuer credit ratings using a semiparametric method ☆

Ruey-Ching Hwang [a,*], Huimin Chung [b], C.K. Chu [c]

[a] Department of Finance, National Dong Hwa University, Hualien, Taiwan
[b] Graduate Institute of Finance, National Chiao Tung University, Hsinchu, Taiwan
[c] Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan

ARTICLE INFO

Article history:
Received 22 October 2008
Received in revised form 29 June 2009
Accepted 27 July 2009
Available online 8 August 2009

JEL classification:
C14
G20

Keywords:
Industry effect
Issuer credit rating
Market-driven variable
Ordered linear probit model
Ordered semiparametric probit model

ABSTRACT

This paper proposes a prediction method based on an ordered semiparametric probit model for credit risk forecast. The proposed prediction model is constructed by replacing the linear regression function in the usual ordered probit model with a semiparametric function, thus it allows for more flexible choice of regression function. The unknown parameters in the proposed prediction model are estimated by maximizing a local (weighted) log-likelihood function, and the resulting estimators are analyzed through their asymptotic biases and variances. A real data example for predicting issuer credit ratings is used to illustrate the proposed prediction method. The empirical result confirms that the new model compares favorably with the usual ordered probit model.

1. Introduction

Credit ratings play an important role in capital markets. Under the New Basel Capital Accord (Basel II), credit ratings will play an even more central role than they have so far. There are two basic types of credit ratings, the bond rating and the issuer credit rating. While the former measures the likelihood of the default or delayed payment of a bond issue, the latter is an overall assessment of the creditworthiness of a company. Currently, there are many widely recognized credit rating agencies, such as Moody's Investors Service and Standard and Poor's Ratings Services (S&P's), etc. They routinely provide credit ratings for bonds and companies.

This study focuses on the S&P's long-term issuer credit rating (LTR). According to the definition given by S&P's, the LTR focuses on the obligor's capacity and willingness to meet its long-term financial commitments. Based on the Compustat North America (COMPUSTAT) database, in year 2007, there were 8010 companies listed on the New York Stock Exchange, American Stock Exchange, or NASDAQ. However, among those 8010 companies, there were only 18.96% (1519) companies having S&P's LTRs. This result indicates that most of companies listed on those stock exchanges do not have S&P's LTRs, which makes their rating predictions quite valuable to practitioners and regulators. Accordingly, the purpose of this paper is to forecast ratings for those companies "without" S&P's LTRs. For two reasons, we do not pursue the issue of rating forecast for companies "with" S&P's LTRs. First, if a company is once rated by S&P's, then it will be rated again unless a special event, for example bankruptcy, happens to the company. Second, the continuously rated companies have relatively unchanged rating categories in general (Galil, 2003; Pettit et al., 2004). Thus it seems less interesting in predicting the ratings of these companies.

☆ The authors thank the reviewer and the editor for their valuable comments and suggestions which greatly improve the presentation of this paper. The research was supported by National Science Council, Taiwan, Republic of China.

* Corresponding author. Tel.: +886 38633139; fax: +886 38633510.
E-mail address: rchwang@mail.ndhu.edu.tw (R.-C. Hwang).

0927-5398/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2009.07.007

There are several well-known statistical techniques for constructing credit rating predictions. These techniques include multiple regression analysis (Horrigan, 1966; Pogue and Soldofsky, 19692; West, 1970), multiple discriminant analysis (Pinches and Mingo, 1973, 1975; Altman and Katz, 1976), ordered linear probit model (OLPM; Kaplan and Urwitz, 1979; Ederington, 1985; Gentry et al., 1988; Hwang et al., 2008), and ordered and unordered linear logit models (Ederington, 1985), etc. Altman et al. (1981) provides a detailed introduction of statistical classification models. The common principal of these approaches is that they are developed using single-period data and parametric models. Credit rating forecasting models based on multiple-period data with independent assumption include, for example, Blume et al. (1998), Poon (2003), and Güttler and Wahrenburg (2007) employing the idea of OLPM. Other approaches based on machine learning techniques, for example, Bayesian networks (Wijayatunga et al., 2006) and support vector machines and neural networks (Huang et al., 2004) were also considered in the literature for credit rating prediction. Basically, the latter approaches based on machine learning techniques are more complicated in computation and interpretation than those based on parametric models.

To forecast S&P's LTRs, the prediction methods based on OLPM and an extension of OLPM will be used in this paper.[1] The OLPM is simply constructed by imposing a linear regression relationship between S&P's LTRs and predictor variables. Its important parameters are determined by maximizing a log-likelihood function. However, if the underlying regression function is not linear, then the advantages of OLPM in explaining and predicting will not be realized. To avoid this potential pitfall, we show in this paper that the idea of semiparametric logit model in Hwang et al. (2007) can be directly extended to OLPM. Specifically, we shall propose an ordered semiparametric probit model (OSPM) for credit rating prediction by replacing the linear regression function in OLPM with a semiparametric function. The proposed OSPM is built on the works of OLPM but needs not assuming any parametric form for the regression function. Thus it is much more flexible in modeling the regression function. Furthermore, the proposed method is developed under the concept of local likelihood, and it turns out that the important parameters in OSPM can be estimated by maximizing a local (weighted) log-likelihood function. Thus the required computation for OSPM is as simple as that for OLPM.

To apply OLPM and OSPM to predict S&P's LTRs, the twenty-four potential predictors in Hwang et al. (2008) for studying important predictors of S&P's LTRs in year 2005 were considered in our data analysis section. These variables include four market-driven variables (Shumway, 2001; Bharath and Shumway, 2008), nineteen accounting variables (Altman, 1968; Poon, 2003; Pettit et al., 2004), and industry effects (Chava and Jarrow, 2004; Pettit et al., 2004). The studied data were collected from COMPUSTAT and Center for Research in Security Prices (CRSP) databases. Our sample consisted of 779 companies receiving S&P's LTRs in April 2007 and having complete values of the twenty-four potential predictors. The sample was further divided into the estimation sample and holdout sample based on the longevity of S&P's LTR (Hwang et al., 2008).[2] According to S&P's Research Insight North America Data Guide (2004, p. 54), S&P's began to use the term LTR on September 1, 1998. Companies receiving S&P's LTRs in consecutive nine years (April 1999–April 2007) were classified into the estimation sample. The rest of the sampled companies were classified into the holdout sample. Based on the division principle, 413 companies were divided into the estimation sample and 366 companies into the holdout sample.

To examine whether our estimation and holdout samples induced selection bias, a procedure based on OLPM with sample selection was performed using LIMDEP 8.0 to test the null hypothesis of no selection bias caused by the above sample division principle. The result of the test shows no rejection of the null hypothesis of interest at 5% level of significance. Before performing the selection bias test, a forward selection procedure based on minimizing classification error rate on the estimation sample (Härdle et al., 2008) was used to objectively determine effective predictors for OLPM. The final list of the selected predictors includes industry effects, two market-driven variables assessing a firm's market capitalization and risk, and two accounting variables measuring a firm's financial leverage and profitability. The values of estimated coefficients of the selected market-driven and accounting variables all agree with their expected signs. This indicates that market-driven variables and industry effects are also important to determine S&P's LTRs. Our variable selection result coincides with that obtained by Hwang (2008) for predicting ratings in year 2005. On the other hand, to study the difference between unsolicited and solicited ratings, Poon (2003) suggested profitability and sovereign credit risk as two major factors in determining S&P's LTRs. Furthermore, to assess biases in credit ratings assigned by Moody's and S&P's for near-to-default issuers, Güttler and Wahrenburg (2007) used accounting and macroeconomic variables as major determinants of issuer credit ratings.

In Section 3, we describe one real data set and provide some summary statistics. The summary statistics show that the predictors under consideration have reasonable power in discriminating the creditworthiness of companies. The real data set was analyzed using methods based on OLPM and OSPM. The prediction performance of each method was measured by the total error rate obtained from the holdout sample. By the error rates summarized in Section 3, we conclude that the prediction method based on OSPM has better performance. The empirical results in Section 3 also demonstrate that the functional form between the S&P's LTR assessment and the selected continuous predictors is nonlinear. This indicates that OLPM may not be adequate in explaining and predicting S&P's LTRs. Thus, by the empirical results, OSPM has potential to be a powerful credit rating forecasting model.

[1] Due to the superiority in explaining and predicting, OLPM has been adopted for multiple-class prediction by a number of studies such as Kaplan and Urwitz (1979) and Gentry et al. (1988), etc. Also, the test procedure of sample selection bias is only available for OLPM (Greene, 2002). On the other hand, it is not suggested using discrete explanatory variables in multiple discriminant analysis (Johnson and Wichern, 2002, p. 641). In this paper, industry effects on S&P's LTRs were estimated through coefficients of six industry indicator variables. Given these industry indicator variables, it is not adequate to use multiple discriminant analysis to predict S&P's LTRs.

[2] Given the pool of companies with S&P's LTRs in April 2007, by comparing the longevity of S&P's LTR, our estimation companies solely correspond to the rated ones, and our holdout companies the newcomers. Their purified composition agrees with our purpose to forecast ratings for companies without S&P's LTRs. On the other hand, one may separate the sampled companies by random allocation. Random allocation has the advantage of eliminating the need to test for selection bias since the resulting estimation and holdout samples have the same composition structure. However, each of the latter samples contains both rated companies and newcomers. Such mixed composition does not agree with our prediction purpose.

The remainder of this paper is organized as follows. In Section 2, our methodology for forecasting credit ratings based on OSPM is developed. Section 3 presents empirical results. Concluding remarks and future research topics are contained in Section 4. Our theoretical results are described in Appendix A. Finally, sketches of the proofs are given in Appendix B.

2. Methodology

In this section, we first describe the basic idea of OSPM. Then we develop the methodology for estimating unknown quantities of OSPM.

2.1. OSPM

The OSPM is defined by imposing a semiparametric regression relationship between S&P's LTRs and predictor variables. It is developed using the estimation sample composed of observations (Y_i, x_i, z_i), for $i=1, \cdots, n$. The value of $Y_i=j$ indicates that the S&P's LTR of the ith company belongs to category j, where $j\in\{1, \cdots, m\}$, $m\geq 2$, and m stands for the total number of categories among S&P's LTRs. The larger the value of Y_i, the better the S&P's LTR category of the ith company. The values of x_i and z_i are collected on the ith company from the $d\times 1$ continuous and $q\times 1$ discrete explanatory variables X and Z, respectively. Our aim is to predict the S&P's LTR category for a given company without S&P's LTR.

Given the estimation sample, the OSPM is defined by

$$\begin{cases} Y_i^* = H(x_i) + \theta z_i + \varepsilon_i, \\ Y_i = j, \text{ if } \tau_{j-1} < Y_i^* \leq \tau_j, \text{ for } j = 1, \cdots, m. \end{cases} \tag{1}$$

Here Y_i^* are latent variables relating to S&P's LTR assessment, $H(x)$ an unknown but smooth function of the value x of the d-dimensional continuous explanatory variable X, θ a $1\times q$ vector of parameters, and ε_i independent standard normal random variables. Also $\tau_0, \cdots, \tau_m$ are threshold parameters discretizing the real line into m intervals, where the values of τ_j are of ascending order and satisfy the conditions $\tau_0=-\infty$, $\tau_1=0$, and $\tau_m=\infty$. Set $\tau=(\tau_2, \cdots, \tau_{m-1})$.

Under model (1), the log-likelihood function of the estimation sample can be expressed by

$$\ell(H, \theta, \tau) = \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_i = j) \ ln(\Phi_{i,j} - \Phi_{i,j-1}), \tag{2}$$

where $\Phi_{i,j}=\Phi\{\tau_j-H(x_i)-\theta z_i\}$, $\Phi(\bullet)$ is the cumulative distribution function of standard normal random variable, and $I(\bullet)$ stands for the indicator function. For a given company with predictor values (x_0, z_0), if $H(x_0)$, θ, and τ can be efficiently estimated by $\hat{H}(x_0)$, $\hat{\theta}$, and $\hat{\tau}$, then, by model (1), its S&P's LTR category can be predicted by

$$\hat{Y}_{\mathrm{OSPM}}(x_0, z_0) = j, \text{ if } \hat{\tau}_{j-1} < \hat{H}(x_0) + \hat{\theta} z_0 \leq \hat{\tau}_j, \text{ for some } j\in\{1, \cdots, m\}. \tag{3}$$

It is of importance to point out that the prediction rule $\hat{Y}_{\mathrm{OSPM}}(x_0, z_0)$ in Eq. (3) is equivalent to basing on cutoff value 1/2:

$$\hat{Y}_{\mathrm{OSPM}}(x_0, z_0) = j, \text{ if } \hat{\Phi}_{0,j-1} < 1/2 \leq \hat{\Phi}_{0,j}, \text{ for some } j\in\{1, \cdots, m\}, \tag{4}$$

where $\hat{\Phi}_{0,0} = 0, \hat{\Phi}_{0,j} = \Phi\{\hat{\tau}_j - \hat{H}(x_0) - \hat{\theta} z_0\}$, for $j=1, \cdots, m-1$, and $\hat{\Phi}_{0,m} = 1$. For improving the performance of the prediction rule based on OSPM, we followed the idea of Altman (1968), Ohlson (1980), Begley et al. (1996), and Hwang et al. (2008), and suggested replacing cutoff value 1/2 in Eq. (4) with some cutoff value $p\in[0,1]$. The resulting prediction rule is denoted by

$$\hat{Y}_{\mathrm{OSPM}}(x_0, z_0) = j, \text{ if } \hat{\Phi}_{0,j-1} < p \leq \hat{\Phi}_{0,j}, \text{ for some } j\in\{1, \cdots, m\}. \tag{5}$$

In order to decide a proper data-based cutoff value for p in Eq. (5), usually one would use the estimation sample to evaluate the performance of the classification scheme. In doing so, there are three types of "in-sample" error rates occurred on the estimation sample:

$$\text{type I error rate } \alpha_{\mathrm{in}}(p) = n^{-1} \sum_{i=1}^{n} I\{\hat{Y}_{\mathrm{OSPM}}(x_i, z_i) > Y_i\},$$
$$\text{type II error rate } \beta_{\mathrm{in}}(p) = n^{-1} \sum_{i=1}^{n} I\{\hat{Y}_{\mathrm{OSPM}}(x_i, z_i) < Y_i\},$$
$$\text{total error rate } \gamma_{\mathrm{in}}(p) = \alpha_{\mathrm{in}}(p) + \beta_{\mathrm{in}}(p).$$

Using the estimation sample, $\alpha_{\mathrm{in}}(p)$ is the rate of misclassifying a company to a higher rating category, and $\beta_{\mathrm{in}}(p)$ the rate of misclassifying a company to a lower rating category. If the type I error would cause severe losses to investors, then it is important to control the magnitude of $\alpha_{\text{in}}(p)$. Thus, a proper data-based cutoff value $\hat{p}_{\text{OSPM}}(u)$ for p may be determined such that

$$\gamma_{\text{in}}\{\hat{p}_{\text{OSPM}}(u)\} = \min_{\alpha_{\text{in}}(p)\le u, p\in[0,1]} \gamma_{\text{in}}(p), \tag{6}$$

for each $u\in[0, 1]$. This approach is to control the magnitude of in-sample type I error rate $\alpha_{\text{in}}(p)$ to be at most u, so that the in-sample total error rate $\gamma_{\text{in}}(p)$ is minimal. On the other hand, if the type II error would cause severe losses to investors, then we may control the magnitude of $\beta_{\text{in}}(p)$ instead. In practice, the value of $u\in[0,1]$ is determined by investors. If there is no restriction on the magnitude of $\alpha_{\text{in}}(p)$ and $\beta_{\text{in}}(p)$, then we simply take $u=1$.

Using the prediction rule based on OSPM with our data-based cutoff value $\hat{p}_{\text{OSPM}}(u)$, the S&P's LTR category of the given company with predictor values (x_0, z_0) is predicted by

$$\hat{Y}_{\text{OSPM}}(x_0, z_0) = j, \text{ if } \hat{\Phi}_{0,j-1} < \hat{p}_{\text{OSPM}}(u) \le \hat{\Phi}_{0,j}, \text{ for some } j\in\{1,\cdots,m\}, \tag{7}$$

for each $u\in[0,1]$. A local likelihood method will be employed in Sections 2.2 and 2.3 to estimate unknown parameters $H(x_0)$, θ, and τ of OSPM.

Note that OLPM is a well-known special case of OSPM. It can be constructed by replacing the semiparametric regression function $H(x)+\theta z$ in OSPM with a linear function of x and z. The same developments (Eqs. (2)–(7)) can be applied for OLPM. Let the data-based cutoff value $\hat{p}_{\text{OLPM}}(u)$ and the prediction rule $\hat{Y}_{\text{OLPM}}(x_0, z_0)$ based on OLPM be similarly defined. The detailed introduction of OLPM can be referred to Kaplan and Urwitz (1979) and Borooah (2002). Both prediction rules based on OSPM and OLPM with their data-based cutoff values will be used in Section 3 to analyze a real data example.

2.2. A local likelihood method for estimating parameters in OSPM

We now apply the local likelihood concept (Tibshirani and Hastie, 1987; Staniswallis, 1989; Fan et al., 1995; Hwang et al., 2007) to develop a procedure for estimating $H(x_0)$, θ, and τ in OSPM, where $x_0=(x_{0,1},\cdots,x_{0,d})^T$ is any given value of the d-dimensional continuous explanatory variable X. The basic idea of the local likelihood method is to center the data around x_0, and weight the log-likelihood such that it places more emphasis on those observations nearest to x_0. It can be simply performed by first considering a neighborhood $S_b(x_0) = \{u = (u_1,\cdots,u_d)^T : |u_k - x_{0,k}| \le b, \text{ for each } k = 1,\cdots,d\}$ of x_0. Here b is some positive constant to be determined later by the estimation sample, and called the bandwidth. If the value of b is small enough and x_i belongs to $S_b(x_0)$, then Taylor's first order expansion says that $H(x_i)\approx H(x_0)+H^{(1)}(x_0)^T(x_i-x_0)$. This result means that such $H(x_i)$ in $\ell(H,\theta,\tau)$ can be approximated by $\eta_0+\eta_1(x_i-x_0)$, where η_0 is a scalar parameter representing $H(x_0)$ and η_1 a $1\times d$ vector of parameters standing for $H^{(1)}(x_0)^T$. Set $\eta=(\eta_0, \eta_1)$.

Based on the above discussion, a local likelihood method for making inference about (η,θ,τ) can be proposed by modifying $\ell(H,\theta,\tau)$ to consider the following "local (weighted)" log-likelihood function

$$\ell_0(\eta,\theta,\tau;x_0) = \sum_{i=1}^{n}\sum_{j=1}^{m} I(Y_i=j) ln(\Phi^{\circ}_{i,j} - \Phi^{\circ}_{i,j-1}) W(x_i),$$

where $\Phi^{\circ}_{i,j}=\Phi\{\tau_j-\eta_0-\eta_1(x_i-x_0)-\theta z_i\}$. The simplest weight $W(x_i)$ assigned to the observation (Y_i, x_i, z_i) in $\ell_0(\eta,\theta,\tau; x_0)$ is the indicator value $I\{x_i\in S_b(x_0)\}$. In this case, it can be mathematically defined as $W(x_i)=K_b(x_i-x_0)=b^{-d}\prod_{k=1}^{d} K\{(x_{i,k}-x_{0,k})/b\}$, where K is the uniform probability density function over $[-1,1]$, also called the kernel function, and $x_i=(x_{i,1},\cdots, x_{i,d})^T$.

Conceptually, a more general weighting scheme can be used for defining weights $W(x_i)$ when constructing the local log-likelihood function. This can be achieved by taking K as a symmetric and unimodal probability density function supported on the interval $[-1,1]$. The results in the literature showed that the choice of bandwidth b plays important role, but the choice of kernel function K is not very important in the weighting scheme. Some discussions of the kernel weighting method can be found in the monographs by Eubank (1988), Müller (1988), Härdle (1990, 1991), Scott (1992), Wand and Jones (1995), Fan and Gijbels (1996), and Simonoff (1996). In this paper, we select $W(x_i)=K_b(x_i-x_0)$ with a general K in all analyses.

Set $(\tilde{\eta},\tilde{\theta},\tilde{\tau})$ as the maximizer of $\ell_0(\eta,\theta,\tau;x_0)$, where $\tilde{\eta} = (\tilde{\eta}_0,\tilde{\eta}_1)$. We define $\tilde{H}(x_0) = \tilde{\eta}_0$ to indicate that it is an estimate of $H(x_0)$. We also point out that θ and τ are global parameters and their corresponding estimates $\tilde{\theta}$ and $\tilde{\tau}$ produced from $\ell_0(\eta,\theta,\tau;x_0)$ may not be efficient, since such estimates are derived by maximizing a local log-likelihood function depending on x_0. In Section 2.3, more efficient estimates of $H(x_0)$, θ, and τ can be achieved.

2.3. More efficient estimates of parameters in OSPM

More efficient estimates of $H(x_0)$, θ, and τ can be derived using the following two-step procedure. We first note that, for each value x_i, an initial estimate $\tilde{H}(x_i)$ of $H(x_i)$ can be obtained by the method outlined in Section 2.2. The two-step procedure includes:

Step 1 θ and τ are estimated by maximizing the pseudo log-likelihood function

$$\ell_1(\theta,\tau) = \sum_{i=1}^{n}\sum_{j=1}^{m} I(Y_i=j) ln(\tilde{\Phi}_{i,j} - \tilde{\Phi}_{i,j-1}),$$

where $\tilde{\Phi}_{i,j} = \Phi\{\tau_j - \tilde{H}(x_i) - \theta z_i\}$. Here $\ell_1(\theta,\tau)$ is obtained by replacing each $H(x_i)$ in $\ell(H,\theta,\tau)$ with its initial estimate $\tilde{H}(x_i)$. Set $(\hat{\theta},\hat{\tau})$ as the maximizer of $\ell_1(\theta,\tau)$. The estimates of (θ,τ) are taken as $(\hat{\theta},\hat{\tau})$.

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

$$\ell_2(\eta; x_0) = \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_i = j) ln(\hat{\Phi}^{\circ}_{i,j} - \hat{\Phi}^{\circ}_{i,j-1}) K_g(x_i - x_0),$$

where $\hat{\Phi}^{\circ}_{i,j} = \Phi\{\hat{\tau}_j - \eta_0 - \eta_1(x_i - x_0) - \hat{\theta} z_i\}$. Here $\ell_2(\eta;\, x_0)$ is obtained by replacing θ and τ in $\ell_0(\eta,\, \theta,\, \tau;\, x_0)$ with their estimates $\hat{\theta}$ and $\hat{\tau}$ produced in Step 1. Set $\hat{\eta} = (\hat{\eta}_0, \hat{\eta}_1)$ as the maximizer of $\ell_2(\eta;\, x_0)$. The estimate of $H(x_0)$ is given by $\hat{H}(x_0) = \hat{\eta}_0$.

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 estimation sample (see our proposal given in Section 2.4). We suggest that the final estimates of $H(x_0)$, θ, and τ be defined by $\hat{H}(x_0)$, $\hat{\theta}$, and $\hat{\tau}$, respectively. Their theoretical properties will be given in Appendix A.

2.4. Choosing the kernel function K and the values of (b, g, p)

Our Theorem 1 in Appendix A shows that $\hat{H}(x_0)$, $\hat{\theta}$, and $\hat{\tau}$ are consistent estimators of $H(x_0)$, θ, and τ, respectively. This result means that the prediction rule $\hat{Y}_{\mathrm{OSPM}}(x_0, z_0)$ in Eq. (7) is reliable. To compute its value, we first need to choose the kernel function K and the values of (b, g, p). Remark 1 in Appendix A points out that, in the sense of yielding smaller asymptotic mean integrated square error of $\hat{H}(x_0)$, the optimal K is the Epanechnikov kernel defined as $K(t) = (3/4)(1 - t^2)I(|t| \le 1)$, and the optimal value of g is of larger order than that of b. Due to the compact support, the Epanechnikov kernel also has the advantage of computational convenience (Härdle, 1991). Thus our kernel function K in applications is taken as the Epanechnikov kernel. But the optimal values of (b, g) are not available in practice for depending on the unknown factors $H(x)$, θ, and τ.

To choose the values of (b, g, p), we suggest considering the in-sample type I, type II, and total error rates of the classification scheme based on OSPM in Eq. (7) as functions of (b, g, p), denoted as $\alpha_{\mathrm{in}}(b, g, p)$, $\beta_{\mathrm{in}}(b, g, p)$, and $\gamma_{\mathrm{in}}(b, g, p)$, respectively. The definition of in-sample error rates has been given in Section 2.1. The proper values for (b, g, p) are then simultaneously determined so that $\gamma_{\mathrm{in}}(b, g, p)$ is minimal, subject to the constraints $p \in [0,1]$, $0 < b < g$, and $\alpha_{\mathrm{in}}(b, g, p) \le u$, for each $u \in [0,1]$. Such selected values for (b, g, p) are denoted as $\{\hat{b}(u), \hat{g}(u), \hat{p}_{\mathrm{OSPM}}(u)\}$.

2.5. Measuring prediction performance

The performance of the prediction rule based on OSPM in Eq. (7) is measured by the "out-of-sample" error rates. These error rates are evaluated on the holdout sample. Suppose that the holdout sample is composed of observations $(\tilde{Y}_i, \tilde{x}_i, \tilde{z}_i)$, for $i = 1, \cdots, n_0$. Using the estimation sample, the Epanechnikov kernel, and the selected values $\{\hat{b}(u), \hat{g}(u), \hat{p}_{\mathrm{OSPM}}(u)\}$, the value of $\hat{H}(\tilde{x}_i)$ for each data point $(\tilde{x}_i, \tilde{z}_i)$ in the holdout sample and those of $\hat{\theta}$ and $\hat{\tau}$ can be computed. The S&P's LTR category for a company with predictor values $(\tilde{x}_i, \tilde{z}_i)$ in the holdout sample is predicted by $\hat{Y}_{\mathrm{OSPM}}(\tilde{x}_i, \tilde{z}_i)$ as defined in Eq. (7). After the evaluation procedure is completed for each company in the holdout sample, the out-of-sample error rates for the prediction rule based on OSPM are defined by

type I error rate $\alpha_{\mathrm{out}}(u) = n_0^{-1} \sum_{i=1}^{n_0} I\{\hat{Y}_{\mathrm{OSPM}}(\tilde{x}_i, \tilde{z}_i) > \tilde{Y}_i\}$,

type II error rate $\beta_{\mathrm{out}}(u) = n_0^{-1} \sum_{i=1}^{n_0} I\{\hat{Y}_{\mathrm{OSPM}}(\tilde{x}_i, \tilde{z}_i) < \tilde{Y}_i\}$,

total error rate $\gamma_{\mathrm{out}}(u) = \alpha_{\mathrm{out}}(u) + \beta_{\mathrm{out}}(u)$,

for each $u \in [0, 1]$. Given the holdout sample, the out-of-sample error rates can be similarly defined for the prediction rule based on OLPM.

3. An empirical study

In this section, an empirical study was performed to investigate the performance of the two prediction rules based on OLPM and OSPM.

3.1. The data

Each sampled company must: (i) be listed on the New York Stock Exchange, American Stock Exchange, or NASDAQ, (ii) adopt calendar fiscal year, (iii) not be a financial services company with SIC code 6000–6999, (iv) have a S&P's LTR in April 2007, and (v) have complete values of the twenty-four potential predictors for studying S&P's LTRs in April 2007. The criterion (i) guarantees that market-driven variables are available. The criterion (ii) synchronizes the timing of predictors in the sense that all market-driven and accounting variables cover the same calendar year. The criterion (iii) excludes the financial services companies since

Table 1
The frequency and percent frequency (in parentheses) distributions of the sampled companies collected from the COMPUSTAT and CRSP databases with complete values of the twenty-four potential predictors for studying S&P's LTRs in April 2007.

	Estimation companies	Holdout companies
Panel A: S&P's LTR		
AAA	4 (0.97%)	2 (0.55%)
AA	18 (4.36%)	3 (0.82%)
A	88 (21.31%)	32 (8.74%)
BBB	156 (37.77%)	78 (21.31%)
BB	89 (21.55%	133 (36.34%)
B	57 (13.80%)	108 (29.51%)
CCC	0 (0%)	9 (2.46%)
CC	1 (0.24%)	0 (0%)
C	0 (0%)	0 (0%)
D	0 (0%)	1 (0.27%)
Total firms	413 (100%)	366 (100%)
Panel B: S&P's LTR category		
{Below BBB}	147 (35.59%)	251 (68.58%)
{BBB}	156 (37.77%)	78 (21.31%)
{AAA, AA, A}	110 (26.63%)	37 (10.11%)
Total firms	413 (100%)	366 (100%)
Panel C: SIC code		
1000–1999	37 (8.96%)	50 (13.66%)
2000–2999	112 (27.12%)	67 (18.31%)
3000–3999	93 (22.52%)	85 (23.22%)
4000–4999	114 (27.60%)	96 (26.23%)
5000–5999	20 (4.84%)	18 (4.92%)
7000–7999	26 (6.30%)	31 (8.47%)
8000–8999	11 (2.66%)	19 (5.19%)
Total firms	413 (100%)	366 (100%)

Panels A, B, and C present those two distributions of the sampled companies according to different S&P's LTR categories, and SIC codes.
Note: The financial services companies with SIC codes 6000–6999 were excluded from the study, since they are subject to regulations and adopt different accounting conventions. There were only three companies having SIC codes less than 1000 and two companies carrying SIC codes greater than 9000. These five companies were also excluded from study. If those industries with few companies are included in the prediction model, then the resulting number of industry indicator variables is increased, thus the estimates of parameters in the corresponding model might become less precise.

they are subject to regulations and adopt different accounting conventions. The criterion (iv) makes sure that the S&P's LTR is available. Finally, the criterion (v) excludes the companies with incomplete potential predictor values. The data were collected from both COMPUSTAT and CRSP databases. Since S&P's considers predictor values as their three-year averages, this study follows the same method (Blume et al., 1998; Poon, 2003). Thus, the value of each of the twenty-four potential predictors for studying S&P's LTRs in April 2007 was taken as the average of its values available from the two databases in years 2004, 2005, and 2006.

Based on the COMPUSTAT database, there were 3785 companies satisfying the selection criteria (i)–(iii), but only 918 companies among them receiving S&P's LTRs in April 2007. However, among those 918 companies, there were 784 companies having complete values of the twenty-four potential predictors for studying S&P's LTRs in April 2007. The missing data problem is not unusual in applications, especially when there are many predictors in the model. However, as long as the missingness occurs at random then the sample will not introduce systematic bias in our analyses (Allison, 2001; Little and Rubin, 2002). We have no reason not to believe that the missingness occurred in COMPUSTAT and CRSP databases is missing at random. On the other hand, in order to study industry effects on S&P's LTRs, the industries were classified by the first digit in the four-digit SIC code. Among the 784 companies, there were only three companies having SIC codes less than 1000 and two companies carrying SIC codes greater than 9000. These five companies were dropped from the sample.[3] Thus our final sample consisted of 779 companies.

For purpose of later analysis, the final sample was further divided into the estimation sample and the holdout sample based on the longevity of S&P's LTR. The reason for adopting this division principle has been given in footnote 2 in Section 1. According to S&P's Research Insight North America Data Guide, S&P's began using the term LTR on September 1, 1998. Companies receiving S&P's LTRs in consecutive nine years (April 1999–April 2007) were classified into the estimation sample. The rest of the sampled companies were classified into the holdout sample. Among our 779 sampled companies, there were 413 and 366 estimation and holdout companies, respectively. Table 1 presents the frequency distributions of the estimation and holdout companies according to different S&P's LTR categories and SIC codes.

[3] If such industries with few companies are included in the prediction model, then the resulting number of industry indicator variables is increased, thus the estimates of parameters in the corresponding model might become less precise.

Table 2
The definitions of the twenty-four potential predictors for studying important predictors of S&P's LTRs in April 2007.

Variable	Definition
Panel A: Four market-driven variables	
EXRET	Monthly return on the firm minus the value-weighted CRSP NYSE/AMEX/NASDAQ index return cumulated to obtain the yearly return
RSIZE	Logarithm of each firm's market equity value divided by the total NYSE/AMEX/NASDAQ market equity value
SIGMA	Standard deviation of each company monthly stock returns
KMV	KMV-Merton default probability
Panel B: Nineteen accounting variables	
Size	
$\log_{10}$ (TA)	Logarithm of Total assets
Financial leverage	
EM	Total assets / Equity
LDC	Long-term debt to capital
TDC	Total debt to capital
SDC	Short-term debt to capital
TDEBITDA	Total debt / (EBIT + DA), DA: depreciation plus amortization
Coverage	
EBITINT	EBIT / Interest expenses
EBITDAINT	(EBIT + DA) / Interest expenses
Cash flow	
FFO	Net income from continuing operations, plus DA, deferred income taxes, and other non-cash expense
INT	Interest expenses
CASHEQ	Total cash and equivalent
Profitability	
OM (%)	Operating margin after depreciation
ROC (%)	Return on capital
ROE (%)	Return on equity
ROA (%)	Return on assets
RETA	Retain earnings / Total assets
Liquidity	
CR	Current ratio
QR	Quick asset ratio
CASHR	Cash ratio
Panel C: Six industry indicator variables for studying industry effects	
SIC_1	1 if SIC code is within 1000–1999, and 0 otherwise
SIC_2	1 if SIC code is within 2000–2999, and 0 otherwise
SIC_3	1 if SIC code is within 3000–3999, and 0 otherwise
SIC_4	1 if SIC code is within 5000–5999, and 0 otherwise
SIC_5	1 if SIC code is within 7000–7999, and 0 otherwise
SIC_6	1 if SIC code is within 8000–8999, and 0 otherwise

Panels A, B, and C present the definitions of market-driven variables, accounting variables, and industry indicator variables, respectively.
Note: The SIC codes 4000–4999 were used as the reference level in studying the industry effects.

S&P's LTR ranges from AAA to D. Panel A of Table 1 presents the frequency distribution of the sampled companies according to their S&P's LTRs. Based on the result from the estimation sample, it seems reasonable to group S&P's LTRs into three categories: {Below BBB} as category 1, {BBB} as category 2, and {AAA, AA, A} as category 3.[4] According to S&P's opinion, firms in the {AAA, AA, A} category mean that they have demonstrated strong capacity to meet their financial obligations. Firms receiving BBB rating mean that they have adequate capacity to meet their financial commitments. However, firms receiving LTR below BBB mean that they are regarded as having speculative characteristics. Panel B of Table 1 gives the frequency distribution of the sampled companies according to the three S&P's LTR categories, and shows that the holdout companies have lower S&P's LTRs on the average. The distribution of the companies in the holdout sample shows that there are about 68.6% companies in the speculative S&P's LTR category. In contrast, there are only about 35.6% companies in the estimation sample with the speculative S&P's LTR category. As discussed in footnote 2 in Section 1, by comparing the longevity of S&P's LTR, the companies in our holdout sample correspond to the newcomers in the pool of companies having S&P's LTRs in April 2007, in contrast to those in the estimation sample. Thus, the result shown in Panel B of Table 1 agrees with the observation reported in Pettit et al. (2004) and Hwang et al. (2008). Blume et al. (1998) also reported the similar observation for bond ratings. Panel C of Table 1 gives the frequency distribution of the sampled companies according to SIC codes. It shows that the companies in the estimation sample are mainly from the industries with SIC codes 2000–4999. The same remark also applies to the companies in the holdout sample.

The twenty-four potential explanatory variables for studying S&P's LTRs in year 2007 include four market-driven variables, nineteen accounting variables, and industry effects. Their definitions are given in Table 2. The four market-driven variables are

[4] The estimation sample was divided into the three categories so that the resulting three cells have approximately equal sizes. On the other hand, if one divides the estimation sample into more categories, then some cells have smaller sizes and the number of threshold parameters increases. Thus the resulting estimates of parameters in the corresponding model might become less precise.

Table 3
Summary statistics and *F*-tests of the estimation sample.

Variable	Mean	Median	Standard deviation	Minimum	Maximum	*p*-value
Panel A: {Below BBB}						
EXRET	0.115	0.070	0.292	−0.422	1.054	0.007**
RSIZE	−4.076	−4.036	0.564	−6.548	−2.898	0.000**
SIGMA	0.099	0.093	0.038	0.044	0.251	0.000**
KMV	0.029	0.000	0.105	0.000	0.966	0.007**
$\log_{10}$ (TA)	3.425	3.420	0.480	2.350	4.478	0.000**
EM	4.749	3.058	13.916	−60.917	139.570	0.096
LDC	0.486	0.468	0.208	0.002	1.013	0.000**
TDC	0.530	0.513	0.205	0.028	1.023	0.000**
SDC	0.044	0.021	0.075	0.000	0.462	0.012*
TDEBITDA	3.633	3.310	9.613	−80.917	69.617	0.026*
EBITINT	5.775	2.637	15.846	−25.763	164.612	0.000**
EBITDAINT	9.724	4.063	33.768	−18.618	394.721	0.000**
FFO	352.271	181.167	496.760	−280.233	3065.000	0.000**
INT	149.658	63.678	244.430	1.770	1915.667	0.000**
CASHEQ	351.692	145.858	566.090	0.179	3641.667	0.000**
OM	11.453	8.450	14.927	−83.099	49.102	0.000**
ROC	3.313	3.464	9.609	−61.724	34.882	0.000**
ROE	3.804	6.582	40.686	−211.23	242.545	0.000**
ROA	2.303	2.491	5.331	−24.334	19.009	0.000**
RETA	0.016	0.070	0.373	−2.016	0.763	0.000**
CR	1.741	1.524	1.098	0.472	11.247	0.026*
QR	1.165	0.980	0.953	0.271	9.855	0.090
CASHR	0.559	0.344	0.901	0.005	9.154	0.055
Panel B: {BBB}						
EXRET	0.067	0.050	0.148	−0.336	0.630	
RSIZE	−3.598	−3.609	0.519	−5.410	−2.325	
SIGMA	0.061	0.059	0.019	0.026	0.114	
KMV	0.003	0.000	0.019	0.000	0.219	
$\log_{10}$ (TA)	3.822	3.785	0.456	2.947	5.100	
EM	3.030	2.643	1.794	1.433	17.591	
LDC	0.350	0.345	0.139	0.002	0.822	
TDC	0.406	0.395	0.150	0.035	0.873	
SDC	0.056	0.045	0.048	0.000	0.225	
TDEBITDA	2.382	1.986	1.698	0.097	17.142	
EBITINT	9.607	5.781	17.731	0.713	171.302	
EBITDAINT	13.342	8.500	24.626	1.602	248.009	
FFO	1046.719	522.681	1619.155	17.967	16,427.302	
INT	209.605	96.369	335.349	1.345	2974.245	
CASHEQ	651.585	285.864	1151.857	4.074	10,287.194	
OM	14.846	12.764	8.828	1.279	52.942	
ROC	8.299	7.638	5.847	−17.410	38.158	
ROE	14.591	13.219	11.126	−22.747	97.052	
ROA	5.217	4.762	3.357	−12.769	14.676	
RETA	0.212	0.219	0.189	−0.747	0.758	
CR	1.492	1.395	0.715	0.279	4.839	
QR	0.982	0.853	0.560	0.159	4.035	
CASHR	0.376	0.220	0.456	0.009	3.487	
Panel C: {AAA, AA, A}						
EXRET	0.032	0.012	0.135	−0.241	0.516	
RSIZE	−3.104	−3.061	0.611	−4.907	−1.678	
SIGMA	0.053	0.052	0.017	0.025	0.106	
KMV	0.010	0.000	0.070	0.000	0.553	
$\log_{10}$ (TA)	4.131	4.121	0.563	2.522	5.317	
EM	2.701	2.314	1.303	1.245	9.300	
LDC	0.271	0.245	0.153	0.001	0.707	
TDC	0.339	0.333	0.174	0.004	0.771	
SDC	0.067	0.058	0.057	0.000	0.286	
TDEBITDA	1.686	1.391	1.415	0.016	9.807	
EBITINT	25.914	11.401	51.392	0.038	369.383	
EBITDAINT	34.493	15.647	71.685	2.208	547.558	
FFO	4140.589	1626.167	6718.647	14.988	45,991.667	
INT	342.962	118.834	576.459	2.992	4176.468	
CASHEQ	2697.290	713.283	4642.380	4.282	29,752.667	
OM	17.992	17.102	8.935	−0.051	52.755	
ROC	13.249	12.470	7.108	−0.328	34.792	

(continued on next page)

Table 3 (*continued*)

Variable	Mean	Median	Standard deviation	Minimum	Maximum	*p*-value
Panel C: {AAA, AA, A}						
ROE	21.083	18.777	13.304	−1.688	96.553	
ROA	8.578	7.940	4.516	−0.211	19.644	
RETA	0.394	0.373	0.252	−0.126	0.989	
CR	1.509	1.337	0.712	0.518	4.329	
QR	1.047	0.933	0.585	0.262	3.488	
CASHR	0.453	0.261	0.503	0.012	2.455	

The *p*-values refer to the *F*-tests of equality of the means among the three S&P's LTR categories based on the estimation sample. Panels A, B, and C present the results for the three S&P's LTR categories {Below BBB}, {BBB}, and {AAA, AA, A}, respectively.
Note: The notation ** and * indicates the significance of the *F*-test at the 1% and 5% levels, respectively.

excess return (EXRET), relative size (RSIZE), standard deviation of monthly returns (SIGMA), and KMV-Merton default probability (KMV).[5] The nineteen accounting variables measure different aspects (size, financial leverage, coverage, cash flow, profitability, and liquidity) of financial health of a company. The industry effects were estimated through the coefficients of six industry indicator variables in the model. Using the data in the estimation sample, Table 3 shows summary statistics and *F*-tests of equality of the means among the three S&P's LTR categories for the market-driven variables and accounting variables. The *p*-values in Table 3 show that testing the null hypothesis of equal means is significant at 5% level for each of the four market-driven variables and are significant for all accounting variables, except EM, QR, and CASHR. This result indicates that most of the variables considered in this paper are effective predictive variables. Table 3 also shows that, on the average, if a company has larger firm size, less risk, lower financial leverage, higher coverage, larger cash flow, or higher profitability, then it has better S&P's LTR category.

3.2. Testing selection bias

To examine whether our criterion for dividing the overall sample into estimation and holdout samples induced selection bias, the procedure of OLPM under our particular sample selection was performed. This procedure was designed by simultaneously applying OLPM to the estimation sample, and applying the two-class linear probit model to the particular sample in which the classes 0 and 1 were assigned respectively to the holdout and estimation companies (see Model 1 of Table 5). The detail of this approach can be found in Greene (2002). Before performing it, a forward selection procedure based on minimizing in-sample total error rate was first used to objectively determine effective predictors for OLPM. The in-sample total error rate was computed using the classification scheme based on OLPM with cutoff value 1/2 for simplicity. The variable selection procedure asked that the newly entered variable must be able to reduce at least one percent of in-sample total error rate of the classification scheme using those already entered variables. Table 4 gives the variable selection result. It shows that the final list of the selected predictors in OLPM includes RSIZE, SIGMA, LDC, ROE, and industry effects. Results in Table 3, by using *F*-test for testing equality of three means (corresponding to the three S&P's LTR categories), show that the four selected continuous predictors are all significant at 1% level.

Using the five selected variables, Table 5 shows the results obtained from performing the procedure of OLPM, under our particular sample selection, based on the application of LIMDEP 8.0. Panel C of Model 1 in Table 5 shows that the null hypothesis of $corr = 0$ was not rejected at 5% level of significance. Here the null hypothesis of $corr = 0$ stands for no sample selection bias caused by our criterion of dividing the overall sample. On the other hand, Model 2 in Table 5 shows the results of OLPM, including the *p*-value of the chi-squared test for model fit. It is also of interest to note that by comparing the parameter estimates of the two OLPMs in Models 1 and 2 of Table 5, we find out that their values are approximately equal. Thus since there is no sample selection bias, the results of Model 2 are adopted for predicting S&P's LTR categories for companies in the holdout sample.

It is important to note that the two selected predictors RSIZE and ROE stand for the market capitalization and profitability of an issuer, respectively. The larger the values of these two predictors (RSIZE and ROE), the better the creditworthiness of an issuer. This implies that RSIZE and ROE should be positively correlated with rating, and the signs of their coefficients should be positive. Another two selected predictors SIGMA and LDC measure various aspects of risk and financial leverage of an issuer, respectively. The larger the values of these two predictors (SIGMA and LDC), the lower the S&P's LTR category. That is, SIGMA and LDC should be negatively correlated with rating, and the signs of their coefficients should be negative. From Model 2 of Table 5, the signs of the values of estimated coefficients for these four selected continuous predictors RSIZE, SIGMA, LDC, and ROE all agree with our expectation. Model 2 also shows that not all industries have the same effect on S&P's LTRs, thus the industry effects are useful for predicting ratings.

Table 6 shows summary statistics and median tests of the estimation and holdout samples for the four selected continuous predictors. For each of these predictors, the two-sample median test was performed to test the null hypothesis of equal magnitude for an estimation company and for a holdout company. The *p*-value in Table 6 shows that the null hypothesis of equal magnitude for companies in our estimation and holdout samples was significant at 5% level for each of the four selected continuous predictors.

[5] Since the computation of KMV-Merton default probability requires the market value of equity, it is treated as a market-driven variable. Its detailed computational procedure can be referred to Bharath and Shumway (2008).

Table 4
The variable selection result.

Selection sequence	Variable	$\gamma_{\text{in}}(1/2)$
1	RSIZE	0.436*
2	SIGMA	0.351*
3	LDC	0.283*
4	Industry effects	0.242*
5	ROE	0.237*
6	$\log_{10}$ (TA)	0.235
7	CR	0.232
8	CASHR	0.228
9	TDEBITDA	0.228
10	QR	0.228
11	EBITDAINT	0.225
12	EXRET	0.225
13	TDC	0.223
14	FFO	0.215
15	SDC	0.215
16	KMV	0.223
17	EM	0.223
18	OM	0.228
19	RETA	0.232
20	INT	0.223
21	EBITINT	0.220
22	ROA	0.230
23	ROC	0.232
24	CASHEQ	0.237

The variables were selected in sequence for OLPM by the forward selection procedure based on minimizing in-sample total error rate. The in-sample total error rate was computed using the classification scheme based on OLPM with cutoff value 1/2. The newly entered variable must be able to reduce at least one percent of in-sample total error rate of the classification scheme using those already entered variables.
Note: The notation * indicates that the variable was selected by the forward selection procedure.

Combining the test results with the magnitude of sample medians in Table 6, we conclude that the holdout companies generally have smaller firm size, larger risk, higher financial leverage, and lower profitability than the estimation companies. The conclusion indicates that the holdout companies have worse credit quality on the average, and coincides with the result that the holdout companies have lower S&P's LTRs shown in Panel B of Table 1.

3.3. Computational procedures

In computing OSPM, the values of the four selected continuous predictors RSIZE, SIGMA, LDC, and ROE were first divided by their sample standard deviations so that they have the same scale. This is important, since the influence of the predictor with very large range in estimating the optimal values of (b, g, p) can be avoided.

As discussed in Section 2.4 and in Remark 1 of Appendix A, the kernel function K was taken as the Epanechnikov kernel. Also, a grid search approach was used in computing the optimal values of (b, g, p) for OSPM. The values of $\gamma_{\text{in}}(b, g, p)$ on the equally spaced logarithmic grid of $51\times 51\times 1001$ values of (b, g, p) in $[0.5, 3]\times[0.5, 3]\times[10^{-5}, 1]$ 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 $\{\hat{b}(u), \hat{g}(u), \hat{p}_{\text{OSPM}}(u)\}$ of $\gamma_{\text{in}}(b, g, p)$ on the grid points with restrictions $p\in[0,1]$, $0<b<g$, and $\alpha_{\text{in}}(b, g, p)\le u$ was taken as the optimal values of (b, g, p).

Using the estimation sample and the values of $\{\hat{b}(1), \hat{g}(1), \hat{p}_{\text{OSPM}}(1)\}$, Fig. 1 shows the plot of $\{x_j, \hat{H}(x)\}$ produced from OSPM for each of the four selected continuous predictors RSIZE, SIGMA, LDC, and ROE. Here x_j stands for the value of the jth selected continuous predictor, and x has the jth component as x_j, but all other components are fixed at their sample median levels. In the plot of $\{x_j, \hat{H}(x)\}$, we have taken the left and the right boundary points of its horizontal axe as the 0.5 and 99.5 percentiles of the values of the jth selected continuous predictor, respectively. These plots are used to visually check the adequacy of the order-one polynomial function assumed for each continuous predictor in the linear regression function of OLPM. From panels (a)–(d) of Fig. 1, the slope of each curve agrees with the expected direction of the corresponding variable effect. However, we also see that the order-one polynomial assumed for each continuous predictor by OLPM is inadequate, since it is clear that there is a nonlinear relationship between x_j and $\hat{H}(x)$ for each of RSIZE, SIGMA, and ROE, except LDC.

3.4. Prediction results

Given the estimation and holdout samples, the performance of the two discussed prediction rules based on OLPM and OSPM is presented in Fig. 2 and Tables 7 and 8.

Table 5
Maximum likelihood estimates of the parameters in Models 1 and 2.

Variable	Model 1		Model 2	
	Coefficient	*p*-value	Coefficient	*p*-value
Panel A: Ordered linear probit specification				
Threshold	1.918	0.000**	1.923	0.000**
Intercept	7.678	0.000**	7.639	0.000**
SIC_1	0.199	0.667	0.207	0.443
SIC_2	0.275	0.466	0.266	0.167
SIC_3	−0.064	0.777	−0.066	0.748
SIC_4	−0.683	0.173	−0.691	0.042*
SIC_5	−0.885	0.016*	−0.883	0.010**
SIC_6	−1.208	0.061	−1.197	0.008**
RSIZE	0.859	0.370	0.834	0.000**
SIGMA	−37.954	0.000**	−37.750	0.000**
LDC	−4.120	0.000**	−4.128	0.000**
ROE	0.015	0.048*	0.015	0.000**
Panel B: Two−class linear probit sample selection specification				
Intercept	2.509	0.000**		
SIC_1	−0.134	0.462		
SIC_2	0.178	0.188		
SIC_3	0.023	0.876		
SIC_4	0.101	0.664		
SIC_5	−0.072	0.728		
SIC_6	−0.270	0.342		
RSIZE	0.505	0.000**		
SIGMA	−5.636	0.000**		
LDC	−0.069	0.734		
ROE	−0.001	0.174		
Panel C: Model fit test				
Chi-squared statistic	0.012	0.914	399.014	0.000**
df	1		10	
corr	0.095	0.981		

Model 1 denotes OLPM with sample selection. Model 2 stands for OLPM. The variables used in each model were selected by the forward selection procedure shown in Table 4. The *p*-values refer to the Wald chi-squared tests for testing the significance of parameters. Panel A shows the results of OLPM under Models 1 and 2. Panel B gives the results of the two-class linear probit model under Model 1. Panel C presents the results for model fit test.
Note: The notation ** and * indicates the significance of the test at the 1% and 5% levels, respectively. The notation *df* and *corr* stands for the degree of freedom and the correlation coefficient between the error term in OLPM and that in the two-class linear probit model, respectively.

Fig. 2 shows the error rates of the two discussed prediction rules. These error rates were derived under the constraint that the in-sample type I error rate was at most u. Panels (a), (c), and (e) of Fig. 2 show the in-sample error rates of the two prediction rules. In the case of $u \leq 0.08$, their in-sample type I error rates are close to the designed bounds, but the in-sample type II and total error rates of OSPM are smaller than those of OLPM. The largest percentage decrease of the in-sample total error rate by OSPM over OLPM is about 21%. On the other hand, panels (b), (d), and (f) of Fig. 2 show the out-of-sample error rates of the two prediction rules. Their out-of-

Table 6
Summary statistics and median tests of the estimation and holdout samples for the four selected continuous predictors.

Variable	Mean	Median	Standard deviation	Minimum	Maximum	*p*-value
Panel A: The estimation sample						
RSIZE	−3.637	−3.663	0.677	−6.548	−1.678	0.000**
SIGMA	0.072	0.066	0.033	0.025	0.251	0.000**
LDC	0.377	0.362	0.191	0.001	1.013	0.041*
ROE	12.480	13.340	26.987	−211.230	242.545	0.001**
Panel B: The holdout sample						
RSIZE	−4.075	−4.122	0.626	−5.826	−2.356	
SIGMA	0.092	0.086	0.041	0.027	0.381	
LDC	0.565	0.393	2.607	−1.465	49.781	
ROE	21.557	10.816	154.852	−514.307	2366.403	

The *p*-values refer to the median tests of equality of the medians between the estimation and holdout samples. Panels A and B present the results for the estimation and holdout samples, respectively.
Note: The notation ** and * indicates the significance of the two-sample median test at the 1% and 5% levels, respectively.

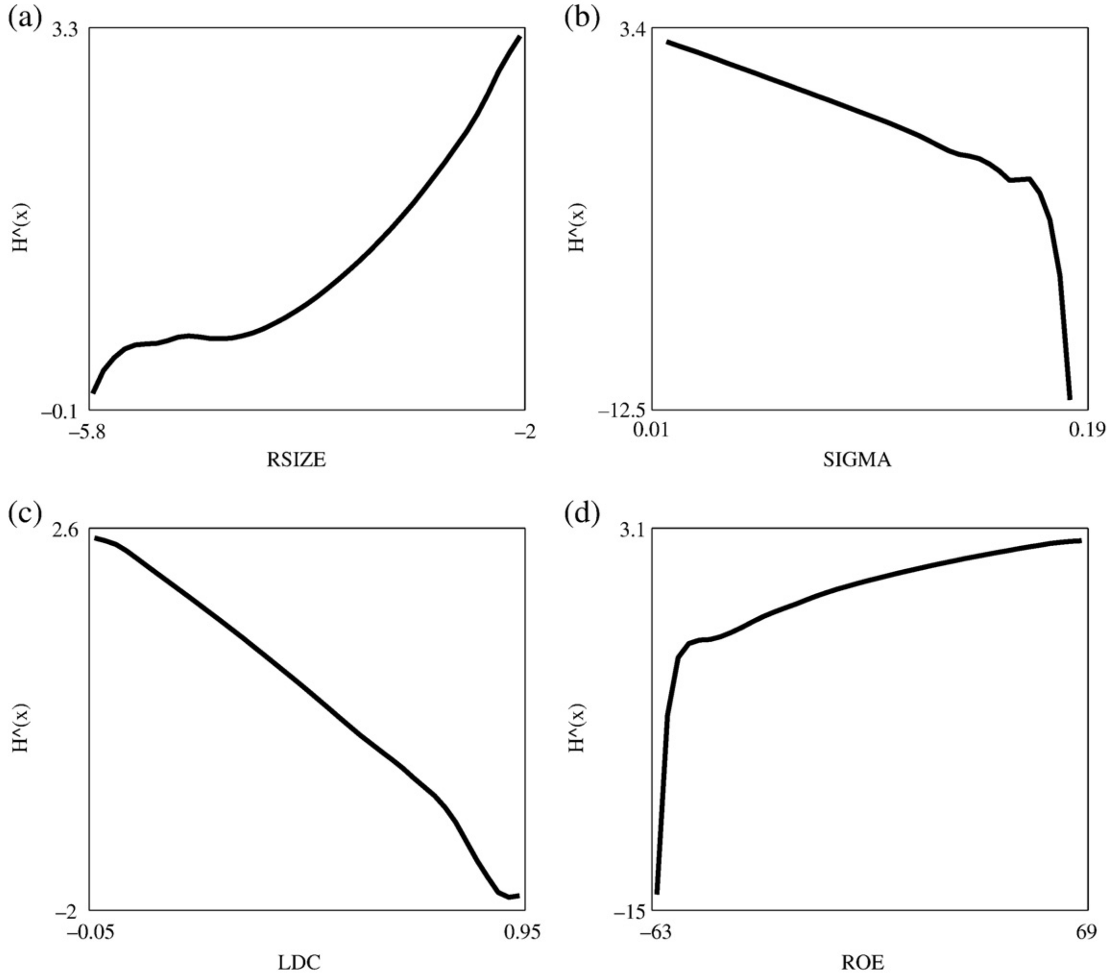

Fig. 1. Plots of marginal relations between the function $\hat{H}(x)$ and the four selected continuous predictors RSIZE, SIGMA, LDC, and ROE. Panels (a)–(d) show the plots of $\{x_j, \hat{H}(x)\}$ resulted from OSPM using $\{\hat{b}(1), \hat{g}(1), \hat{p}_{OSPM}(1)\}$. Here x_j in the plot of $\{x_j, \hat{H}(x)\}$ stands for the value of the j-th selected continuous predictor, and x has the jth component as x_j, but all other components are fixed at their sample median levels. From panels (a)–(d), the slope of each curve agrees with the expected direction of the corresponding variable effect. But these plots show clearly that there is a nonlinear relationship between x_j and for each of RSIZE, SIGMA, and ROE, except LDC.

sample type II error rates are very similar, for all $u \in [0,1]$. However, the out-of-sample type I and total error rates of OSPM are in general smaller than those of OLPM, for all $u \in [0,1]$. The largest percentage decrease of the out-of-sample total error rate by OSPM over OLPM is 22%. Considering the out-of-sample type I and total error rates, OSPM clearly outperforms OLPM, for all $u \in [0,1]$.

In the case of $u = 1$, the results of classifying the estimation companies and predicting the holdout companies are given in Tables 7 and 8, respectively. All error rates shown in these two tables were produced under no constraint on the magnitude of in-sample type I or type II error rate. Table 7 shows that OSPM has smaller in-sample type I and total error rates, but larger in-sample type II error rate. Table 8 shows that OSPM has better prediction performance than OLPM, since the out-of-sample total error rates of their predictions are equal to 18.9% and 24.0%, respectively. Also, OSPM has better ability in predicting the speculative grade {Below BBB}. This is important, since misclassifying speculative grade to investment grade ({BBB} or {AAA, AA, A}) might cause severe losses to investors.

4. Conclusion remarks and future research topics

In this paper, a multiple-class prediction method based on OSPM is proposed. Our OSPM is developed by replacing the linear regression function in OLPM with a semiparametric regression function $H(x) + \theta z$. Here $H(x)$ is an unknown but smooth function of the value x of the d-dimensional continuous explanatory variable X, and z is the value of q-dimensional discrete explanatory variable Z. Hence OSPM is more flexible than OLPM in modeling regression function. The estimators of unknown quantities in OSPM are developed from the local likelihood method, and computed by maximizing a weighted log-likelihood function. Thus the required computation for OSPM is as simple as that for OLPM. The large sample properties of these estimators are studied through their asymptotic biases and variances. Theoretical results show that the computed regression function value using OSPM consistently estimates the true regression function value. Thus OSPM is a reliable prediction model.

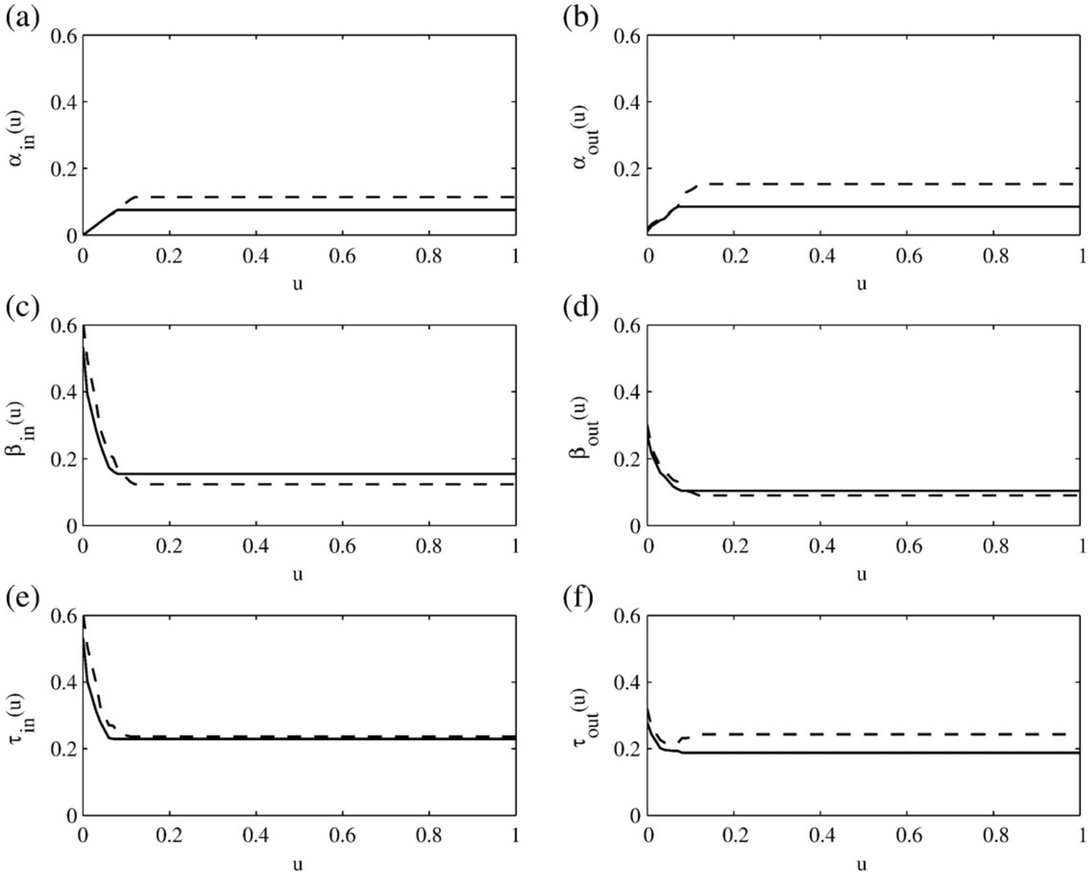

Fig. 2. The performance of the two prediction rules based on OLPM (dashed curve) and OSPM (solid curve) using the five selected variables RSIZE, SIGMA, LDC, ROE, and industry effects. Panels (a), (c), and (e) show respectively the in-sample type I, type II, and total error rates obtained from the 413 estimation companies by the two prediction rules. Panels (b), (d), and (f) show respectively the out-of-sample type I, type II, and total error rates derived from the 366 holdout companies by the two prediction rules.

To decide the optimal prediction rule, we propose to control the magnitude of in-sample type I error rate to be at most u, where $u \in [0,1]$, so that the sum of in-sample type I and II error rates is minimal. Based on the estimation sample, the type I error rate denotes the rate of misclassifying a company to a higher rating category, and type II error rate stands for the rate of misclassifying a company to a lower rating category. Controlling the magnitude of type I error rate is important, since the type I error might cause severe losses to investors.

Table 7
Classification results obtained from the 413 estimation companies.

True category	Classified category		
	{Below BBB}	{BBB}	{AAA, AA, A}
Panel A: OLPM with $u=1$			
{Below BBB}	123	23	1
{BBB}	16	122	18
{AAA, AA, A}	4	35	71
$\alpha_{\text{in}}=(23+1+18)/413=0.102$, $\beta_{\text{in}}=(16+4+35)/413=0.133$, $\gamma_{\text{in}}=0.102+0.133=0.235$			
Panel B: OSPM with $u=1$			
{Below BBB}	126	21	0
{BBB}	17	129	10
{AAA, AA, A}	4	43	63
$\alpha_{\text{in}}=(21+0+10)/413=0.075$, $\beta_{\text{in}}=(17+4+43)/413=0.155$, $\gamma_{\text{in}}=0.075+0.155=0.230$			

Panels A and B show the results obtained by the two prediction rules based on OLPM and OSPM in the case of $u=1$, respectively.

Table 8
Prediction results obtained from the 366 holdout companies.

True category	Predicted category		
	{Below BBB}	{BBB}	{AAA, AA, A}
Panel A: OLPM with $u=1$			
{Below BBB}	207	35	9
{BBB}	16	54	8
{AAA, AA, A}	2	18	17
$\alpha_{\text{out}}=(35+9+8)/366=0.142$, $\beta_{\text{out}}=(16+2+18)/366=0.098$, $\gamma_{\text{out}}=0.142+0.098=0.240$			
Panel B: OSPM with $u=1$			
{Below BBB}	223	26	2
{BBB}	15	60	3
{AAA, AA, A}	4	19	14
$\alpha_{\text{out}}=(26+2+3)/366=0.085$, $\beta_{\text{out}}=(15+4+19)/366=0.104$, $\gamma_{\text{out}}=0.085+0.104=0.189$			

Panels A and B show the results obtained by the two prediction rules based on OLPM and OSPM in the case of $u=1$, respectively.

One additional advantage of using OSPM is that the relation between $H(x)$ and the value x of the d-dimensional continuous predictive variable X can be obtained from the plots of $\{x_j, \hat{H}(x)\}$, for $j=1,\cdots,d$. Here x_j denotes the value of the jth continuous predictor, and x in $\hat{H}(x)$ has the jth component as x_j, but all other components are fixed at some values, for example their sample medians. Using these plots of $\{x_j, \hat{H}(x)\}$, the adequacy of the order-one polynomial function assumed for each continuous predictor in the linear regression function of OLPM can be visually checked. If the linear regression function of OLPM is not proper, the plots of $\{x_j, \hat{H}(x)\}$ may guide us on how to make a better selection of parametric regression function for ordered probit model. For example, if the plot of $\{x_j, \hat{H}(x)\}$, for some j, presents a quadratic relation, then the relation between the regression function and x_j should be an order-two polynomial. Sometime, using a parametric regression function in ordered probit model is important, particularly when one has many predictors to be considered simultaneously and does not have enough sample data to estimate the regression function nonparametrically.

One real data example for predicting S&P's LTRs in year 2007 has been used to illustrate OSPM. To find important predictors of S&P's LTRs, we have considered twenty-four potential predictors used in previous studies. The twenty-four potential predictors include four market-driven variables, nineteen accounting variables, and industry effects. A data set containing 779 companies (413 estimation companies and 366 holdout companies) having complete values of the twenty-four potential predictors for studying S&P's LTRs in April 2007 was collected from COMPUSTAT and CRSP databases. The result obtained by a forward selection procedure shows that the final list of the selected predictors for OLPM contains industry effects, two market-driven variables, and two accounting variables. Given the estimation sample, our empirical results demonstrate that the functional form between the S&P's LTR assessment and the selected continuous predictors is nonlinear. This indicates that OLPM may not be adequate in explaining and predicting S&P's LTRs. Given the holdout sample, our empirical results also show that the prediction method based on OSPM has better performance than that based on OLPM, in the sense of yielding smaller out-of-sample total error rate. Thus, by the empirical results, OSPM has potential to be a powerful credit rating forecasting model.

In order to estimate OSPM in practice, we need to decide proper values of bandwidth parameters b and g. In this paper, we suggest using a grid search approach to find those proper values. However, this approach suffers from heavy computational burden. One possible remedy for this drawback is to use the plug-in method to estimate those proper values. For example, we may determine those values by minimizing the estimated mean square errors of $\hat{H}(x)$ and $\hat{\theta}$. For more discussion of the plug-in approach, see for example Härdle et al. (1992) and Jones et al. (1996).

There are some possible extensions for the methods considered in this paper. First, in this paper, we only used cross-sectional data to study the performance of OSPM. The model can be directly applied to panel data with independence assumption (Blume et al., 1998; Poon, 2003; Güttler and Wahrenburg, 2007). Second, to account for the autocorrelations among panel data, we may introduce a dynamic OLPM or OSPM with autocorrelation structure (Lipsitz et al., 1994; Müller and Czado, 2005). Third, our OSPM can be applied to panel data with cross-sectional dependence by introducing unobservable macroeconomic factors whose presence creates cross-sectional correlations among credit ratings (Feng et al., 2008; Stefanescu et al., 2009). Fourth, the performance of OSPM for predicting credit ratings was only studied in this paper using firm-specific variables including market-driven variables, accounting variables, and industry effects. The macroeconomic variables such as the change in GDP, unemployment rate, return on the S&P 500 index, and Chicago Fed National Activities Index have been considered in Güttler and Wahrenburg (2007) and Stefanescu et al. (2009) as major determinants of credit ratings. It is of interest to study the effects of macroeconomic variables on OSPM for predicting credit ratings. Fifth, industry effects on S&P's LTRs were studied in this paper by introducing industry indicator variables in each of OLPM and OSPM. They could also be studied using frailty factors to describe the unobservable heterogeneity (Duffie et al., 2006; Chava et al., 2008). Finally, the performance of both classification and prediction schemes based on each of OLPM and OSPM was measured in this paper by classification error rates. These error rates do not account for the degree of classification errors. To correct for the drawback, we may consider misclassification costs (Johnson and Wichern, 2002) among the given S&P's LTR categories, that is, the cost incurred when a company of category j is erroneously classified as belonging to category i. A proper classifier should be able to minimize the total (average) misclassification cost.

Appendix A. Theoretical results

In Appendix A, asymptotic properties of $\hat{\theta}$,$\hat{\tau}$,$\hat{H}(x_0)$, and $\tilde{H}(x_0)$ are studied. For doing it, we need the following conditions:

(C1) The function $H(x)$ is defined on $[0,1]^d$, and each of its second order partial derivative is Lipschitz continuous on $[0,1]^d$.
(C2) The frequency function $f(x, z)$ of (X, Z) is Lipschitz continuous and bounded above zero on $[0,1]^d$ with respect to x, for each z. Also, the conditional probability density function $f(z|x)$ of Z given $X=x$ can not be zero or one for each x, and is Lipschitz continuous with respect to x.
(C3) The kernel function K is a symmetric and Lipschitz continuous probability density function supported on $[-1,1]$.
(C4) The values of b and g are selected on the interval $[sn^{-1+s}, s^{-1}n^{-s}]$, where s is an arbitrarily small positive constant. The values of b and g satisfy $nb^{d+2} \gg 1 \gg g \gg b$. The notation $a_n \gg b_n$ means that $b_n/a_n \to 0$, as $n \to \infty$.

In order to present asymptotic properties of $\hat{\theta}$, $\hat{\tau}$, $\hat{H}(x_0)$, and $\tilde{H}(x_0)$, we need more notation. Set $\varphi_k = \max\{-1,(x_{0,k}-1)/b\}$, $\rho_k = \min\{1,x_{0,k}/b\}$, $H^{(2)}(u) = \{H_{i,j}(u)\}_{d\times d}$ as the Hessian matrix of $H(u)$, and $K^{\#}(u) = \prod_{k=1}^{d} K(u_k)$, for $k=1,\cdots,d$ and $u=(u_1,\cdots,u_d)^T$. Let

$$\kappa_0 = \int_{\varphi_1}^{\rho_1}\cdots\int_{\varphi_d}^{\rho_d} K^{\#}(u)du,\ \kappa_1 = \int_{\varphi_1}^{\rho_1}\cdots\int_{\varphi_d}^{\rho_d} uK^{\#}(u)du,\ \kappa_2 = \int_{\varphi_1}^{\rho_1}\cdots\int_{\varphi_d}^{\rho_d} (uu^T)K^{\#}(u)du,$$
$$\kappa_2^* = \int_{\varphi_1}^{\rho_1}\cdots\int_{\varphi_d}^{\rho_d} \{u^T H^{(2)}(x_0)u\}K^{\#}(u)du,\ \kappa_3^* = \int_{\varphi_1}^{\rho_1}\cdots\int_{\varphi_d}^{\rho_d} u\{u^T H^{(2)}(x_0)u\}K^{\#}(u)du.$$

Set Q as the collection of all values of the q-dimensional discrete explanatory variable Z, $d_j(u,z) = \phi_j(u,z) - \phi_{j-1}(u,z)$, and $\psi_j(u,z) = \{\phi_j(u, z)e_j - \phi_{j-1}(u, z)e_{j-1}\}/d_j(u, z)$, for $j=1,\cdots, m$. Here e_j is an $(m-2)\times 1$ vector with the $(j-1)$th component as 1 and all other components as 0, $\phi_j(u,z) = \phi\{\tau_j - H(u) - \theta z\}$, and ϕ is the probability density function of the standard normal random variable. Define $\Gamma_j = \int_0^1\cdots\int_0^1 T_j(u)du$, for $j=0,1,2$, where $T_0(u) = \sum_{z\in Q}\sum_{j=1}^{m} d_j(u,z)^2\psi_j(u,z)\, f(u,z)\,/\,D_j(u,z)$, $T_1(u)$ and $T_2(u)$ are $T_0(u)$ with $f(u, z)$ replaced respectively by $z^T f(u,z)$ and $\psi_j(u,z)^T f(u,z)$, and $D_j(u,z)$ is $d_j(u,z)$ with ϕ replaced by Φ. Similarly define ξ_0, ξ_1, and ξ_2 as κ_0, κ_1, and κ_2 with $K^{\#}(u)$ replaced by $K^{\#}(u)^2$, and $\Sigma_j = \int_0^1\cdots\int_0^1 S_j(u)du$, for $j=0,1,2$, where $S_0(u)$, $S_1(u)$, and $S_2(u)$ are $T_0(u)$, $T_1(u)$, and $T_2(u)$ with $\psi_j(u, z)$, $\psi_j(u, z)z^T$, and $\psi_j(u, z)\psi_j(u, z)^T$ replaced by 1, z, and zz^T, respectively. Let

$$G = \Sigma_2 - \Gamma_1^T\Gamma_2^{-1}\Gamma_1,\ \lambda_0 = (1/2)\left\{\int_{-1}^{1} t^2K(t)dt\right\}\int_0^1\cdots\int_0^1\left\{\sum_{j=1}^{d} H_{j,j}(u)\right\}T_0(u)du,$$

and λ_1 be λ_0 with $T_0(u)$ replaced by $S_1(u)$. Finally, define quantities related to asymptotic biases and variances of $\hat{\theta}$, $\hat{\tau}$, and $\hat{H}(x_0)$:

$$b_\theta = G^{-1}(\Gamma_1^T\Gamma_2^{-1}\lambda_0 - \lambda_1),\ V_\theta = G^{-1},$$
$$b_\tau = \Gamma_2^{-1}(\Gamma_1 b_\theta + \lambda_0),\ V_\tau = \Gamma_2^{-1}\Gamma_1 G^{-1}\Gamma_1^T\Gamma_2^{-1} + \Gamma_2^{-1},$$
$$b_H(x_0) = \frac{(\kappa_2^*)^{1-d}det(\kappa_2^*\kappa_2 - \kappa_3^*\kappa_1^T)}{2\kappa_0^{1-d}det(\kappa_0\kappa_2 - \kappa_1\kappa_1^T)},\ V_H(x_0) = \frac{\xi_0 - 2\kappa_1^T\kappa_2^{-1}\xi_1 + \kappa_1^T\kappa_2^{-1}\xi_2\kappa_2^{-1}\kappa_1}{S_0(x_0)(\kappa_0 - \kappa_1^T\kappa_2^{-1}\kappa_1)^2}.$$

Asymptotic biases and variances of $\hat{\theta}$, $\hat{\tau}$, and $\hat{H}(x_0)$ are shown in Theorem 1, and those of $\tilde{H}(x_0)$ in Eq. (A10) of Appendix B. Their proofs are given in Appendix B.

Theorem 1. Assume model (1) and let conditions (C1)–(C4) be satisfied. Asymptotic biases and variances of $\hat{\theta}$, $\hat{\tau}$, and $\hat{H}(x_0)$ can be expressed as

$$Bias(\hat{\theta}) = b^2 b_\theta\{1 + o(1)\},\ Var(\hat{\theta}) = n^{-1}V_\theta\{1 + o(1)\}, \tag{A1}$$

$$Bias(\hat{\tau}) = b^2 b_\tau\{1 + o(1)\},\ Var(\hat{\tau}) = n^{-1}V_\tau\{1 + o(1)\}, \tag{A2}$$

$$Bias\{\hat{H}(x_0)\} = g^2 b_H(x_0)\{1 + o(1)\},\ Var\{\hat{H}(x_0)\} = n^{-1}g^{-d}V_H(x_0)\{1 + o(1)\}, \tag{A3}$$

for each $x_0 \in [0, 1]^d$.

Remark 1. By Theorem 2.1 of Ruppert and Wand (1994) and our Eq. (A3), the optimal K satisfying the conditions in (C3) for constructing $\hat{H}(x_0)$, for each $x_0 \in [0, 1]^d$, is the Epanechnikov kernel defined as $K(t) = (3/4)(1-t^2)I(|t| \le 1)$, in the sense of having smaller asymptotic mean integrated square error. On the other hand, by Eq. (A3), the optimal choice of the value of g, in terms of having smaller asymptotic mean integrated square error of $\hat{H}(x_0)$, is $g^* = c_0^{1/(d+4)}n^{-1/(d+4)}$, where $c_0 = \left\{d\int_0^1\cdots\int_0^1 V_H(u)du\right\}\left\{4\int_0^1\cdots\int_0^1 b_H(u)^2du\right\}^{-1}$. However, the optimal value g^* is not available in practice since it depends on the unknown factors $H(x)$, θ, τ, and $f(x, z)$. Similarly, by Eqs. (A1)–(A3), (A10), and (C4), the optimal value b^* of b for constructing $\tilde{H}(x_0)$, in terms of having smaller asymptotic mean square error of $\hat{\theta}$ and $\hat{\tau}$, satisfies the condition $n^{-1/(d+4)} \gg b \gg n^{-1/(d+2)}$. The result indicates that the value of g^* is of larger order than that of b^*, and that the asymptotic mean integrated square error of $\hat{H}(x_0)$ using g^* is of smaller order in magnitude than that of $\tilde{H}(x_0)$ using b^*.

Appendix B. Sketches of the proofs

In Appendix B, sketches of the proof for Theorem 1 are given. The following notation will be used throughout Appendix B. Set $\omega=(\eta,\theta,\tau),\tilde{\omega}=(\tilde{\eta},\tilde{\theta},\tilde{\tau})$, $\zeta=(\theta,\tau)$, and $\hat{\zeta}=(\hat{\theta},\hat{\tau})$. Let $\ell_j^{(1)}$ and $\ell_j^{(2)}$ be the gradient vector and the Hessian matrix of ℓ_j given in Sections 2.2–2.3, for each $j=0,1,2$. Let f_X be the marginal probability density function of the d-dimensional continuous explanatory variable X. Set P_0 as the event that the number of x_i falling into $\prod_{j=1}^{d}[x_{0,j}-b/2,x_{0,j}+b/2]$ is less than $\rho_0 n\int_{x_{0,1}-b/2}^{x_{0,1}+b/2}\cdots\int_{x_{0,d}-b/2}^{x_{0,d}+b/2}f_X(u)du$, where ρ_0 is a positive constant satisfying $\rho_0\le 1/4$, and Q_0 the event that the number of x_i falling into $\prod_{j=1}^{d}[x_{0,j}-b,x_{0,j}+b]$ is greater than $\varphi_0 n\int_{x_{0,1}-b}^{x_{0,1}+b}\cdots\int_{x_{0,d}-b}^{x_{0,d}+b}f_X(u)du$, where φ_0 is a positive constant satisfying $\varphi_0\ge\exp(1)$.

Proof of the asymptotic bias and variance of $\tilde{H}(x_0)$. By the first order Taylor theorem, we have

$$0=\ell_0^{(1)}(\tilde{\omega};x_0)=\ell_0^{(1)}(\omega;x_0)+\ell_0^{(2)}(\omega^*;x_0)(\tilde{\omega}-\omega), \tag{A4}$$

for each $x_0\in[0,1]^d$, where ω^* lies in the line segment connecting ω and $\tilde{\omega}$. Using (C1)–(C4), the large deviation theorem in Section 10.3.1 of Serfling (1980), and approximations to standard errors of functions of random variables in Section 10.5 of Stuart and Ord (1987), a straightforward calculation leads to the following asymptotic results: as $n\to\infty$,

$$\Pr(P_0\cup Q_0)=O\{exp(-nb)\}, \tag{A5}$$

$$E\{\ell_0^{(1)}(\omega;x_0)\}=(1/2)nb^2A_0\{1+o(1)\}, \tag{A6}$$

$$E\{\ell_0^{(2)}(\omega;x_0)\}=(-1)nB_0\{1+o(1)\}, \tag{A7}$$

$$Var\{\ell_0^{(1)}(\omega;x_0)\}=nb^{-d}C_0\{1+o(1)\}, \tag{A8}$$

for each ω, where the little-o terms in Eqs. (A6)–(A8) all tend to zero uniformly in x_0,

$$A_0=\begin{bmatrix}\kappa_2^*S_0\\ b\kappa_3^*S_0\\ \kappa_2^*S_1\\ (-1)\kappa_2^*T_0\end{bmatrix},\ B_0=\begin{bmatrix}\kappa_0S_0, & bS_0\kappa_1^T, & \kappa_0S_1^T, & (-1)\kappa_0T_0^T\\ bS_0\kappa_1, & b^2S_0\kappa_2, & b\kappa_1S_1^T, & (-b)\kappa_1T_0^T\\ \kappa_0S_1, & bS_1\kappa_1^T, & \kappa_0S_2, & (-1)\kappa_0T_1^T\\ (-1)\kappa_0T_0, & (-b)T_0\kappa_1^T, & (-1)\kappa_0T_1, & \kappa_0T_2\end{bmatrix},$$

and C_0 is B_0 with κ_j replaced by ξ_j, for $j\ge 0$. Using the results of Eqs. (A4)–(A8) and comparing the magnitudes of $\ell_0^{(1)}(\omega;x_0)=O_p(nb^2+n^{1/2}b^{-d/2})$ and $\ell_0^{(2)}(\omega^*;x_0)=O_p(n)$ in Eq. (A4), we have

$$\tilde{\omega}-\omega=O_p(b^2+n^{-1/2}b^{-d/2}). \tag{A9}$$

By Eqs. (A4)–(A9), approximations to standard errors of functions of random variables, and Theorem A.3 of Anderson (2003), the asymptotic bias and variance of $\tilde{H}(x_0)$ can be expressed as

$$Bias\{\tilde{H}(x_0)\}=b^2b_H(x_0)\{1+o(1)\},\ Var\{\tilde{H}(x_0)\}=n^{-1}b^{-d}V_0(x_0)\{1+o(1)\}, \tag{A10}$$

for each $x_0\in[0,1]^d$. Here the little-o terms in Eq. (A10) all tend to zero uniformly in x_0, and $V_0(x_0)$ is the (1,1)th component of the matrix $B_0^{-1}C_0B_0^{-1}$. The proof of the asymptotic bias and variance of $\tilde{H}(x_0)$ is complete.

Proof of Eqs. (A1) and (A2). By the first order Taylor theorem, we have

$$0=\ell_1^{(1)}(\hat{\zeta})=\ell_1^{(1)}(\zeta)+\ell_1^{(2)}(\zeta^*)(\hat{\zeta}-\zeta), \tag{A11}$$

where ζ^* lies in the line segment connecting ζ and $\hat{\zeta}$. Using Eq. (A10), (C1)–(C4), and approximations to standard errors of functions of random variables, a straightforward calculation leads to the following asymptotic results: as $n\to\infty$,

$$E\{\ell_1^{(1)}(\zeta)\}=nb^2A_1\{1+o(1)\}, \tag{A12}$$

$$E\{\ell_1^{(2)}(\zeta)\}=(-1)nC_1\{1+o(1)\}, \tag{A13}$$

$$Var\{\ell_1^{(1)}(\zeta)\}=nC_1\{1+o(1)\}, \tag{A14}$$

for each ζ, where

$$A_1=\begin{bmatrix}(-1)\lambda_1\\ \lambda_0\end{bmatrix},\ C_1=\begin{bmatrix}\Sigma_2, & (-1)\Gamma_1^T\\ (-1)\Gamma_1, & \Gamma_2\end{bmatrix}.$$

Following the same argument of Eqs. (A4)–(A9) and using Eqs. (A11)–(A14), we have

$$\hat{\xi} - \zeta = O_p(b^2 + n^{-1/2}). \tag{A15}$$

Combining this result with Eqs. (A12)–(A14), and using approximations to standard errors of functions of random variables and Theorem A.3 of Anderson (2003), the results of Eqs. (A1) and (A2) follow.

Proof of Eq. (A3). Using (C1)–(C4), Eqs. (A1)–(A2), and approximations to standard errors of functions of random variables, a straightforward calculation leads to the following asymptotic results: as $n \to \infty$,

$$E\{\ell_2^{(1)}(\eta; x_0)\} = (1/2)ng^2 A_2\{1 + o(1)\}, \tag{A16}$$

$$E\{\ell_2^{(2)}(\eta; x_0)\} = (-1)nB_2\{1 + o(1)\}, \tag{A17}$$

$$Var\{\ell_2^{(2)}(\eta; x_0)\} = ng^{-d}C_2\{1 + o(1)\}, \tag{A18}$$

for each η, where the little-o terms in Eqs. (A16)–(A18) all tend to zero uniformly in x_0,

$$A_2 = \begin{bmatrix} \kappa_2^* S_0 \\ g\kappa_3^* S_0 \end{bmatrix}, \; B_2 = \begin{bmatrix} \kappa_0 S_0, & gS_0\kappa_1^T \\ gS_0\kappa_1, & g^2 S_0 \kappa_2 \end{bmatrix},$$

and C_2 is B_2 with κ_j replaced by ξ_j, for $j \geq 0$. Following the same argument of Eqs. (A4)–(A9) and using Eqs. (A16)–(A18), we have $\hat{\eta} - \eta = O_p(g^2 + n^{-1/2}g^{-d/2})$. Combining this result with Eqs. (A16)–(A18), and using approximations to standard errors of functions of random variables and Theorem A.3 of Anderson (2003), the result of Eq. (A3) follows.

References

Allison, P.D., 2001. Missing Data. SAGE Publications, London.

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

Altman, E.I., Katz, S., 1976. Statistical bond rating classification using financial and accounting data. In: Schiff, M., Sorter, G. (Eds.), Proceedings of the Conference on Topical Research in Accounting. New York University Press, New York, pp. 205–239.

Altman, E.I., Avery, R.B., Eisenbeis, R.A., Sinkey, J.F., 1981. Application of classification techniques in business. JAI Press, Connecticut, Banking and Finance.

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis. Wiley, New Jersey.

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, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. Review of Financial Studies 21, 1339–1369.

Blume, M.E., Lim, F., MacKinlay, A.C., 1998. The declining credit quality of U.S. corporate debt: myth or reality? Journal of Finance 53, 1389–1414.

Borooah, V.K., 2002. Logit and Probit. SAGE Publications, London.

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

Chava, S., Stefanescu, C., Turnbull, S., 2008. Modeling the loss distribution. Available at http://faculty.london.edu/cstefanescu/research.html.

Duffie, D., Eckner, A., Horel, G., Saita, L., 2006. Frailty correlated default. Available at http://www.stanford.edu/~duffie/working.htm.

Ederington, L.H., 1985. Classification models and bond ratings. The Financial Review 20, 237–262.

Eubank, R.L., 1988. Spline smoothing and nonparametric regression. Marcel Dekker, New York.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and Its Application — Theory and Methodologies. Chapman and Hall, New York.

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

Feng, D., Gourieroux, C., Jasiak, J., 2008. The ordered qualitative model for credit rating transitions. Journal of Empirical Finance 15, 111–130.

Galil, K., 2003. The quality of corporate credit rating: an empirical investigation. Available at http://ssrn.com/abstract=406681.

Gentry, J.A., Whitford, D.T., Newbold, P., 1988. Predicting industrial bond ratings with a probit model and fund flow components. The Financial Review 23, 269–286.

Greene, W.H., 2002. LIMDEP Version 8.0 Econometric Modeling Guide. Econometric Software Inc, New York.

Güttler, A., Wahrenburg, M., 2007. The adjustment of credit ratings in advance of defaults. Journal of Banking and Finance 31, 751–767.

Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Härdle, W., 1991. Smoothing Techniques: With Implementation in S. Springer, Berlin.

Härdle, W., Hart, J., Marron, J.S., Tsybakov, A.B., 1992. Bandwidth selection for average derivative estimation. Journal of the American Statistical Association 87, 218–226.

Härdle, W., Moro, R.A., Schäfer, D., 2008. Graphical data representation in bankruptcy analysis. In: Chen, C., Härdle, W., Unwin, A. (Eds.), Handbook of Data Visualization. Springer, Berlin, pp. 853–873.

Horrigan, J.O., 1966. The determination of long-term credit standing with financial ratios. Journal of Accounting Research 4, 44–62.

Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems 37, 543–558.

Hwang, R.C., Cheng, K.F., Lee, J.C., 2007. A semiparametric method for predicting bankruptcy. Journal of Forecasting 26, 317–342.

Hwang, R.C., Cheng, K.F., Lee, C.F., 2008. On multiple-class prediction of issuer credit ratings. To appear in Applied Stochastic Models in Business and Industry. Available at: http://www3.interscience.wiley.com/cgi-bin/fulltext/121392086/PDFSTART.

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

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

Kaplan, R.S., Urwitz, G., 1979. Statistical models of bond ratings: a methodological inquiry. Journal of Business 52, 231–261.

Lipsitz, S.R., Kim, K., Zhao, L., 1994. Analysis of repeated categorical data using generalized estimating equations. Statistics in Medicine 13, 1149–1163.

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

Marron, J.S., Wand, M.P., 1992. Exact mean integrated square error. Annals of Statistics 20, 712–736.

Müller, H.G., 1988. Nonparametric Regression Analysis of Longitudinal Data. Springer, Berlin.

Müller, G., Czado, C., 2005. An autoregressive ordered probit model with application to high frequency financial data. Journal of Computational and Graphical Statistics 14, 320–338.
Ohlson, J., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 18, 109–131.
Pettit, J., Fitt, C., Orlov, S., Kalsekar, A., 2004. The new world of credit ratings. Available at http://ssrn.com/abstract=593522.
Pinches, G.E., Mingo, K.A., 1973. A multivariate analysis of industrial bond ratings. Journal of Finance 28, 1–18.
Pinches, G.E., Mingo, K.A., 1975. A note on the role of subordination in determining industrial bond ratings. Journal of Finance 30, 201–206.
Pogue, T.F., Soldofsky, R.M., 1969. What's in a bond rating? Journal of Financial and Quantitative Analysis 4, 201–228.
Poon, W.P.H., 2003. Are unsolicited credit ratings biased downward? Journal of Banking and Finance 27, 593–614.
Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. Annals of Statistics 22, 1346–1370.
Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.
Serfling, R.J., 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.
Simonoff, J.S., 1996. Smoothing Methods in Statistics. Springer, New York.
Standard & Poor's, 2004. Standard & Poor's Research Insight North America Data Guide. McGraw-Hill, New York.
Staniswallis, J.G., 1989. The kernel estimate of a regression function in likelihood-based models. Journal of the American Statistical Association 84, 276–283.
Stefanescu, C., Tunaru, R., Turnbull, S., 2009. The credit rating process and estimation of transition probabilities: a Bayesian approach. Journal of Empirical Finance 16, 216–234.
Stuart, A., Ord, J.K., 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.
Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman and Hall, London.
West, R.R., 1970. An alternative approach to predicting corporate bond ratings. Journal of Accounting Research 8, 118–125.
Wijayatunga, P., Mase, S., Nakamura, M., 2006. Appraisal of companies with Bayesian networks. International Journal of Business Intelligence and Data Mining 1, 329–346.