Routledge
Taylor & Francis Group

Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: http://www.tandfonline.com/loi/rquf20

Modeling default risk with support vector machines

Shiyi Chen , W. K. Härdle & R. A. Moro

To cite this article: Shiyi Chen , W. K. Härdle & R. A. Moro (2011) Modeling default risk with support vector machines, Quantitative Finance, 11:1, 135-154, DOI: 10.1080/14697680903410015

To link to this article: https://doi.org/10.1080/14697680903410015

Published online: 20 Apr 2010.

Quantitative Finance, Vol. 11, No. 1, January 2011, 135–154

Modeling default risk with support vector machines

SHIYI CHEN*†, W. K. HÄRDLE‡ and R. A. MORO§¶

†China Center for Economic Studies (CCES), Fudan University, 220 Handan Road, 200433 Shanghai, PR China
‡Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany
§Department of Economics and Finance, Brunel University, Uxbridge, Middlesex UB8 3PH, UK
¶DIW econ, Mohrenstr. 58, 10117 Berlin, Germany

(*Received 10 January 2007; in final form 12 January 2009*)

Predicting default risk is important for firms and banks to operate successfully. There are many reasons to use nonlinear techniques for predicting bankruptcy from financial ratios. Here we propose the so-called Support Vector Machine (SVM) to predict the default risk of German firms. Our analysis is based on the Creditreform database. In all tests performed in this paper the nonlinear model classified by SVM exceeds the benchmark logit model, based on the same predictors, in terms of the performance metric, AR. The empirical evidence is in favor of the SVM for classification, especially in the linear non-separable case. The sensitivity investigation and a corresponding visualization tool reveal that the classifying ability of SVM appears to be superior over a wide range of SVM parameters. In terms of the empirical results obtained by SVM, the eight most important predictors related to bankruptcy for these German firms belong to the ratios of activity, profitability, liquidity, leverage and the percentage of incremental inventories. Some of the financial ratios selected by the SVM model are new because they have a strong nonlinear dependence on the default risk but a weak linear dependence that therefore cannot be captured by the usual linear models such as the DA and logit models.

Keywords: Statistical learning theory; Applications to default risk; Capital asset pricing; Economics of risk

1. Introduction

Predicting default probabilities and deducing the corresponding risk classification is becoming more and more important in order for firms to operate successfully and for banks to clearly grasp their clients' specific risk class. In particular, the implementation of the Basel II capital accord will further exert pressure on firms and banks. As both the risk premium and the credit costs are determined by the default risk, the firms' ratings will have a deeper economic impact on banks as well as on the firms themselves than ever before. Thus, from a risk management perspective, the choice of a correct rating model that can capture consistent predictive information concerning the probabilities of default over some successive time periods is of crucial importance.

There are strands of the literature that deal with the statistical and stochastic analysis of default risk (Burnham and Anderson 1998, Caouette *et al.* 1998, Shumway 1998, Sobehart *et al.* 2000, Saunders and Allen 2002, Gaeta 2003, Chakrabarti and Varadachari 2004, Giesecke 2004, Zagst and Hocht 2006). One models default events using accounting data, whereas other models recommend using market information. Market-based models can be further classified into structural models and reduced form models. There is also a hybrid approach that uses accounting data as well as market information to predict the probability of default. The market-based approach relies on the time series of company market data. Unfortunately, time series long enough to reliably estimate the risk is not available for most companies. Moreover, the majority of German firms are not listed and, therefore, their market price is unknown. This justifies the choice of a model for which only cross-sectional or pooled accounting data would be required. For this study, accounting data for bankrupt and operating German companies was provided by Creditreform.

Among the accounting-based models, the first attempts to identify the difference between the financial ratios of

*Corresponding author. Email: shiyichen@fudan.edu.cn

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2011 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/14697680903410015

solvent and insolvent firms were the studies of Ramser and Foster (1931), Fitzpatrick (1932), Winakor and Smith (1935) and Merwin (1942). These studies settled the fundamentals for bankruptcy prediction research. It was not until the 1960s that the traditional research was changed. Beaver (1966) pioneeringly presented the univariate approach to discriminant analysis (DA) for bankruptcy prediction. Altman (1968) expanded this analysis to multivariate analysis. Up to the 1980s, DA was the dominant method in bankruptcy prediction. However, there are obvious modeling restrictions of this approach, some of which are the assumptions of normality, homoscedasticity of the disturbances, fulfillment of conditional expectation of the dependent variable between 0 and 1, and no adjustment for multicollinearity. During the 1980s the DA method was replaced by logistic analysis, which fits the logistic regression model for binary or ordinal response data by the method of maximum likelihood estimation (MLE). In fact, the logit model uses the logistic cumulative distribution function in modeling the default probability. Among the first users of logit analysis in the context of bankruptcy were Ohlson (1980), Collins and Green (1982), Lo (1986) and Platt *et al.* (1994). The advantage of the logit model is that it does not assume multivariate normality and equal variance disturbance, and its probability lies between 0 and 1 (Härdle and Simar 2003). However, the logit model is also sensitive to the collinearity among the variables. In addition, the key assumption behind the logit model is that the logarithm of odds is linear in the underlying random variable; therefore, common to DA and logit modeling is a linear classifying hyperplane that separates insolvent and solvent firms. This works well if the data are linearly separable. A linear separating hyperplane is, however, not suitable if there is doubt that the separation mechanism is of a nonlinear kind. There are good reasons to take the linear non-separability case seriously (Falkenstein *et al.* 2000).

Many nonlinear numerical methodologies have been developed to solve the linear non-separability problem: Maximum Expected Utility (MEU), Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The MEU model was proposed at Standard & Poor's Risk Solutions Group, which allows models to incorporate the nonlinearity, non-monotonicity, and interactions present in the data, reducing the risk of overfitting. Friedman and Sandow (2003a, b) and Friedman and Huang (2003) demonstrated how the MEU method outperforms the Logit model. ANN was introduced to analyse bankrupt firms in the 1990s (see Hertz *et al.* (1991), Refenes (1995) and Härdle *et al.* (2004) for more details). This method also discards the assumption of linearity and mutual independence of explanatory variables for the default prediction function (Serrano *et al.* 1993, Back *et al.* 1994, 1996, Wilson and Sharda 1994). ANN models built using *K*-fold cross-validation techniques can be very robust and reduce over-fitting. Although the nonlinear ANN can classify a dataset much better than the linear models, it has often been criticized to be vulnerable to the multiple minima problem. Common to the OLS and MLE for linear models, ANN also makes use of the principle of minimizing empirical risk, which usually leads to a poor level of classification for out-of-sample data (Haykin 1999).

Based on statistical learning theory, an alternative nonlinear separation method, the Support Vector Machine (SVM), was recently introduced in default risk analysis. The SVM yields a single minimum without undesirable local fits as often produced by ANN. This property results from the minimized target function that is convex quadratic and linearly restricted. In addition, the SVM is also able to handle the interactions between the ratios and does not need any parameter restrictions and prior assumptions such as that concerning the distribution for latent errors. Furthermore, the biggest advantage of SVM among all the alternatives is its ability to minimize the risk associated with model misspecification, which endows SVM with an excellent separating ability. The current literature in statistical learning theory has produced strong evidence that SVM systematically outperforms standard pattern recognition/classification, function regression and data analysis techniques (Vapnik 1995, Haykin 1999). The application of SVM to company default analysis is less reported in the management science and finance literature. Härdle *et al.* (2005, 2007) report that, compared with the traditional DA and logit models in predicting the probabilities of default and rating firms, the SVM has a superior performance. Gestel *et al.* (2005) combined SVM and the logistic regression model to capture the multivariate nonlinear relations. This combination technique balances the interpretability and predictability required to rating banks.

In this study, we investigate the applicability of this new technique to predicting the risk scores and the probabilities of defaults (PDs) of German firms from the Creditreform database spanning from 1996 through 2002. The aim is to investigate (1) which of the accounting ratios are meaningful and have predictive character for bankruptcy, and (2) does a well-specified SVM-based nonlinear model consistently outperform the benchmark logit model in predicting PDs as predicted by theory?

The rest of the paper is organized as follows. In the next section we give a theoretical introduction to the Support Vector Machine (SVM) for classification. Section 3 describes the Creditreform database and the variables and ratios used in this study. In section 4, we present the validation procedures, re-sampling technique, performance measures and the ratios selection methods. Section 5 analyses the empirical results, including the predictors related to bankruptcy, the sensitivity analysis of SVM parameters, and a comparison of the predictive performance between SVM and the logit model. Section 7 offers conclusions.

2. The Support Vector Machine

The term Support Vector Machine (SVM) originates from Vapnik's statistical learning theory (Vapnik 1995, 1997), which formulates the classification problem as a quadratic programming (QP) problem. The principles on which the SVM is based, especially the regularization principle for solving ill-posed problems, are also described by Tikhonov (1963), Tikhonov and Arsenin (1977) and Vapnik (1979). The SVM transforms by nonlinear mapping the input space (of covariates) into a high-dimensional feature space and then solves a linear separable classification problem in this feature space. Thus, linear separable classification in the feature space corresponds to linearly non-separable classification in the lower-dimensional input space. As the name implies, the design of the SVM hinges on the extraction of a subset of the training data that serves as support vectors and that represents a stable characteristic of the data.

Given a training data set $\{x_i, y_i\}_{i=1}^n\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, with input vector $\mathbf{x}_i \in R^d$ (company financial ratios in this study) $x_i \in R^d$ and output scalar $y_i \in \{+1, -1\}$ $y_i = \{+1, -1\} \in R^1$ ($-1 =$ 'successful', $+1 =$ 'bankrupt'), we aim to find a classifying (score) function $f(\mathbf{x})$ to approximate the latent, unknown decision function $g(\mathbf{x})$. In the logistic and the DA case, this is simply a linear function. In the SVM case, the classifying function is

$$f(\mathbf{x}) = \sum_{l=1}^{l} w_l \phi_l(\mathbf{x}) + b = \mathbf{w}^T \phi(\mathbf{x}) + b, \tag{1}$$

where $\phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \ldots, \phi_l(\mathbf{x})]^T$ and $\mathbf{w} = [w_1, \ldots, w_l]^T$.

The nonlinear functions $\phi(\mathbf{x})$ are the transformation functions from the input space to the feature space that represent the features of the input space. A simple example of features for a quadratic function in a two-dimensional space is $\phi_1 = x_1^2$, $\phi_2 = \sqrt{2}x_1x_2$ and $\phi_3 = x_2^2$. The dimension of the feature space is l, which is directly related to the capacity of the SVM to approximate a smooth input–output mapping; the higher the dimension of the feature space, the more accurate, at the cost of variability, the approximation will be. Parameter $\mathbf{w}$ denotes a set of linear weights connecting the feature space to the output space, and b is the bias or threshold. The optimal solution $\mathbf{w}^*$ and b^* can be used to construct the optimal hyperplane $\mathbf{w}^{*T}\phi(\mathbf{x}) + b^* = 0$ and the classification function $f(\mathbf{x}) = \mathbf{w}^{*T}\phi(\mathbf{x}) + b^*$. We can predict solvent and insolvent companies using the estimated function $f(\mathbf{x})$.

2.1. *Advantage of SVM for classification in theory*

The main superiority of nonlinear non-parametric SVM over the benchmarking methods in predicting company credit risk results from its special theoretical device in two ways: (1) it takes linearly non-separable situations into account, whereas the DA and logit models only work well if the data are linear separable; and (2) it adopts the principle of structural risk minimization rather than empirical risk minimization employed by the OLS, MLE, ANN (and other) models. We illustrate the principle in figure 1 using the simplest classifying function $f(\mathbf{x}) = -x_1 - 2x_2 + 2$, where $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{w} = (-1, -2)$ and $b = 2$.

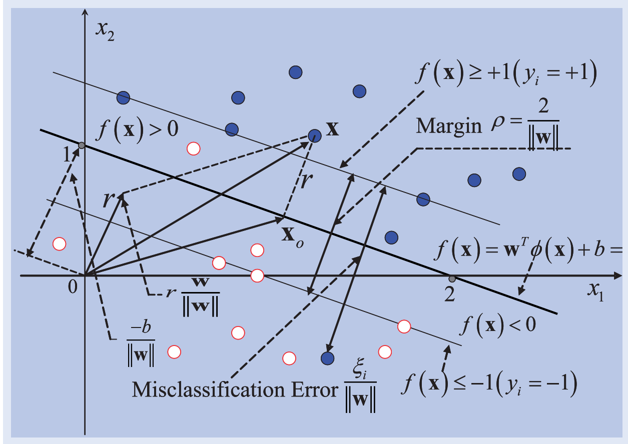

Figure 1. Separation margin, misclassification error and structural risk minimization for the SVM in two-dimensional input space.

The statistical problem is how to construct a classifying hyperplane (hypersurface) and obtain the classifying function $f(\mathbf{x})$. If the data set is linearly separable, the perfect classification hyperplane does exist. The function $f(\mathbf{x})$ gives an algebraic measure of the distance from $\mathbf{x}$ to the optimal hyperplane. Perhaps the easiest way to see this is to express $\mathbf{x}$ as $\mathbf{x} = \mathbf{x}_0 + r(\mathbf{w}/||\mathbf{w}||)$, where $\mathbf{x_0}$ is the normal projection of $\mathbf{x}$ onto the optimal hyperplane, r is the desired algebra distance from any point $\mathbf{x}$ to the optimal hyperplane (positive if $\mathbf{x}$ is on the positive side of the optimal hyperplane and negative otherwise), and $||\mathbf{w}||$ is the Euclidean norm of the weight vector $\mathbf{w}$. Since, by definition, $f(\mathbf{x}_0) = 0$, it follows that

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}_0 + \mathbf{w}^T r \frac{\mathbf{w}}{||\mathbf{w}||} + b = f(\mathbf{x}_0) + r||\mathbf{w}|| = r||\mathbf{w}||$$

or

$$r = \frac{f(\mathbf{x})}{||\mathbf{w}||}.$$

Because of the values of y_i being ± 1, the parameters $(\mathbf{w}, b)$ for the optimal hyperplane must satisfy the constraints $f(\mathbf{x}) \geq 1$ for $y_i = +1$ (insolvent) or $f(\mathbf{x}) \leq -1$ for $y_i = -1$ (solvent), that is $y_i \cdot f(\mathbf{x}) \geq 1$. The particular data points for which the constraint is satisfied with the equality sign are called *support vectors*, hence the name 'Support Vector Machine'. In conceptual terms, the support vectors are those data points that lie closest to the decision surface and are therefore the most difficult to classify. As such, they have a direct bearing on the optimum location of the classification hyperplane and play a prominent role in the operation of SVM. Now consider the support vectors; they are located on the upper and lower separation band for which $f(\mathbf{x}) = \pm 1$. Therefore, the algebraic distance from the support vectors to the optimal hyperplane is

$$r = \frac{f(\mathbf{x})}{||\mathbf{w}||} = \frac{\pm 1}{||\mathbf{w}||}.$$

Let ρ denote the optimum value of the margin of separation between solvent and insolvent companies.

Then it follows that $\rho = 2r = 2/||\mathbf{w}||$, which states that maximizing the margin of separation between classes is equivalent to minimizing the Euclidean norm of $\mathbf{w}$, $\|\mathbf{w}\|$. Thus, the classifying function in the linear separable case can be derived from maximizing the separation margin directly. Likewise, the distance from the origin to the optimal hyperplane is given by $-b/||\mathbf{w}||$, as shown in figure 1.

If the training set is linearly non-separable, the hyperplane that can correctly classify the training set no longer exists and, naturally, we need to find a hypersurface instead. For the hypersurface, however, we know less about the concept of the geometrical margin that is particular for the hyperplane; therefore, it is more difficult to find a hypersurface than a hyperplane. The transformation from the input space into higher-dimensional feature space, i.e. $\mathbf{x} \mapsto \phi(\mathbf{x})$, is then introduced in the SVM. It is possible that the new training set in the feature space $\{\phi(\mathbf{x}_i), y_i\}_{i=1}^n$ becomes linearly separable. Accordingly, the problem of finding a hypersurface in the input space is transformed into finding a hyperplane in the feature space and letting its margin or the 'safe' distance between classes, where in the perfectly separable case no observation can lie, be maximized.

It is not possible to construct a separating hyperplane without encountering classification errors. The margin of separation between classes is said to be soft if a data point violates the condition $y_i \cdot f(\mathbf{x}) \geq 1$. This violation can arise in one of two ways: (1) the data point falls inside the region of separation but on the right side of the decision surface; and (2) the data points falls on the wrong side of the decision surface. Note that we have correct classification in case (1), but misclassification in case (2). Therefore, a new set of non-negative slack variables $\{\xi_i\}_{i=1}^n$ are introduced and the condition is softened to $y_i \cdot f(\mathbf{x}) \geq 1 - \xi_i$. Note $0 < \xi_i \leq 1$ for case (1), $\xi_i \geq 1$ for case (2), and $\xi_i = 0$ for the linearly separable case. The support vectors are those particular data points that satisfy the soft condition precisely even if $\xi_i > 0$. The support vectors are thus defined in exactly the same way for both linearly separable and non-separable cases. In fact, using the soft constraints and the condition $\xi_i \geq 0$, the slack variables ξ_i can be represented as a hinge loss function which is the tightest convex upper bound of the misclassification loss and special and preferred to the loss function of the logit model because it allows a sparse solution, in the sense that some observations of the training set, if they are classified correctly, may not be necessary to construct the separating boundary. Sparseness of the solution also greatly simplifies the computation of SVM because then usually only few observations, so-called support vectors, are required to restore the solution, while for the logit regression, all observations are necessary.

The algebraic distance from the misclassification point to the optimal hyperplane is $r = [(1 - \xi_i)/\|\mathbf{w}\|]$, which can be derived making use of the same algebraic manipulation as in the linear separable case. Thus, the distance between the misclassification point and the upper band, the case in figure 1, is $\xi_i/\|\mathbf{w}\|$ and the tolerance to misclassification errors on the training set can be measured by $\sum_{i=1}^n \xi_i/||\mathbf{w}||$. Our goal is to find a separating hyperplane for which the misclassification error, averaged on the training set, is minimized, which is similar to minimize the sum of residual squares, the empirical risk in OLS and MLE estimation.

Thus, two targets exist for SVM in the linear non-separable case: still maximize the separation margin $2/||\mathbf{w}||$ and simultaneously minimize the misclassification distance $\sum_{i=1}^n \xi_i/||\mathbf{w}||$. The most intuitive form of the objective function to be minimized is

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2}||\mathbf{w}|| + C \sum_{i=1}^{n} \frac{\xi_i}{||\mathbf{w}||}. \tag{2}$$

As shown above, the second term is the margin-based loss function, which is the sum of errors measured as the distance from a misclassified observation to the hyperplane boundary, its class weighted with the parameter C. Equation (2) exhibits the so-called structural risk minimizing principle held by the SVM method. The benchmark models such as the DA and logit estimated by OLS and MLE, and simple ANN-based nonlinear models with no constraints usually employ the principle of minimizing error functions calculated on the training sample. Therefore, SVM not only minimizes the traditional empirical risk, but also maximizes the separating margin, and finally obtains a trade-off between two targets. It is this kind of special design of minimizing the structural risk that endows SVM with stronger classifying ability than the benchmark methods.

2.2. SVM algorithm

To minimize the cost function (2), an equivalent quadratic cost function, $(1/2)||\mathbf{w}||^2 + C\sum_{i=1}^n \xi_i$, can be obtained from equation (2) multiplied by $||\mathbf{w}||$ ($||\mathbf{w}|| > 0$). Thus, the primary problem of the SVM for the non-separable case is expressed as

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{n} \xi_i, \tag{3}$$

s.t.

$$y_i \times \{\mathbf{w}^T \phi(\mathbf{x}_i) + b\} + \xi_i \geq 1, \tag{4}$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, n. \tag{5}$$

As before, minimizing the first term of equation (3) is equivalent to maximizing the separation margin. The scaling factor 1/2 is included here for convenience of presentation. As for the second term, it is an upper bound on the number of misclassification errors. The formulation of the cost function in equation (3) is also therefore in perfect accord with the principle of structural risk minimization. The penalty parameter $C > 0$ is introduced to integrate the weights of two targets. It controls the trade-off between the complexity of the machine and the number of non-separable points; that is, the penalty parameter C controls the extent of penalization (or the tolerance) to misclassification errors on the training set. Partially the optimization function is derived from the problem of separating the population of defaulters from non-defaulters. However, it contains a second part responsible for margin maximization that is introduced artificially. Although it introduces a bias to the original optimization problem, it reduces the complexity of the SVM and increases its accuracy on out-of-sample data. The value of parameter C has to be selected by the user (Haykin 1999). The optimization problem for non-separable patterns stated above includes the optimization problem for linearly separable patterns as a special case. Specifically, setting $\xi_i = 0$ for all i in both equations (3) and (4) reduces them to the corresponding forms for the linearly separable case.

The corresponding dual problem of SVM for non-separable patterns can be derived using the Karush–Kuhn–Tucker conditions (Fletcher 1987, Bertsekas 1995) as follows:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^{n} \alpha_j, \tag{6}$$

s.t

$$\sum_{i=1}^{n} y_i \alpha_i = 0, \tag{7}$$

$$0 \le \alpha_i \le C, \quad i = 1, 2, \ldots, n, \tag{8}$$

where α_i and α_j are Lagrange multipliers. Note that neither the slack variables ξ_i nor their Lagrange multipliers appear in the dual problem. Thus, the objective function (6) to be minimized is the same in both the linear separable and non-separable cases. Deng and Tian (2004) demonstrate that the dual problem is easier to solve than the primal problem. We can then use the optimal solution α_i^* to obtain the solution of the primal problem:

$$\mathbf{w}^* = \sum_{i=1}^{n} y_i \alpha_i^* \phi(\mathbf{x}_i), \tag{9}$$

$$b^* = y_j - \sum_{i=1}^{n} y_i \alpha_i^* K(\mathbf{x}_i, \mathbf{x}_j), \quad \forall j \in \{j \mid 0 < \alpha_j^* < C\}. \tag{10}$$

By substitution, the nonlinear classifying (score) function can be obtained:

$$\begin{aligned} f(\mathbf{x}_j) &= \mathbf{w}^{*T} \phi(\mathbf{x}_j) + b^* = \sum_{i=1}^{n} y_i \alpha_i^* \phi^T(\mathbf{x}_i) \phi(\mathbf{x}_j) + b^* \\ &= \sum_{i=1}^{n} y_i \alpha_i^* K(\mathbf{x}_i, \mathbf{x}_j) + b^*, \end{aligned} \tag{11}$$

where $K(\mathbf{x}_i, \mathbf{x}_j) = \phi^T(\mathbf{x}_i)\phi(\mathbf{x}_j)$ is the inner product kernel function in which $\mathbf{x}_i$ belongs to the training set and $\mathbf{x}_j$ is the new company financial ratio, either in the training set or validating and forecasting set. For the classification problem, the decision function (11) is constructed to help us deduce in what kind of category, say $+1$ or -1, the new output $f(\mathbf{x}_j)$ corresponding to $\mathbf{x}_j$ is located. To the end, the intuitive way is to compare $\mathbf{x}_j$ with $\mathbf{x}_i$ pairwise; if $\mathbf{x}_j$ is closer to $\mathbf{x}_i$ on the positive side, then the new output $f(\mathbf{x}_j)$ nears $+1$, if $\mathbf{x}_j$ is closer to $\mathbf{x}_i$ on the negative side $f(\mathbf{x}_j)$ falls into the category -1. This is reasonable because a similar input should lead to the same output. Therefore, the decision function only depends on the proximity between two observations and the classification is in fact a proximity problem. In SVM, the inner product kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ is the key tool to measure this kind of proximity. In addition, the SVM theory considers the form of $K(\mathbf{x}_i, \mathbf{x}_j)$ in the Hilbert space without specifying $\phi(\cdot)$ explicitly and without computing all corresponding inner products, which provides the flexibility of the high-dimensional Hilbert space for low computational costs and greatly reduces the computational complexity. Thus, the kernel becomes the crucial part of SVM.

It is necessary to find an appropriate kernel in order to solve the optimization problem of SVM. The requirement on the kernel function is to satisfy Mercer's theorem (Mercer 1908, Courant and Hilbert 1970), such that the Kernel matrix, $\{K(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^{n}$, is symmetric and semi-positive definite. Mercer's theorem tells us whether or not a candidate kernel is actually an inner-product kernel in some space and therefore admissible for use in a support vector machine. Within this requirement there is some freedom in how it is chosen. The usual chosen kernels are linear, polynomial and Gaussian kernel functions. A different kernel requires estimating the extent of proximity based on a different metric criterion. In this study, we choose an anisotropic Gaussian kernel for the SVM:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-(\mathbf{x}_i - \mathbf{x}_j)^T r^{-2} \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j)/2), \tag{12}$$

where Σ is the variance–covariance matrix of the data and r is the Gaussian, also known as the radial basis kernel coefficient which implicitly controls the complexity of the feature space and the solution—the larger r, the less the complexity. Therefore, based on expression (11), for any new company $\mathbf{x}_j$, those companies from the training sample $\mathbf{x}_i$ will have a greater impact on $f(\mathbf{x}_j)$ if $\mathbf{x}_j$ are closer to $\mathbf{x}_i$. The anisotropic Gaussian kernel offers a way of measuring the proximity between two companies; it is higher when the companies are close and smaller when they are far from each other.

3. Data and financial ratios

3.1. Data description

The data used in this study is the Creditreform database. It contains a random sample of 20,000 solvent and 1000 insolvent firms in Germany and spans the period from 1996 to 2002, although the data are concentrated in 2001 and 2002 with approximately 50% of the observations coming from this period. Most firms appear in the database several times in different years. Each firm is described by a set of financial statement variables such as those in balance sheets and

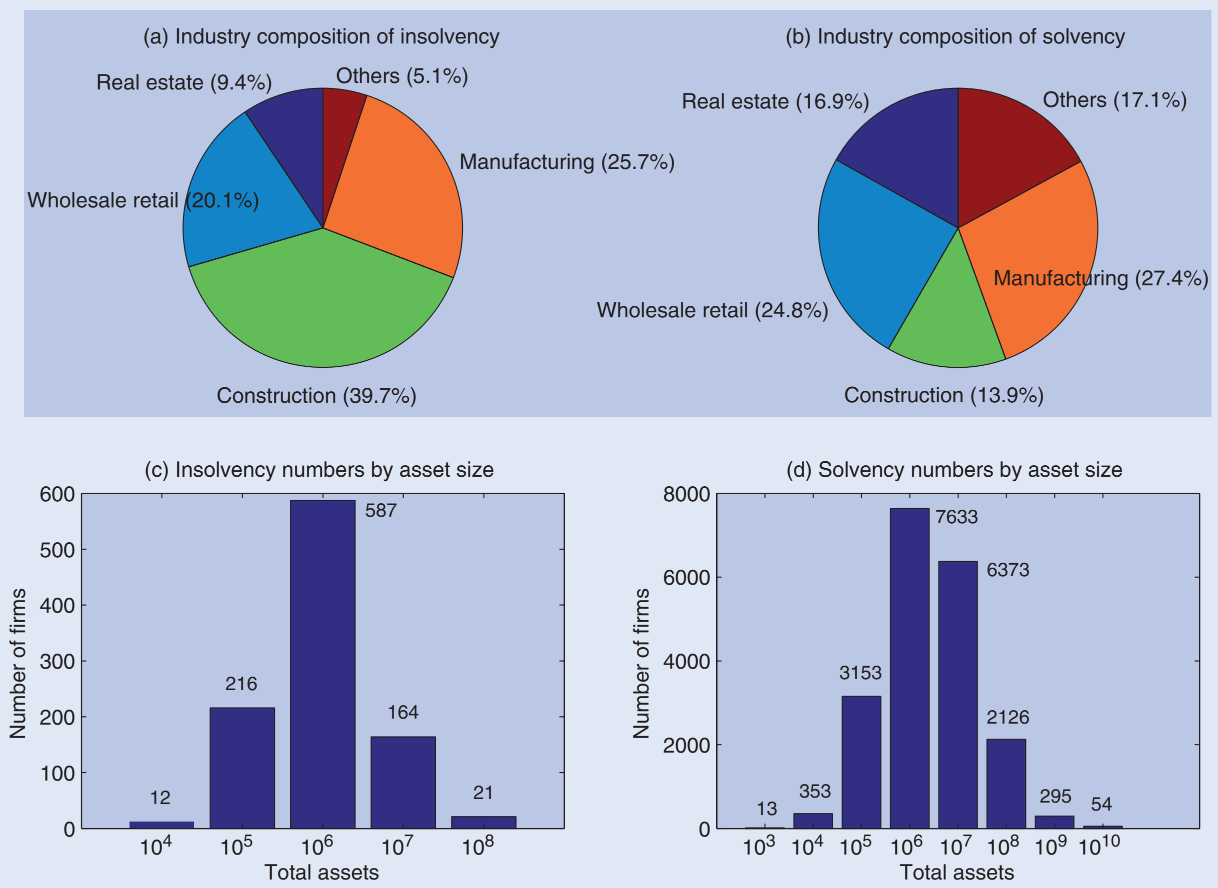

Figure 2. Industry composition and size distribution of the companies in the Creditreform database.

income statements. The data for the insolvent firms were collected two years prior to insolvency.

Figure 2 shows the industry composition and size distribution of the database. The industries to which each firm belongs can be systematically classified according to an internationally recognized system—Classification of Economic Activities, Edition 1993 (WZ 93)—published by the German Federal Statistical Office. WZ 93 uses a hierarchy of five different levels. The higher the level, the more precise the description of the main activity. In terms of the classification industry codes of WZ 93, as shown in figures 2(a) and (b), the 1000 insolvent firms consist of about 39.7% construction, 25.7% manufacturing, 20.1% the wholesale and retail trade, 9.4% real estate and 5.1% others. The others among the 1000 insolvent firms include agriculture, mining, electricity, gas and water supply, hotels and restaurants, transport and communication, financial intermediation and social service activities. The industries of the 20,000 solvent firms are manufacturing (27.4%), wholesale and retail trade (24.8%), real estate (16.9%), construction (13.9%) and others (17.1%). Different from the 'others' of insolvent firms, the others in solvency contain additional industries such as publishing, administration and defense, education and health.

The distribution of total assets can be regarded as being representative of the distribution of the firm size. In figures 2(c) and (d), the 1000 insolvent sample comprises 12 firms located in the size category 10^4 EUR, 216 in 10^5 EUR, 587 in 10^6 EUR, 164 in 10^7 EUR and 21 in 10^8 EUR. (Here, 10^4 EUR represents one category of asset size in which the firms have total assets of between 10,000 and 99,999 EUR. The definition of the other size categories is similar to that for 10^4 EUR.) The number of firms corresponding to each asset size category of the 20,000 solvent firms is 13 (10^3 EUR and below), 353 (10^4 EUR), 3153 (10^5 EUR), 7633 (10^6 EUR), 6373 (10^7 EUR), 2126 (10^8 EUR), 295 (10^9 EUR) and 54 (10^{10} EUR and above).

In an attempt to obtain a more homogeneous company sample, we cleaned the database of companies whose characteristics are very different from the others. That is to say, we do not attempt to cover all firms in the database for our study because of the very different nature of some firms. Thus, in focusing on predicting the PDs of German firms we eliminated the following types of firms from the whole sample.

- *Firms with a small percentage composition of industry*—that is, we eliminate the firms that belong to the 'other' industries in the insolvent and solvent databases, for example financial intermediation and public institutions. Thus only four main types of industry (Construction, Manufacturing, Wholesale & Retail Trade and Real Estate) remain in the study.
- *Smallest and largest firms*—that is, we exclude those firms that, because of their asset size,

Table 1. Variables used in the study.

Abbreviation	Variable	Abbreviation	Variable
CASH	Cash and cash equivalents	DEBT	Debt
INV	Inventories	AP	Accounts payable
CA	Current assets	SALE	Sales
ITGA	Intangible assets	AD	Amortization and depreciation
TA	Total assets	INTE	Interest expense
QA	Quick assets (=CA-INV)	EBIT	Earnings before interest and tax
AR	Accounts receivable	OI	Operating income
LB	Lands and buildings	NI	Net income
OF	Own funds	IDINV	Increase (decrease) inventories
CL	Current liabilities	IDL	Increase (decrease) liabilities
TL	Total liabilities	IDCASH	Increase (decrease) cash
WC	Working capital (=CA-CL)		

are not located in the categories 10^5, 10^6 and 10^7 EUR. As Khandani *et al.* (2001) noted, the credit quality of the smallest firms is often as dependent on the finances of a key individual as on the firm itself; the number of largest firms that go bankrupt is usually very small in Germany.

We further clean the database to ensure that the value of some variables, such as the denominator when calculating the ratios, should not be zero. We also exclude the firms solvent in 1996 because of missing insolvency values for this year.

Thus, 783 insolvent firms and 9583 solvent firms were chosen and analysed. The bankrupt firms are paired with non-bankrupt firms with a similar industry and total asset size. Correspondingly, the predicted default probabilities and rating results in this study are only suitable for German firms from four main industry sectors (Construction, Manufacturing, Wholesale & Retail Trade and Real Estate) and with medium asset size (lying within the categories 10^5, 10^6, and 10^7 EUR).

3.2. Ratio definitions

The Creditreform database provides many financial statement variables for each firm. In accordance with the existing literature, 28 ratios were selected for the bankruptcy analysis. In summary, there are 28 financial ratios (including one size variable) and a binary response, which records whether the firm went bankrupt within two years of the financial statements or not. There is also information on the industry distribution and on the year of the accounts. There are no missing values. These ratios can be grouped into the following six broad categories (factors): profitability, leverage, liquidity, activity, firm size and the percentage change for some variables. The variables applied to calculate these ratios are shown in table 1. Table 2 describes these ratios and how they were calculated. For simplicity, we provide short names for some ratios that capture the essence of what they measure. Table 3 summarizes the descriptive statistics of the 28 ratios for both the insolvency and solvency sample.

In previous studies, profitability ratios have appeared to be strong predictors related to bankruptcy. In addition, among all the potential risk factors, there are more profitability ratios than any other factor. The profitability ratios employed in our study are return on assets (ROA, NI/TA), net profit margin (NI/SALE), OI/TA, operating profit margin (OI/SALE), EBIT/TA, EBITDA and EBIT/SALE, denoted respectively as x1, x2, x3, x4, x5, x6 and x7.

The ROA figure gives investors an idea of how effectively the firm is deploying its assets to generate income. The higher the ROA number, the better, because the firm is earning more money on less investment. Net profit margin measures how much of every dollar of sales a firm actually keeps in earnings. A higher profit margin indicates a more profitable firm that has better control over its costs compared with its competitors. Some investors add extraordinary items back into net income when performing this calculation because they would like to use operating returns on assets, which represent a firm's true operating performance. Operating income is also required to calculate operating profit margin, which describes a firm's operating efficiency and pricing strategy. EBIT is all profits before taking into account interest payments and income taxes. An important factor contributing to the widespread use of EBIT is the way in which it nullifies the effects of different capital structures and tax rates used by different firms. By excluding both taxes and interest expenses the figure homes in on the firm's ability to profit and thus makes for easier cross-firm comparisons. EBIT is the precursor to EBITDA, which takes the process further by removing two non-cash items from the equation (depreciation and amortization). Thus, defaulting firms usually have lower profitability values; however, firms with extremely large and volatile profitability may also be likely to translate into higher default probabilities. We will try to capture this kind of complex nonlinear dependence in our database.

Leverage is also a key measure of firm risk. In this study, seven leverage ratios are analysed. They are simple and adjusted own funds ratio, CL/TA, net indebtedness, TL/TA, debt ratio (DEBT/TA) and interest coverage ratio (EBIT/INTE), represented by x8 through x14.

The own funds ratio measures the ratio of a firm's internal capital to its assets. The simple version is widely used in credit models, which is basically the mirror image

Table 2. Definitions of accounting ratios.

Ratio No.	Definition	Ratio	Category
x1	NI/TA	Return on assets (ROA)	Profitability
x2	NI/SALE	Net profit margin	Profitability
x3	OI/TA		Profitability
x4	OI/SALE	Operating profit margin	Profitability
x5	EBIT/TA		Profitability
x6	(EBIT + AD)/TA	EBITDA	Profitability
x7	EBIT/SALE		Profitability
x8	OF/TA	Own funds ratio (simple)	Leverage
x9	(OF-ITGA)/(TA-ITGA-CASH-LB)	Own funds ratio (adjusted)	Leverage
x10	CL/TA		Leverage
x11	(CL-CASH)/TA	Net indebtedness	Leverage
x12	TL/TA		Leverage
x13	DEBT/TA	Debt ratio	Leverage
x14	EBIT/INTE	Interest coverage ratio	Leverage
x15	CASH/TA		Liquidity
x16	CASH/CL	Cash ratio	Liquidity
x17	QA/CL	Quick ratio	Liquidity
x18	CA/CL	Current ratio	Liquidity
x19	WC/TA		Liquidity
x20	CL/TL		Liquidity
x21	TA/SALE	Asset turnover	Activity
x22	INV/SALE	Inventory turnover	Activity
x23	AR/SALE	Account receivable turnover	Activity
x24	AP/SALE	Account payable turnover	Activity
x25	Log(TA)		Size
x26	IDINV/INV	Percentage of incremental inventories	Percentage
x27	IDL/TL	Percentage of incremental Liabilities	Percentage
x28	IDCASH/CASH	Percentage of incremental cash flow	Percentage

Table 3. Descriptive statistics of the 28 accounting ratios. IQR is the interquartile range.

	Insolvent				Solvent			
Ratio	q0.05	Med.	q0.95	IQR	q0.05	Med.	q0.95	IQR
NI/TA	−0.19	0.00	0.09	0.04	−0.09	0.02	0.19	0.06
NI/SALE	−0.15	0.00	0.06	0.03	−0.07	0.01	0.10	0.03
OI/TA	−0.22	0.00	0.10	0.06	−0.11	0.03	0.27	0.09
OI/SALE	−0.16	0.00	0.07	0.04	−0.08	0.02	0.13	0.04
EBIT/TA	−0.19	0.02	0.13	0.07	−0.09	0.05	0.27	0.09
EBITDA	−0.13	0.07	0.21	0.08	−0.04	0.11	0.35	0.12
EBIT/SALE	−0.14	0.01	0.10	0.04	−0.07	0.02	0.14	0.05
OF/TA	0.00	0.05	0.40	0.13	0.00	0.14	0.60	0.23
(OF-ITGA) / (TA-ITGA-CASH-LB)	−0.01	0.05	0.56	0.17	0.00	0.16	0.95	0.32
CL/TA	0.18	0.52	0.91	0.36	0.09	0.42	0.88	0.39
(CL-CASH)/TA	0.12	0.49	0.89	0.36	−0.05	0.36	0.83	0.41
TL/TA	0.29	0.76	0.98	0.35	0.16	0.65	0.96	0.40
DEBT/TA	0.00	0.21	0.61	0.29	0.00	0.15	0.59	0.31
EBIT/INTE	−7.90	1.05	7.20	2.47	−6.78	2.16	73.95	5.69
CASH/TA	0.00	0.02	0.16	0.05	0.00	0.03	0.32	0.10
CASH/CL	0.00	0.03	0.43	0.11	0.00	0.08	1.40	0.29
QA/CL	0.18	0.68	1.90	0.54	0.25	0.94	4.55	1.00
CA/CL	0.56	1.26	3.73	0.84	0.64	1.58	7.15	1.56
WC/TA	−0.32	0.15	0.63	0.36	−0.22	0.25	0.73	0.41
CL/TL	0.34	0.84	1.00	0.37	0.22	0.85	1.00	0.44
SALE/TA	0.43	1.63	4.15	1.41	0.50	2.08	6.19	1.76
INV/SALE	0.02	0.16	0.89	0.26	0.01	0.11	0.56	0.16
AR/SALE	0.02	0.12	0.33	0.11	0.00	0.09	0.25	0.09
AP/SALE	0.03	0.14	0.36	0.10	0.01	0.07	0.24	0.08
Log(TA)	13.01	14.87	17.16	1.69	12.82	15.41	17.95	2.37
IDINV/INV	−1.20	0.00	0.75	0.34	−0.81	0.00	0.56	0.07
IDL/TL	−0.44	0.00	0.48	0.15	−0.53	0.00	0.94	0.14
IDCASH/CASH	−12.71	0.00	0.94	0.79	−7.13	0.00	0.91	0.52

of TL/TA, as expected: they are mathematical complements. We have made some adjustments to the simple own funds ratio to counter creative accounting practices, and to try to generate a better measure of firm credit strength. The adjustments are also used by Khandani *et al.* (2001). Net indebtedness measures the level of short-term liabilities not covered by the firm's most liquid assets as a proportion of its total assets. Thus, in addition to measuring the short-term leverage of a firm, it also provides a measure of the liquidity of a firm. While the debt ratio performs about as well as TL/TA for public firms, it does considerably worse for private firms, which makes TL/TA preferred. The difference between debt and liabilities is that liabilities is a more inclusive term that includes debt, deferred taxes, minority interest, accounts payable, and other liabilities. The interest coverage ratio is highly predictive. Falkenstein *et al.* (2000) argue that the interest coverage ratio turns out to be one of the most valuable explanatory variables in the public firm dataset in a multivariate context, although in the private firm database its relative power decreases significantly.

Six liquidity ratios, CASH/TA, cash ratio, quick ratio, current ratio, WC/TA and CL/TA (x15 through x20), are analysed in this paper. Liquidity is a common variable in most credit decisions and represents the ability to convert an asset into cash quickly. In the private dataset, CASH/TA is the most important single variable relative to default. Quick ratio is an indicator of a firm's short-term liquidity and measures a firm's ability to meet its short-term obligations with its most liquid assets. The larger the quick ratio, the better the position of the firm. The quick ratio is more conservative than the current ratio because it excludes inventory from current assets. Current ratio is mainly used to give an idea of the firm's ability to pay back its short-term liabilities (debt and payables) with its short-term assets (cash, inventory, receivables). If a firm is in default, its current ratio must be low. Yet, just as the cash in your wallet does not necessarily imply wealth, a high current ratio does not necessarily imply health. Working capital measures both a firm's efficiency and its short-term financial health. Altman (1968) reported that the WC/TA ratio is a measure of the net liquid assets of the firm relative to the total capitalization and proved to be more valuable than the current ratio and the quick ratio. Falkenstein *et al.* (2000) showed that, firstly, the CL/TL ratio appears of little use in forecasting, second that the quick ratio appears slightly more powerful than the WC/TA ratio, and third, the quick ratio and current ratio carry roughly similar information.

Activity ratios also capture important bankruptcy information and are frequently used when performing fundamental analysis for different firms. We analyse four different activity ratios: the asset turnover (TA/SALE, x21), the inventory turnover (INV/SALE, x22), the account receivable and payable turnover (AR/SALE, x23; AP/SALE, x24).

The asset turnover ratio is a standard financial ratio illustrating the sales-generating ability of the firm's assets. Usually, the asset turnover is non-monotonic and very flat. Note that some studies report that the asset turnover degrades model predictability, for example the Z-score that reduces the asset turnover performs better than the one that keeps it. The reciprocal of the inventory turnover shows how many times a firm's inventory is sold and replaced over a period. A high turnover implies poor sales and, therefore, excess inventory. High inventory levels are unhealthy because they represent an investment with a rate of return of zero. Accounts payable and receivable turnover ratios are more powerful predictors, the reciprocals of which also display how many times the firm's accounts are converted into sales over a period. The former is a short-term liquidity measure used to quantify the rate at which a firm pays off its suppliers. The latter is a measure used to quantify a firm's effectiveness in extending credit as well as collecting debts. By maintaining accounts receivable, firms are indirectly extending interest-free loans to their clients. The above description of the activity ratios is usually true in the manufacturing industry but is not the case for other industries. For instance, service firms may have no inventory to turn over.

Sales or total assets are almost indistinguishable as indicators of size risk, which makes the choice between the two measures arbitrary. In this study, we use the natural logarithm of total assets (log(TA), x25) to represent the firm size to investigate the default risk of small, medium (SMEs) and large firms. For example, access to capital for these firms is very different and may affect the prediction ability of some financial ratios and, consequently, the performance of the SVM model. Due to the available variables provided by the Creditreform database, we also compute three ratios of the percentage of incremental inventories, liabilities and cash flow (x26, x27, x28), respectively. For example, the increased (decreased) cash flow is the additional operating cash flow that an organization receives from taking on a new project. A positive incremental cash flow means that the firm's cash flow will increase with the acceptance of the project, the ratio of which is a good indication that an organization should spend some time and money investing in the project.

Previous empirical research has found that a firm is more likely to go bankrupt if it is unprofitable, highly leveraged, and suffers cashflow difficulties (Myers 1977, Aghion and Bolton 1992, Lennox 1999). Moreover, large firms are less likely to encounter credit constraints because of reputation effects. This is clearly demonstrated by the statistical description of financial ratios in table 3, which shows that insolvent firms are typically small, have poor profitability and liquidity, and are highly leveraged, compared with solvent firms, with only a few exceptions such as EBIT/SALE, OF/TA and EBIT/INTE. In addition, the firms that go on to default have higher values for the activity ratio. Except for the last three, all ratios for insolvent firms vary less than for solvent firms because of the smaller number of observations.

The statistics described in table 3 reveal that several of the ratios are highly skewed and there are many outliers; this may affect whether they can be of much help in identifying insolvent and solvent firms. It is also possible that many of these outliers are errors of some kind. Therefore, the ratios used in the following analysis are processed as follows: if $x_i < q_{0.05}(x_i)$, then $x_i = q_{0.05}(x_i)$, and if $x_i > q_{0.95}(x_i)$, then $x_i = q_{0.95}(x_i)$, $i = 1, 2, \ldots, 28$. $q_\alpha(x_i)$ is an α quantile of x_i. Thus, the discriminating results obtained from both the SVM and the logit model are robust and not sensitive to outliers.

4. Prediction framework

4.1. The validation procedure

To compare the SVM and the logit models in a setting most close to the real situation in which these models are used in practice, the holdout method is chosen in this study for cross validation, namely training of the model on all available data up to the present period and the forecasting of default events for the next period. In this study, the training data are chosen from 1997 through 1999, and the validating set are selected from 2000 through 2002. Then the model is first estimated using the training data; once the model form and parameters are established, the model is used to identify insolvencies among all the firms available during the holdout period (2000–2002). Note that the predicted outputs for 2000 through 2002 are out of time for firms existing in the previous three years, and out of sample for all the firms whose data become available only after 2000. Such out-of-sample and out-of-time tests are the most appropriate way to compare model performance. The validation result set is the collection of all the out-of-sample and out-of-time model predictions that can then be used to analyse the performance of the model in more detail. For an introduction to the validation framework, see Sobehart *et al.* (2001).

Following the holdout validation procedure, we construct a training set containing 387 insolvent and 3534 solvent companies and a validation set containing 396 default events and 6049 non-defaulters. Note that the training and validation sets are themselves a subsample of the population and, therefore, may yield spurious model performance differences based only on data anomalies. A common approach to overcome this problem is to use the re-sampling techniques to leverage the available data and reduce the dependency on the particular sample at hand (Efron and Tibshirani 1993, Herrity *et al.* 1999, Horowitz 2001). Re-sampling approaches provide two related benefits (Sobehart *et al.* 2001). First, they give an estimate of the variability around the actual reported model performance. This variability can be used to determine whether differences in model performance are statistically significant, using familiar statistical tests. Second, because of the low numbers of defaults, re-sampling approaches decrease the likelihood that individual defaults (or non-defaults) will overly influence the chances of a particular model being ranked higher or lower than another model. Similar to previous bankruptcy studies, this paper also adopts a matched pairs approach for drawing subsamples for both the training and validation set. The advantage of the matching procedure is that it helps to cut the cost of data collection, as the proportion of insolvent firms in the population is very small. The problem that the use of relatively small samples could lead to over-fitting can be avoided by the re-sample techniques.

The re-sampling technique employed in this analysis is the bootstrap, which proceeds as follows. We use all insolvent firms, 387 in the training set and 396 in the validation set, and randomly select a subsample with the same number of solvencies from the 3534 solvencies in the training set and the 6049 solvencies in the validation set, respectively.

For the selected validation subset the performance measure is calculated and recorded. Then we perform a Monte Carlo experiment: another subsample is drawn, and the process is repeated. This continues for many repetitions until a distribution for each performance measure is established. In this paper the process will be repeated 30 times.

4.2. Performance measures

We now introduce two metrics for measuring and comparing the performance of credit risk models: the Accuracy Ratio (AR) and the misclassification error. These two measures aim to determine the power of discrimination that a model exhibits in warning of default risk. These techniques are quite general and can be used to compare different types of models even when the model outputs differ and are difficult to compare directly.

AR is a valuable and simple tool to determine the discriminative power of risk models. AR can be derived from the Cumulative Accuracy Profile (CAP) curve, which is particularly useful in that it simultaneously measures Type I and Type II errors (Herrity *et al.* 1999, Engelmann *et al.* 2003, Basle Committee on Banking Supervision 2005). In statistical terms, the CAP curve represents the cumulative probability distribution of default events for different percentiles of the risk score scale. To obtain CAP curves, firms are first ordered by their risk scores. For a given fraction x% of the total number of firms, a CAP curve is constructed by calculating the percentage $y(x)$ of the defaulters whose risk score is equal to or smaller than that for fraction x. In other words, for a given x, $y(x)$ measures the fraction of defaulters (of the total defaulters) whose risk scores are equal to or smaller than those of fraction x (of the total firms). One would expect a concentration of non-defaulters at the highest scores and defaulters at the lowest scores.

Figure 3 shows a CAP plot. The random CAP represents the case of zero information (which is equivalent to a random assignment of scores). The ideal CAP represents the case in which the model is able to discriminate perfectly, and all defaults are caught at the lowest model output. The actual CAP shows the performance of the model being evaluated. It depicts the percentage of defaults captured by the model.

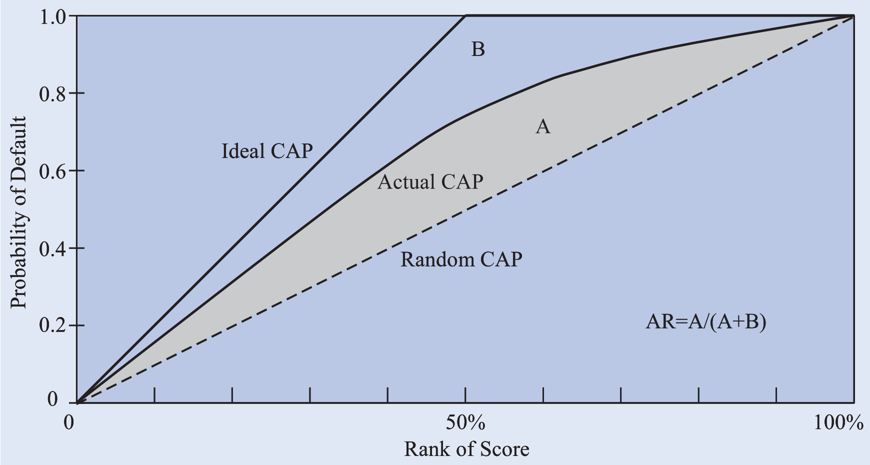

Figure 3. Cumulative accuracy profile (CAP) curve.

Therefore, AR is defined as the ratio of the area between a model's CAP curve and the random CAP curve to the area between the perfect CAP curve and the random CAP curve (see figure 3). The AR value is a fraction between zero and one. Risk measures with AR that approach zero have little advantage over a random assignment of risk scores, whereas those close to one display good predictive power. Mathematically, the AR value is defined as

$$AR = \frac{\int_0^1 y(x)\,\mathrm{d}x - (1/2)}{\int_0^1 y_{\text{ideal}}(x)\,\mathrm{d}x - (1/2)}. \tag{13}$$

If the number of bankruptcies equals the number of operating companies in the sample, then the AR becomes

$$AR \approx 2\int_0^1 y(x)\,\mathrm{d}x - 1. \tag{14}$$

In addition, when evaluating the explanatory power of the bankruptcy models, it is helpful to define two types of prediction error: a type I error, which indicates low default risk when in fact the risk is high, and a type II error, which conversely indicates a high default risk when in fact the risk is low. Usually, minimizing one type of error comes at the expense of increasing the other type of error. Clearly, the type I and type II error rates depend on the number of firms predicted to fail. The higher (lower) the number of firms predicted to go bankrupt, the smaller (larger) is the type I error rate and the larger (smaller) is the type II error rate. The number of predicted bankruptcies depends on the cut-off probability, which is equal to 0.5 in our study. From a supervisory viewpoint, type I errors are more problematic as they produce higher costs. Usually, the cost of a default is higher than the loss of prospective profits. Altman *et al*. (1977) estimated the relative costs of type I and type II errors for commercial bank loans as being 7:1. Sobehart *et al*. (2001) also described the cost scenarios schematically.

For more details on the performance measures, we refer to DeLong *et al*. (1988), Swets (1998), Keenan and Sobehart (1999), Swets *et al*. (2000), Sobehart *et al*. (2001) and Sobehart and Keenan (2004).

4.3. Predictor selection

In this study, the benchmark linear parametric probability model is the conditional logit model estimated by MLE, which is described as follows:

$$\Pr(y_i = 1|\mathbf{x}_{i1},\ldots,\mathbf{x}_{id}) = \frac{1}{1+\mathrm{e}^{-(\beta_0+\beta_1\mathbf{x}_{i1}+\cdots+\beta_d\mathbf{x}_{id}+\varepsilon_i)}}. \tag{15}$$

Based on equation (1) or (11), the target nonlinear non-parametric probability model estimated by the SVM can also be expressed in the following form:

$$\Pr(y_i = 1|\mathbf{x}_{i1},\ldots,\mathbf{x}_{id}) = f(\mathbf{x}_{i1},\ldots,\mathbf{x}_{id}) + \varepsilon_i, \tag{16}$$

where $y_i = 1$ indicates the bankrupt company, and $y_i = 0$ for the logit case and $y_i = -1$ for SVM represent the successful firm; the input vectors $\mathbf{x}_i$ are the relevant company financial ratios explaining the probability of bankrupcy. Before we begin to estimate the models, the process of predictor selection is illustrated.

For a parametric model we can estimate the distribution of the coefficients of the predictors and their confidence intervals. However, we cannot do so for non-parametric models. Instead, we can use the bootstrap technique, as described in the subsection on the validation procedure, to empirically estimate the distribution of the AR on many subsamples. In this study we randomly select 30 subsamples and compute the corresponding ARs 30 times. The median AR provides a robust measure to compare different ratios as predictors.

There are so many possible financial ratios that can be used as explanatory variables in credit scoring models that selection criteria are needed to obtain a parsimonious model. There are two main methods for selecting the appropriate ratios (Falkenstein *et al*. 2000). The first is forward stepwise selection. Start with the predictor that has the highest performance accuracy and then sequentially add the next predictor that also has the highest accuracy in the group and higher than the former until additional predictors have no additional improvement.

Table 4. Median of the AR measure for a univariate SVM model. Accounts payable turnover (AP/SALE, x24) produces the highest AR median.

No.	Ratio	AR median	No.	Ratio	AR median
x1	NI/TA	28.428	x15	CASH/TA	22.140
x2	NI/SALE	22.985	x16	CASH/CL	25.821
x3	OI/TA	36.358	x17	QA/CL	28.746
x4	OI/SALE	31.413	x18	CA/CL	16.983
x5	EBIT/TA	29.941	x19	WC/TA	14.264
x6	EBITDA	29.155	x20	CL/TL	−7.608
x7	EBIT/SALE	19.447	x21	SALE/TA	17.414
x8	OF/TA	32.941	x22	INV/SALE	24.764
x9	(OF-ITGA) / (TA-ITGA-CASH-LB)	31.938	x23	AR/SALE	17.468
x10	CL/TA	18.020	**x24**	**AP/SALE**	**49.174**
x11	(CL-CASH)/TA	23.319	x25	Log(TA)	23.816
x12	TL/TA	22.477	x26	IDINV/INV	15.493
x13	DEBT/TA	16.528	x27	IDL/TL	−9.528
x14	EBIT/INTE	28.270	x28	IDCASH/CASH	−6.562

The second is backward elimination in which one starts with all predictors, then reduces all of the poor variables. In this study, forward selection is preferred for the SVM method due to its relatively lower computational cost. The logit model, with forward selection, together with the investigation of the statistical significance and correct sign of the individual parameters of the predictors, is likely to choose different explanatory variables than the SVM. To compute and compare each method more conveniently, we will only report the results of the logit model with the same predictors as the SVM-based model. The discriminating power of each ratio is assessed using the median of the AR performance measures.

5. Empirical results

This section discusses the empirical results for each stage of the analysis of the German bankruptcy data using an SVM model. The prediction horizon in each case is two years, i.e. the data were recorded two years prior to bankruptcy for the companies that would become bankrupt. The balance sheet and income statement data for 20,000 solvent and 1000 insolvent firms in Germany were selected randomly by Creditreform. These data are represented as the financial ratios listed in table 4. They cover the period from 1996 to 2002. Each company may appear several times in different years.

5.1. *Selection of the first predictor and the sensitivity of the SVM parameters*

The first stage of analysing default risk is the selection of the first best predictor related to bankruptcy among the 28 ratios using the median of the AR metric in which the SVM model has one input. It is often argued that the SVM lacks interpretability of the results as is the case for the logit model. Most importantly, since there are no distributional assumptions underlying the SVM modeling, it is impossible to test the significance of variables within the SVM framework. Therefore, we will identify the most significant variable in an additional procedure before analysing the SVM model.

Based on table 4 we can see that Accounts Payable Turnover (AP/SALE, x24) provides the highest median AR of 49.17%. We can also see that CL/TL (x20), IDL/TL (x27) and IDCASH/CASH (x28) have a very low accuracy: their median AR values are below zero. For the next step we will select Accounts Payable Turnover (x24) as the first best single predictor related to German default firms, which is somewhat different from previous studies in which it was usually argued that the most significant predictors were profitability or leverage ratios. In fact, the SVM-based nonlinear model is able to search the nonlinear dependence of the data automatically as opposed to the logit model and it is Accounts Payable Turnover selected by SVM as the first predictor that greatly improves the classifying performance of SVM by more than 10%. Using most of the other ratios as the first predictor, the SVM-based model does not exceed the logit model by much in modeling the default risk.

The accounts payable turnover ratio is calculated by taking the average accounts payable and dividing it by the total sales during the same period. Its reciprocal shows investors how many times per period the firm pays its average payable amount. If the turnover ratio increases from one period to another, this is a sign that it takes the firm longer to pay off its suppliers than before. The opposite is true when the turnover ratio is falling, which means that the firm is paying off suppliers at a faster rate. Therefore, the firms with higher accounts payable turnover values will have less ability to convert their accounts into sales, have lower revenues, and go bankrupt more readily.

The SVM model has two control parameters, the influence of which was investigated in this study: the penalty parameter C and the Gaussian kernel coefficient r. C controls the tolerance to misclassification errors on the training set, while r represents the complexity of classifying functions. The possibility of fine-tuning SVM using these parameters, besides the flexibility of its classification function, further contributed to the higher performance of the SVM compared with the logit model,

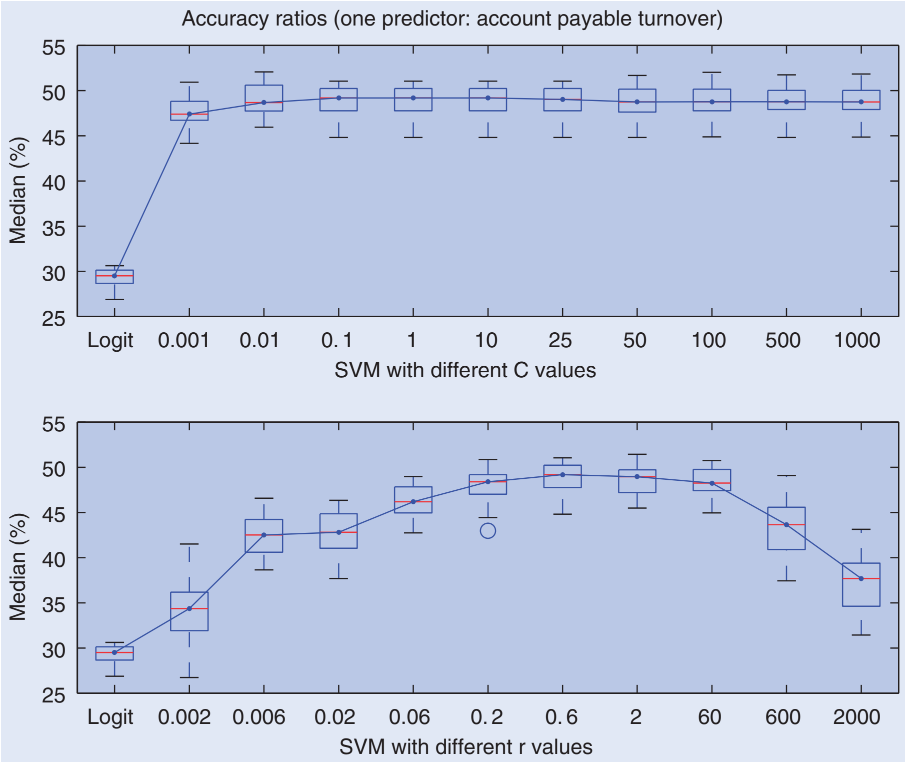

Figure 4. Sensitivity of the SVM to different parameters.

Table 5. Misclassification error (30 randomly selected samples; one predictor AP/SALE, x24).

	Parameter		Type I error		Type II error		Total error	
Model	C	r	Mean	Std	Mean	Std	Mean	Std
SVM	0.001	0.6	40.57	0.1167	23.43	0.9812	32.01	0.5723
	0.1	0.6	38.42	0.5125	24.45	1.1938	31.44	0.7014
	10	0.6	34.43	1.2126	27.86	1.637	31.15	0.9433
	100	0.6	25.22	0.6176	34.66	1.3541	29.94	0.8086
	1000	0.6	25.76	0.7705	34.26	1.3805	30.01	0.8712
	10	0.002	37.2	2.4512	32.79	2.5753	34.99	1.7611
	10	0.06	31.86	3.1527	29.25	2.2887	30.56	1.1405
	10	0.6	34.43	1.2126	27.86	1.637	31.15	0.9433
	10	60	37.27	0.5112	25.87	1.2134	31.57	0.7798
	10	2000	41.09	0.0791	24.85	0.3265	32.97	0.1123
Logit			38.15	0.5625	32.77	1.1888	35.46	0.7151

which has no similar adjustment parameters. Moreover, a grater SVM performance is a consequence of the SVM loss function, which is a tighter upper bound on the {0,1} step loss function. For univariate models, as figure 4 illustrates, the gain in performance of the SVM over the logit model is substantial and greater than for multivariate models since the former intrinsically has a larger number of degrees of freedom than the latter, which is limited by the number of variables.

The results in table 4 were obtained from the SVM with parameters $C=10$ and $r=0.6$, which were chosen according to the following sensitivity investigation of the SVM parameters (see box plot in figure 4 and table 5). That is to say, the values of parameters C and r could be determined experimentally via the standard use of a re-sampling training data set. Obviously, the SVM differs in different values of the penalty parameter C and the Gaussian kernel coefficient r. The ratio AP/SALE (x24) is exemplified here and the result for the benchmark logit model is also reported.

Here the median ARs are also estimated on 30 bootstrapped subsamples. On the whole, the discriminating ability of the SVM seems to be more sensitive to the value of r rather than to that of C. In figure 4(top), with fixed $r=0.6$, the median of the AR starts from 47.4% for $C=0.001$ and reaches the highest value 49.2% for $C=10$ and slightly decreases to 48.7% when $C=1000$. The varying range of AR is very small. Figure 4(bottom) illustrates the AR of the SVM versus r with fixed $C=10$. Within the interval, r is found to have a strong impact on the AR value, which starts at 34.4% when $r=0.002$ and drastically increases to the highest value 49.2%

when $r = 0.6$ and then decreases to 37.7% when $r = 2000$. In both parts of the figure the discriminating performance of the logit model is inferior to that of the SVM-based model with different parameter values.

As we have seen, $C = 10$ and $r = 0.6$ seem to be the best choice of parameter combination for the study in this paper. Thus, if we do not mention it particularly, the results of the SVM in the remaining part of this paper are all obtained using these parameter values. Note that this is not the case for the other data sample. The appropriate values of the C and r parameters will vary from sample to sample, therefore the sensitivity investigation of the SVM parameters should be carried out before classifying different data samples.

Table 5 shows the percentage of misclassified out-of-sample observations for the logit model and the SVM-based model with different parameters using a single predictor, the Account Payable Turnover. These errors are also obtained by bootstrap, and are all significant according to the standard deviations listed in table 5. Smaller values indicate better model accuracy. As shown in the table, the logit model has higher type I, type II and total error rates than the SVM-based model with only a few exceptions, suggesting that a well-specified SVM-based nonlinear model is superior to a logit model. For the SVM, with an increase of C from 0.001 to 1000, type II errors also increase, but type I errors decrease, and the total errors first decrease and then increase slightly. With increasing r values, type I and total errors also follow a U-shaped trend and type II errors have a monotonic negative relation with the r value. Therefore, $C = 10$ and $r = 0.6$ also appear to be the appropriate trade-off choice for our study in the following part of this paper. They produce only 34.43% type I errors, 27.86% type II errors and 31.15% total errors, whereas logit analysis produces 38.15% type I errors, 32.77% type II errors and 35.46% total errors.

As is evident from figure 5, which shows a univariate dependence of PD on AP/SALE, this dependence is not monotonously increasing or following any distinctive pattern, e.g. a logistic function. The SVM, being a more flexible non-parametric approach, is better suited for describing a broader class of dependence, such as this one, than the logit model. Another advantage of the SVM is its smaller bias in the estimation of the boundary between the solvent and insolvent companies in a situation when the number of the former is much larger than the number of the latter, as is almost always the case. The score of the logit model, which is interpreted as a PD, can be significantly biased for score values much lower or higher than 0.5. Subsequently, the threshold score for the boundary between solvent and insolvent companies is also biased. This is one reason for the substantial improvement in accuracy of the SVM compared with the logit model, as illustrated in figure 4. Because of this feature the SVM gains an additional improvement over the logit model if instead of subsamples with a 50/50 ratio of insolvent versus solvent companies we use subsamples where solvent companies prevail.

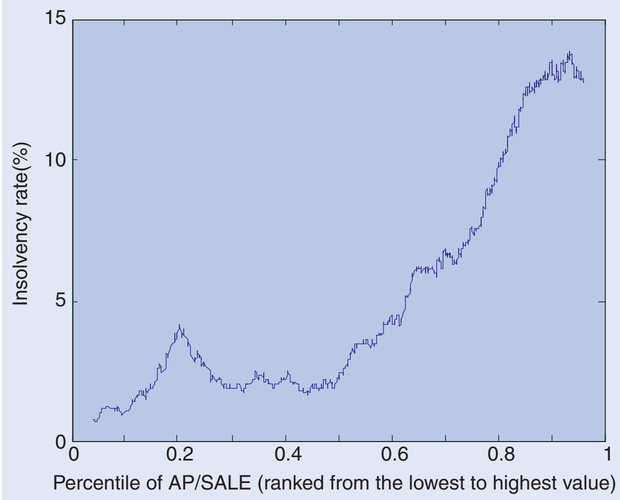

Figure 5. Insolvency rate evaluated for the financial ratio AP/SALE (x24) from the German Creditreform database. The k-nearest-neighbors procedure was used with the size of the window around 1/12 of 18,800 observations (the observations with zero values of sales used as the denominator to calculate the ratios were deleted from all 21,000 observations).

5.2. *Comparison of models with two predictors and PD visualization*

Table 6 shows the identifying performance of bivariate SVM-based models using the best predictor from the univariate model (AP/SALE) and one other. The values of the median of the AR direct us to the profitability ratio OI/TA (x3), the value of which increases to the highest of 56.46%, which indicates that OI/TA (x3) is the best choice for the second predictor.

Therefore, different from the usual result that NI/TA dominates other profitability ratios related to default risk, our study reveals that OI/TA performs better than the others in identifying bankrupt German firms. As the operating income does not include items such as investments in other firms, taxes, interest expenses and depreciation, the ratio represents a firm's true operating performance.

For two dimensions (i.e. two predictors), graphs are obviously an extremely useful tool for studying the data and assessing the quality of different default risk models. In addition, because of its nonlinearity it is more necessary for the SVM-based model to use visual tools than for the logit model to represent classification results. We demonstrate an application of visualization techniques for default analysis and parameter sensitivity investigation based on the SVM in figure 6. In the case of the logit model, the scores can be directly explained as the default probabilities, whereas for the SVM-based model the probabilities of default need to be calculated using the risk scores predicted by the estimated classifying function. Making use of the monotonic logistic cumulative distribution function, the default probabilities of German companies by SVM are calculated from the scores and then plotted as the background contour in figure 6 (corresponding to the right-hand bar in each sub-figure). The two predictors are the ratios AP/SALE (x24) and OI/TA (x3). These graphs are a subset of those used in

Table 6. Median of AR measure for a bivariate SVM model. AP/SALE (x24) and OI/TA (x3) produce the highest AR median.

No.	Ratio	AR median	No.	Ratio	AR median
x1	NI/TA	54.362	x15	CASH/TA	53.011
x2	NI/SALE	53.809	x16	CASH/CL	52.233
x3	**OI/TA**	**56.460**	x17	QA/CL	50.553
x4	OI/SALE	55.652	x18	CA/CL	44.678
x5	EBIT/TA	54.409	x19	WC/TA	48.676
x6	EBITDA	53.847	x20	CL/TL	49.725
x7	EBIT/SALE	52.948	x21	SALE/TA	49.624
x8	OF/TA	51.907	x22	INV/SALE	51.305
x9	(OF-ITGA) / (TA-ITGA-CASH-LB)	51.316	x23	AR/SALE	49.604
x10	CL/TA	48.197	**x24**	**AP/SALE**	
x11	(CL-CASH)/TA	49.680	x25	Log(TA)	51.545
x12	TL/TA	51.080	x26	IDINV/INV	49.904
x13	DEBT/TA	52.231	x27	IDL/TL	49.013
x14	EBIT/INTE	46.517	x28	IDCASH/CASH	46.617

the study. White and black points represent the 396 insolvent and 396 solvent firms from one random subsample of the validation set. The outliers were capped at the 5% and 95% quantiles as described in section 3.2 and kept in the subsample. In most panels of figure 6 they appear at the border. The classifying decision function (optimal hyperplane) is represented by the line denoted 0.5, along which the default probability is 0.5 and the risk scores are zero for SVM. The lines denoted 0.3 and 0.7 (or, more accurately, 0.27 and 0.73) are the lower and upper boundaries of the separation margin corresponding to scores of -1 and $+1$ in SVM. As shown in figure 6, clearly most successful firms lying in the blue area have positive profitability (OI/TA) and relatively lower account payable turnover (AP/SALE), while a majority of bankrupt firms is located in the opposite area. As known, low profitability usually indicates a high default risk, but extremely high profitability may also indicate a high cash flow volatility that is likely to translate into a higher default probability. Although the SVM-based model is sufficiently flexible to reveal a nonlinear dependence between profitability and PD, different from the logit model, for the Creditreform data in this study, the dependence could be too weak to be captured by SVM. Also, the sensitivity investigation results of the free parameters, C and r, of SVM could easily be determined from the figure.

Figure 6(a) shows the classification results for the logit model. Because the disadvantage of the logit model is the linearity of its solution, we see a straight classification line that is the linear combination of two predictors. Figure 6(b) shows the discriminating results obtained with the SVM-based model using a classifying function of moderate complexity ($r = 0.6$) and $C = 10$. This nonlinear classifying line (score 0 and PD 0.5) seems to identify the two types of firms very well with the areas in which solvent and insolvent firms are localized.

Fix $r = 0.6$. If the penalty is too low (C decreases to 0.01 and 0.1 as in figures 6(c) and (d)), the discriminating curve becomes flatter than that in figure 6(b). The calculated default probabilities are too small to display the two boundaries. That is, most of the firms fall inside the separation region but the insolvent and solvent firms are still clustered in their own areas. If the penalty increases, for example $C = 500$ as in figure 6(e), the identifying ability of SVM cannot be increased further than shown in figure 6(b).

Fix $C = 10$. If the complexity of the classifying functions increases (the r value decreases to 0.06 as illustrated in figure 6(f)), the SVM will try to capture each observation, although the majority of the insolvent firms still lie inside the band (0.5, 0.7) and above, with the solvent firms inside (0.5, 03) and below. The complexity in this case is too high for the given sample. If the r value increases to 60 (figure 6(g)), the classifying curve becomes flatter than that with $r = 0.6$; if r increases further to 2000 (figure 6(h)), the discriminating curve can be approximated as a linear combination of two predictors and is similar to the benchmark logit model, although the coefficients of the predictors may be different. The calculated default probabilities are also very small. The complexity here is too low to obtain a more detailed picture.

Although two cases of high complexity clearly demonstrate overfitting, (f) when $C = 10$ and $r = 0.06$, and (e) when $C = 500$ and $r = 0.6$, in all other cases the separating line is moderately nonlinear and for the case of a virtually linear SVM (h) with $C = 10$ and $r = 2000$ the separating line resembles that for the logit regression (a), with a different slope. Perfect separation for out-of-sample observations is not possible in any case. Nevertheless, comparing panel (a) for the logit with panel (f) for the SVM that achieved the maximum separation power, we observe that the most important difference between the two is in the area where the density of observations is the highest and even a small change in shape can lead to a substantial change in the classification ability.

The sensitivity analysis information obtained from this graphical analysis is similar to Härdle *et al.* (2005) and also confirms the choice combination of parameters as described in the sensitivity investigation of section 5.1. A set of alternative random subsamples as extracted from the validation set also display similar findings using the same visualization technique.

While the analysis here has been restricted to only two classes, namely bankruptcy and solvency, it can easily be

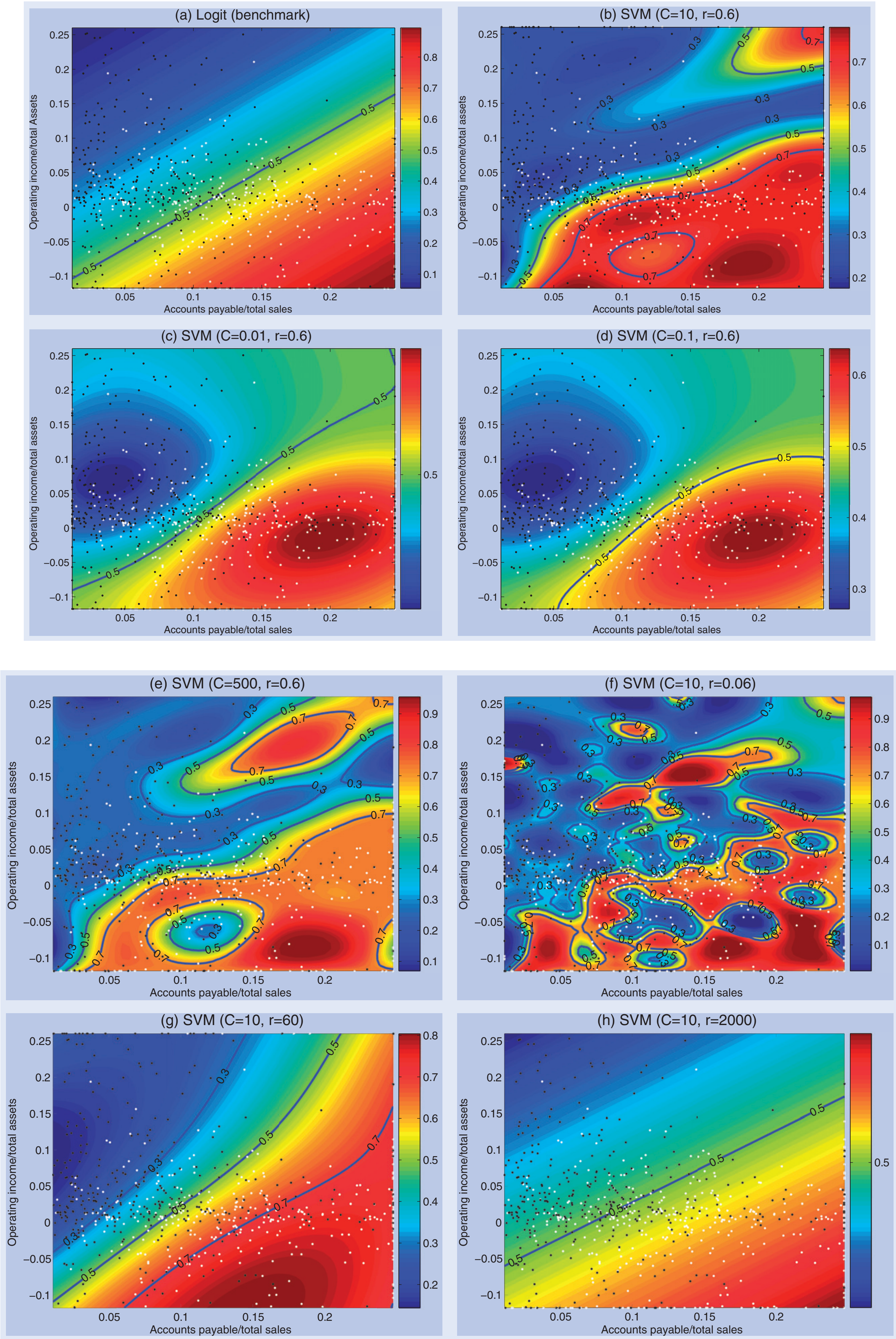

Figure 6. Default probabilities predicted for one random subsample and sensitivity analysis for the SVM.

Table 7. Median of AR measure for the best SVM model with eight important financial ratios calculated on 50/50 subsamples.

No.	Ratio	AR median Logit	AR median SVM	Predictors	No.	Ratio	AR median Logit	AR median SVM	Predictors
x1	NI/TA	35.12	59.93		**x15**	**CASH/TA**			**3**
x2	**NI/SALE**	**35.15**	**60.51**	**8**	x16	CASH/CL	34.87	59.42	
x3	**OI/TA**			**2**	x17	QA/CL	34.66	55.62	
x4	OI/SALE	35.06	60.44		x18	CA/CL	34.41	54.93	
x5	**EBIT/TA**			**7**	x19	WC/TA	34.72	59.48	
x6	EBITDA	34.93	59.85		x20	CL/TL	33.91	57.45	
x7	EBIT/SALE	35.14	60.4		x21	SALE/TA	35.05	56.61	
x8	OF/TA	35.04	59.64		**x22**	**INV/SALE**			**6**
x9	(OF-ITGA)/(TA-ITGA-CASH-LB)	34.94	59.42		x23	AR/SALE	35.15	59.81	
x10	CL/TA	33.94	58.19		**x24**	**AP/SALE**			**1**
x11	(CL-CASH)/TA	34.01	57.76		x25	Log(TA)	36.14	55.77	
x12	**TL/TA**			**4**	**x26**	**IDINV/INV**			**5**
x13	DEBT/TA	34.97	59.07		x27	IDL/TL	35.22	58.88	
x14	EBIT/INTE	35.03	54.37		x28	IDCASH/CASH	35.06	55.08	

generalized to multiple classes. In a multiple class case, financial analysts usually pre-specify rating classes (i.e. AAA, A, BB, C, etc.). A certain range of scores and default probabilities is associated with each rating class. The ranges are computed on the basis of historical data. According to the similarity of the scores, a new firm is assigned to one particular class. Therefore, we can draw more than one classifying function in the figure above to separate different rating classes.

5.3. *Powerful predictors related to insolvent German firms*

The selection procedure will be repeated for each new ratio added. The values of the AR increase until the model includes eight ratios, then they slowly decline. The medians of the AR for the models with eight ratios are shown in table 7. Most of the models tested here had AR values in the range 43.50–60.51% for out-of-sample and out-of-time tests. The results reported here are the product of the bootstrap approach described in the previous section. Obviously, the SVM-based model including ratios AP/SALE (x24), OI/TA (x3), CASH/TA (x15), TL/TA (x12), IDINV/INV (x26), INV/SALE (x22), EBIT/TA (x5) and NI/SALE (x2) attains the highest median AR, 60.51%. For comparison, we also report the median AR for the benchmark logit model with the same ratios. We can see that, for models containing the former seven ratios and one of the remaining, the medians of the AR are always higher for the SVM. This clearly reveals that the SVM-based model is always consistently superior to the benchmark logit model in identifying bankrupt firms and confirms the theoretical advantage of SVM for classification in the linear non-separable case. With respect to the percentage of correctly classified out-of-sample observations, a similar result is achieved (71.85% for the SVM-based model vs. 67.24% for the logit model).

It is noteworthy that, because the insolvency data was collected two years prior to insolvency, the predicted risk scores and calculated performance metrics in this study measure the model's ability to identify the firms that are going to default within the next two years. For example, the predicted default probability for 2002 denotes the probability that a firm defaults in 2003 or 2004.

We could not significantly improve upon our results by adding more ratios, and no model with fewer ratios performed as well. The eight selected predictors related to bankrupt German firms are AP/SALE (account payable turnover, x24), OI/TA (x3), CASH/TA (x15), TL/TA (x12), IDINV/INV (percentage of changing inventories, x26), INV/SALE (inventory turnover, x22), EBIT/TA (x5) and NI/SALE (net profit margin, x2). The size of the company was controlled in the analysis by the logarithm of the total assets (log(TA), x25). This can serve as a proxy for the cost of capital. In contrast to other studies, firm size has been shown to have no important effects on the probability of bankruptcy, which could be the result of pre-selecting only medium-sized companies.

Among the powerful predictors in identifying bankrupt German firms, there are two activity ratios (Account Payable Turnover and Inventory Turnover), three profitability ratios (OI/TA, EBIT/TA and Net Profit Margin), one liquidity ratio (CASH/TA), one leverage ratio (TL/TA) and one percentage of change ratio (Percentage of Incremental Inventories). It seems that activity ratios play the most important role in predicting the default probabilities of German firms. The activity ratio measures a firm's ability to convert different positions of their balance sheets into cash or sales. German firms will typically try to turn their accounts payable and inventories into sales as fast as possible because these will actually lead to higher revenues. Instead of ROA, EBIT/TA has a more powerful impact on insolvent German firms. In essence, it measures the operating performance and true productivity of firm assets on whose earning power the existence of the firm is based. Of course, the earnings of a firm only cannot tell the entire story. High earnings are good, but an increase in earnings does not mean that the net profit margin of a firm is improving.

For instance, if a firm has costs that have increased at a greater rate than sales, it leads to a lower profit margin. This is an indication that costs need to be under better control. Therefore, net profit margin is also very useful when analysing German bankruptcy data. In our study the liquidity ratio CASH/TA is only inferior to activity and profitability ratios when explaining German bankruptcies. Its strong explanatory power may result because the sample used in this study is mainly composed of private firms and this might not be true for public firms used in previous studies. The leverage ratio TL/TA also has a powerful influence on the identification of German bankruptcies. This metric is used to measure a firm's financial risk by determining how much of its assets have been financed by debt. This is a very broad ratio as it includes short- and long-term liabilities (debt) as well as all types of both tangible and intangible assets. The higher a firm's degree of leverage, the more the firm is considered risky. A firm with high leverage is more vulnerable to downturns in the business cycle because the firm must continue to service its debt regardless of how bad sales are. The incremental inventories provided by the Creditreform database also contain useful information for studying insolvent German firms.

To summarize our results, a German firm is most likely to go bankrupt when it has high turnover, low profits, low cash flows, is highly leveraged and has a high percentage of changing inventories. Although these results are similar to those of previous studies, the discovery of significant effects of the activity ratio and incremental inventories for predicting defaults in Germany is new.

6. Conclusions

We use a discrimination technique, the Support Vector Machine for classification, to analyse the German bankrupt company database spanning from 1996 through 2002. The identifying ability of an SVM-based nonlinear and non-parametric model is compared with that of the benchmark logit model with regard to two performance metrics (AR and misclassification error) on the basis of bootstrapped subsamples. The evidence from empirical results consistently shows that a credit risk model based on SVM significantly outperforms the benchmark linear parametric model in modeling the default risk of German firms out of sample and out of time. The sensitivity of the SVM to the penalty parameter C and Gaussian kernel coefficient r is examined according to the median of the AR using box plots (see figure 4), classification errors (see table 5) and two-dimensional visualization tools (figure 6). It is found that the discriminating ability of the SVM seems to be more sensitive to the values of r than C. Thus, appropriate trade-off values of parameters C and r should be chosen for bankruptcy analysis; for example, $C=10$ and $r=0.6$ in this study for the formal empirical analysis.

In addition to the unique minimum, no prior assumptions and it not being necessary to adjust the collinearity between the ratios, in particular the principle of structural risk minimization, endows the SVM approach with the most excellent classifying ability among all alternatives. Also, the SVM-based model is good at searching the linear non-separable hypersurface, which the logit model cannot do. As shown in table 4, the ratio Account Payable Turnover was selected by SVM among 28 candidates as the first best predictor to model the risk, which drastically upgrades the classifying accuracy, AR, of SVM by more than 10% as opposed to most of the other ratios selected. Otherwise, the performance gap between the SVM-based and logit model would not be so great, as shown in table 7. If the data are nonlinear, e.g. the Creditreform database, no linear model is able to separate the populations optimally, regardless of the DA, and the logit and probit models. The SVM method (as well as other pattern-recognition techniques) provides a more consistent way of finding the nonlinearities in the data, as opposed to performing an *ad-hoc* search of all possible combinations of the logit model. The holdout validation method, the most appropriate for modeling the real risk in practice, and the bootstrap re-sampling technique, guarantee the robustness and stability of the SVM approach. Due to the application of a kernel function and the sparseness of the algorithm, the achievement of such an improvement by SVM is not at a cost of much computational time, just a few seconds. Therefore, the empirical evidence confirms the theoretical advantage of SVM for classification and justifies it as applicable in practice. Of course, the non-parametric nature behind the SVM will come at the expense of understanding and insight; that is, the impact (the magnitude and direction and its significance) of the predictors on the default probabilities cannot be interpreted explicitly, in contrast to the parametric logit model. What the SVM is good at is capturing the nonlinearities better and forecasting the default probabilities more accurately than the benchmark.

As described in section 5.3, there are eight accounting ratios that are powerful predictors related to the bankruptcy of German companies. It turns out that activity ratios such as Account Payable and Inventory Turnover play the most important role in predicting the default probabilities. The percentage of incremental inventories provided by the Creditreform database also contains useful information for German bankruptcy analysis. These findings are new and somewhat different from the other default risk studies. The ability to automatically find the nonlinear dependence of the SVM model and the application of a widely accepted forward stepwise selection procedure in our case provides adequate selection that cannot be done by the usual linear classifying techniques such as the DA, logit model. That is to say, for German companies, Account Payable and Inventory Turnover, the percentage of incremental inventories selected have a strong nonlinear dependence on PDs, but a weak linear dependence that may lead to their unpopularity. Consistent with previous research, the profitability ratios, e.g. OI/TA, EBIT/TA and NI/SALE (net profit margin), are also powerful predictors related to German insolvency. Other results are similar to published research, e.g. that liquidity and leverage ratios also have important effects on the probability of default for German companies. But, in contrast to the others, firm size (log(TA), x25) was not chosen by the forward selection procedure as a predictor, which could be the result of pre-selecting only medium-sized companies.

Acknowledgements

The authors thank the Editors, Philip Angell, David Burgoyne, Beth Cawte, Collette Teasdale, and two referees for their constructive suggestions that significantly improved the paper. This work was supported by Deutsche Forschungsgemeinschaft through SFB 649 'Economic Risk'. Shiyi Chen was also sponsored by the Shanghai Pujiang Program, the Shanghai Leading Academic Discipline Project (No. B101) and the State Innovative Institute of Project 985 at Fudan University. W.K. Härdle was also partially supported by the National Center for Theoretical Sciences (South), Taiwan. R.A. Moro was supported by the German Academic Exchange Service (DAAD).

References

Aghion, P. and Bolton, P., An 'incomplete contracts' approach to financial contracting. *Rev. Econ. Stud.*, 1992, **59**(3), 473–494.

Altman, E.I., Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance*, 1968, **23**(4), 589–609.

Altman, E.I., Haldeman, R. and Narayanan, P., Zeta analysis: a new model to identify bankruptcy risk of corporations. *J. Bank. Finance*, 1977, **1**(1), 29–54.

Back, B., Laitinen, T. and Sere, K., Neural networks and bankruptcy prediction, in *17th Annual Congress of the European Accounting Association*, Venice, Italy, 1994. Abstract in *Collected Abstracts of the 17th Annual Congress of the European Accounting Association* 116.

Back, B., Laitinen, T., Sere, K. and Wezel, M., Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms. Technical Report 40, TUCS Research Group, 1996.

Basle Committee on Banking Supervision, Studies on the validation of internal rating systems. AIG/RTF BIS Working Paper No. 14, 2005.

Beaver, W., Financial ratios as predictors of failures. Empirical research in accounting: Selected studies. *J. Account. Res.*, 1966, **5**(suppl.), 71–111.

Bertsekas, D.P., *Nonlinear Programming*, 1995 (Athenas Science: Belmont, MA).

Burnham, K.P. and Anderson, D.R., *Model Selection and Inference*, 1998 (Springer: New York).

Caouette, J.B., Altman, E.I. and Narayanan, P., *Managing Credit Risk: The Next Great Financial Challenge*, 1998 (Wiley: New York).

Chakrabarti, B. and Varadachari, R., Quantitative methods for default probability estimation – a first step towards Basel II. i-flex solutions, 2004.

Collins, R. and Green, R., Statistical methods for bankruptcy prediction. *J. Econ. Business*, 1982, **34**(4), 349–354.

Courant, R. and Hilbert, D., *Methods of Mathematical Physics*, Vol. I and II, 1970 (Wiley Interscience: New York).

DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L., Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics*, 1988, **44**(3), 837–845.

Deng, N.Y. and Tian, Y.J., *New Methods in Data Mining: Support Vector Machine*, 2004 (Science Press: Beijing).

Efron, B. and Tibshirani, R.J., *An Introduction to the Bootstrap*, 1993 (Chapman & Hall: New York).

Engelmann, B., Hayden, E. and Tasche, D., Testing rating accuracy. *Risk*, 2003, **January**, 82–86.

Falkenstein, E., Boral, A. and Carty, L., Riskcalc for private companies: Moody's default model, Report Number: 56402, Moody's Investors Service, Inc., New York, 2000.

Fitzpatrick, P., *A Comparison of the Ratios of Successful Industrial Enterprises With Those of Failed Companies,* 1932 (The Accountants Publishing Company: Washington, DC).

Fletcher, R., *Practical Methods of Optimization*, 2nd ed., 1987 (Wiley: New York).

Friedman, C. and Sandow, S., Model performance measures for expected utility maximizing investors. *Int. J. Theor. Appl. Finance*, 2003a, **6**(4), 355–401.

Friedman, C. and Sandow, S., Learning probabilistic models: an expected utility maximization approach. *J. Mach. Learn. Res.*, 2003b, **4**, 257–291.

Friedman, C. and Huang, J., Default probability modeling: a maximum expected utility approach. Standard & Poor's Risk Solutions Group, New York, 2003.

Gaeta, G., editor, *The Certainty of Credit Risk: Its Measurement and Management*, 2003 (Wiley Finance (Asia): Singapore).

Gestel, T.V., Baesens, B., Dijcke, P.V., Suykens, J., Garcia, J. and Alderweireld, T., Linear and nonlinear credit scoring by combining logistic regression and support vector machines. *J. Credit Risk*, 2005, **1**(4), 31–60.

Giesecke, K., Credit risk modeling and valuation: An introduction. In *Credit Risk: Modeling and Management*, 2nd ed., edited by D. Shimko, pp. 487–526, 2004 (Risk Books: London).

Hanley, A. and McNeil, B., The meaning and use of the area under a receiver operating characteristics (ROC) curve. *Diagn. Radiol.*, 1982, **143**(1), 29–36.

Härdle, W., Moro, R.A. and Schäfer, D., Predicting bankruptcy with support vector machines. In *Statistical Tools for Finance and Insurance*, edited by P. Cizek, W. Härdle, and R. Weron, pp. 225–248, 2005 (Springer: Berlin).

Härdle, W., Moro, R.A. and Schäfer, D., Graphical data representation in bankruptcy analysis. In *Handbook for Data Visualization*, edited by Ch.-H. Chen, W. Härdle, and A. Unwin, pp. 853–872, 2007 (Springer: Berlin).

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A., *Nonparametric and Semiparametric Models*, 2004 (Springer: Heidelberg).

Härdle, W. and Simar, L., *Applied Multivariate Statistical Analysis*, 2003 (Springer: Berlin).

Haykin, S., *Neural Networks: A Comprehensive Foundation*, 1999 (Prentice-Hall: Engelwood Cliffs, NJ).

Herrity, J.V., Keenan, S.C., Sobehart, J.R., Carty, L.V. and Falkenstein, E.G., Measuring private firm default risk. Moody's Investors Service Special Comment, 1999.

Hertz, J., Krogh, A. and Palmer, R.G., *The Theory of Neural Network Computation*, 1991 (Addison Welsey: Redwood, CA).

Horowitz, J.L., *The Bootstrap*, Vol. 5, 2001 (Elsevier: Amsterdam).

Keenan, S.C. and Sobehart, J.R., Performance measures for credit risk models. Research report #1-10-10-99, Moody's Risk Management Services, 1999.

Khandani, B., Lozano, M. and Carty, L., Moody's riskcalc for private companies: The German model. Rating Methodology, Moody's Investors Service, 2001.

Lennox, C., Identifying failing companies: A re-evaluation of the logit, probit and DA approaches. *J. Econ. Business*, 1999, **51**, 347–364.

Lo, A.W., Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *J. Econometr.*, 1986, **31**(2), 151–178.

Mercer, J., Functions of positive and negative type, and their connection with the theory of integral equations. *Trans. London Philos. Soc. A*, 1908, **209**, 415–446.

Merwin, C., Financing small corporations in five manufacturing industries, 1926–36. National Bureau of Economic Research, 1942.

Myers, S., Determinants of corporate borrowing. *J. Financial Econ.*, 1977, **5**(2), 147–175.

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

Platt, H., Platt, M. and Pedersen, J., Bankruptcy discrimination with real variables. *J. Business, Finance Account.*, 1994, **21**(4), 491–510.

Ramser, J. and Foster, L., A demonstration of ratio analysis. Bulletin No. 40, Bureau of Business Research, University of Illinois, 1931.

Refenes, A.P., *Neural Networks in the Capital Markets*, 1995 (Wiley: Chichester).

Saunders, A. and Allen, L., *Credit Risk Measurement*, 2nd ed., 2002 (Wiley: New York).

Serrano, C., Martin, B. and Gallizo, J.L., Artificial neural networks in financial statement analysis: Ratios versus accounting data. Technical report, paper presented at the 16th Annual Congress of the European Accounting Association, Turku, Finland, April 28–30, 1993.

Shumway, T., Forecasting bankruptcy more accurately: a simple hazard model. Working Paper, University of Michigan Business School, 1998.

Sobehart, J.R., Stein, R.M., Mikityanskaya, V. and Li, L., Moody's public firm risk model: a hybrid approach to modeling default risk. Moody's Investors Service Rating Methodology, 2000.

Sobehart, J., Keenan, S. and Stein, R., Benchmarking quantitative default risk models: A validation methodology. *Algo Res. Q.*, 2001, **4**(1/2), 57–72.

Sobehart, J.R. and Keenan, S.C., Performance evaluation for credit spread and default risk models. In *Credit Risk: Models and Management*, 2nd ed., edited by D. Shimko, pp. 275–305, 2004 (Risk Books: London).

Swets, J.A., Measuring the accuracy of diagnostic systems. *Science*, 1998, **240**(4857), 1285–1293.

Swets, J.A., Dawes, R.M. and Monahan, J., Better decisions through science. *Sci. Am.*, 2000, **October**, 82–87.

Tikhonov, A.N., On solving ill-posed problem and method regularization. *Dokl. Akad. Nauk USSR*, 1963, **153**, 501–504.

Tikhonov, A.N. and Arsenin, V.Y., *Solution of Ill-posed Problems*, 1977 (W.H. Winston: Washington, DC).

Vapnik, V., *Estimation of Dependencies Based on Empirical Data*, 1979 (Nauka: Moscow).

Vapnik, V., *The Nature of Statistical Learning Theory*, 1995 (Springer: New York).

Vapnik, V., *Statistical Learning Theory*, 1997 (Wiley: New York).

Wilson, R.L. and Sharda, R., Bankruptcy prediction using neural networks. *Decis. Supp. Syst.*, 1994, **11**, 545–557.

Winakor, A. and Smith, R., Changes in the financial structure of unsuccessful industrial corporations. Bulletin No. 51, Bureau of Business Research, University of Illinois, 1935.

Zagst, R. and Hocht, S., Comparing default probability models. Working Paper, Munich University of Technology, 2006.