Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

Discussion Papers

811

Laura Auria • Rouslan A. Moro

Support Vector Machines (SVM) as a Technique for Solvency Analysis

Berlin, August 2008

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Available for free downloading from the DIW Berlin website.

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN.
Papers can be downloaded free of charge from the following websites:

http://www.diw.de/english/products/publications/discussion_papers/27539.html
http://ideas.repec.org/s/diw/diwwpp.html
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=1079991

Support Vector Machines (SVM) as a Technique for Solvency Analysis

by

Laura Auria[1] and Rousian A. Moro[2]

Abstract

This paper introduces a statistical technique, Support Vector Machines (SVM), which is considered by the Deutsche Bundesbank as an alternative for company rating. A special attention is paid to the features of the SVM which provide a higher accuracy of company classification into solvent and insolvent. The advantages and disadvantages of the method are discussed. The comparison of the SVM with more traditional approaches such as logistic regression (Logit) and discriminant analysis (DA) is made on the Deutsche Bundesbank data of annual income statements and balance sheets of German companies. The out-of-sample accuracy tests confirm that the SVM outperforms both DA and Logit on bootstrapped samples.

Keywords: company rating, bankruptcy analysis, support vector machines
JEL Classification: C13, G33, C45

Acknowledgements: the work of R. Moro was supported by Deutsche Bank and its foundation Geld und Währung. Additionally R. Moro acknowledges the support of the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk". All analysis was done on the premises of Deutsche Bank in Hannover and Frankfurt.

1. Introduction

There is a plenty of statistical techniques, which aim at solving binary classification tasks such as the assessment of the credit standing of enterprises. The most popular techniques include traditional statistical methods like linear Discriminant Analysis (DA) and Logit or Probit Models and non-parametric statistical models like Neural Networks. SVMs are a new promising non-linear, non-parametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields. Applied to solvency analysis, the common objective of all these clas-

[1] Deutsche Bundesbank, Georgplatz 5, 30159 Hannover.
[2] German Institute for Economic Research, Mohrenstr. 58, 10117 Berlin.

sification techniques is to develop a function, which can accurately separate the space of solvent and insolvent companies, by benchmarking their score value. The score reduces the information contained in the balance sheet of a company to a one-dimensional summary indicator, which is a function of some predictors, usually financial ratios. Another aim of solvency analysis is to match the different score values with the related probability of default (PD) within a certain period. This aspect is especially important in the Eurosystem, when credit scoring is performed with the target of classifying the eligibility of company credit liabilities as a collateral for central bank refinancing operations, since the concept of eligibility is related to a benchmark value in terms of the annual PD.

The selection of a classification technique for credit scoring is a challenging problem, because an appropriate choice given the available data can significantly help improving the accuracy in credit scoring practice. On the other hand, this decision should not be seen as an “either / or” choice, since different classification techniques can be integrated, thus enhancing the performance of a whole credit scoring system. In the following paper SVMs are presented as a possible classification technique for credit scoring. After a review of the basics of SVMs and of their advantages and disadvantages on a theoretical basis, the empirical results of an SVM model for credit scoring are presented.

2. Basics of SVMs

SVMs are a new technique suitable for binary classification tasks, which is related to and contains elements of non-parametric applied statistics, neural networks and machine learning. Like classical techniques, SVMs also classify a company as solvent or insolvent according to its score value, which is a function of selected financial ratios. But this function is neither linear nor parametric. The formal basics of SVMs will be subsequently briefly explained. The case of a linear SVM, where the score function is still linear and parametric, will first be introduced, in order to clarify the concept of margin maximisation in a simplified context. Afterwards the SVM will be made non-linear and non-parametric by introducing a kernel. As explained further, it is this characteristic that makes SVMs a useful tool for credit scoring, in the case the distribution assumptions about available input data can not be made or their relation to the PD is non-monotone.

Margin Maximization

Assume, there is a new company j, which has to be classified as solvent or insolvent according to the SVM score. In the **case of a linear SVM** the score looks like a DA or Logit score, which is a linear combination of relevant financial ratios $x_j = (x_{j1}, x_{j2}, \ldots x_{jd})$, where x_j is a vector with d financial ratios and x_{jk} is the value of the financial ratio number k for company j, $k=1,\ldots,d$. So z_j , the score of company j, can be expressed as:

$$z_j = w_1 x_{j1} + w_2 x_{j2} + \ldots + w_d x_{jd} + b \quad (1)$$

In a compact form:

$$z_j = x_j^T w + b \qquad (1')$$

where w is a vector which contains the weights of the d financial ratios and b is a constant. The comparison of the score with a benchmark value (which is equal to zero for a balanced sample) delivers the "forecast" of the class – solvent or insolvent – for company j.

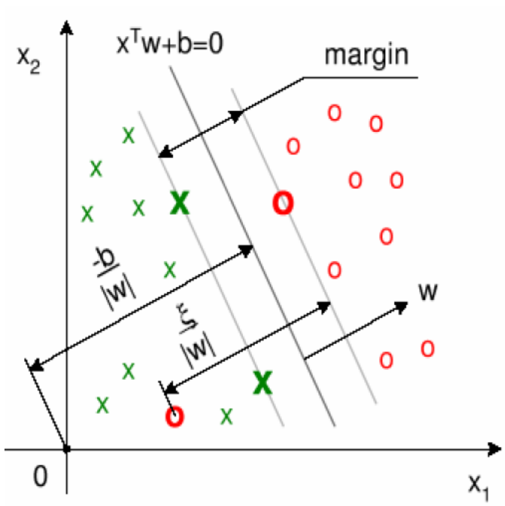
In order to be able to use this decision rule for the classification of company j, the SVM has to learn the values of the score parameters w and b on a training sample. Assume this consists of a set of n companies $i = 1, 2, ..., n$. From a geometric point of view, calculating the value of the parameters w and b means looking for a hyperplane that best separates solvent from insolvent companies according to some criterion. The criterion used by SVMs is based on **margin maximization** between the two data classes of solvent and insolvent companies. The margin is the distance between the hyperplanes bounding each class, where in the hypothetical **perfectly separable** case no observation may lie. By maximising the margin, we search for the classification function that can most safely separate the classes of solvent and insolvent companies. The graph below represents a binary space with two input variables. Here crosses represent the solvent companies of the training sample and circles the insolvent ones. The threshold separating solvent and insolvent companies is the line in the middle between the two **margin boundaries**, which are canonically represented as $x^T w+b=1$ and $x^T w+b=-1$. Then the margin is $2 / ||w||$, where $||w||$ is the norm of the vector w.

In a **non-perfectly separable** case the margin is "**soft**". This means that in-sample classification errors occur and also have to be minimized. Let ξ_i be a non-negative slack variable for in-sample misclassifications. In most cases $\xi_i =0$, that means companies are being correctly classified. In the case of a positive ξ_i the company i of the training sample is being misclassified. A further criterion used by SVMs for calculating w and b is that all misclassifications of the training sample have to be minimized.

Let y_i be an indicator of the state of the company, where in the case of solvency $y_i =-1$ and in the case of insolvency $y_i =1$. By imposing the constraint that **no observation may lie within the margin except some classification errors**, SVMs require that either $x_i^T w+b \geq 1-\xi_i$ or $x_i^T w+b \leq -1+\xi_i$, which can be summarized with:

$$y_i \left(x_i^T w + b\right) \geq 1 - \xi_i, \ \forall \ i = 1,...,n. \qquad (3)$$

Figure 1. Geometrical Representation of the SVM Margin



Source: W. Härdle, R.A. Moro, D. Schäfer, March 2004, *Rating Companies with Support Vector Machines*, Discussion Paper Nr. 416, DIW Berlin.

The optimization problem for the calculation of w and b can thus be expressed by:

$$\min_w \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \quad (2)$$

$$s.t. \quad y_i\left(x_i^T w + b\right) \geq 1 - \xi_i \ , \quad (3)$$

$$\xi_i \geq 0 \quad (4)$$

In the first part of (2) we maximise the margin $2 / ||w||$ by minimizing $||w||^2 / 2$, where the square in the norm of w comes from the second term, which originally is the sum of in-sample misclassification errors $\xi_i \ / \ ||w||$ times the parameter C. Thus SVMs maximize the margin width while minimizing errors. This problem is quadratic i.e. convex.

C = “capacity” is a tuning parameter, which weights in-sample classification errors and thus controls the generalisation ability of an SVM. The higher is C, the higher is the weight given to in-sample misclassifications, the lower is the generalization of the machine. Low generalisation means that the machine may work well on the training set but would perform miserably on a new sample. Bad generalisation may be a result of overfitting on the training sample, for example, in the case that this sample shows some untypical and non-repeating data structure. By choosing a low C, the risk of overfitting an SVM on the training sample is reduced. It can be demonstrated that C is linked to the width of the margin. The smaller is C, the wider is the margin, the more and larger in-sample classification errors are permitted.

Solving the above mentioned constrained optimization problem of calibrating an SVM means searching for the minimum of the following Lagrange function:

$$L(w,b,\xi;\alpha,\nu)=\frac{1}{2}w^T w+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left\{y_i\left(w^T x_i+b\right)-1+\xi_i\right\}-\sum_{i=1}^{n}\nu_i\xi_i, \tag{5}$$

where $\alpha_i \geq 0$ are the Lagrange multipliers for the inequality constraint (3) and $\nu_i \geq 0$ are the Lagrange multipliers for the condition (4). This is a convex optimization problem with inequality constraints, which is solved my means of classical non-linear programming tools and the application of the **Kuhn-Tucker Sufficiency Theorem**. The solution of this optimisation problem is given by the saddle-point of the Lagrangian, minimised with respect to w, b, and ξ and maximised with respect to α and ν. The entire task can be reduced to a convex quadratic programming problem in α_i. Thus, by calculating α_i, we solve our classifier construction problem and are able to calculate the parameters of the linear SVM model according to the following formulas:

$$w=\sum_{i=1}^{n}y_i\alpha_i x_i \tag{6}$$

$$b=\frac{1}{2}\left(x_{+1}^{\mathrm{T}}+x_{-1}^{\mathrm{T}}\right)\cdot w \tag{7}$$

As can be seen from (6), $\alpha_{i,}$ which must be non-negative, weighs different companies of the training sample. The companies, whose α_i are not equal to zero, are called **support vectors** and are the relevant ones for the calculation of w. Support vectors lie on the margin boundaries or, for non-perfectly separable data, within the margin. By this way, the complexity of calculations does not depend on the dimension of the input space but on the number of support vectors. Here x_{+1} and x_{-1} are any two support vectors belonging to different classes, which lie on the margin boundaries.

By substituting (6) into the score (1'), we obtain the score z_j as a function of the scalar product of the financial ratios of the company to be classified and the financial ratios of the support vectors in the training sample, of α_i, and of y_i. By comparing z_j with a benchmark value, we are able to estimate if a company has to be classified as solvent or insolvent.

$$\Rightarrow z_j=\sum_{i=1}^{n}y_i\alpha_i\left\langle x_i,x_j\right\rangle+b \tag{8}$$

Kernel-transformation

In the **case of a non-linear SVM**, the score of a company is computed by substituting the scalar product of the financial ratios with a kernel function.

$$z_j = \sum_{i=1}^{n} y_i \alpha_i \langle x_i, x_j \rangle + b \rightarrow z_j = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b \,, \qquad (8')$$

Kernels are symmetric, semi-positive definite functions satisfying the Mercer theorem. If this theorem is satisfied, this ensures that there exists a (possibly) non-linear map Φ from the input space into some feature space, such that its inner product equals the kernel. The non-linear transformation Φ is only **implicitly** defined through the use of a kernel, since it only appears as an inner product.

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \,. \qquad (9)$$

This explains how non-linear SVMs solve the classification problem: the input space is transformed by Φ into a feature space of a higher dimension, where it is easier to find a separating hyperplane. Thus the kernel can side-step the problem that data are non-linearly separable by **implicitly** mapping them into a feature space, in which the linear threshold can be used. Using a kernel is equivalent to solving a linear SVM in some new higher-dimensional feature space. The non-linear SVM score is thus a linear combination, but with new variables, which are derived through a kernel transformation of the prior financial ratios. The score function does not have a compact functional form, depending on the financial ratios but on some transformation of them, which we do not know, since it is only implicitly defined. It can be shown that the solution of the constrained optimisation problem for non-linear SVM is given by:

$$w = \sum_{i=1}^{n} y_i \alpha_i \Phi(x_i) \qquad (6')$$

$$b = -\frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_{i,} x_{+1}) + \sum_{i=1}^{n} \alpha_i y_i K(x_{i,} x_{-1}) \right) \qquad (7')$$

But, according to (7') and (8'), we do not need to know the form of the function Φ, in order to be able to calculate the score. Since for the calculation of the score (8) the input variables are used as a product, only the kernel function is needed in (8'). As a consequence, **Φ and w are not required for the solution of a non-linear SVM**.

One can choose among many types of kernel functions. In practice, many SVM models work with **stationary Gaussian kernels with an anisotropic radial basis.** The reason why is that they are very flexible and can build fast all possible relations between the financial ratios. For example linear transformations are a special case of Gaussian kernels.

$$K(x_i, x_j) = e^{-(x_j - x_i)^T r^{-2} \Sigma^{-1} (x_j - x_i)/2} \qquad (10)$$

Here Σ is the variance-covariance matrix of all financial ratios of the training set. This kernel first transforms the "anisotropic" data to the same scale for all variables. This is the meaning of "isotropic". So there is no risk that financial ratios with greater numeric ranges dominate those with smaller ranges. The only parameter which has to be chosen when using Gaussian kernels is *r*, which controls the radial basis of the kernel. This reduces the complexity of model selection. The higher is *r*, the smoother is the threshold which separates solvent from insolvent companies.[3]

Gaussian kernels non-linearly map the data space into a higher dimensional space. Actually the definition of a Gaussian process by specifying the covariance function (depending on the distance of the company to be evaluated from each company of the training sample) **avoids explicit definition of the function class of the transformation**. There are many possible decompositions of this covariance and thus also many possible transformation functions of the input financial ratios. Moreover each company shows its own co-variance function, depending on its relative position within the training sample. That is why the kernel operates locally. The value of the kernel function depends on the distance between the financial ratios of the company *j* to be classified and respectively one company *i* of the training sample. This kernel is a normal density function up to a constant multiplier. x_i is the center of this kernel, like the mean is the center of a normal density function.

3. What Is the Point in Using SVMs as a Classification Technique?

All classification techniques have advantages and disadvantages, which are more or less important according to the data which are being analysed, and thus have a relative relevance. SVMs can be a useful tool for insolvency analysis, in the case of non-regularity in the data, for example when the data are not regularly distributed or have an unknown distribution. It can help evaluate information, i.e. financial ratios which should be transformed prior to entering the score of classical classification techniques. The **advantages of the SVM technique** can be summarised as follows:

1. By introducing the kernel, SVMs gain flexibility in the choice of the form of the threshold separating solvent from insolvent companies, which needs not be linear and even needs not have the same functional form for all data, since its function is non-parametric and operates locally. As a consequence they can work with financial ratios, which show a non-monotone relation to the score and to the probability of default, or which are non-linearly dependent, and this without needing any specific work on each non-monotone variable.
2. Since the kernel **implicitly** contains a non-linear transformation, no assumptions about the functional form of the transformation, which makes data linearly separable, is necessary. The transformation occurs implicitly on a robust theoretical basis and human expertise judgement beforehand is not needed.
3. SVMs provide a good out-of-sample generalization, if the parameters *C* and *r* (in the case of a Gaussian kernel) are appropriately chosen. This means that, by choosing an appropriate generalization grade, SVMs can be robust, even when the training sample has some bias.

3 By choosing different *r* values for different input values, it is possible to rescale outliers.

4. SVMs deliver a unique solution, since the optimality problem is convex. This is an advantage compared to Neural Networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples.
5. With the choice of an appropriate kernel, such as the Gaussian kernel, one can put more stress on the similarity between companies, because the more similar the financial structure of two companies is, the higher is the value of the kernel. Thus when classifying a new company, the values of its financial ratios are compared with the ones of the support vectors of the training sample which are more similar to this new company. This company is then classified according to with which group it has the greatest similarity.

Here are some examples where the SVM can help coping with non-linearity and non-monotonicity. One case is, when the coefficients of some financial ratios in equation (1), estimated with a linear parametric model, show a sign that does not correspond to the expected one according to theoretical economic reasoning. The reason for that may be that these financial ratios have a non-monotone relation to the PD and to the score. The unexpected sign of the coefficients depends on the fact, that data dominate or cover the part of the range, where the relation to the PD has the opposite sign. One of these financial ratios is typically the growth rate of a company, as pointed out by [10]. Also leverage may show non-monotonicity, since if a company primary works with its own capital, it may not exploit all its external financing opportunities properly. Another example may be the size of a company: small companies are expected to be more financially instable; but if a company has grown too fast or if it has become too static because of its dimension, the big size may become a disadvantage. Because of these characteristics, the above mentioned financial ratios are often sorted out, when selecting the risk assessment model according to a linear classification technique. Alternatively an appropriate evaluation of this information in linear techniques requires a transformation of the input variables, in order to make them monotone and linearly separable.[4]

A common **disadvantage of non-parametric techniques** such as SVMs is the lack of transparency of results. SVMs cannot represent the score of all companies as a simple parametric function of the financial ratios, since its dimension may be very high. It is neither a linear combination of single financial ratios nor has it another simple functional form. The weights of the financial ratios are not constant. Thus **the marginal contribution of each financial ratio to the score is variable**. Using a Gaussian kernel each company has its own weights according to the difference between the value of their own financial ratios and those of the support vectors of the training data sample.

Interpretation of results is however possible and can rely on graphical visualization, as well as on a local linear approximation of the score. The SVM threshold can be represented within a bi-dimensional graph for each pair of financial ratios. This visualization technique cuts and projects the multidimensional feature space as well as the multivariate threshold function separating solvent and insolvent companies on a bi-dimensional one, by fixing the values of the other financial ratios equal to the values of the company, which has to be classified. By this way, different companies will have different threshold projections.

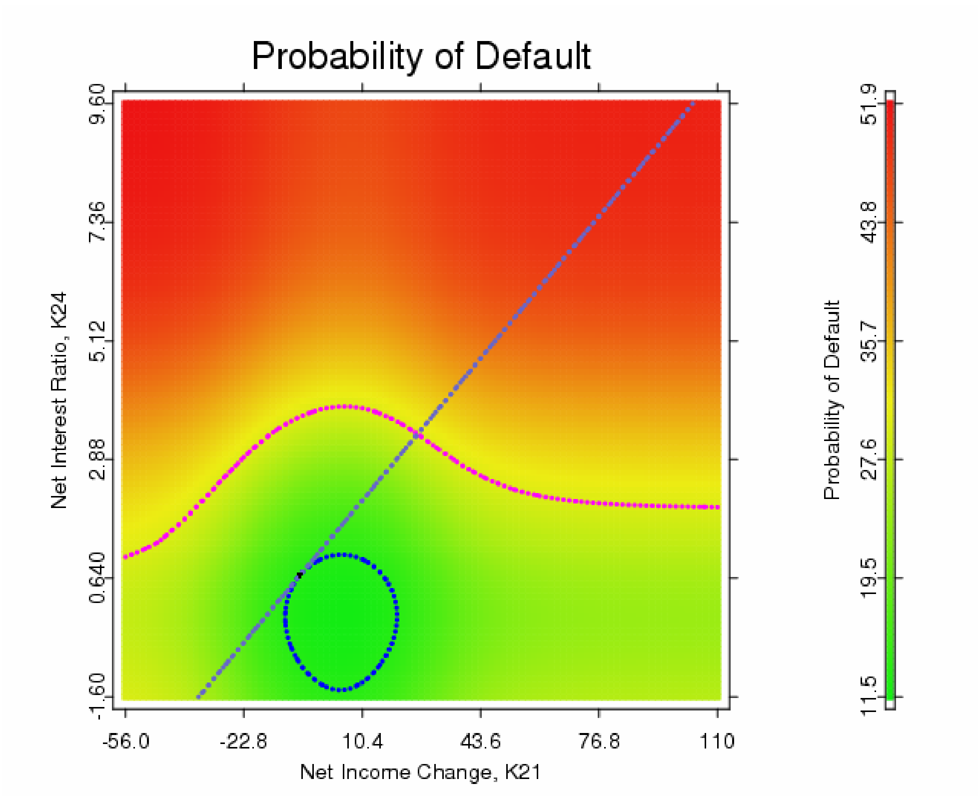
[4] See [6] for an analysis of the univariate relation between the PD and single financial ratios as well as for possible transformations of input financial ratios in order to reach linearity.

However, an analysis of these graphs gives an important input about the direction towards which the financial ratios of non-eligible companies should change, in order to reach eligibility.

The PD can represent a third dimension of the graph, by means of isoquants and colour coding. The approach chosen for the estimation of the PD can be based on empirical estimates or on a theoretical model. Since the relation between score and PD is monotone, a local linearization of the PD can be calculated for single companies by estimating the tangent curve to the isoquant of the score. For single companies this can offer interesting information about the factors influencing their financial solidity.

In the figure below the PD is estimated by means of a Gaussian kernel[5] on data belonging to the trade sector and then smoothed and monotonized by means of a Pool Adjacent Violator algorithm.[6] The pink curve represents the projection of the SVM threshold on a binary space with the two variables K21 (net income change) and K24 (net interest ratio), whereas all other variables are fixed at the level of company *j*. The blue curve represents the isoquant for the PD of company *j*, whose coordinates are marked by a triangle.

Figure 2. Graphical Visualization of the SVM Threshold and of a Local Linearization of the Score Function: Example of a Projection on a Bi-dimensional Graph with PD Colour Coding



[5] This methodology is based on a non-parametric estimation of the PD and has the advantage that it delivers an individual PD for each company based on a continuous, smooth and monotonic function. This PD-function is computed on an empirical basis, so there is no need for a theoretical assumption about the form of a link function.

[6] See [11].

The grey line corresponds to the linear approximation of the score or PD function projection for company *j*. One interesting result of this graphical analysis is that successful companies with a low PD often lie in a closed space. This implies that there exists an optimal combination area for the financial ratios being considered, outside of which the PD gets higher. If we consider the net income change, we notice that its influence on the PD is non-monotone. Both too low or too high growth rates imply a higher PD. This may indicate the existence of the optimal growth rate and suggest that above a certain rate a company may get into trouble; especially if the cost structure of the company is not optimal i.e. the net interest ratio is too high. But if a company lies in the optimal growth zone, it can also afford a higher net interest ratio.

4. An Empirical SVM Model for Solvency Analysis

In the following chapter, an empirical SVM model for solvency analysis on German data is being presented.[7] The estimation of score functions and their validation are based on balance sheets of solvent and insolvent companies, whereas a **company is classified as insolvent if it is the subject of failure judicial proceeding**. The study is conducted over a long period, in order to construct durable scores that are resistant, as far as possible, to cyclical fluctuations. So the original data set consists of about 150.000 firm-year observations, spanning the time period from 1999 to 2005. The **forecast horizon** is three and a half years. That is, in each period a company is considered insolvent, if it has been the subject of legal proceedings within the three and a half years since the observation date. Solvent companies are those that have not gone bankrupt within three and a half years after the observation date. With shorter term forecast horizons, such as one-year, data quality would be poor, since most companies do not file a balance sheet, if they are on the point of failure. Moreover, companies that go insolvent already show weakness three years before failure. In order to improve the accuracy of analysis, a different model was developed for each of the following three **sectors**: manufacturing, wholesale/retail trade and other companies. The three models for the different sectors were trained on data over the time period 1999-2001 and then validated out-of-time on data over the time period 2002-2005.

Two important points for the selection of an accurate SVM model are the choice of the input variables, i.e. of the financial ratios, which are being considered in the score, as well as of the tuning parameters C and r (once a Gaussian kernel has been chosen).

Table 1. Training and Validation Data Set Size – Without Missing Values

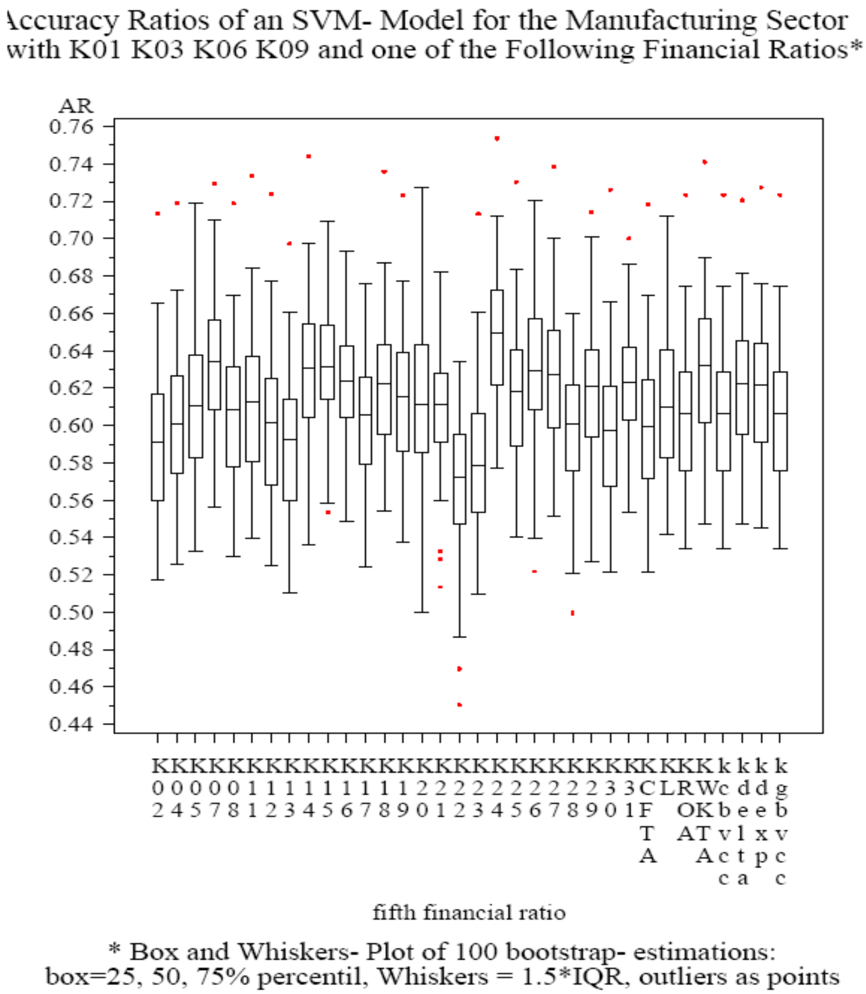
sector	year							total	
	1999	2000	2001	2002	2003	2004	2005	solv.	ins.
manufacturing	6015	5436	4661	5202	5066	4513	698	30899	692
wholesale / retail trade	12806	11230	9209	8867	8016	7103	996	57210	1017
other	6596	6234	5252	5807	5646	5169	650	34643	711

[7] The database belongs to the balance sheet pool of the "Deutsche Bundesbank".

The choice of the input variables has a decisive influence on the performance results and is not independent from the choice of the classification technique. These variables normally have to comply with the assumptions of the applied classification technique. Since the SVM needs no restrictions on the quality of input variables, it is free to choose them only according to the model accuracy performance. The **input variables selection methodology** applied in this paper is based on the following empirical tools.

The discriminative power of the models is measured on the basis of their accuracy ratio (AR) and percentage of correctly classified observations, which is a compact performance indicator, complementary to their error quotes. Since there is no assumption on the density distribution of the financial ratios, a robust comparison of these performance indicators has to be constructed on the basis of bootstrapping. The different SVM models are estimated 100 times on 100 randomly selected training samples, which include all insolvent companies of the data pool and the same number of randomly selected solvent ones. Afterwards they are validated on 100 similarly selected validation samples. The model, which delivers the best median results over all training and validation samples, is the one which is chosen for the final calibration. A similar methodology is used for **choosing the optimal capacity *C* and the kernel-radius *r* of the SVM model**. That combination of C and r values is chosen, which delivers the highest median AR on 100 randomly selected training and validation samples.

Figure 3. Choice of the Financial Ratios of an SVM Model for the Manufacturing Sector: An Example for the Choice of the Fifth Input Variable



Our analysis first started by estimating the three SVM models on the basis of four financial ratios, which are presently being used by the “Bundesbank” for DA and which are expected to comply with its assumptions on linearity and monotonicity. By integrating the model with further non-linearly separable variables a significant performance improvement in the SVM model was recorded. The new input variables were chosen out of a catalogue, which is summarized in Table 3, on the basis of a bootstrapping procedure by means of forward selection with an SVM model. Variables were added to the model sequentially until none of the remaining ones would improve the median AR of the model. Figure 3 shows the AR distributions of different SVM models with 5 variables. According to these graphical results one should choose K24 as the fifth variable. As a result of this selection procedure, the median AR peaked with ten input variables (10FR) and then fell gradually.

Table 2. Final Choice of the Input Variables Forward Selection Procedure

Sector		
Manufacturing	**Wholesale/Retail Trade**	**Other**
K01: pre-tax profit margin	K01: pre-tax profit margin	K02: operating profit margin
K03: cash flow ratio	K04: capital recovery ratio,	K05: debt cover
K06: days receivable	K06: days receivable	K06: days receivable
		K07: days payable
K09: equity ratio adj.	K09: equity ratio adj.	K08: equity ratio
K11: net income ratio	K17: liquidity 3 (current assets to short debt)	K12:guarantee a.o. obligation ratio (leverage 1)
K15: liquidity 1		
K18: short term debt ratio	K18: short term debt ratio	K18: short term debt ratio
K24: net interest ratio	K21: net income change	K19: inventories ratio
	K24: net interest ratio	K21: net income change
K26: tangible asset growth	K31: days of inventories	K31: days of inventories
KWKTA: working capital to total assets	KL: leverage	KL: leverage

A univariate analysis of the relation between the single variables and the PD showed that most of these variables actually have a non-monotone relation to the PD, so that considering them in a linear score would require the aforementioned transformation. Especially growth variables as well as leverage and net interest ratio showed a typical non-monotone behaviour and were at the same time very helpful in enhancing the predictive power of the SVM.

Figure 4 summarizes the predictive results of the three final models, according to the above mentioned bootstrap procedure. Based on the procedure outlined above, the following values of the kernel tuning parameters were selected: $r = 4$ for the manufacturing and trade sector and $r = 2.5$ for other companies. This suggests that this sector is less homogeneous than the other two. The capacity of the SVM model was chosen as $C = 10$ for all the three sectors. It is interesting to notice, that the robustness of the results, measured by the spread of the ARs over different samples, became lower, when the number of financial ratios being considered grew. So there is a trade-off between the accuracy of the model and its robustness.

Table 3. The Catalogue of Financial Ratios – Univariate Summary Statistics and Relation to the PD[8]

Variable	Name	Aspect	Q 0.01	median	Q 0.99	IQR	Relation to the PD
K01	Pre-tax profit (income) margin	profitability	-57.1	2.3	140.1	6.5	- n.m.
K02	Operating profit margin	profitability	-53	3.6	80.3	7.2	-
K03	Cash flow ratio (net income ratio)	liquidity	-38.1	5.1	173.8	10	-
K04	Capital recovery ratio	liquidity	-29.4	9.6	85.1	15	-
K05	Debt cover (debt repayment capability)	liquidity	-42	16	584	33	-
K06	Days receivable (accounts receiv-able collection period)	activity	0	29	222	34	+ n.m.
K07	Days payable (accounts payable collection period)	activity	0	20	274	30	+ n.m.
K08	Equity (capital) ratio	financing	-57	16.4	95.4	27.7	-
K09	Equity ratio adj. (own funds ratio)	financing	-55.8	20.7	96.3	31.1	-
K11	Net income ratio	profitability	-57.1	2.3	133.3	6.4	+/- n.m.
K12	guarantee a.o. obligation ratio (leverage 1)	leverage	0	0	279.2	11	-/+ n.m.
K13	Debt ratio	liquidity	-57.5	2.4	89.6	18.8	-/+ n.m.
K14	Liquidity ratio	liquidity	0	1.9	55.6	7.2	-
K15	Liquidity 1	liquidity	0	3.9	316.7	16.7	-
K16	Liquidity 2	liquidity	1	63.2	1200	65.8	- n.m.
K17	Liquidity 3	liquidity	2.3	116.1	1400	74.9	- n.m.
K18	Short term debt ratio	financing	0.2	44.3	98.4	40.4	+
K19	Inventories ratio	investment	0	23.8	82.6	35.6	+
K20	Fixed assets ownership ratio	leverage	-232.1	46.6	518.4	73.2	-/+ n.m.
K21	Net income change	growth	-60	1	133	17	-/+/- n.m.
K22	Own funds yield	profitability	-413.3	22.4	1578.6	55.2	+/- n.m.
K23	Capital yield	profitability	-24.7	7.1	61.8	10.2	-
K24	Net interest ratio	cost. structure	-11	1	50	1.9	+ n.m.
K25	Own funds/pension provision r.	financing	-56.6	20.3	96.1	32.4	-
K26	Tangible assets growth	growth	-0.2	13.9	100	23	-/+ n.m.
K27	Own funds/provisions ratio	financing	-53.6	27.3	98.8	36.9	-
K28	Tangible asset retirement	growth	0.1	19.3	98.7	18.7	-/+ n.m.
K29	Interest coverage ratio	cost structure	-2364	149.5	39274.3	551.3	n.m.
K30	Cash flow ratio	liquidity	-27.9	5.2	168	9.7	-
K31	Days of inventories	activity	0	41	376	59	+
K32	Current liabilities ratio	financing	0.2	59	96.9	47.1	+
KL	Leverage	leverage	1.4	67.2	100	39.3	+ n.m.
KWKTA	Working capital to total assets	liquidity	565.9	255430	51845562.1	865913	+/- n.m.
KROA	Return on assets	profitability	-42.1	0	51.7	4.8	n.m.
KCFTA	Cash flow to total assets	liquidity	-26.4	9	67.6	13.6	-
KGBVCC	Accounting practice, cut		-2	0	1.6	0	n.m.
KCBVCC	Accounting practice		-2.4	0	1.6	0	n.m.
KDEXP	Result of fuzzy expert system, cut		-2	0.8	2	2.8	-
KDELTA	Result of fuzzy expert system		-7.9	0.8	8.8	3.5	-

n.m.= non-monotone

+ = positive relation

- = negative relation

+ n.m.= non monotone relation, mostly positive

- n.m.= non monotone relation, mostly negative

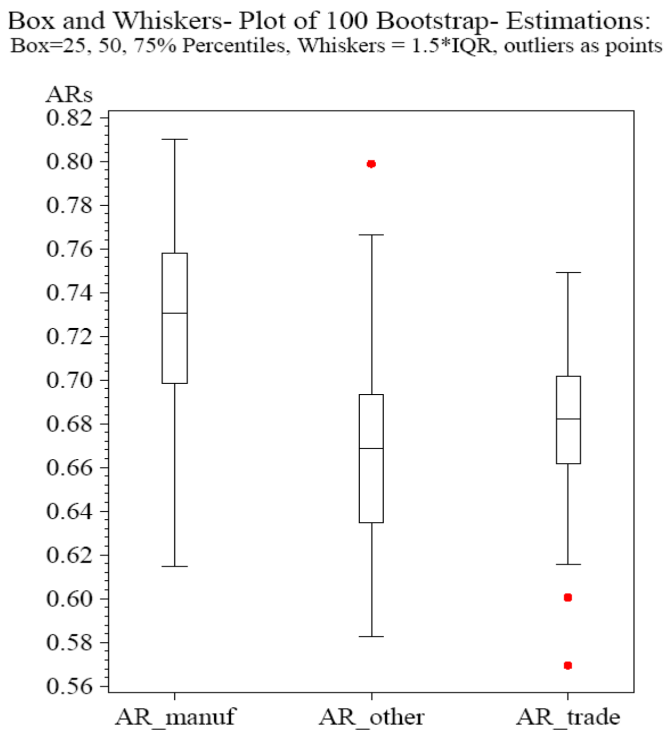
+/- n.m. = non-monotone relation, first positive then negative

-/+ n.m. = non-monotone relation, first negative then positive

-/+/- n.m. = non-monotone relation, first negative, then positive then again negative

[8] K1-K32 as well as KGBVCC and KDEXP are financial ratios belonging to the catalogue of the "Deutsche Bundesbank". See [4].

Figure 4. Predictive Results: ARs of the Final SVM Model after Bootstrapping



5. Conclusions

SVMs can produce accurate and robust classification results on a sound theoretical basis, even when input data are non-monotone and non-linearly separable. So they can help to evaluate more relevant information in a convenient way. Since they linearize data on an implicit basis by means of kernel transformation, the accuracy of results does not rely on the quality of human expertise judgement for the optimal choice of the linearization function of non-linear input data. SVMs operate locally, so they are able to reflect in their score the features of single companies, comparing their input variables with the ones of companies in the training sample showing similar constellations of financial ratios. Although SVMs do not deliver a parametric score function, its local linear approximation can offer an important support for recognising the mechanisms linking different financial ratios with the final score of a company. For these reasons SVMs are regarded as a useful tool for effectively complementing the information gained from classical linear classification techniques.

References

[1] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens and J. Vanthienen, 2003, *Benchmarking State-of-the-art Classification Algorithms for Credit Scoring*, Journal of the Operational Research Society (2003), 0, 1-9.

[2] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, *A Practical Guide to Support Vector Classification*, http://www.csie.ntu.edu.tw.

[3] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Repr. 2006, Cambridge University Press, 2000.

[4] Deutsche Bundesbank, *How the Deutsche Bundesbank Assesses the Credit Standing of Enterprises in the Context of Refinancing German Credit Institutions*, Markets Department, June 2004, http://www.bundesbank.de/download/gm/gm_broschuere_bonitaetunternehmen_en.pdf

[5] B. Engelmann, E. Hayden, D. Tasche, 2003, *Measuring the Discriminative Power of Rating Systems*, Deutsche Bundesbank Discussion Paper, Series 2: Banking and Financial Supervision, No 01/2003.

[6] E. Falkenstein, 2000, *Riskcalc for Private Companies: Moody's Default Model*, Moody's Investor Service.

[7] T. Van Gestel, B. Baesens, J. Garcia, P. Van Dijcke, *A Support Vector Machine Approach to Credit Scoring*, http://www.defaultrisk.com/pp_score_25.htm.

[8] W. K. Härdle, R. A. Moro., D. Schäfer, *Rating Companies with Support Vector Machines*, DIW Discussion Paper No. 416, Berlin, 2004.

[9] W. K. Härdle, R. A. Moro., D. Schäfer, *Support Vector Machines – Eine neue Methode zum Rating von Unternehmen*, DIW Wochenbericht No. 49/04, Berlin, 2004.

[10] E. Hayden, *Modeling an Accounting-Based Rating System for Austrian Firms*, Dissertation, Fakultät für Wirtschaftwissenschaften und Informatik, Universität Wien, Juni 2002.

[11] E. Mammen, *Estimating a Smooth Monotone Regression Function*, The Annals of Statistics, Vol. 19, No. 2, June 1991, Pp. 724-740.

[12] B. Schölkopf, A. Smola, *Learning with Kernels -Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, Cambridge, MA, 2002, http://www.learning-with-kernels.org.

[13] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 2000.