RATING COMPANIES – A SUPPORT VECTOR MACHINE ALTERNATIVE

W. HÄRDLE [2,3]

R. A. MORO [1,2,3]

D. SCHÄFER [1]

[1] Deutsches Institut für Wirtschaftsforschung (DIW); [2] Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin;
[3] MD*Tech

DIW Berlin

HUMBOLDT-UNIVERSITÄT ZU BERLIN

DEUTSCHE BUNDESBANK

C.A.S.E. CENTER FOR APPLIED STATISTICS AND ECONOMICS

Bundesbank, 29th November 2005

MD*TECH Method & Data Technologies

Classical Rating Methods

Most rating methods implemented by European central banks are **linear methods** (discriminant analysis and logit/probit regression). They evaluate the score as:

$$Z = a_1x_1 + a_2x_2 + ... + a_dx_d$$

where $x_1, x_2, \ldots, x_d$ are financial ratios

Linear Discriminant Analysis (DA)

Fisher (1936); company scoring: Beaver (1966), Altman (1968)

Z-score:

$$Z_i = a_1 x_{i1} + a_2 x_{i2} + ... + a_d x_{id} = a^\top x_i,$$

where $x_i = (x_{i1}, ..., x_{id})^\top$ are financial ratios for the i-th company.

The classification rule:
$Z_i \geq z$: successful company
$Z_i < z$: failure

Logit/Probit Regression

Probit model, Martin (1977), Ohlson (1980)

$$\mathrm{E}[y_i|x_i] = \Phi\left(a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_d x_{id}\right), \quad y_i = \{0, 1\}$$

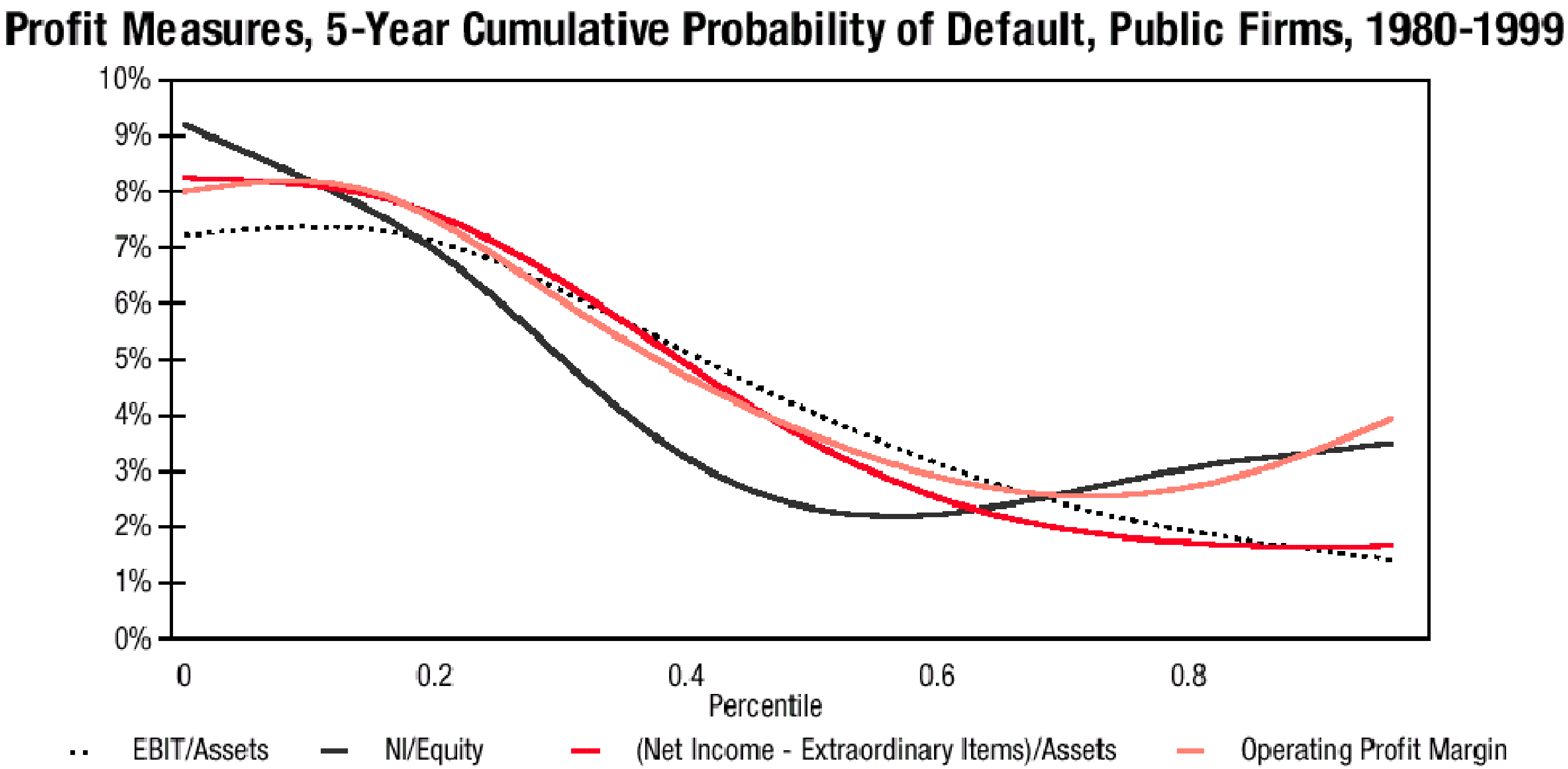
Logit model

$$\mathrm{E}[y_i|x_i] = \frac{1}{1 + \exp(-a_0 - a_1 x_{i1} - \ldots - a_d x_{id})}$$

The score function looks the same as for DA

$$Z_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_d x_{id} = a^\top x_i,$$

Probability of Default (Company Data)



Source: Falkenstein et al. (2000)

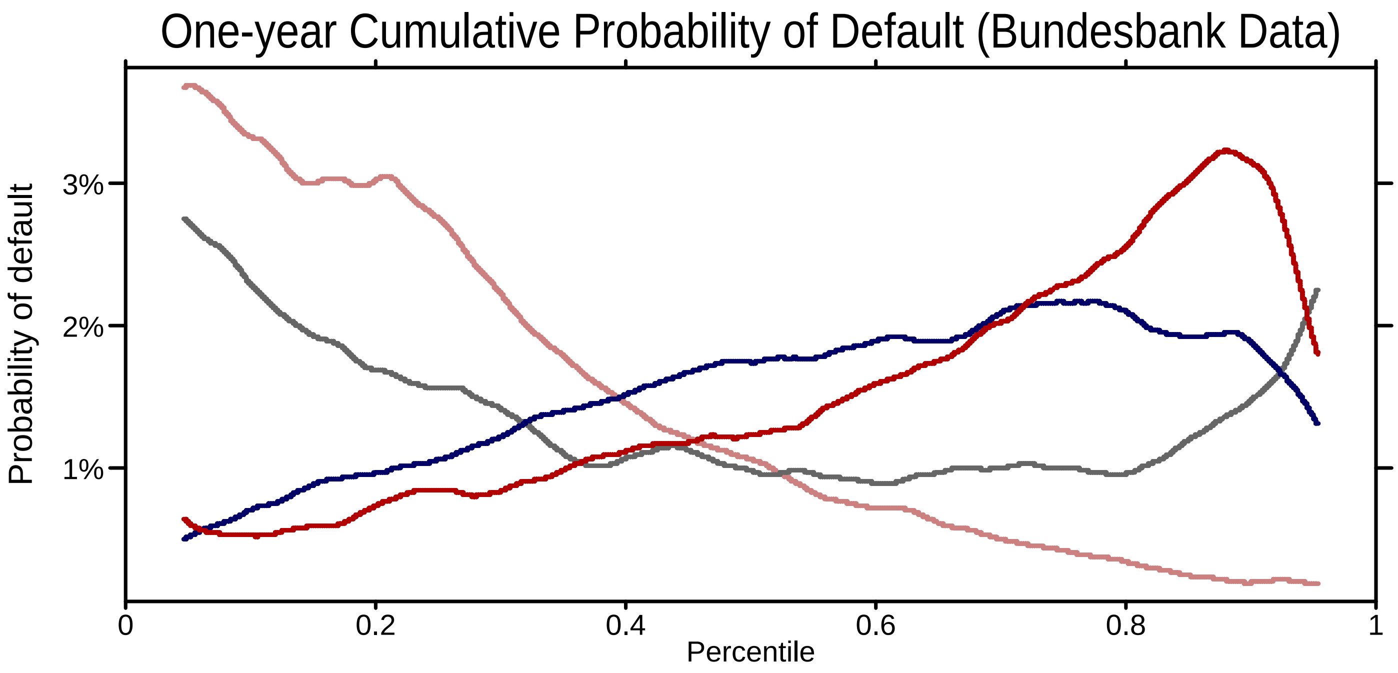
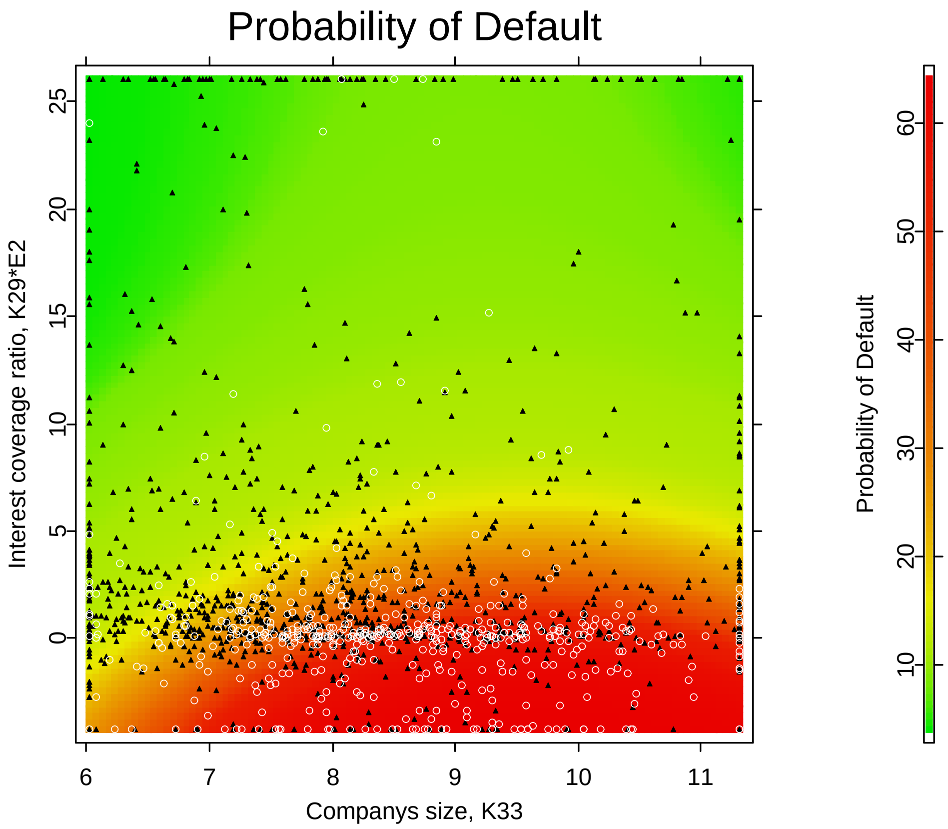


Figure 1: Four of eight financial ratios included in the model with the highest prediction power. The ratios are K21, K24, K29 and K33.

Linearly Non-separable Classification Problem



Outline

Classification Set Up

The **training set** $\{x_i, y_i\}$, $i = 1, 2, \ldots, n$ represents information about companies

$y_i = 1$ for insolvent; $y_i = -1$ for solvent firms

x_i is a vector of financial ratios

We estimate the class y of some unknown firm described with x

This is done with a classifier function $f : X \mapsto \{+1; -1\}$, so that the error rate be low

Support Vector Machine (SVM)

SVMs are a group of methods for classification (and regression)

- SVMs possess a flexible structure which is not chosen a priori
- The properties of SVMs can be derived from statistical learning theory
- SVMs do not rely on asymptotic properties; they are especially useful when d/n is big, i.e. in most practically significant cases
- SVMs give a unique solution and often outperform Neural Networks

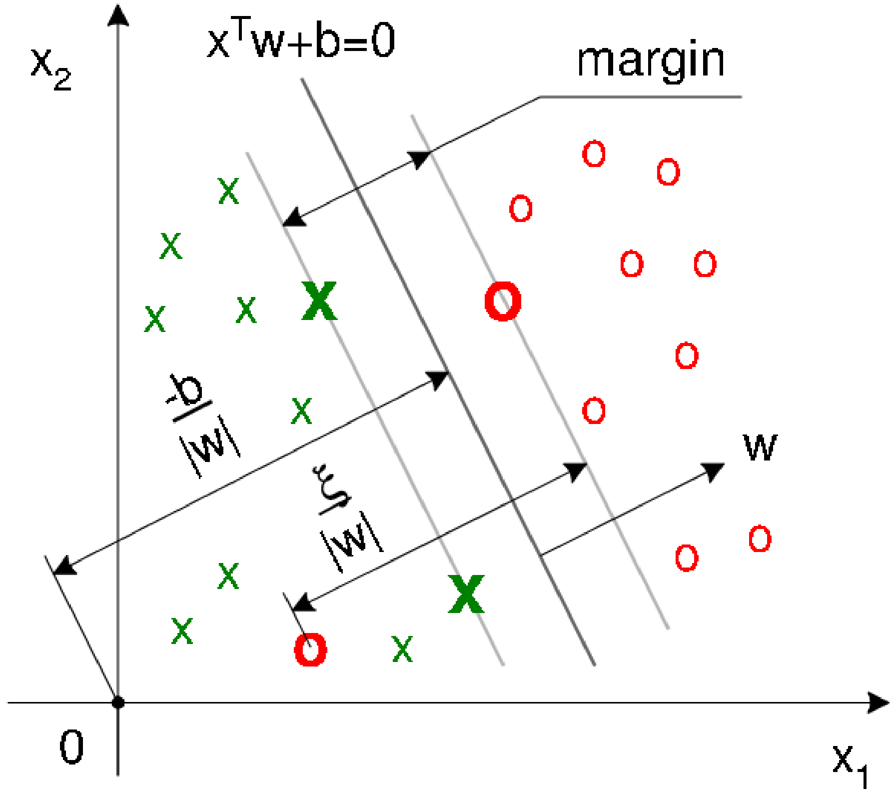
SVM Basics

The training set: $\{x_i, y_i\}$, $i = 1, 2, \ldots, n$; $y_i = \{+1; -1\}$.

Find the classification function that can most safely separate two classes, i.e. when the distance between classes is the highest

The gap between parallel hyperplanes separating two classes where with separable data the vectors of neither class can lie is called **margin**

Linear SVM. Non-separable Case



The inequality below guarantees that the data of one class would lie on the same side of the margin zone if corrected with positive slack variables ξ_i, $i = 1, 2, \ldots, n$

$$y_i(x_i^\top w + b) \geq 1 - \xi_i$$

The objective function subject to constrained minimisation:

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

where C ("capacity") is a bandwidth parameter. Under such a formulation the problem has a **unique** solution

The score is: $f(x) = x^\top w + b$

Classification rule: $g(x) = \mathrm{sign}(f) = \mathrm{sign}(x^\top w + b)$

Non-linear SVM

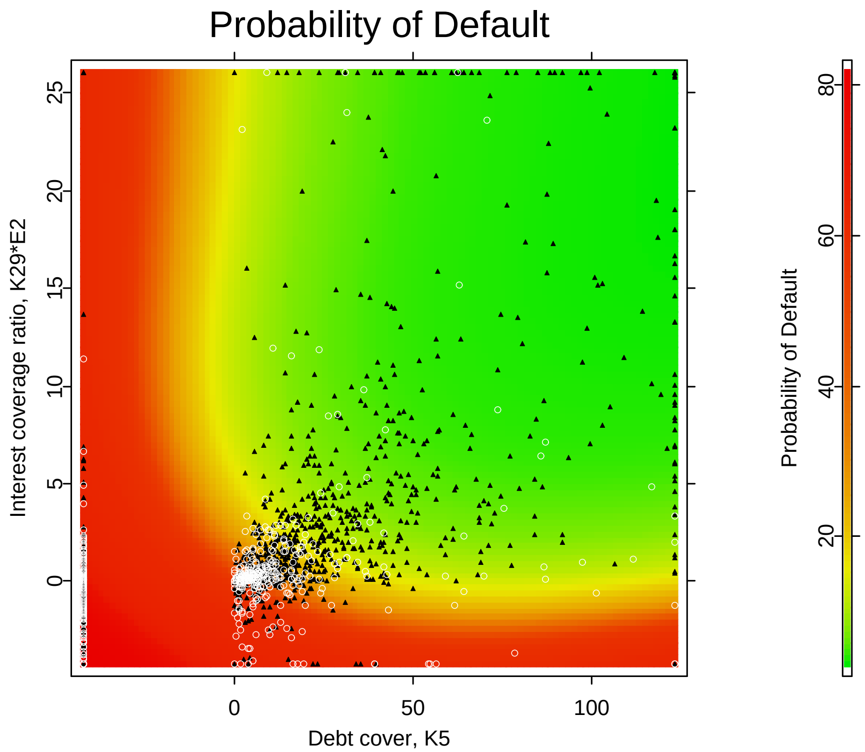


Figure 2: Extension of SVMs to a non-linear case via kernel techniques is possible due to their specific properties

Control Parameters of an SVM

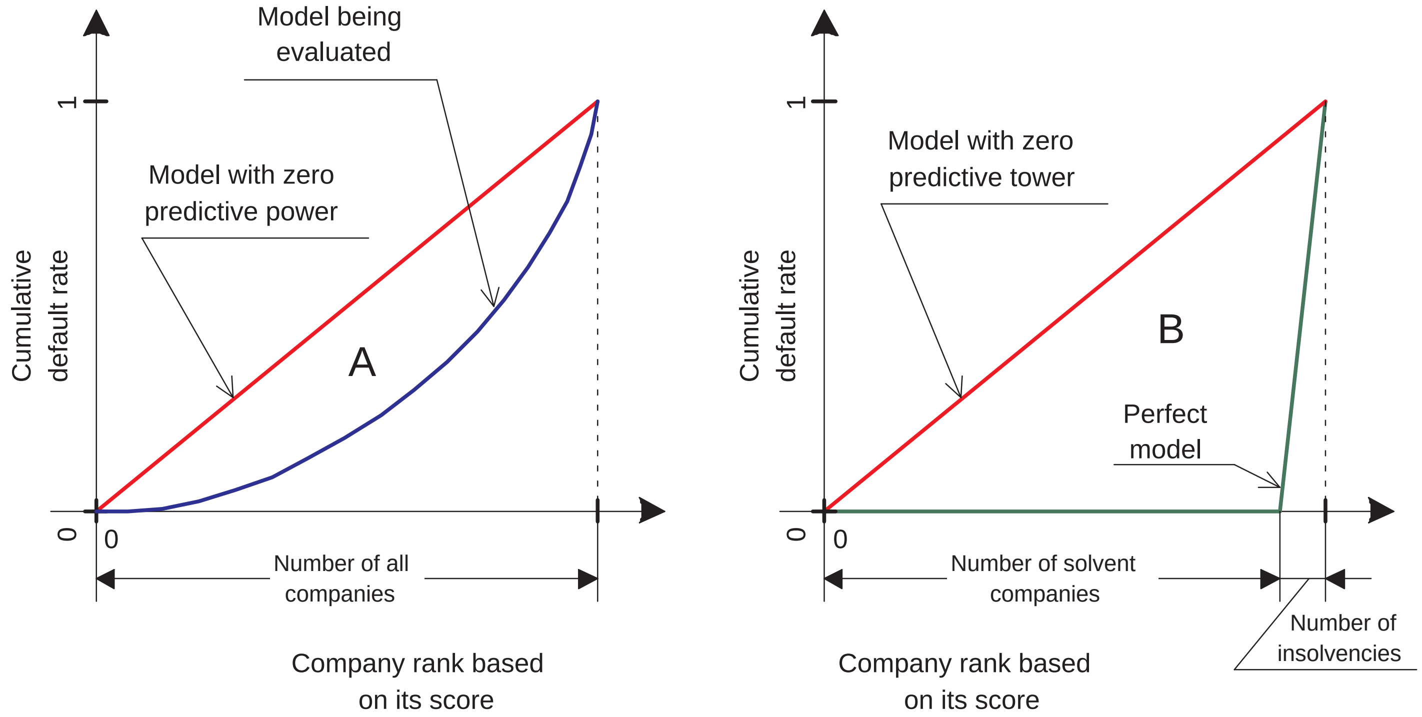
An SVM is defined by

1. Type of its kernel function
2. Capacity C that controls the complexity of the model. It is optimised to achieve the highest accuracy (accuracy ratio or prediction accuracy)

Out-of-Sample Accuracy Measures

- Percentage of correctly cross-validated observations
- Percentage of correctly validated out-of-sample observations, α- and β-errors
- Power curve (PC) aka Lorenz curve or cumulative accuracy profile. PC for a real model lies between PCs for the perfect and zero predictive power models
- Accuracy ratio (AR)

Accuracy Ratio



$$\text{Accuracy Ratio (AR)} = A/B$$

Data Description

Source: Bundesbank's Central Corporate Database

Around 553000 balance sheets, 8150 belong to insolvent companies

Selected were private companies with turnover >36000 EUR a year, also satisfying a number of minor criteria

All bankruptcies took place in 1997-2004 no later than three years and no sooner than three months after the last report was submitted

Data Description

- selection of variables was performed on subsamples of 1000 bankrupt companies and 1000 solvent ones. From those subsamples a training and validation sets were constructed, each including 500 solvent and 500 insolvent companies
- the procedure of the random selection of the training and validation sets was repeated 100 time. Each time accuracy ratio and forecasting accuracy was computed and their distribution represented as a box plot
- each observation can appear only in one set
- 32 financial ratios and one random variable were analysed

Variables and Their Predictive Power

No.	Name (Eng.)	Name (Ger.)	med. AR
K1	Pre-tax profit margin	Umsatzrendite	0.388
K2	Operating profit margin	Betriebsrendite	0.273
K3	Cash flow ratio	Einnahmenüberschussquote	0.361
K4	Capital recovery ratio	Kapitalrückflussquote	0.435
K5	Debt cover	Schuldentilgungsfähigkeit	0.455
K6	Days receivable	Debitorenumschlag	0.235
K7	Days payable	Kreditorenumschlag	0.346
K8	Equity ratio	Eigenkapitalquote	0.323
K9	Equity ratio (adj.)	Eigenmittelquote	0.336

No.	Name (Eng.)	Name (Ger.)	med. AR
K10	Random variable	Zufallsvariable	-0.003
K11	Net income ratio	Umsatzrendite ohne a.E.	0.404
K12	Leverage ratio	Quote aus Haftungsverhltnissen	0.113
K13	Debt ratio	Finanzbedarfsquote	0.250
K14	Liquidity ratio	Liquidittsquote	0.211
K15	Liquidity 1	Liquiditätsgrad 1	0.263
K16	Liquidity 2	Liquiditätsgrad 2	0.189
K17	Liquidity 3	Liquiditätsgrad 3	0.168
K18	Short term debt ratio	kurzfr. Fremdkapitalquote	0.296
K19	Inventories ratio	Vorratsquote	0.176

No.	Name (Eng.)	Name (Ger.)	med. AR
K20	Fixed assets ownership r.	Deckungsgrad Anlagevermgen	0.166
K21	Net income change	Umsatzveränderungen	0.195
K22	Own funds yield	Eigenkapitalrendite	0.264
K23	Capital yield	Gesamtkapitalrendite	0.362
K24	Net interest ratio	Nettozinsquote	0.281
K25	Own funds/pension prov. r.	Pensionsrückstellungsquote	0.306
K26	Tangible asset growth	Investitionsquote	0.033
K27	Own funds/provisions ratio	Eigenkapitalrückstellungsq.	0.321
K28	Tangible asset retirement	Abschreibungsquote	0.046
K29	Interest coverage ratio	Zinsdeckung	0.449

No.	Name (Eng.)	Name (Ger.)	med. AR
K30	Cash flow ratio	Einnahmenüberschußquote	0.300
K31	Days of inventories	Lagedauer	0.305
K32	Current liabilities ratio	Fremdkapitalstruktur	0.181
K33	Log of total assets	Log vom Gesamtkapital	0.175

Summary Statistics

Predictor	Group	$q_{0.01}$	$q_{0.99}$	Median	IQR
K1	Profitability	-26.9	78.5	2.3	5.9
K2	Profitability	-24.6	64.8	3.8	6.3
K3	Liquidity	-22.6	120.7	5.0	9.4
K4	Liquidity	-24.4	85.1	11.0	17.1
K5	Liquidity	-42.0	507.8	17.1	34.8
K6	Activity	0.0	184.0	31.1	32.7
K7	Activity	0.0	248.2	23.2	33.2
K8	Financing	0.3	82.0	14.2	21.4
K9	Financing	0.5	86.0	19.3	26.2

Predictor	Group	$q_{0.01}$	$q_{0.99}$	Median	IQR
K10	Random	-2.3	2.3	0.0	1.4
K11	Profitability	-29.2	76.5	2.3	5.9
K12	Leverage	0.0	164.3	0.0	4.1
K13	Liquidity	-54.8	80.5	1.0	21.6
K14	Liquidity	0.0	47.9	2.0	7.1
K15	Liquidity	0.0	184.4	3.8	14.8
K16	Liquidity	2.7	503.2	63.5	58.3
K17	Liquidity	8.4	696.2	116.9	60.8
K18	Financing	2.4	95.3	47.8	38.4
K19	Investment	0.0	83.3	28.0	34.3

Predictor	Group	$q_{0.01}$	$q_{0.99}$	Median	IQR
K20	Leverage	1.1	3750.0	60.6	110.3
K21	Growth	-50.6	165.6	3.9	20.1
K22	Profitability	-510.5	1998.5	32.7	81.9
K23	Profitability	-16.7	63.1	8.4	11.0
K24	Cost structure	-3.7	36.0	1.1	1.9
K25	Financing	0.4	84.0	17.6	25.4
K26	Growth	0.0	108.5	24.2	32.6
K27	Financing	1.7	89.6	24.7	30.0
K28	Growth	1.0	77.8	21.8	18.1
K29	Cost structure	-1338.6	34350.0	159.0	563.2

Predictor	Group	$q_{0.01}$	$q_{0.99}$	Median	IQR
K30	Liquidity	-14.1	116.4	5.2	8.9
K31	Activity	0.0	342.0	42.9	55.8
K32	Financing	0.3	98.5	58.4	48.4
K33	Other	4.9	13.0	7.9	2.1

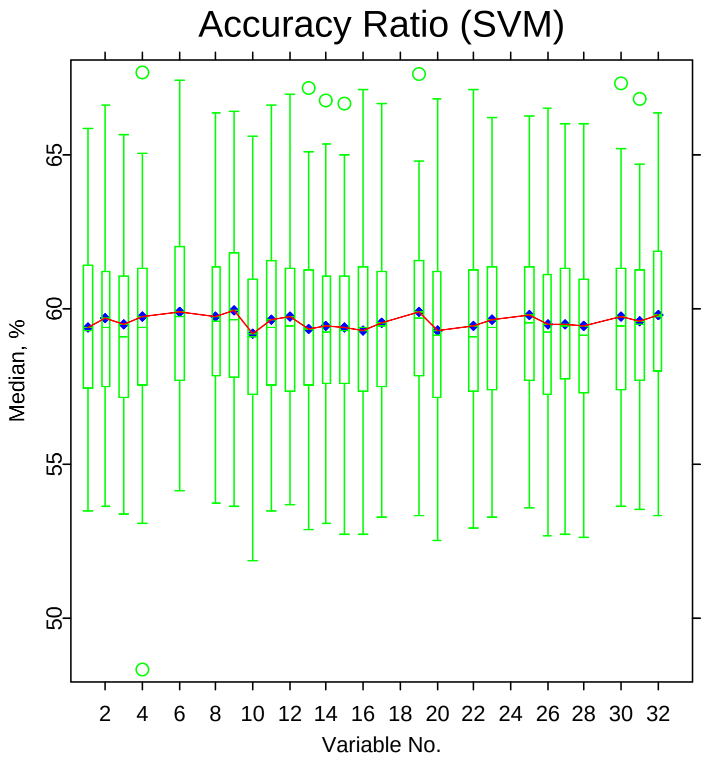


Figure 3: AR for several models. The SVM model with the highest AR including variables K5, K29, K7, K33, K18, K21, K24 and alternatively one of the remaining variables.

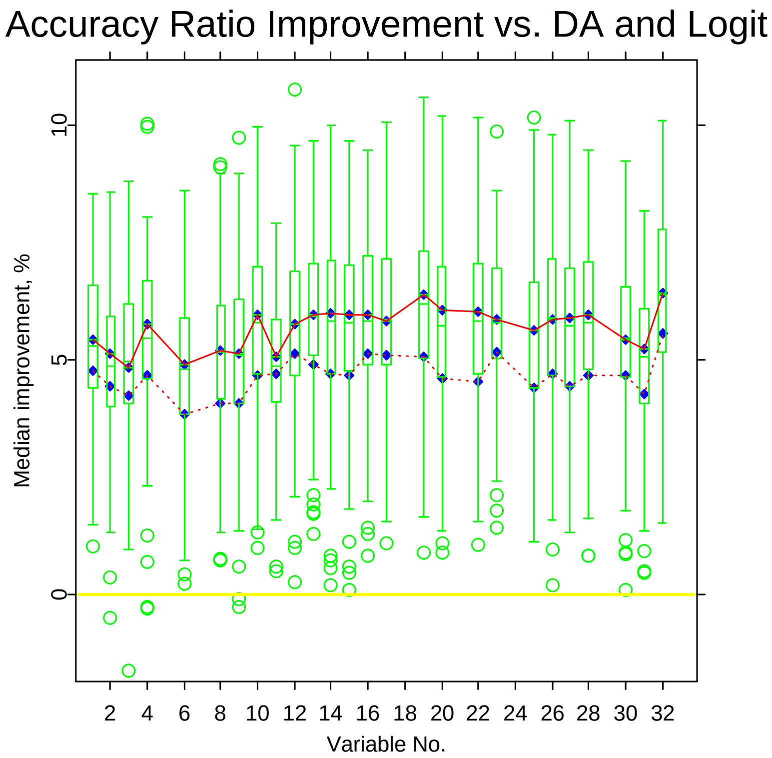


Figure 4: Improvement in AR of SVM vs. robust DA and Logit. Variables included are K5, K29, K7, K33, K18, K21, K24 and alternatively one of the remaining variables.

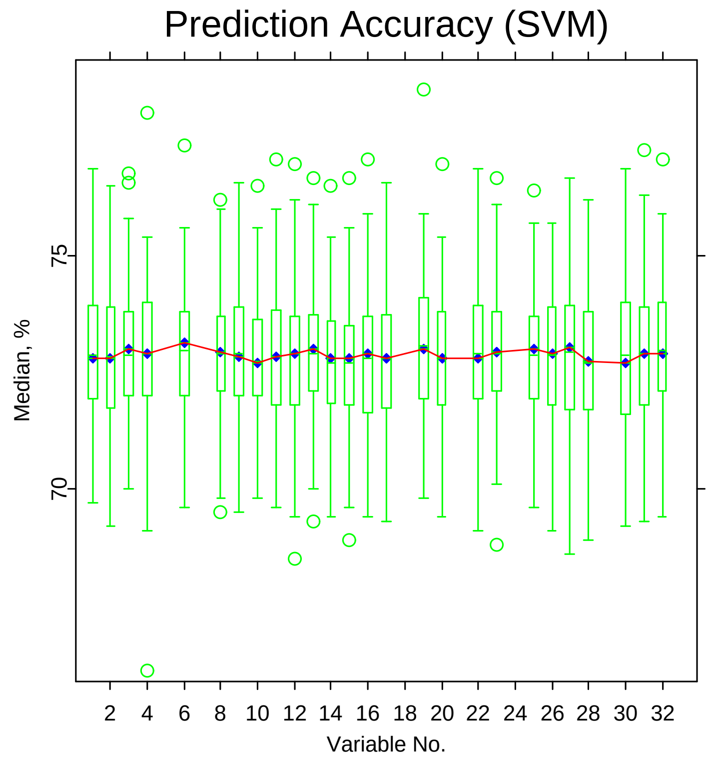


Figure 5: Prediction accuracy for several models. The SVM model with the highest AR including variables K5, K29, K7, K33, K18, K21, K24 and alternatively one of the remaining variables.

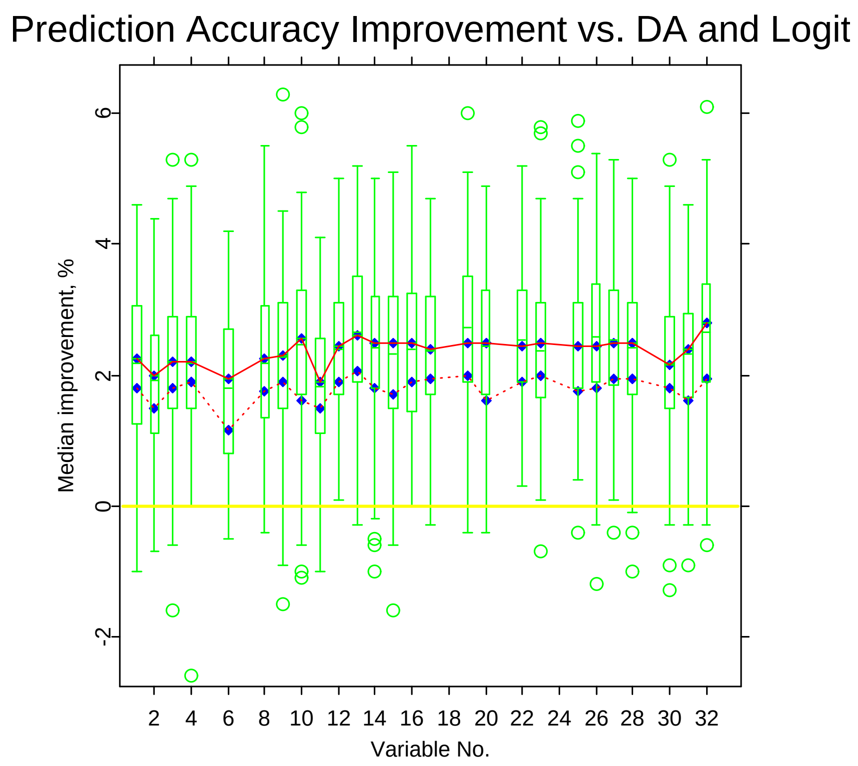


Figure 6: Improvement in prediction accuracy of SVM vs. robust DA and Logit. Variables included are K5, K29, K7, K33, K18, K21, K24 and alternatively one of the remaining variables.

Out-of-sample Classification Results

The model for which the highest AR is obtained is analysed. It includes:

K5: debt cover

K29: interest coverage ratio

K7: days payable

K33: company size

K18: short term debt ratio

K21: net income change

K24: net interest ratio

K9: equity ratio (adj.)

All 8150 observations of bankrupt companies are included

Comparison Procedure

The data used with DA and logit regressions were first cleared of outliers:

$$\text{if } x_i < q_{0.05} \text{ then } x = q_{0.05}$$

$$\text{if } x_i > q_{0.95} \text{ then } x = q_{0.95}$$

SVM did not require any data preprocessing

All estimations were repeated on 100 subsamples of all 8150 insolvent and the same number of solvent company observations selected randomly. Each subsample was evenly divided into a training and validation set.

All estimates are medians, i.e. robust measures.

Support Vector Machines

		Estimated median	
		Bankrupt	Non-bankrupt
Data	Bankrupt	79.0%	21.0%
	Non-bankrupt	31.3%	68.7%

Accuracy Ratio: 62.0%

Prediction Accuracy: 73.8%

SVM vs. DA Improvement

		Estimated median	
		Bankrupt	Non-bankrupt
Data	Bankrupt		0.8%
	Non-bankrupt	4.6%	

Accuracy Ratio Improvement: 5.2%

Prediction Accuracy Improvement: 2.7%

SVM vs. Logit Improvement

		Estimated median	
		Bankrupt	Non-bankrupt
Data	Bankrupt		1.3%
	Non-bankrupt	2.9%	

Accuracy Ratio Improvement: 5.2%

Prediction Accuracy Improvement: 2.0%

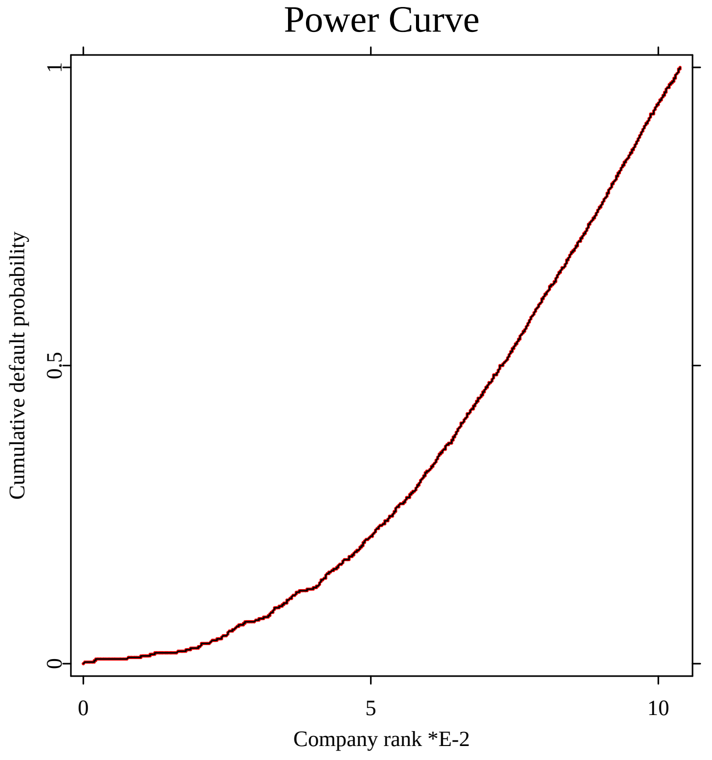


Figure 7: Power (Lorenz) curve for an SVM.

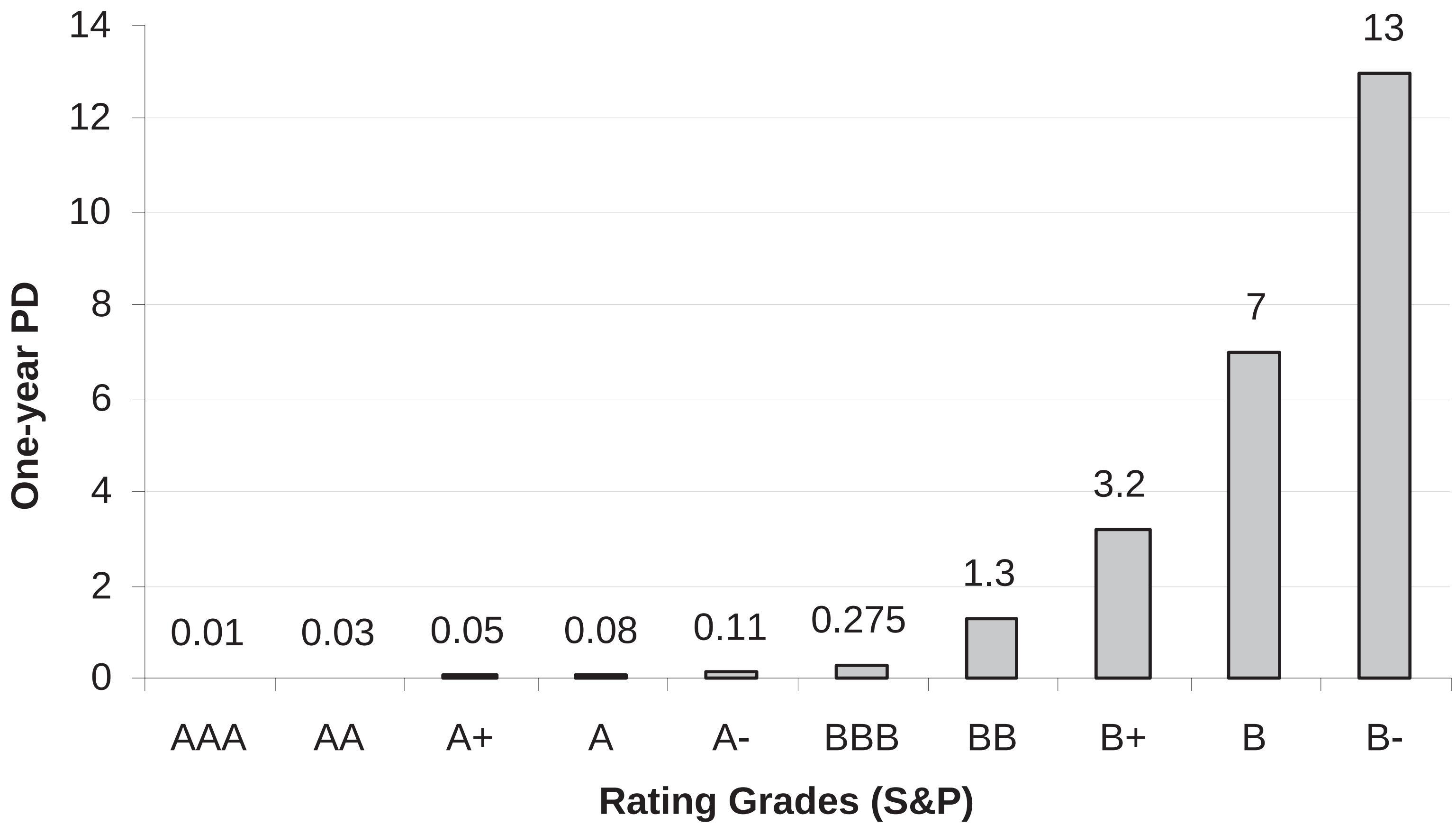
Economic Effects of Introducing SVMs

On the Bundesbank data (8150 bankruptcies) SVM can deliver forecasting accuracy 2% better than DA and logistic regression. Around 500 bankruptcies happen each year out of 20000 companies.

This is translated into

- ca. 10 avoided bankruptcy losses a year or one a month and
- 400 more companies becoming eligible for credit a year

Rating Grades and Probabilities of Default



Convertion of Scores into PDs

The score values $f = x^\top w + b$ estimated by an SVM correspond to default probabilities:

$$f \mapsto PD$$

The only assumption: the higher f the higher is PD

The mapping procedure:

1. Estimate PDs for companies of the training set: select $2 * h + 1$ nearest neighbours including the observation itself in terms of score; compute empirical PD for the observation i as

$$PD_i = \frac{\#Insolvencies(i-h, i+h)}{\#all(i-h, i+h)}$$

Convertion of Scores into PDs

2. Monotonise the PDs so that the dependence of PD from score be monotonical using the Pool Adjacent Violator algorithm

3. Compute a PD for any other company as a weighted average of neighbouring points of the training set in terms of score using kernels

$$PD(x) = \sum_{i=1}^{n} w_i(x) PD_i$$

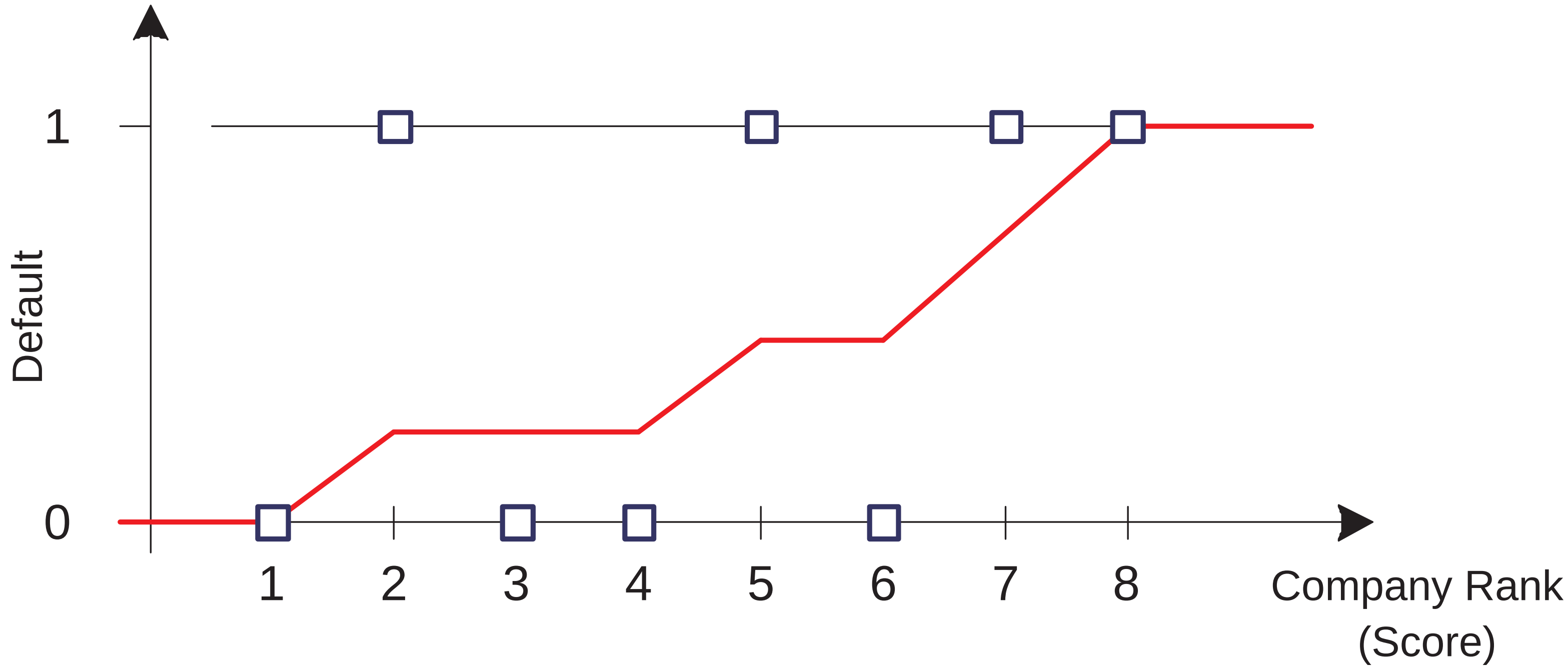


Figure 8: Cumulative default rate as a function of score.

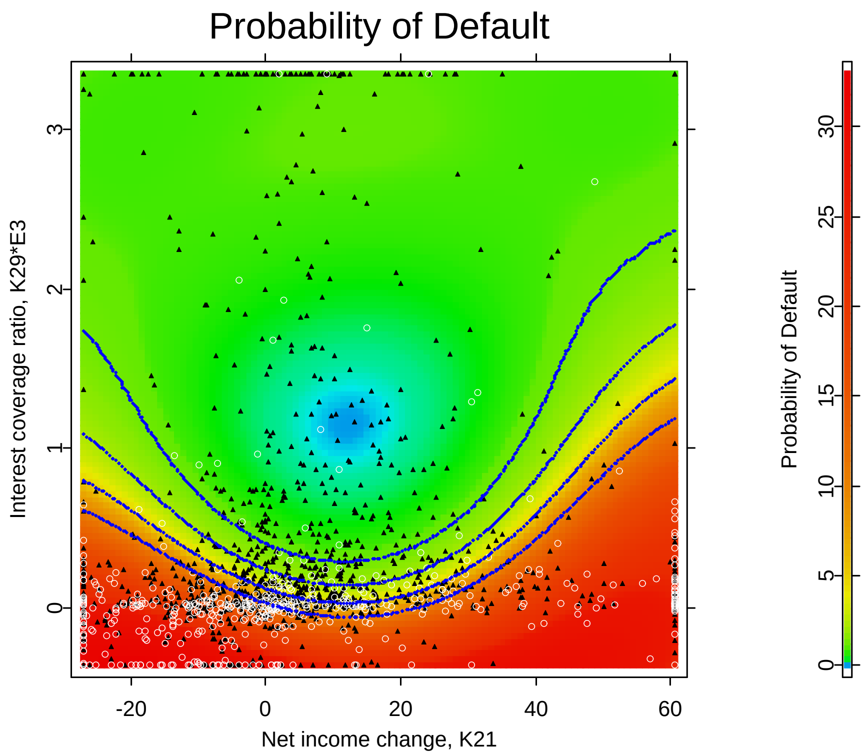


Figure 9: Estimation of PDs. The boundaries of six risk classes are shown, which correspond to the rating classes: BBB and above (investment grade), BB, B+, B, B- and lower.

Conclusions

- **The rating method must** *be suitable for a great number of evaluated companies...*

  The SVM was extensively tested with the complete Bundesbank data set in 50000 different data and variable configurations.

- *...have a systematic inner structure, be reproducible (reliable) and produce comparable (stable) results in time...*

  The SVM delivers a stable and unique solution, the model is not changed unless crucially different information arrives in time.

- *...be robust with a high generalisation ability...*

  The SVM produces consistent estimates with different data; generalisation ability is *optimised* to achieve the highest accuracy.

Conclusions

- **The rating method must** *have a high forecasting accuracy (low misclassification rate)...*

  SVM reliably exceeds both DA and Logit in forecasting accuracy (2% lower misclassification rate, 6% higher AR). The improvement is highly significant even for small data sets.

- *...deliver results free from economic inconsistencies...*

  The flexibility of the SVM structure allows to avoid models not supported with economic data.

- *...provide a comprehensive and well-balanced analysis of the core operating areas (capital structure, liquidity, profitability)...*

  The SVM offers more types of analysis including the analysis of complex non-linear interdependencies between operating areas.

Conclusions

- **The rating method must** *be transparent in producing the results, be practically convenient for credit departments and acceptable by companies...*

  The SVM is based on widely accepted principles; its solution can be representable in an easily understandable traditional form.

- *...be suitable for practical implementations...*

  The SVM is easily implementable and controlled without any special skills. Besides PDs it is well suitable for evaluating LGDs and effects of monetary policy.

- *...be applicable for creating multiple rating classes...*

  The PDs estimated with an SVM form a basis for building rating classes.

References

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *The Journal of Finance*, September: 589-609.

Basel Committee on Banking Supervision (2003). The New Basel Capital Accord, third consultative paper, *http://www.bis.org/bcbs/cp3full.pdf*.

Beaver, W. (1966). Financial Ratios as Predictors of Failures. Empirical Research in Accounting: Selected Studies, *Journal of Accounting Research*, supplement to vol. 5: 71-111.

Falkenstein, E. (2000). *RiskCalc for Private Companies: Moody's Default Model*, Moody's Investors Service.

Füser, K. (2002). Basel II – was muß der Mittelstand tun?, *http://www.ey.com/global/download.nsf/Germany/ Mittelstandsrating/$file/Mittelstandsrating.pdf*.

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

Martin, D. (1977). Early Warning of Bank Failure: A Logit Regression Approach, *The Journal of Banking and Finance*, 249-276.

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *The Journal of Finance*, **29**: 449-470.

Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, Spring: 109-131.

Platt, J.C. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, *Technical Report MSR-TR-98-14*, April.

Division of Corporate Finance of the Securities and Exchange Commission (2004). Standard industrial classification (SIC) code list, *http://www.sec.gov/info/edgar/siccodes.htm*.

Securities and Exchange Commission (2004). Archive of Historical Documents, *http://www.sec.gov/cgi-bin/srch-edgar*.

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

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