Graphical Data Representation in Bankruptcy Analysis

Wolfgang K. Härdle*
Rouslan A. Moro**
Dorothea Schäfer***

* C.A.S.E. - Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin, Germany
** C.A.S.E. - Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin, Germany, and
German Institute for Economic Research, Berlin, Germany
*** German Institute for Economic Research, Berlin, Germany

Graphical Data Representation in Bankruptcy Analysis

W. K. Härdle[1], R. A. Moro[2], and D. Schäfer[3]

[1] C.A.S.E., Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin
   `stat@wiwi.hu-berlin.de`
[2] C.A.S.E., Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin and
   German Institute for Economic Research, Königin-Luise-Straße 5, 14195 Berlin
   `rmoro@diw.de`
[3] German Institute for Economic Research, Königin-Luise-Straße 5, 14195 Berlin
   `dschaefer@diw.de`

Graphical data representation is an important tool for model selection in bankruptcy analysis since the problem is highly non-linear and its numerical representation is much less transparent. In classical rating models a convenient representation of ratings in a closed form is possible reducing the need for graphical tools. In contrast to that non-linear non-parametric models achieving better accuracy often rely on visualisation. We demonstrate an application of visualisation techniques at different stages of corporate default analysis based on Support Vector Machines (SVM). These stages are the selection of variables (predictors), probability of default (PD) estimation and the representation of PDs for two and higher dimensional models with colour coding. It is at this stage when the selection of a proper colour scheme becomes essential for a correct visualisation of PDs. The mapping of scores into PDs is done as a non-parametric regression with monotonisation. The SVM learns a non-parametric score function that is, in its turn, non-parametrically transformed into PDs. Since PDs cannot be represented in a closed form, some other ways of displaying them must be found. Graphical tools give this possibility.

Keywords: company rating, default probability, support vector machines, colour coding

JEL classification: C14, G33, C45

1 Company Rating Methodology

Application of statistical techniques to corporate bankruptcy started in the 60's with the development of computers.[1] The first technique introduced was discriminant analysis (DA) for univariate [3] and multivariate models [1]. After DA the logit and probit models were introduced in [14] and [16]. Nowadays these models are widely used in practice, i.e. they are at the core of the rating solutions at most European central banks. The solution in the traditional framework is a linear function (a hyperplane in a multidimensional feature space) separating successful and failing companies. A company score is computed as a value of that function. In the case of the probit and logit models the score can be directly transformed into a probability of default (PD), which denotes the probability with which a company can go bankrupt within a certain period. The major disadvantages of these popular approaches is the linearity of the solution and, in the case of logit and probit models, the prespecified form of the link function between PDs and the linear combination of predictors (Figure 1).

In Figure 1 successful and failing companies are denoted with black triangles and white quadrangles respectively. There is an equal number of companies of both classes in the sample. Following the DA and logit classification rule, which give virtually the same result, we are more likely to find a failing company above and to the right from the straight line. This may lead to a conclusion that companies with significantly negative values of operating profit margin and equity ratio can be classified as successful. This, for example, allows for companies with liabilities much greater than total assets to be classified as successful. Such a situation is avoided by using a non-linear classification method, such as the SVM, which produces a non-linear boundary.

Following a traditional approach we would expect a monotonic relationship between predictors and PDs, like the falling relation for the interest coverage ratio (Figure 2). However, in reality this dependence is often non-monotonic as for such important indicators as the company size or net income change. In the latter case companies that grow too fast or too slow have a higher probability of default. That is the reason for contemplating non-linear techniques as alternatives. Two prominent examples are recursive partitioning [4] and neural networks [17]. Despite the strength of the two approaches they have visible drawbacks: orthogonal division of the data space in recursive partitioning that is usually not justified and heuristic model specification in neural networks.
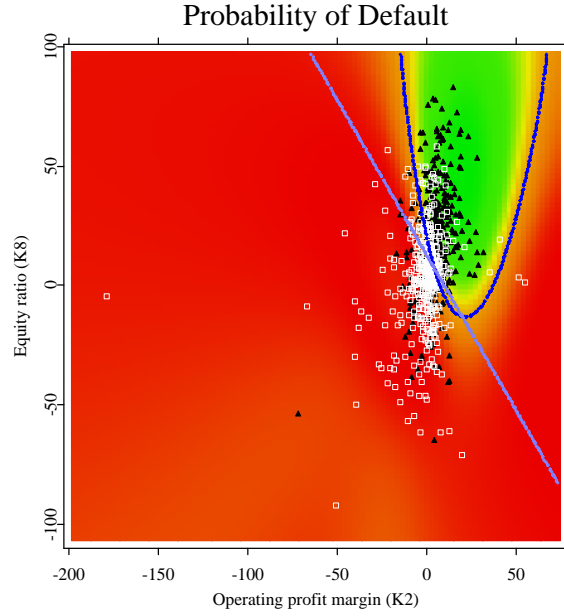

Fig. 1. A classification example. The boundary between the classes of solvent and insolvent companies was estimated using DA and logit regression (two indistinguishable linear boundaries) and an SVM (a non-linear boundary).
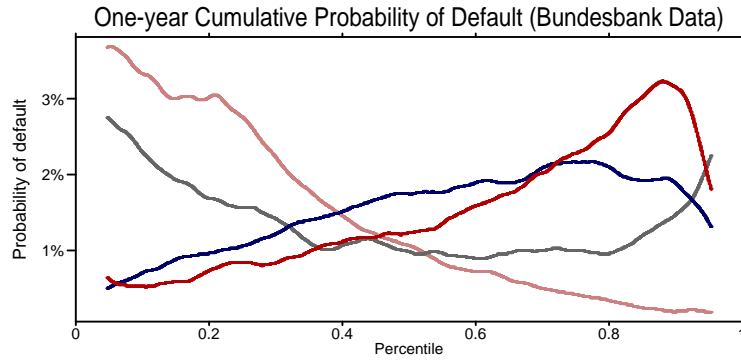

Fig. 2. One year cumulative PDs evaluated for several financial ratios on the German Bundesbank data. The ratios are net income change, K21 (gray), net interest ratio, K24 (red), interest coverage ratio, K29 (pink) and logarithm of total assets, K33 (blue). The k-nearest-neighbours procedure was used with the size of the window being around 8% of all observations. The total number of observations is 553500.

Recursive partitioning, also known as classification and regression trees (**CART**) performs classification by orthogonally dividing the data space. At each step only a division (split) along one of the axes is possible. The axis is chosen such, that a split along it reduces the variance in each of the subspaces and maximises the variance between them. Entropy based criteria can also be used. The visible drawback is the orthogonal division itself which imposes severe restrictions on the smoothness of the classifying function and may not adequately capture the correlation structure between the variables. Orthogonal division means that the separating hyperplane can only consist of orthogonal segments parallel to the coordinate grid, whereas the boundary between the classes has a smoothly changing gradient.

The **neural network** (NN) is a network of linnear classifiers (neurons) that are connected with one another in a prespecified way. The outputs of some of the neurons are inputs for others. The performance of a NN greatly depends on its structure that must be adapted for solving different problems. The network must be designed manually that requires a substantial experience from the operator. Moreover, NNs mostly do not povide a global solution but only a local one. This feature, as well as too much heuristics create many obstacles on the way of using NNs at the rating departments of banks.

We would like to have a model that is able to select a classifying function based on very general criteria. The SVM is a statistical technique that in many applications, such as optical character recognition and medical diagnostics, showed very good performance. It has a flexible solution and is controlled by adjusting only few parameters. Its overall good performance and flexibility make the SVM a suitable candidate [9].

Within a rating methodology each company is described by a set of variables x, such as financial ratios. Financial ratios, such as debt ratio (leverage) or interest coverage (earnings before interest and taxes to interest) characterise different sides of company operation. They are constructed on the basis of balance sheets and income statements. For example, the Bundesbank uses 32 ratios (predictors) computed using the company statements from its corporate bankruptcy data base. The predictors and basic statistics are given in Table 1. The whole Bundesbank data base covers the period 1987–2005 and consists of 553500 anonymised statements of solvent and insolvent companies. Most companies appear in the database several times in different years.

The class y of a company can be either $y = -1$ ('successful') or $y = 1$ ('bankrupt'). Initially, an unknown classifier function $f : x \rightarrow y$ is estimated on a training set of companies (x_i, y_i), $i = 1, ..., n$. The training set represents the data for companies which are known to have survived or gone bankrupt. In order to obtain PDs from the estimated scores f, rating practitioners usually rely on prespecified rating classes (i.e. BBB, C, AA, etc.). A certain range of scores and PDs belong to each rating class. The ranges are computed on the basis of historical data. To derive a PD for a newly scored company its score f is compared with the historical values of f's for each class. Basing on the

similarity of the scores a company is assigned to one particular class. The PD of this class becomes the PD of the company.
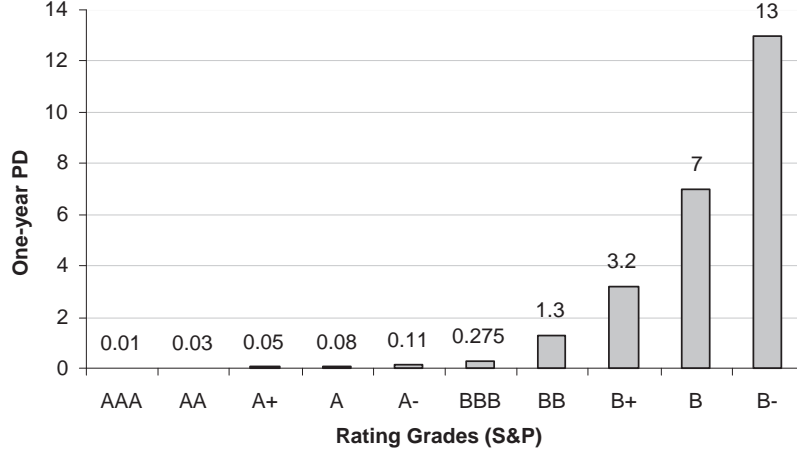

Fig. 3. One year probabilities of default for different rating grades [5].

Company bond ratings play an important role in determining the cost of debt refinancing since they reflect the probability of defaulting on the debt (Figure 3). One can notice that the differences between the classes in terms of PDs are not the same. For example, the PD increases by 6.7% or 24 times between classes BBB and B, but only by 0.07 or 8 times between classes AAA und A. The colours for coding PDs must be selected so that perceptually the classes would look equidistant, no matter what their absolute PD is. This can be achieved by using an appropriate colour scheme and colour distance scaling. The use of the HLS colour scheme in combination with a logarithmic colour scaling will be demonstarated in section 6.

2 The SVM Approach

The SVM [18] is a regression (and classification) technique that is based on margin maximisation (Figure 4) between two data classes. The margin is the distance between the hyperplanes bounding each class where in a linear perfectly separable case no observation may lie. The classifier function used by the linear SVM is a hyperplane symmetrically surrounded with a margin zone. It can be shown [9] that by maximising the margin one reduces the complexity of such a classifier. By applying kernel techniques the SVM can be extended to learn non-linear classifying functions (Figure 5).

In Figure 4 misclassifications are unavoidable when using linear classifying functions (linearly non-separable case). To account for misclassifications
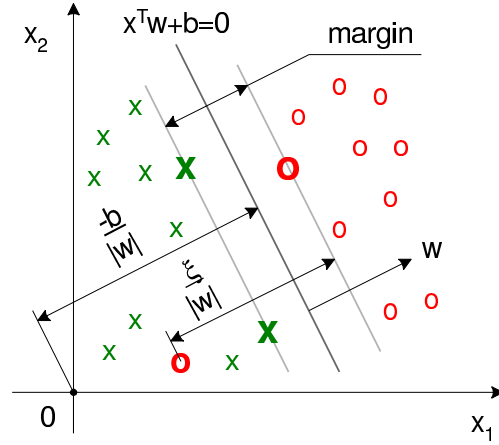
Fig. 4. The separating hyperplane $x^\top w + b = 0$ and the margin in a non-separable case. The observations marked with bold crosses and zeros are support vectors. The hyperplanes bounding the margin zone equidistant from the separating hyperplane are represented as $x^\top w + b = 1$ and $x^\top w + b = -1$.
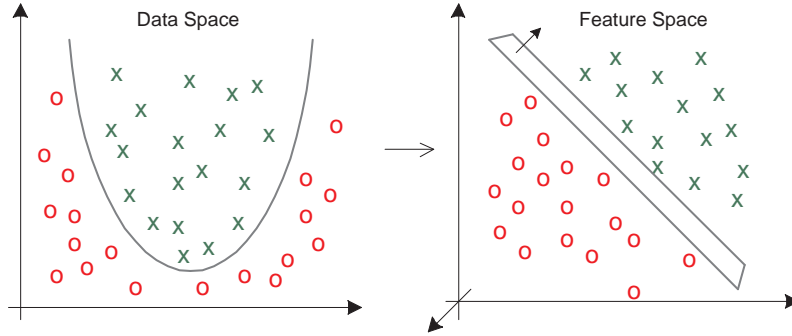

Fig. 5. Mapping from a two-dimensional data space into a three-dimensional space of features $\mathbb{R}^2 \mapsto \mathbb{R}^3$ using a quadratic kernel function $K(x_i, x_j) = (x_i^\top x_j)^2$. The three features correspond to the three components of a quadratic form: $\tilde{x}_1 = x_1^2$, $\tilde{x}_2 = \sqrt{2}x_1 x_2$ and $\tilde{x}_3 = x_2^2$, thus, the transformation is $\Psi(x_1, x_2) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2)^\top$. The data separable in the data space with a quadratic function will be separable in the feature space with a linear function. A non-linear SVM in the data space is equivalent to a linear SVM in the feature space. The number of features will grow fast with d and the degree of the polynomial kernel p, which equals 2 in our example, making the closed-form representation of Ψ such as here practically impossible

the penalty ξ_i is introduced, which is related to the distance from the hyperplane bounding observations of the same class to observation i. $\xi_i > 0$ if a misclassification occurs. All observations satisfy the following two constraints:

$$y_i(x_i^\top w + b) \geq 1 - \xi_i, \tag{1}$$

$$\xi_i \geq 0. \tag{2}$$

With the normalisation of w, b and ξ_i as in (1) the margin equals to $2/\|w\|$. The convex objective function to be minimised given the constraints (1) and (2) is:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i. \tag{3}$$

The parameter C called capacity is related to the width of the margin zone. The smaller the C is, the bigger margins are possible. Using well established theory for optimisation of convex functions [6] we can derive the dual Lagrangian:

$$L_D = \frac{1}{2}w(\alpha)^\top w(\alpha) - \sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\delta_i\alpha_i + \sum_{i=1}^{n}\gamma_i(\alpha_i - C) - \beta\sum_{i=1}^{n}\alpha_i y_i \tag{4}$$

for the dual problem:

$$\min_{\alpha_i,\delta_i,\gamma_i,\beta}\max_{w_k,b,\xi_i} L_D, \tag{5}$$

Here for a linear SVM:

$$w(\alpha)^\top w(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^\top x_j. \tag{6}$$

For obtaining non-linear classifying functions in the data space a more general form is applicable:

$$w(\alpha)^\top w(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j). \tag{7}$$

The function $K(x_i, x_j)$ is called a kernel function. Since it has a closed form representation, the kernel is a convenient way of mapping low dimensional data into a highly dimensional (often infinitely dimensional) space of features. It must satisfy the Mercer conditions [15], i.e. be symmetric and semipositive definite or, in other words, represent a scalar product in some Hilbert space [19].

In our study we applied an SVM with an anisotropic Gaussian kernel

$$K(x_i, x_j) = \exp\left\{-(x_i - x_j)^\top r^{-2}\Sigma^{-1}(x_i - x_j)/2\right\}, \tag{8}$$

where r is a coefficient and Σ is a variance-covariance matrix. The coefficient r is related to the complexity of classifying functions: the hgher the r is, the lower is the complexity. If kernel functions allow for sufficiently rich feature spaces, the performances of SVMs are comparable in terms of out-of-sample forecasting accuracy [18].

3 Company Score Evaluation

The company score is computed as:

$$f(x) = x^\top w + b, \tag{9}$$

where $w = \sum_{i=1}^{n} \alpha_i y_i x_i$ and $b = \frac{1}{2}(x_+ + x_-)^\top w$; x_+ and x_- are the observations from the opposite classes for which constraint (1) becomes equality. By substituting the scalar product with a kernel function we will derive a non-linear score function:

$$f(x) = \sum_{i=1}^{n} K(x_i, x)\alpha_i y_i + b. \tag{10}$$

The non-parametric score function (10) does not have a compact closed form representation. This necessitates the use of graphical tools for its visualisation.

4 Variable Selection

In this section we describe the procedure and the graphical tools for selecting the variables of the SVM model used in forecasts. We have two most important criteria of model accuracy: the accuracy ratio (AR), which will be used here as a criterion for model selection, (Figure 6) and the percentage of correctly classified out-of-sample observations. Higher values indicate better model accuracy.
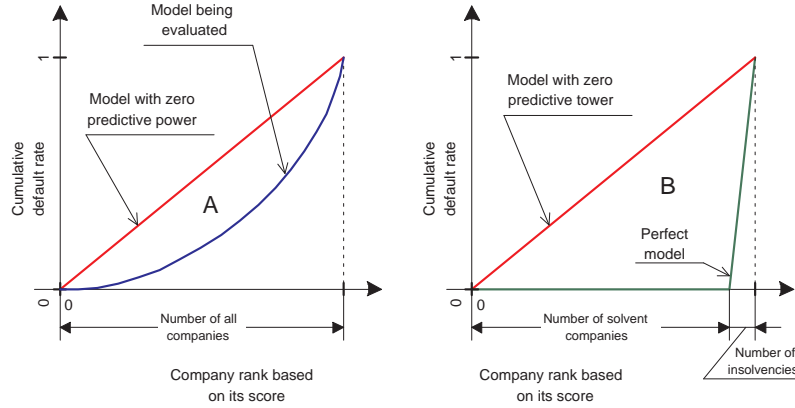

Fig. 6. The power curves for a perfect (green), random (red) and some real (blue) classification models. The AR is the ratio of two areas A/B. It lies between 0 for a random model with no predictive power and 1 for a perfect model.

We start model selection from the simplest, i.e. univariate models and then pick up the one with the highest AR. The problem that arises is how to determine the variable which provides the highest AR across possible data samples. For a parametric model we would need to estimate the distribution of the coefficients at the variables and, hence, their confidence intervals. This approach, however, is practically irrelevant for non-parametric models.

Instead we can compare goodness of models with respect to some accuracy measure, in our case AR. Firstly we will estimate the distributions of AR for different models. This can be done using bootstrapping [12]. We randomly select training and validation sets as subsamples of 500 solvent and 500 insolvent companies each. We used the 50/50 ratio since this is the worst case with the minimum AR. The two sets are not overlapping, i.e. do not contain common observations. For each of these sets we apply the SVM with parameters that provide the highest AR for bivariate models (Figure 7) and estimate ARs. Then we perform a Monte Carlo experiment: repeat the generation of subsamples and computing of ARs 100 times. Each time we will record the ARs and then estimate their distribution.

At the end of this procedure we obtain an empirically estimated distribution of AR on bootstrapped subsamples. The median AR provides a robust measure to compare different variables as predictors. The same approach can be used for comparing SVM with DA and logit regression in terms of predictive power. We compute AR for the same subsamples with the SVM, DA and logit models. The median improvements in AR for the SVM over DA and the SVM over the logistic regression are also reported below.

We will start this procedure with all univariate models with 33 variables K1-K9, K11-K33 as they are denoted at the Bundesbank and variable K10, which is a standard normal random variable used as a reference (Table 1). For each model the resulting distribution of ARs will be represented as box plots (Figure 8). The red line depicts medians. The box within each box plot shows the interquartile range (IQR), while the whiskers span to the distance of 3/2 IQR in each direction from the median. Outliers beyond that range are denoted with circles.

Basing on Figure 8 we can conclude that variables K5 (Debt Cover) and K29 (Interest Coverage Ratio) provide the highest median AR around 50%. We can also notice that variables K12, K26 and K28 have a very low accuracy: their median ARs do not exceed 11.5%. The model based on random variable K10 has AR equal zero, in other words, it has no predictive power whatsoever. For the next step we will select variable K5 that was included in the best univariate model.

For bivariate models we will select the best predictor from the univariate models (K5) and one of the rest that delivers the highest AR (K29) (Figure 9). This procedure will be repeated for each new variable added. The AR is growing until the model has eight variables, then it slowly declines. Median ARs for the models with eight variables are shown in Figure 10.
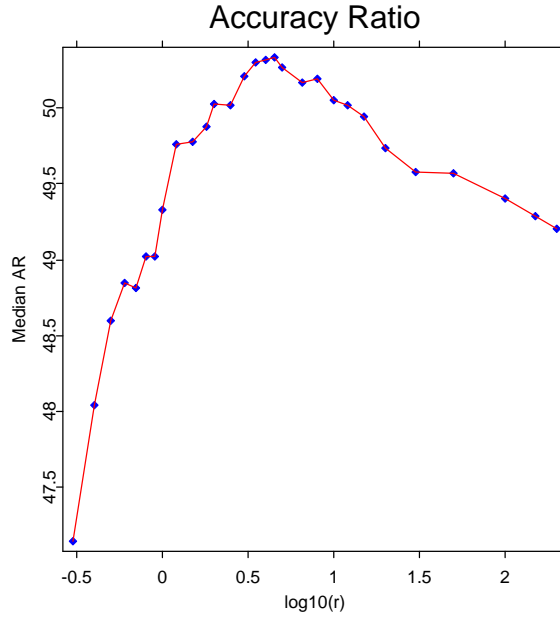
Fig. 7. The relationship between an accuracy measure (AR) and the coefficient r in the SVM formulation. Higher r's correspond to less complex models. The median ARs were estimated on 100 bootstrapped subsamples of 500 solvent and 500 insolvent companies both in the training and validation sets. A bivariate SVM with the variables K5 and K29 was used. We will be using $r = 4$ in all SVMs used in this chapter.

We have also conducted experiments with subsamples of the size of 5000 observations. The change of median was extremely small (one–two orders of magnitude smaller than the interquartile range). The interquartile range got narrower as it was expected, i.e. the difference between models with bigger samples is only more statistically significant. Thus, proving that if the difference is significant on a sample of 1000 observations, it can be guaranteed that this will remain so for bigger samples.

The SVM based on variables K5, K29, K7, K33, K18, K21, K24, K33 and K9 attains the highest median AR of around 60.0%. For comparison we plot an improvement in AR for the SVM vs. DA and logit regression on the same 100 subsamples. The data used in the DA and logit models were processed as following: if $x_i < q_{0.05}(x_i)$ then $x_i = q_{0.05}(x)$ and if $x_i > q_{0.95}(x_i)$ then $x_i = q_{0.95}(x_i)$; $i = 1, 2, \ldots, 8$; $q_\alpha(x_i)$ is an α quantile of x_i. Thus, the DA and logit regression applied were *robust versions* not sensitive to outliers. Without such a procedure the improvement would be much higher.

Table 1. Summary Statistics. q_α is an α quantile. IQR is the interquartile range.

Var.	Name	Group	$q_{0.01}$	Median	$q_{0.99}$	IQR
K1	Pre-tax profit margin	Profitability	-26.9	2.3	78.5	5.9
K2	Operating profit margin	Profitability	-24.6	3.8	64.8	6.3
K3	Cash flow ratio	Liquidity	-22.6	5.0	120.7	9.4
K4	Capital recovery ratio	Liquidity	-24.4	11.0	85.1	17.1
K5	Debt cover	Liquidity	-42.0	17.1	507.8	34.8
K6	Days receivable	Activity	0.0	31.1	184.0	32.7
K7	Days payable	Activity	0.0	23.2	248.2	33.2
K8	Equity ratio	Financing	0.3	14.2	82.0	21.4
K9	Equity ratio (adj.)	Financing	0.5	19.3	86.0	26.2
K10	Random Variable	Test	-2.3	0.0	2.3	1.4
K11	Net income ratio	Profitability	-29.2	2.3	76.5	5.9
K12	Leverage ratio	Leverage	0.0	0.0	164.3	4.1
K13	Debt ratio	Liquidity	-54.8	1.0	80.5	21.6
K14	Liquidity ratio	Liquidity	0.0	2.0	47.9	7.1
K15	Liquidity 1	Liquidity	0.0	3.8	184.4	14.8
K16	Liquidity 2	Liquidity	2.7	63.5	503.2	58.3
K17	Liquidity 3	Liquidity	8.4	116.9	696.2	60.8
K18	Short term debt ratio	Financing	2.4	47.8	95.3	38.4
K19	Inventories ratio	Investment	0.0	28.0	83.3	34.3
K20	Fixed assets ownership r.	Leverage	1.1	60.6	3750.0	110.3
K21	Net income change	Growth	-50.6	3.9	165.6	20.1
K22	Own funds yield	Profitability	-510.5	32.7	1998.5	81.9
K23	Capital yield	Profitability	-16.7	8.4	63.1	11.0
K24	Net interest ratio	Cost struct.	-3.7	1.1	36.0	1.9
K25	Own funds/pension prov. r.	Financing	0.4	17.6	84.0	25.4
K26	Tangible asset growth	Growth	0.0	24.2	108.5	32.6
K27	Own funds/provisions ratio	Financing	1.7	24.7	89.6	30.0
K28	Tangible asset retirement	Growth	1.0	21.8	77.8	18.1
K29	Interest coverage ratio	Cost struct.	-1338.6	159.0	34350.0	563.2
K30	Cash flow ratio	Liquidity	-14.1	5.2	116.4	8.9
K31	Days of inventories	Activity	0.0	42.9	342.0	55.8
K32	Current liabilities ratio	Financing	0.3	58.4	98.5	48.4
K33	Log of total assets	Other	4.9	7.9	13.0	2.1

Figure 11 represents the absolute improvement for SVM over robust DA (upper line) and SVM over robust logit regression (lower line). We can see that for all models containing variables K5, K29, K7, K33, K18, K21, K24 and one of the remaining variables the median AR was always higher for the SVM. Thus, the SVM model is always dominating in accuracy DA and logit regression with regard to AR.
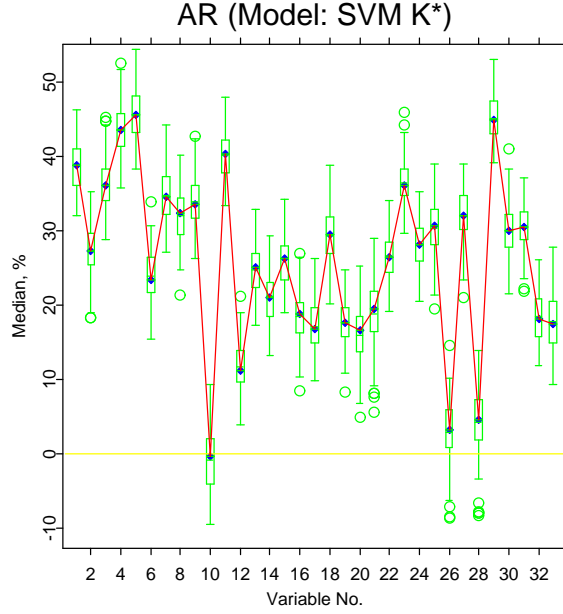
Fig. 8. Accuracy ratios for univariate SVM models. Box-plots are estimated basing on 100 random subsamples. The AR for the model containing only random variable K10 is zero.

5 Conversion of Scores into PDs

There is another way to look at the company score. It defines the distance between companies in terms of the distance to the boundary between the classes. The lower is the score, the farther is a company from the class of bankrupt companies, therefore, we can assume, the lower PD it must have. This means that the dependence between scores and PDs is assumed to be monotonous. This is the only kind of dependence that was assumed in all rating models mentioned in this chapter and the only one we use for PD calibration.

The conversion procedure consists of the estimation of PDs for the observation of the training set with a subsequent monotonisation (step one and two) and the computation of a PD for some new company (step three).

Step one is the estimation of PDs for the companies of the training set. We used kernel techniques to preliminary evaluate PDs for observation i from the training set, $i = 1, 2, \ldots, n$:

$$\widetilde{PD}(x_i) = \frac{\sum_{j=1}^{n} K_h(x_i, x_j) I_{\{y_j=1\}}}{\sum_{j=1}^{n} K_h(x_i, x_j)} \tag{11}$$
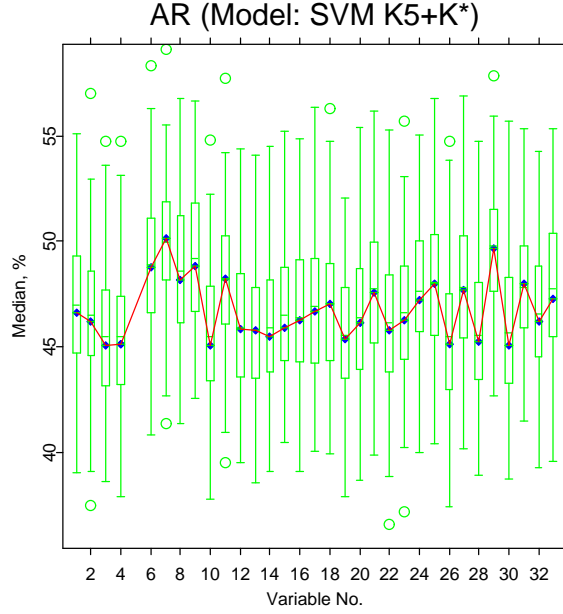
Fig. 9. Accuracy ratios for bivariate SVM models. Each model includes variable K5 and one of the remaining. Box-plots are estimated basing on 100 random subsamples.

Here a k-nearest-neighbour Gaussian kernel was used. h is the kernel bandwidth.

The preliminary PDs evaluated in this way are not necessarily a monotonical function of the score. The monotonisation of $\widetilde{PD}_i$, $i = 1, 2, \ldots, n$ is achieved at step two using the Pool Adjacent Violator (PAV) algorithm ([2] and [13]). As a result we obtain monotonised probabilities of default $PD(x_i)$ for the observations of the training set.

Finally, at step three the PDs are computed for any observation described with x as an interpolation between two PDs of the neighbouring, in terms of the score, observations from the training set, x_i and x_{i-1}, $i = 2, 3, \ldots, n$:

$$PD(x) = PD(x_i) + \frac{f(x) - f(x_{i-1})}{f(x_i) - f(x_{i-1})} \left\{ PD(x_i) - PD(x_{i-1}) \right\}. \qquad (12)$$

If the score for an observation x lies beyond the range of scores for the training set, then $PD(x)$ equals to the score of the first neighbouring obseration of the training set.

Figure 12 is an example of the cumulative PD curve (power curve) and estimated PDs for a subsample of 200 companies. The PD curve has a plateau area for the observations with a high score. Default probabilities can change from 15% to 80% depending on the score.
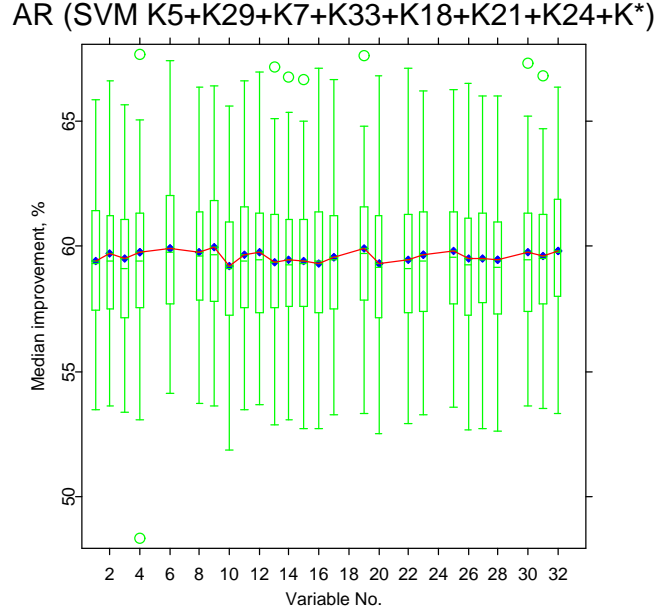
Fig. 10. Accuracy ratios for SVM models with eight variables. Each model includes variables K5, K29, K7, K33, K18, K21, K24 and one of the remaining. Box-plots are estimated basing on 100 random subsamples.

6 Colour Coding of PDs

The RGB colour space is based on three primary colours, red, green and blue, that are mixed to produce other ones. It is the colour coding scheme that is used in monitors and TV. It is, however, inconvenient for colour coding since we would like to make adjustments only to the channel responsible for colour while keeping lightness and saturation constant. This can be achieved with the HLS colour space.

We will represent the probability of default (PD) estimated with the SVM as two-dimensional plots where each colour represents a specific PD. The PD is a number that can be represented on a gray scale, e.g. in the RGB encoding as (i, i, i) with i changing from 0 to 255, e.g. the colour R=255, G=0, B=0 corresponds to red, and R=255, G=0, B=255 to violet, etc..

The HLS stands for hue, lightness (or luminance) and saturation. By adjusting only hues and keeping luminance and saturation fixed we can generate simulated colours from the range shown in Figure 14. The pure red colour corresponds to H=0 or 360, the pure green to H=120 and the pure blue to H=240.

Fig. 11. Median improvement in AR. SVM vs. DA (the upper line) and SVM vs. logit regression (the lower line). Box-plots are estimated basing on 100 random subsamples for the case of DA. Each model includes variables K5, K29, K7, K33, K18, K21, K24 and one of the remaining

Red colour is often used in finance to highlight negative information, while green and blue are used to convey positive information. Therefore, we would like to code PDs with colours ranging from red for the highest PD to blue-green for the most solvent company. For this end we normalise PDs so, that the lowest PD corresponds to the hue equal 180 (green-blue) or 120 (green) while the highest PD corresponds to the hue equal 0 (red). The resulting graph that shows the data and PDs in the dimensions of variables K33 and K29 is shown in Figures 15–17. The three figures correspond consequently to three SVMs with high, average and high complexity. The saturation was fixed at 0.85 to make colours look more noble and the luminance was fixed at 0.46, the maximum possible value for the chosen saturation. The HLS colours obtained in this way were transformed into RGB ones and plotted by the XploRe ([8], [7]) as a contour plot. The outliers that lie beyond the 5% and 95% quantiles are plotted at the rand.

To produce the plot a grid was generated with 101 steps both in horizontal and vertical directions. For each point of the grid a PD was estimated and represented in the HLS colour encoding. Then the HLS colour was converted
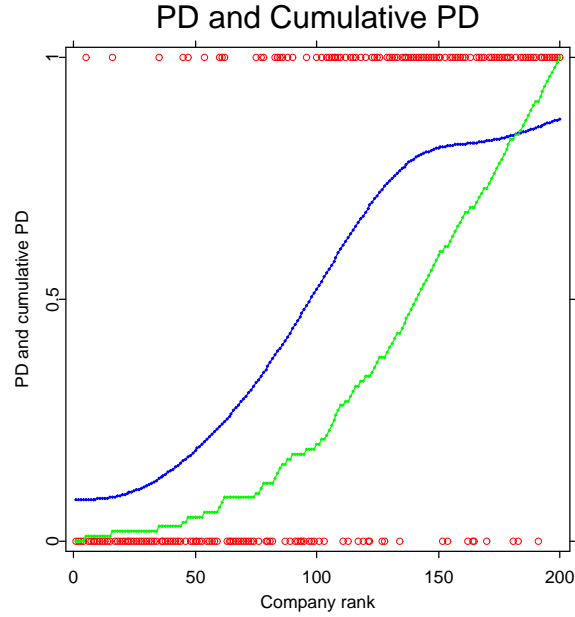
Fig. 12. PD (blue line) and cumulative PD (green line) estimated with the SVM for a subsample of 200 observation from the Bundesbank data. The variables were included into the model that achieved the highest AR: K5, K29, K7, K33, K18, K21, K24 and K9. The higher is the score, the higher is the rank of a company.

into an RGB colour and plotted by the XploRe as a small filled quadrangle. The quadranges evenly cover the whole area giving a continuous PD representation. The contour lines can be also added to the graph as illustrated by Figure 18.

7 Conclusion

In this chapter we demonstrated the application of graphical tools for variable selection, data visualisation and financial information representation and discussed such essential aspects of graphical analysis as colour coding. We believe that graphical analysis will have an increased importance as non-parametric models, such as SVM, are becoming more and more popular. On the other hand graphical representation can facilitate the acceptance of non-parametric models in various areas, e.g. finance, medicine, sound and image processing, etc. This will contribute to the development of those areas since non-linear non-parametric models better represent reality and provide higher forecasting accuracy.
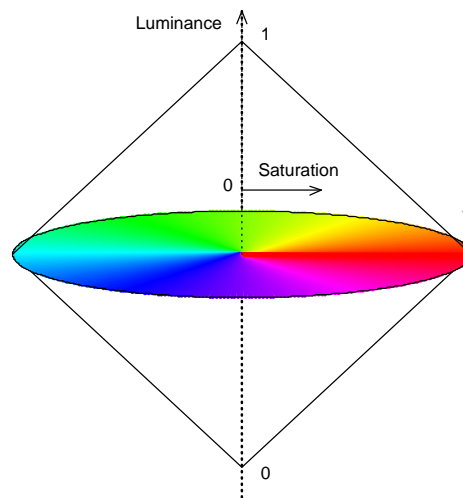
Fig. 13. The luminance and saturation dimensions of the HLS colour space. We will keep luminance and saturation constant and encode the information about PDs with hue.

References

1. Altman, E.: 1968, 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy'. *The Journal of Finance* pp. 589–609.
2. Barlow, R. E., J. M. Bartholomew, J. M. Bremmer, and H. D. Brunk: 1972, *Statistical Inference Under Order Restrictions.* John Wiley & Sons, New York, NY.
3. Beaver, W.: 1966, 'Financial Ratios as Predictors of Failures. Empirical Research in Accounting: Selected Studies'. *Journal of Accounting Research* pp. 71–111. supplement to vol. 5.
4. Frydman, H., E. Altman, and D.-L. Kao: 1985, 'Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress'. *The Journal of Finance* **40**(1), 269–291.
5. Füser, K.: 2002, 'Basel II – was muß der Mittelstand tun?'. *http://www.ey.com/global/download.nsf/Germany/Mittelstandsrating/ $file/Mittelstandsrating.pdf.*
6. Gale, D., H. W. Kuhn, and A. W. Tucker: 1951, *Linear Programming and the Theory of Games, in Activity Analysis of Production and Allocation, T. C. Koopmans (ed.).* John Wiley & Sons, New York, NY.
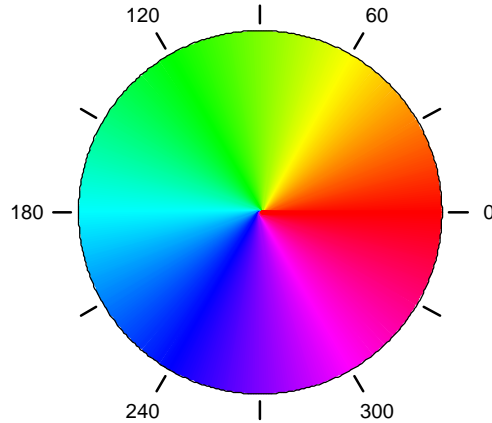7. Härdle, W., Z. Hlávka, and S. Klinke: 2000a, *XploRe Application Guide.* Springer Verlag, Berlin.

Fig. 14. The hue dimension of the HLS colour space.

8. Härdle, W., S. Klinke, and M. Müller: 2000b, *XploRe Learning Guide*. Springer Verlag, Berlin.
9. Härdle, W., R. A. Moro, and D. Schäfer: 2004a, *Predicting Bankruptcy with Support Vector Machines, W. Härdle (ed.)*. Springer Verlag, Berlin.
10. Härdle, W., M. Müller, S. Sperlich, and A. Werwatz: 2004b, *Nonparametric and Semiparametric Models*. Springer Verlag, Berlin.
11. Härdle, W. and L. Simar: 2003, *Applied Multivariate Statistical Analysis*. Springer Verlag.
12. Horowitz, J. L.: 2001, *The Bootstrap, J. J. Heckman and E. E. Leamer (eds.)*, Vol. 5. Elsevier Science B.V.
13. Mammen, E.: 1991, 'Estimating a Smooth Monotone Regression Function'. *Annals of Statistics* **19**, 724–740.
14. Martin, D.: 1977, 'Early Warning of Bank Failure: A Logit Regression Approach'. *Journal of Banking and Finance* (1), 249–276.
15. Mercer, J.: 1909, 'Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations'. *Philosophical Transactions of the Royal Society of London* **209**, 415–446.
16. Ohlson, J.: 1980, 'Financial Ratios and the Probabilistic Prediction of Bankruptcy'. *Journal of Accounting Research* pp. 109–131.
17. Tam, K. and M. Kiang: 1992, 'Managerial Application of Neural Networks: the Case of Bank Failure Prediction'. *Management Science* **38**(7), 926–947.
18. Vapnik, V. N.: 1995, *The Nature of Statistical Learning Theory*. Springer, New York.
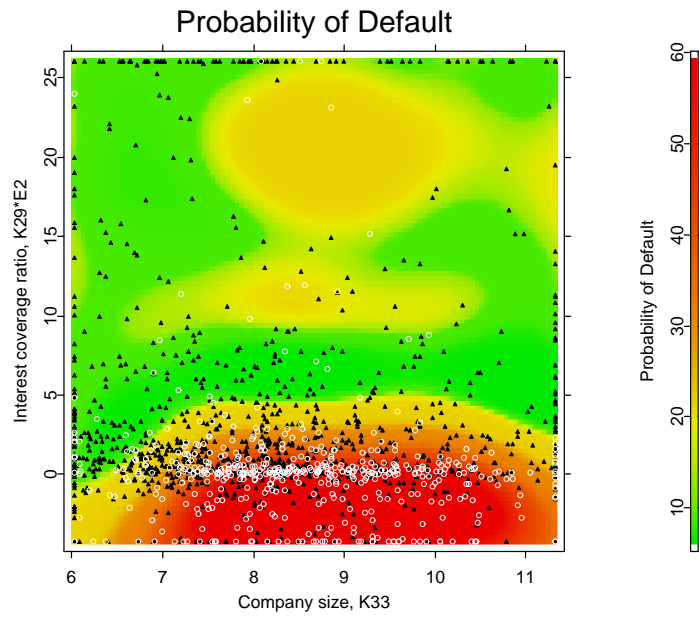
Probability of Default

Fig. 15. Probability of default estimated for a random subsample of 500 failing and 500 surviving companies plotted for the variables K33 and K29. An SVM of high complexity with the radial basis kernel $0.5\Sigma^{1/2}$ was used.

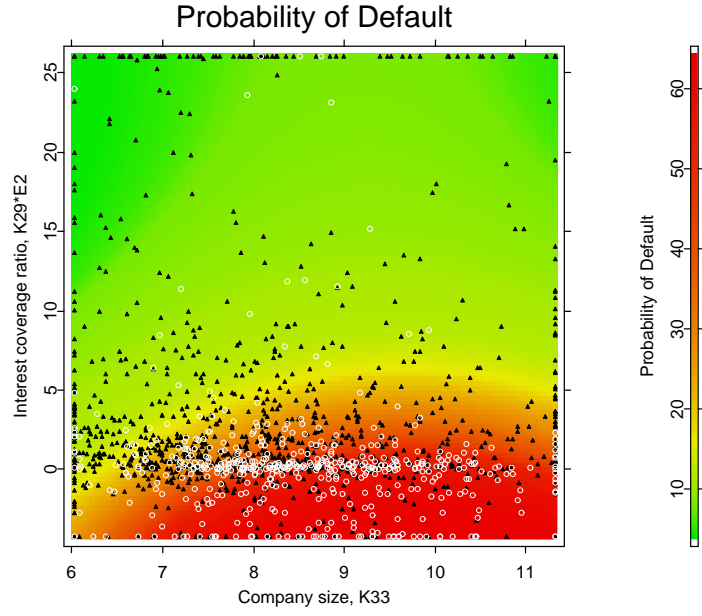19. Weyl, H.: 1928, *Gruppentheorie und Quantenmechanik*. Hirzel, Leipzig.

Fig. 16. Probability of default estimated for a random subsample of 500 failing and 500 surviving companies plotted for the variables K33 and K29. An SVM of average complexity with the radial basis kernel $4\Sigma^{1/2}$ was used.
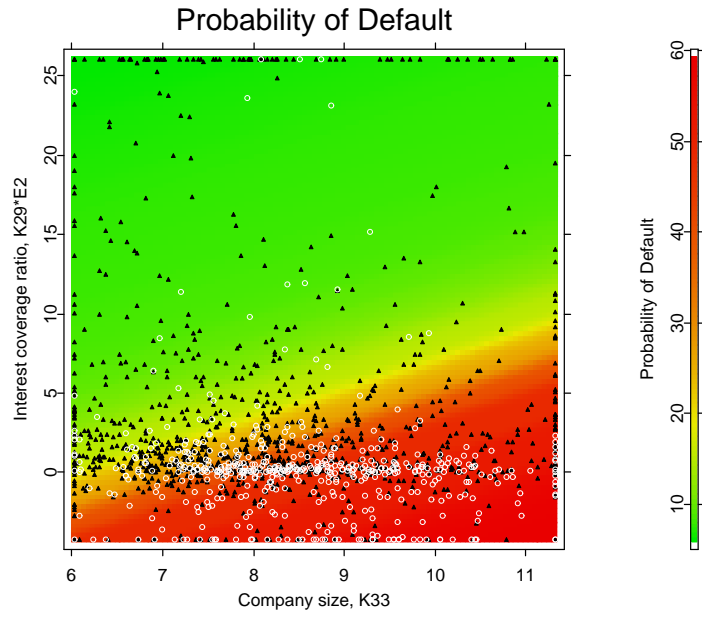
Fig. 17. Probability of default estimated for a random subsample of 500 failing and 500 surviving companies plotted for the variables K33 and K29. An SVM of low complexity with the radial basis kernel $100\Sigma^{1/2}$ was used.
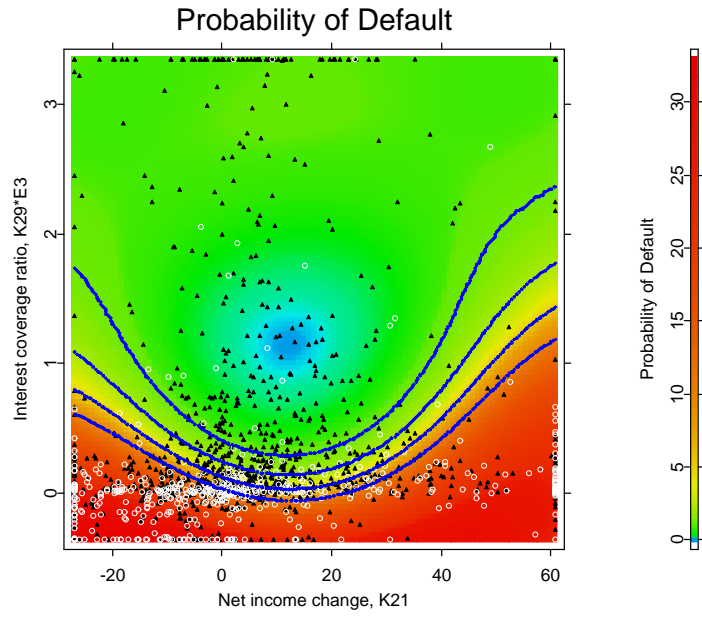
Fig. 18. Probability of default plotted for the variables K21 and K29. The boundaries of five risk classes are shown in blue, which correspond to the rating classes: BBB and above (investment grade), BB, B+, B, B- and lower.

SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.

001   "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K.
       Härdle, January 2006.
002   "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and
       Wolfgang K. Härdle, January 2006.
003   "On the Appropriateness of Inappropriate VaR Models" by Wolfgang
       Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.
004   "Regional Labor Markets, Network Externalities and Migration: The Case
       of German Reunification" by Harald Uhlig, January/February 2006.
005   "British Interest Rate Convergence between the US and Europe: A
       Recursive Cointegration Analysis" by Enzo Weber, January 2006.
006   "A Combined Approach for Segment-Specific Analysis of Market Basket
       Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.
007   "Robust utility maximization in a stochastic factor model" by Daniel
       Hernández–Hernández and Alexander Schied, January 2006.
008   "Economic Growth of Agglomerations and Geographic Concentration of
       Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and
       Axel Werwatz, January 2006.
009   "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental
       and Dominique Demougin, January 2006.
010   "Common Functional Principal Components" by Michal Benko, Wolfgang
       Härdle and Alois Kneip, Jauary 2006.
011   "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings"
       by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten
       Trenkler, February 2006.
012   "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for
       Deterministic Terms" by Carsten Trenkler, February 2006.
013   "Penalties and Optimality in Financial Contracts: Taking Stock" by
       Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February
       2006.
014   "Core Labour Standards and FDI: Friends or Foes? The Case of Child
       Labour" by Sebastian Braun, February 2006.
015   "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang
       Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.