Banco de Portugal
EUROSISTEMA

Estudos e Documentos de Trabalho

Working Papers

22 | 2008

ANALYSIS OF THE PREDICTORS OF DEFAULT FOR PORTUGUESE FIRMS

Ana I. Lacerda
Russ A. Moro

December 2008

Please address correspondence to
Ana I. Lacerda
Economics and Research Department
Banco de Portugal, Av. Almirante Reis no. 71, 1150-012 Lisboa, Portugal;
Tel.: 351 21 313 0211, Email: ailacerda@bportugal.pt

BANCO DE PORTUGAL

Edition

Economics and Research Department

Av. Almirante Reis, 71-6th

1150-012 Lisboa

www.bportugal.pt

Pre-press and Distribution

Administrative Services Department

Documentation, Editing and Museum Division

Editing and Publishing Unit

Av. Almirante Reis, 71-2nd

1150-012 Lisboa

Printing

Administrative Services Department

Logistics Division

Lisbon, December 2008

Number of copies

170

ISSN 0870-0117

ISBN 978-989-8061-61-4

Legal Deposit No 3664/83

Analysis of the Predictors of Default for Portuguese Firms

Ana I. Lacerda* and Russ A. Moro†

The paper presents an insolvency risk analysis of Portuguese companies with three techniques: logistic regression, discriminant analysis and support vector machines (SVM). It identifies the most critical predictors of default based on the accounting, employee and debt concentration data. A comparison of the three methods reveals a superiority of SVM. Non-financial information such as employee data and the debt concentration index appear to be strong predictors of default.

KEYWORDS: bankruptcy, company rating, probability of default, support vector machines

JEL CLASSIFICATION: C14, G33, C45

*Banco de Portugal, Av. Almirante Reis nº71, 1150-165 Lisboa, Portugal; ISEG, Rua do Quelhas, 1200-781 Lisboa and Universidade Nova de Lisboa, Campus de Campolida, 1099-032 Lisboa; e-mail: ailacerda@bportugal.pt; phone: +351 21 3130211

†Brunel University, Department of Economics, School of Social Sciences, Uxbridge, Middx UB8 3PH, United Kingdom and DIW Econ, Mohrenstr. 58, 10117 Berlin, Germany; e-mail: rmoro@diw.de; phone: +49 30 89789-262

1 Introduction

The concept of credit risk plays a central role in the development of the financial system and credit risk assessment is indispensable for its stability. A notable step in the attempts to understand what distinguishes solid and potentially bankrupt firms was the financial ratio analysis (Winakor & Smith, 1935), which became a standard tool. With the rapid development of computers statistical methods, which can identify patterns in the data, became to draw attention of economists. Since the main problem was to identify the patterns of failing in contrast to successful companies, the choice of classification techniques was natural. In 1966 a paper of Beaver (1966) appeared which introduced a univariate version of discriminant analysis (Fisher, 1936) for rating companies.

The discriminant analysis (DA) in its multivariant form was first introduced by Altman (1968). The author combined five financial ratios in a linear way with weights to produce what he called the Z-score (see also Altman, Haldeman & Narayanan (1977)). Despite being forty years old the model and its modifications are still very popular and are frequently used in practice. The assumptions of the model – the normality and the equality of the variance-covariance matrices for defaulting and non-defaulting companies – were often criticised though.

A development of linear rating techniques was the logistic regression (Logit) (Martin (1977), Ohlson (1980)). The Logit model is widely used and is particularly attractive because its score is calibrated as probability of default (PD). In DA as well as in the majority of other methods the score has to be recomputed into PD using historical observations.

Despite its simplicity, the application of linear DA and Logit has a substantial drawback. A financial analyst would have to conclude that, as in the Z-score model, each financial ratio x_i has the same qualitative effect on the Z-score either increasing or reducing PD (see expression 1.1).

$$Z = a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_d x_d. \tag{1.1}$$

This contradicts both economic theory, which among other suggests the existance of the optimal capital structure and growth rates and the evidence of the non-monotonic dependence between some financial ratios and the score or PD accumulated in the literature (Falkenstein, Boral & Carty (2000), Manning (2004), Fernandes (2005) and Härdle, Moro & Schäfer (2007)). For example, growth rates of such an important financial indicator as net income (NI) display a non-linear dependence from PD (Härdle, Moro & Schäfer, 2007). A declining or very slowly growing net income may create problems with paying company debt obligations. On the other hand, excessively high net income growth rates are likely to be non-sustainable in the long run and may indicate a high volatility of income. Both situations would lead to higher PD. The increase of PD due to high volatility is completely in accordance with the literature, e.g. the Merton's model (Merton (1974), Bharath & Shumway (2008)) or the gambler's ruin model (Wilcox, 1971). We, however, do not apply the latter two approaches directly since they require long time series unavailable for private as well as for small and medium size firms which comprise the majority of firms.

Another problem is the identification of the shape of the dependence between PD and financial ratios even it is known to be monotonic. In the Logit model a logistic transformation of financial ratios is used which may not be true in reality.

The non-linear and non-monotonic nature of the dependence between some financial ratios and PD lead to the application of non-linear models such as recursive partitioning, a.k.a. classification and regression trees (Frydman, Altman & Kao, 1985) and neural networks (Tam & Kiang, 1992), and, more recently, Support Vector Machines (SVM) (Friedman (2002), Härdle, Moro & Schäfer (2005) and Martens, Baesens, van Gestel & Vanthienen (2006)).

In this work using Logit, DA and SVM we will characterise the credit risk of Portuguese firms, establishing the main indicators of default of Portuguese firms. Using different methodologies previous studies have already examined the determinants of corporate credit default. Antunes, Ribeiro & Antão (2005) estimate the probability of default of non-financial corporations using information about the loan, economic sector of the debtor and the macroeconomic environment. Soares (2006) and Bonfim (2008) use firm specific level information as well as macroeconomic information. In the former study selected ratios related to leverage, liquidity, profitability as well as asset financing structure are identified as having a significant contribution to the discriminant function. In the latter, besides measures of liquidity, leverage and profitability, measures of recent investment and sales performance were also found valuable in explaining default probabilities. In this study probit regression supported with duration models is used as a tool. The aim of the work is to assess the main indicators of default for Portuguese firms using firm level information, which includes employee data, comparing the results with the ones obtained with the different models.

The purpose of this paper is to relax the monotonicity assumption of the dependence of PD on firms' characteristics and bridge the gap between modeling capabilities and the predictions of theory. The paper is organised as follows. In section 2 a description is presented of the SVM, a technique for modelling complex non-linear dependencies in the data. In section 3, the firm level data of the Portuguese economy as well as the information about Portuguese credit is described. Additionally, a summary statistics for the ratios and indicators used in the paper is reported. Section 4 analyses individual predictors and their effect on PD. The most critical predictors identified with SVM, as well as DA and Logit, are presented in section 5. Section 6 examines the performance of SVM, Logit and DA as tools for estimating PD and, finally, section 7 concludes.

2 The Support Vector Machine as a Classification Method

The Support Vector Machine (SVM) is a non-linear classification method based on the separation of two classes of observations with the maximum margin or gap between the observation classes, in our case non-defaulting and defaulting companies. Non-linear

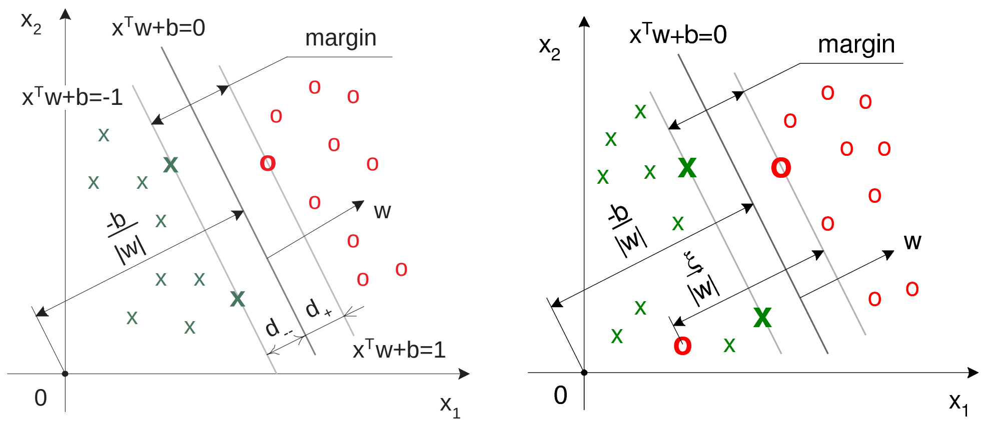

Figure 1: The separating hyperplane $x^\top w + b = 0$ and the margin in a linearly separable (left) and non-separable (right) case. Crosses denote non-defaulting companies, zeros are the defaulting ones. The hyperplanes bounding the margin zone equidistant from the separating hyperplane are represented as $x^\top w + b = 1$ and $x^\top w + b = -1$. The misclassification penalty in the non-separable case is proportional to the distance $\xi / \|w\|$.

smooth separating surface appears as a result of applying (i) kernel transformations (Vapnik, 1995) and (ii) Tikhonov regularisation (Tikhonov (1963) and Tikhonov & Arsenin (1977)).

Each company in the dataset is characterised with a $d \times 1$ vector of predictors x, usually financial ratios, and an indicator of default, $y = -1$ for non-defaulting companies and $y = 1$ for defaulting companies.

Figure 1 illustrates the principle of the separation with the maximum margin for a linear SVM in a separable and a non-separable case. The margin is the gap between the opposite classes of observations. In the linearly separable case (the left panel) the margin $d_- + d_+ = 2/\|w\|$. The separating rule here is linear and similar to that of DA and Logit:

$$x^\top w + b = 0.$$

Here w is a $d \times 1$ vector of weights that define the slope of the separating hyperplane and b is a scalar called the threshold.

In a linearly perfectly separable case no observations can be located inside the margin gap, i.e. all observations $i = 1, 2, \ldots, n$ must satisfy the constraints

$$x_i^\top w + b \geq 1 \quad \text{for} \quad y_i = 1, \tag{2.1}$$

$$x_i^\top w + b \leq -1 \quad \text{for} \quad y_i = -1. \tag{2.2}$$

In a linearly non-separable case (figure 1, right panel) a so called slack variable $\xi \geq 0$ is introduced to account for misclassifications. It is proportional to the distance from a misclassified observation to the boundary of its class and modifies constraints (2.1) and

(2.2) in such a way that they still hold for all observations $i = 1, 2, \dots, n$:

$$x_i^\top w + b \geq 1 - \xi_i \quad \text{for} \quad y_i = 1, \tag{2.3}$$
$$x_i^\top w + b \leq -1 + \xi_i \quad \text{for} \quad y_i = -1. \tag{2.4}$$

The observations marked with bold crosses and zeros in figure 1 are called support vectors because only they describe or "support" the separating hyperplane, hence the name of the method.

The SVM minimisation problem is a trade-off between margin $2/\|w\|$ maximisation and the minimisation of the cumulative error $\sum_{i=1}^n \frac{\xi_i}{\|w\|}$ subject to constraints (2.3) and (2.4) which can be rewritten as one:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\| + C \sum_{i=1}^n \frac{\xi_i}{\|w\|} \tag{2.5}$$
$$\text{s.t.} \quad y_i(x_i^\top w + b) \geq 1 - \xi_i, \tag{2.6}$$
$$\xi_i \geq 0. \tag{2.7}$$

The coefficient C that assignes the priority to each of the maximisation and minimisation tasks is called the capacity. If it is large, the minimisation of in-sample classification errors becomes more important at the expense of the generalisation ability.

If the minimised expression (2.5) is multiplied by $\|w\|$ which is strictly positive, it becomes clear that the SVM problem is convex. This implies the uniqueness and stability of the solution which is a very valuable property in practice since it guarantees that small changes in the data will not lead to a large change in the rating. For example, neural networks do not possess such a property.

The primal Lagrangian formulation of the problem (2.5)–(2.7) with Lagrange multipliers α_i and μ_i for each observation $i = 1, 2, \dots, n$ is

$$\min_{w,b,\xi_i} \max_{\alpha_i,\mu_i} L_P = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \left\{y_i(x_i^\top w + b) - 1 + \xi_i\right\} - \sum_{i=1}^n \xi_i \mu_i.$$

The Karush-Kuhn-Tucker (KKT) optimality conditions or first order optimality conditions (Gale, Kuhn & Tucker, 1951) are derived by equating the first derivative of the primal Lagrangian with respect to each optimised parameter to zero:

$$
\begin{aligned}
\nabla_w L_P &= 0 \quad \Leftrightarrow \quad w = \sum_{i=1}^{n} \alpha_i y_i x_i, & (2.8)\\
\frac{\partial L_P}{\partial b} &= 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0, & (2.9)\\
\frac{\partial L_P}{\partial \xi_i} &= 0 \quad \Leftrightarrow \quad C - \alpha_i - \mu_i = 0, & (2.10)\\
\alpha_i\{y_i(x_i^\top w + b) - 1 + \xi_i\} &= 0,\\
\mu_i \xi_i &= 0,\\
y_i(x_i^\top w + b) - 1 + \xi_i &\geq 0,\\
\alpha_i &\geq 0,\\
\mu_i &\geq 0,\\
\xi_i &\geq 0,
\end{aligned}
$$

The dual problem is derived by substituting KKT conditions (2.8)–(2.10) into the primal problem in the Lagrangian formulation. The primal and dual problems are equivalent because of their convexity (Gale, Kuhn & Tucker, 1951). The dual problem is

$$
\begin{aligned}
\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j, \\
\text{s.t.} \quad 0 \leq \alpha_i \leq C, \\
\sum_{i=1}^{n} \alpha_i y_i = 0.
\end{aligned} \tag{2.11}
$$

Here α_i, $i = 1, 2, \ldots, n$, are Lagrange multipliers which are the solution of the optimisation problem (2.11). The support vectors are those which have $\alpha_i > 0$. For the rest observations $\alpha_i = 0$, hence, the solution depends only on support vectors. The dual problem (2.11) has an equivalent matrix representation which can be convenient for solving the problem with a software packet such as Matlab as a classical constrained quadratic optimisation problem (Fletcher, 1987):

$$
\begin{aligned}
\max_{\alpha} \iota^\top \alpha - \alpha^\top H \alpha, \\
\text{s.t.} \quad 0 \leq \alpha \leq C, \\
y^\top \alpha = 0.
\end{aligned} \tag{2.12}
$$

In (2.12) α is a $n \times 1$ vector of Lagrange multipliers, ι is a $n \times 1$ vector of ones, C is a $n \times 1$ vector of the capacity coefficients, y is a $n \times 1$ vector of company classes $\{+1, -1\}$. The element i, j of the matrix of inner products a.k.a. the Gramm matrix H is $h_{i,j} = y_i y_j x_i^\top x_j$. In this paper the SVM quadratic constrained optimisation problem was solved using the Sequential Minimal Optimisation (SMO) method (Platt, 1998) programmed in C++. The SMO method uses the constraint $y^\top \alpha = 0$ in order to convert the problem into an unconstrained optimisation problem only for two observations. In this way the matrix of inner products H which has the dimension $n \times n$ does not have to be stored in the computer memory which allows processing data sets with a very large number of observations n.

The score function and the classification rule are

$$f(x) = x^\top w + b = x^\top \sum_{i=1}^{n} \alpha_i y_i x_i + b = \sum_{i=1}^{n} \alpha_i y_i x_i^\top x + b \tag{2.13}$$

$$\begin{cases} f(x) < 0 & \Rightarrow x \text{ is not in default,} \\ f(x) \geq 0 & \Rightarrow x \text{ is in default.} \end{cases}$$

Here $b = -\frac{1}{2}\left(x_+^\top + x_-^\top\right) w = -\frac{1}{2}\sum_{i=1}^{n} \alpha_i y_i \left(x_+^\top + x_-^\top\right) x_i$ where x_+ and x_- are any support vectors belonging to the opposite classes $y = +1$ and $y = -1$.

One of the advantages of the linear SVM with the dual problem (2.11) or (2.12) is the ease with which it can be transformed into a non-linear one using kernel techniques. The SVM dual problem depends on the observations only in form of the inner product $x_i^\top x_j$. The inner product is a scalar and makes it possible to condense the information from many variables in it even when the number of variables is large. The inner product of transformed variables is as easy to construct and handle as the inner product of the original ones:

$$K(x_i, x_j) = \Psi(x_i)^\top \Psi(x_j),$$

where Ψ is the transformation applied to observations. $K(x_i, x_j)$ is called a kernel function and must satisfy the properties of an inner product, i.e. be symmetrical and semipositive definite (Mercer, 1909). An obvious adavantage of such an approach compared to a direct application of Ψ is that the transformation can be very complex, even infinitely complex, but the inner product would still be given by a scalar kernel function $K(x_i, x_j)$. In this paper a Gaussian kernel function is used:

$$K(x_i, x_j) = \exp\left\{-(x_i - x_j)^\top r^{-2}(x_i - x_j)/2\right\}. \tag{2.14}$$

Gaussian kernels are known for their ability to represent infinitely complex transformations Ψ. This, for a certain radial basis coefficient r, allows the separation of any groups of observations without an error provided that the identical observations always have the same label y.

The non-linear kernel SVM firstly implicitly maps the data into a so called feature space and then performs linear classification there, which is equivalent to a non-linear classification in the original data space. In the example in figure 2 a quadratic kernel function $K(x_i, x_j) = \Psi(x_i)^\top \Psi(x_j) = (x_i^\top x_j)^2$ maps the original two-dimensional data x into a three-dimensional feature space produced by the transformations $\tilde{x_1} = x_1^2$, $\tilde{x_2} = \sqrt{2}x_1 x_2$ and $\tilde{x_3} = x_2^2$. The transformation represented with the quadratic kernel in this case is $\Psi(x) = \Psi(x_1, x_2) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2) = \tilde{x}$.

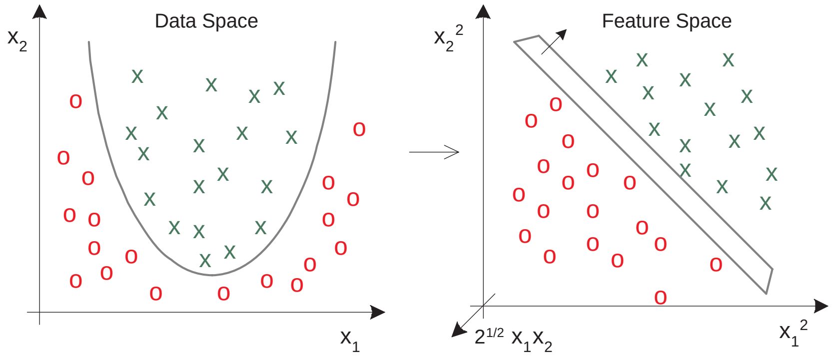

Figure 2: Mapping from a two-dimensional data space into a three-dimensional space of features $\mathbb{R}^2 \mapsto \mathbb{R}^3$ and performing a linear classification there.

The dual SVM problem (2.11) after a kernel transformation becomes

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \tag{2.15}$$
$$\text{s.t.} \quad 0 \leq \alpha_i \leq C,$$
$$\sum_{i=1}^{n} \alpha_i y_i = 0.$$

The inner product matrix H in (2.12) for a non-linear kernel SVM has elements $h_{i,j} = y_i y_j K(x_i, x_j)$. It is often referred to as the kernel matrix. For a Gaussian kernel it has the full rank.

The non-linear SVM score function resembles (2.13) and is

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b. \tag{2.16}$$

Here the threshold $b = -\frac{1}{2} \sum_{i=1}^{n} \alpha_i y_i \{K(x_+, x_i) + K(x_-, x_i)\}$.

3 Data Description

The presented analysis was performed for the year 2005. The data used in this paper originate from these three sources: the Central Balance Sheet Database, the Employee Database, and the Central Credit Register.

Central Balance Sheet Database

This data set is a joint project of the Bank of Portugal, Ministry of Justice, Ministry of Finance and the Portuguese Institute of Statistics, among others. The participation of all Portuguese firms is mandatory and firms must provide accounting, fiscal and statistical information. For the year 2005 this database has information for 340 303 firms.

Employee Database (*Quadros de Pessoal*)

This data source is an annual survey covering all firms employing paid labour in the economy. Firms must report general information such as their sales, the ownership of the capital (public, foreign or private), the number of employees as well as specific information about each employee. Each employee is characterised with age, seniority, wage level, the level of education, the relation to the firm, namely if the employee is also an owner, and the date of the first employment in the firm. For the year 2005 this database has information for 340 782 firms and 3 084 711 employees.

Central Credit Register (*Central de Responsabilidades de Crédito*)
This data set is managed by the Bank of Portugal and records information about any credit provided by a financial institution operating in Portugal to any non-financial firm. In this data set it is possible to identify the firm that receives the credit, the amount of the credit as well as the type of credit being considered. For instance, it is possible to identify if a short-term credit or a long-term credit is being considered. Moreover, information about credit overdue is also registered. The reports must be submitted on a monthly basis. In December 2005 and 2006 there is credit information for 540 491 and 648 206 firms, respectively.

In this study, in the spirit of the Basel II accord, we intend to analyse the future behaviour of firms conditional on their present state concerning the payment of their obligations. Hence, as 2005 is the only year when all three data bases overlap, we will consider firms operating in December 2005, which have honoured their payment obligations, and will observe their behaviour in 2006. Moreover, given the specificities of this study, it was necessary to enforce several selection criteria for the group of firms to be analysed. In addition to eliminating the incoherent or incomplete observations reported, firms that presented sales below 15 000 euros as well as the firms with zero total assets were excluded. Moreover, account was not taken of self-employed persons, firms owned by the government as well as firms whose primary social purpose is financial holding.

Our definition of default is in line with the specifications of the Basel II capital accord. A firm was considered as at default during a given year if in that year the *Central Credit Register* reports the existence of credit overdue for more than three months. After applying the above criteria, a dataset of 84 506 firms non-defaulting firms on 2005 was obtained. From these firms we identified 2 900 defaulting firms against 81 606 non-defaulting firms in the year 2006. It should be noted that default exhibits persistence over time. Hence, from the defualting firms in the year 2005, 72% also exhibit default in 2006.

Finally, it should be stressed that our sample covers a high share of corporate activity. In sectorial terms manufacturing (19%), construction (19%), retail and wholesale (21%), and real estate (17%) are the sectors with the highest representation in terms of firm number. In terms of total credit, these four economic sectors have 79% of total credit (manufacturing 19%, construction 14%, retail and wholesale 35% and real estate 11%).

In our analysis we used as predictors measures of the concentration of debt among different financial institutions, measures of profitability of a firm, leverage, cost structure, liquidity, activity and considered the volatility of sales and sales growth. Additionally, other measures characterising employment, the age of a firm and the size were also taken into account. Specifically, the predictors used are the following:

Profitability

1. NI/TA – return on assets, the ratio of net income to total assets.
2. NI/S – net profit margin, the ratio of net income to sales.
3. OI/TA – operating return on assets, the ratio of operating income to total assets.
4. OI/S – operating profit margin, the ratio of operating income to sales.
5. SM/S – historical sales measured as the mean of (real) sales for the last (max) 10 years to sales.
6. EBIT/TA – gross return on assets, the ratio of earnings before interest and taxes to total assets.
7. EBIT/S – gross profit margin, the ratio of earnings before interest and taxes to sales.

Leverage

1. OK/TA – own capital ratio, the ratio of own capital to total assets.
2. CL/TA – current debt ratio, the ratio of current liabilities to total assets.
3. TD/TA – bank debt ratio, the ratio of total bank debt to total assets.

Cost Structure

1. INTpaid/FD – average cost of financial debt, the ratio of interest paid to financial debt.
2. INT/D – average cost of debt, the ratio of interest payments to debt.
3. EBIT/INTpaid – interest coverage ratio 1, the ratio of earnings before interest and taxes to interest paid.
4. EBIT/INT – interest coverage ratio 2, the ratio of earnings before interest and taxes to net interest payments.

Liquidity

1. STD/D – fraction of debt which is short term debt (liquidity).
2. CASH/TA – the ratio of cash and cash equivalents to total assets.
3. CASH/CL – cash ratio, the ratio of cash and cash equivalents to current liabilities.
4. CRL/D – the ratio of open credit lines to debt.
5. CRL/CL – the ratio of open credit lines to current liabilities.
6. QA/CL – quick ratio, the ratio of quick assets (current assets minus inventories) to current liabilities.
7. CA/CL – current ratio, the ratio of current assets to current liabilities.
8. WC/TA – the ratio of working capital (current assets minus current liabilities) to total assets.
9. CL/TL – the ratio of current liabilities to total liabilities.
10. CL/TL2 – the ratio of current liabilities to liabilities, where liabilities do not include provisions and accruals.

Activity

1. TA/S – asset turnover, the ratio of total assets over sales.
2. INV/S – inventory turnover, the ratio of inventories over sales.
3. AR/S – account receivable turnover, the ratio of account receivables over sales.
4. AP/S – account payable turnover, the ratio of account payable over sales.

Dynamics

1. SGM/SGV – historical sales growth measured as the mean of (real) sales growth for the last (max) 10 years to sales growth volatility measured as the volatility of (real) sales growth for the last (max) 10 years.
2. SM/SV – historical sales measured as the mean of (real) sales for the last (max) 10 years to sales volatility measured as volatility of (real) sales for the last (max) 10 years.
3. SG05 – last year (real) sales growth.

Size

1. LogSM – logarithm of historical sales measured as the mean of (real) sales for the last (max) 10 years.
2. LogTA – company size, logarithm of total assets.
3. LogS – logarithm of sales.

Non-accounting Characteristics

1. AGE – company age.
2. EN – number of employees, end of year number of employees.
3. EMA – employee age, median age of employees (coded from 1 to 9).
4. EE – employee education, median education level of employees (coded from 1 to 9).
5. EE2 – employee education, mean education level of employees (coded from 1 to 9).
6. MS – proportion of women among employees (coded 1 if all employees are men and 2 if all are women).
7. HW – wage concentration, the highest wage as a proportion of the median wage.
8. MEL – seniority, the median of employment length.

Diversification

1. NB – number of banking relationships.
2. HER – Herfindahl Index of debt concentration.

Test and Dependent Variables

1. RandV – an artificially generated random normally distributed variable containing no information about bankruptcies.
2. BANKR – the bankruptcy indicator, +1 for bankrupt and -1 for non-bankrupt companies.

In table 1 the descriptive statistics are presented, separating defaulting firms from the non-defaulting ones.

Table 1: Summary statistics for defaulting firms (first five columns) and non-defaulting firms (last five columns). N is the number of observations which contain the variable. $q_{0.05}$ and $q_{0.95}$ are 5% and 95% quantiles. IQR is the interquartile range.

Var.	N	$q_{0.05}$	Median	$q_{0.95}$	IQR	N	$q_{0.05}$	Median	$q_{0.95}$	IQR
Profitability										
NI/TA	2900	-0.48	0.00	0.09	0.12	81606	-0.26	0.01	0.14	0.05
NI/S	2900	-0.76	0.00	0.11	0.18	81606	-0.35	0.01	0.14	0.05
OI/TA	2900	-0.26	0.07	0.35	0.17	81606	-0.18	0.11	0.45	0.18
OI/S	2900	-0.50	0.09	0.40	0.22	81606	-0.35	0.11	0.39	0.16
SM/S	2900	0.48	0.87	2.64	0.52	81606	0.47	0.83	1.85	0.36
EBIT/TA	2900	-0.43	0.01	0.14	0.13	81606	-0.24	0.03	0.19	0.08
EBIT/S	2900	-0.69	0.01	0.20	0.19	81606	-0.33	0.03	0.20	0.08
Leverage										
OK/TA	2900	-0.80	0.09	0.47	0.25	81606	-0.36	0.20	0.64	0.29
CL/TA	2900	0.15	0.74	1.56	0.46	81606	0.10	0.60	1.17	0.45
TD/TA	2900	0.01	0.20	0.74	0.29	81606	0.01	0.17	0.65	0.25
Cost Structure										
INTpaid/FD	2900	0.01	0.12	1.47	0.18	81606	0.00	0.08	0.94	0.12
EBIT/INTpaid	2821	-22.86	0.64	8.99	5.07	78947	-21.65	1.87	35.82	5.80
EBIT/INT	2830	-24.04	0.66	9.48	5.26	79367	-27.92	1.86	36.30	6.13
INT/D	2900	0.01	0.08	0.30	0.07	81606	0.00	0.06	0.25	0.06
Liquidity										
STD/D	2900	0.00	0.31	1.00	0.61	81606	0.00	0.47	1.00	0.88
CASH/TA	2900	0.00	0.03	0.50	0.10	81606	0.00	0.05	0.47	0.13
CASH/CL	2851	0.00	0.04	0.89	0.15	80555	0.00	0.10	1.38	0.27
CRL/D	2900	0.00	0.01	0.56	0.07	81606	0.00	0.02	1.70	0.20
CRL/CL	2851	0.00	0.00	0.30	0.03	80555	0.00	0.01	0.57	0.09
QA/CL	2851	0.05	0.56	2.20	0.69	80555	0.07	0.73	3.28	0.88
CA/CL	2851	0.17	0.95	3.23	0.72	80555	0.26	1.16	4.93	0.97
WC/TA	2900	-1.07	-0.02	0.59	0.50	81606	-0.71	0.10	0.68	0.45
CL/TL	2900	0.18	0.91	1.00	0.37	81606	0.15	0.88	1.00	0.40
CL/TL2	2900	0.18	1.00	1.00	0.35	81606	0.16	1.00	1.00	0.37
Activity										
TA/S	2900	0.41	1.30	6.40	1.33	81606	0.30	0.94	4.87	0.95
INV/S	2900	0.00	0.15	2.36	0.51	81606	0.00	0.12	1.99	0.38
AR/S	2900	0.00	0.20	1.11	0.41	81606	0.00	0.17	0.77	0.34
AP/S	2900	0.03	0.39	1.77	0.47	81606	0.01	0.22	1.11	0.32
Dynamics										
SGM/SGV	2504	-0.84	0.29	1.13	0.61	71265	-0.65	0.32	1.16	0.53
SM/SV	2801	1.04	2.61	10.75	2.40	78876	1.12	3.08	13.48	3.31
SG05	2801	-0.56	-0.05	1.43	0.45	78876	-0.43	0.00	1.29	0.34
Size										
LogSM	2900	10.58	12.42	14.94	1.73	81606	10.60	12.54	15.27	1.92
LogTA	2900	10.83	12.79	15.46	1.79	81606	10.66	12.72	15.58	1.97
LogS	2900	10.58	12.45	15.04	1.71	81606	10.68	12.67	15.52	1.98
Non-accounting Characteristics										
AGE	2898	1.00	7.00	31.00	11.00	81547	1.00	9.00	35.00	13.00
EN	2900	1.00	5.00	43.50	9.00	81606	1.00	6.00	51.00	9.00
EMA	2899	26.75	37.50	51.17	10.58	81573	27.08	37.92	53.08	10.92
EE	2894	2.00	3.50	5.00	1.00	81409	2.00	4.00	6.00	1.50
EE2	2894	2.15	3.50	5.33	1.27	81409	2.14	3.67	5.80	1.50
MS	2900	1.00	1.29	2.00	0.53	81593	1.00	1.33	2.00	0.53
HW	2541	1.00	1.60	4.41	0.95	72371	1.00	1.62	5.03	1.04
MEL	2896	0.71	3.94	15.66	4.92	81531	0.84	4.58	16.25	5.41
Diversification										
NB	2900	1.00	3.00	6.00	2.00	81606	1.00	2.00	5.00	2.00
HER	2900	0.27	0.58	1.00	0.49	81606	0.32	0.88	1.00	0.47

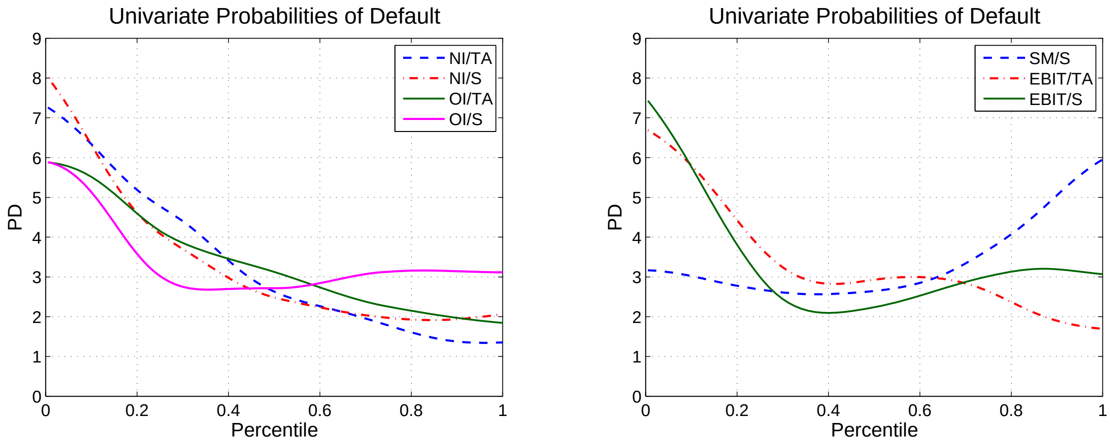

Figure 3: Univariate probabilities of default for the *profitability* financial ratios.

4 Analysis of the Predictors of Default

The analysis of financial ratios and other characteristics of a company starts with considering their individual power as predictors of default. This is done best by estimating univariate dependencies of PD from each variable. Since the range of each predictor can change significantly, we represent all predictors with their percentiles. Univariately estimated PDs are reported in figures 3 – 11. They were obtained as k nearest neighbor estimates (k-NN) with Gaussian weights:

$$PD(q) = \frac{\sum_{i=1}^{n} \mathbf{I}_{\{y_i=1\}} e^{-\frac{(q-q_i)^2}{2\sigma^2}}}{\sum_{i=1}^{n} e^{-\frac{(q-q_i)^2}{2\sigma^2}}}, \tag{4.1}$$

where $0 \leq q \leq 1$ is a percentile, q_i is the percentile of company i of the data set and the smoothing parameter σ is set to 0.08. $\mathbf{I}_{\{y_i=1\}}$ is the indicator function which equals 1 if $y_i = 1$ i.e. when company i is in default and 0 otherwise.

The variables differ a lot in their predictive power. For example, such variables as NI/TA, EBIT/TA or EBIT/INT and EBIT/INTpaid are traditionally reported in the literature as strong predictors, the result which we also confirm. On the contrary, some variables have no discriminating power whatsoever, e.g. EN, EMA or HW.

Another observation is that some predictors, many of which have a high discriminating power, display a non-monotonous dependence with PD, e.g. SM/S, EBIT/S, EBIT/INT, EBIT/INTpaid, CASH/TA, CRL/D, CRL/CL, CL/TL, INV/S, LogTA or SG05. Some of non-monotonous dependencies, most notably similar to SG05, have been reported in the literature but some such as SM/S, CRL/D or CRL/CL are reported in this paper for the first time.

Below we are offering comments of the estimated dependencies. The predictors are discussed in the same order as they appear in the list and table 1.

STD/D displays a weak dependence with PD, an exception is an area around the 30% percentile, where PD increases, a result requiring further investigation. PD is rising

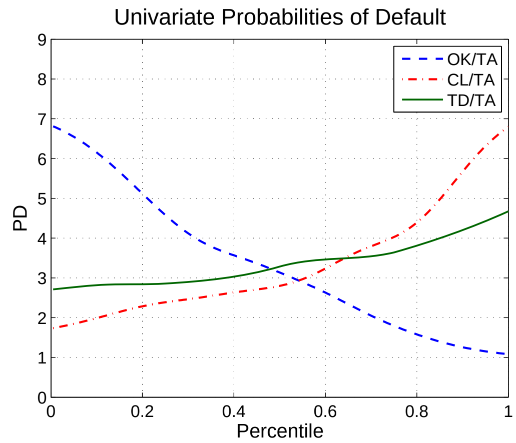

Figure 4: Univariate probabilities of default for the *leverage* financial ratios.

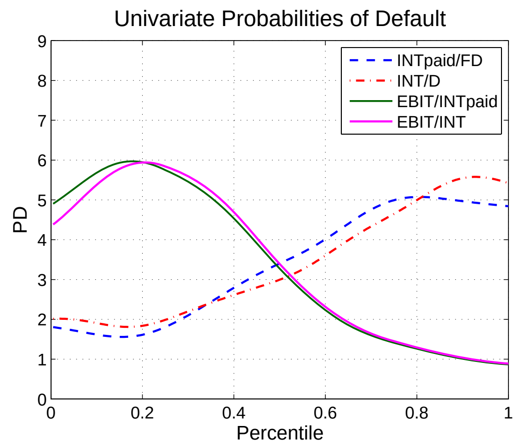

Figure 5: Univariate probabilities of default for the *cost structure* financial ratios.

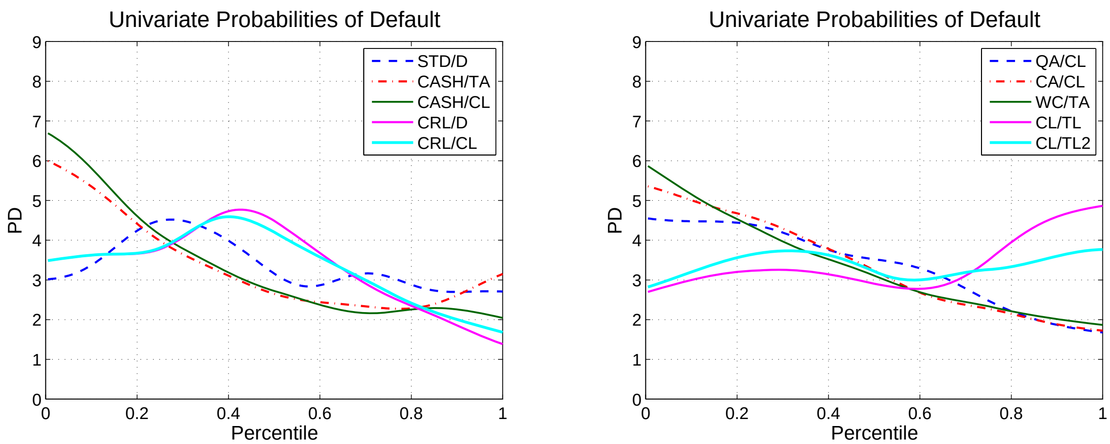

Figure 6: Univariate probabilities of default for the *liquidity* financial ratios.

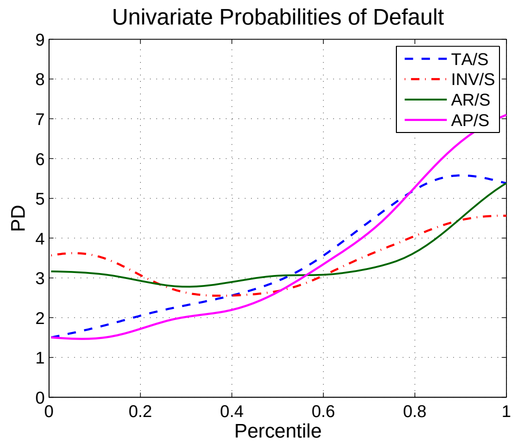

Figure 7: Univariate probabilities of default for the *activity* financial ratios.

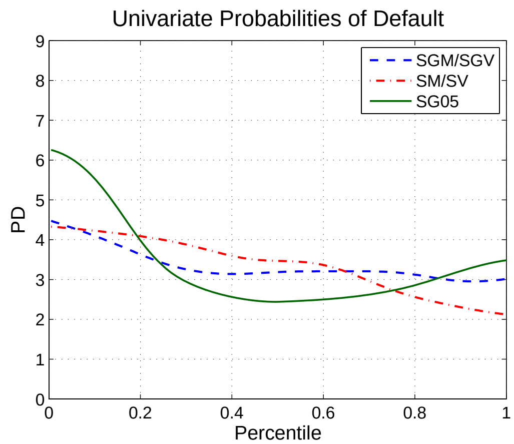

Figure 8: Univariate probabilities of default for the *dynamics* financial indicators.

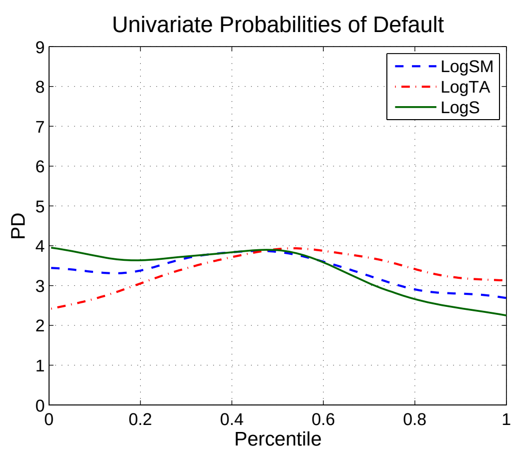

Figure 9: Univariate probabilities of default for the measures of firm *size*.

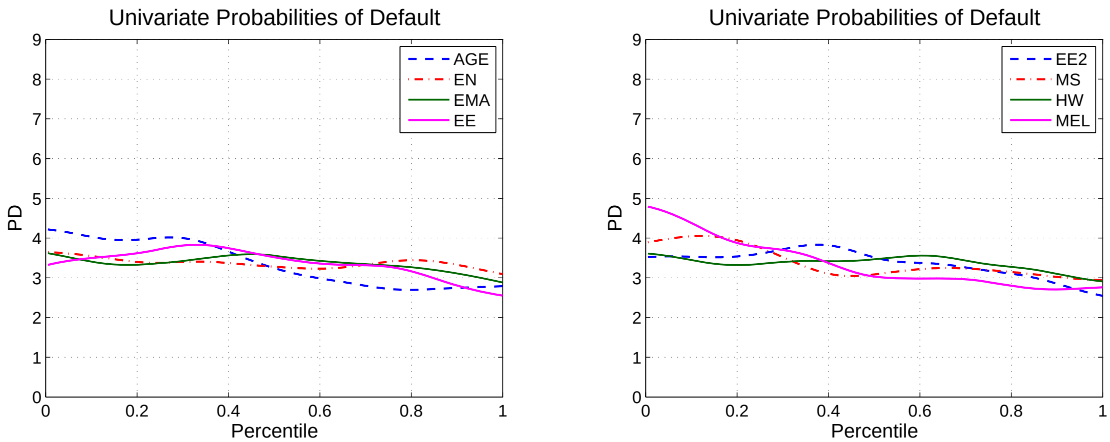

Figure 10: Univariate probabilities of default for *non-accounting* characteristics.

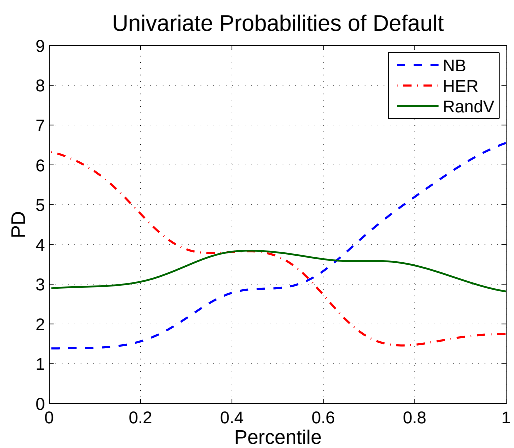

Figure 11: Univariate probabilities of default for the measures of diversification. The dependence from a random variable is reported for a comparison.

when the number of banking relations NB is increasing and falling when the Herfindahl concentration index is rising. Both dependencies are strong predictors of default meaning that risky firms are looking for financing from multiple sources that indicates financial constraints on their operation.

NI/TA, EBIT/TA and OI/TA display a monotonic or near monotonic behaviour: PD is falling when profitability, which they measure, grows. EBIT/S and OI/S and to a lesser degree NI/S have a peculiar behaviour in such a way that first PD is falling to around 2% and then remains approximately constant or slightly rising, as it is the case with EBIT/S. This means that when the profit margins are low, their increase has a strong impact on PD. Later, however, a further increase of margins by reducing costs does not give any effect or even leads to an increase of PD. Thus, a reduction of labour or capital costs when they are already low does not affect margins significantly but is rather an indicator of underinvestment in labour and capital.

SM/S is first modestly decreasing and then rising rapidly. When sales are low compared to the sales average over the past years, PD is high. It is notable that PD is also increasing when sales are much higher than the historic average. This may indicate a high volatility of sales which negatively affects PD.

OK/TA, CL/TA and TD/TA are related to PD in a monotonic way: lower leverage reduces PD while higher leverage increases it. CL/TA is more sensitive than TD/TA. A higher average costs of financial debt INTpaid/FD and debt INT/D increase PD from around 2% to 5% and 5.5% respectively.

EBIT/INTpaid and a similar variable EBIT/INT have a non-monotonic dependence with PD. PD first rises for negative coverage ratios and then significantly drops from 6% to 1%. A low minimum PD achieved with these ratios – the lowest among all ratios – makes them an important predictor when the identification of the most solvent companies is considered.

CASH/TA has a proclaimed non-monotonic dependence with PD. An increase in cash reserved first significantly reduces insolvency risks and PD falls from 6% to 2-2.5% but then rises again as cash constitutes non-productive assets which do not generate profits. CASH/CL has a similar behaviour as CASH/TA except for the area of high cash reserves when PD remains flat since cash constitutes a substantial portion of current liabilities and increase of the numerator is followed by a likewise increase of the denominator.

CRL/D has the maximum around the 40% percentile with PD% slightly decreasing for low values of the ratio. This can be attributed to the rise of indebtness. For high values PD is falling indicating the easying up of financial constraints with the increase of the credit line with financial institutions. A similar behaviour has CRL/CL.

The increase of QA/CL, CA/CL and WC/TA has the effect of monotonously reducing PD from 4.5–6% to 2%. CL/TL has no visible effect on PD until the 60% percentile with PD remaining almost constant around 3%. Then PD sharply rises to 5% indicating problems with financing when long term debt constitutes a smaller and smaller fraction of total liabilities. CL/TL2 does not significantly influence PD.

TA/S which represents a sales generating ability of assets has an increasing dependence with PD except for the 90% percentile which corresponds to the companies which hardly need any new financing with associated insolvency risks since their sales are low. INV/S has a U-shaped dependence with the minimum PD around the 40% percentile at 2.5%. Keeping inventories, which are non-productive assets, high relative to sales increases PD up to 4.5%. On the other hand, too low inventories also lead to a higher PD (3.5%).

There is little connection between AR/S and PD until the 80% percentile when the ratio shoots up from 3.5%, which is around the average PD for all companies, up to 5.5%. This increase can be attributed to the inability of a firm to collect payments from its customers on time and a weak position of a firm. On the contrary, AP/S belongs to the strongest predictors with PD increasing steadily from 1.5% to 7%.

PD is slightly decreasing with SGM/SGV and SM/SV, confirming a hypothesis that the significance of the trend in sales or the stability of sales contribute to reducing PD. LogSM has only a very weak effect on PD, LogTA has a much more proclaimed non-monotonic one. PD is the highest for the companies of the median size (4%) and is lower for small (2.5%) and large (3%) companies. The smaller companies may rely less on external financing while large companies are more diversified.

The sales growth over the last year measured in real terms SG05 has a characteristic U-shaped non-monotonic dependence with PD. There is an optimal area of growth rates between 40% and 60% percentile where PD is the lowest around 2.4%. This area corresponds to the growth rate of a median company in the market. Any abberation from the optimum has a negative impact on company performance. Low or negative growth rates are detrimental for company operation and sharply increase PD up to 6–6.5%. However, higher than median growth rates put a firm at a disadvantage as well and increase PD up to 3.5%. Excessively rapid growth is not sustainable in the long run and may be a result of high volatility in sales.

The increase of LogS reduces PD from 4% to 2.5%, this reduction taking place only for the companies of a bigger than the median size meaning that big companies are able to diversify away their insolvency risks. Older companies (AGE) are less risky. The education level of the employees EE and EE2 has only a marginal effect on PD reducing it. The number of employees, employee age and wage concentration EN, EMA and HW have likewise hardly any discernible effect on PD. MS has an influence on PD: the firms with a predominantly male employees have PD 1% higher that the firms with predominantly female employees (4% vs. 3%). The increase in the median employment length MEL reduces PD from 4.5–5% to 3%, however, its positive influence is detectable only for the companies of the size below median which rely more on an experience of individual workers than on production organisation.

For comparison purposes we report univariate PDs estimated for a random standard normally distributed variable RandV that contains no economic information. The PD oscillates around the average PD of 3.5% in the range ±0.5%. The shape of the curve is, however, less stable with respect to the smoothing parameter σ. Although RandV and, for example, MS have the same range of 1%, the latter has a very robust stable shape which allows to distinguish it from a random variable.

One conclusion that we can draw from the analysis is that predictors have very different relationships with PD which is summarised in table 2. Moreover, many strong predictors have a non-monotonic relationship. This in most cases has a fundamental economic explanation. To perform the analysis statistically for multivariate data in order to

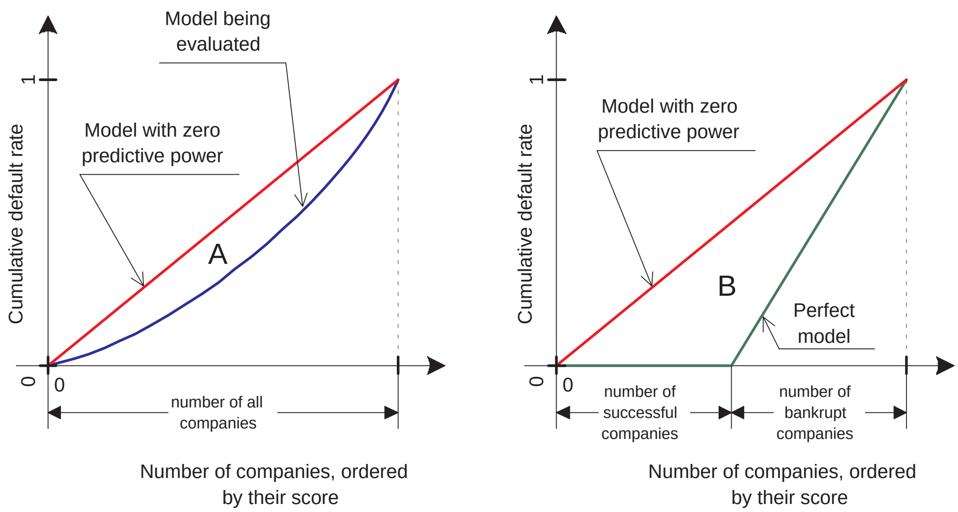

Figure 12: Accuracy Ratio (AR) is the ratio of two areas A and B.

estimate PD we need a flexible model capable of capturing a non-monotonic and non-linear character of the dependence. In the next chapter we are analysing the variables in a multivariate settings with such a model, SVM, and compare the results with Logit and DA.

5 Selection of the Predictors of Default and Rating Model Comparison

The criterion for comparing different models is a robust accuracy measure, the median Accuracy Ratio (AR) estimated on bootstrapped subsamples. AR is the ratio of two areas (i) between the cumulative default curves for the model being evaluated and the model with a zero predictive power and (ii) between the cumulative default curves for the ideal model and the model with a zero predictive power (figure 12). AR is used since it is not sensitive to a monotonic transformation of a score in contrast to other accuracy measures such as hit rate or α and β errors.

The bootstrap procedure (Efron & Tibshirani, 1993) is performed by selecting from the pool of data for the year 2005 of two non-overlapping random subsamples of 800 observations (400 non-defaulting and 400 defaulting firms). One of those subsamples is used as a training set and the other one as a testing set. A classification model is trained on the former and its AR is estimated on the latter. The procedure is repeated 100 times creating a set of 100 estimates of AR from which the median is computed and used for the comparison of models. The model with the highest median AR is preferred.

All data were first cleaned from outliers by capping them: if $x < q_{inf}(x)$ then $x = q_{inf}(x)$ and if $x > q_{sup}(x)$ then $x = q_{sup}(x)$. Here $q_{inf}(x) = Median(x) - 1.5IQR(x)$ and $q_{sup}(x) = Median(x) + 1.5IQR(x)$. Then all data were standardised as $x_{new} = (x - median(x))/\sigma(x)$. This was done to avoid an excessive influence of the variables

Table 2: The dependence of PD from indvidual company characteristics.

Var.	Type of Dependence
	Profitability
NI/TA	strongly falling
NI/S	strongly falling
OI/TA	falling
OI/S	strongly falling, then flat
SM/S	flat, then strongly rising
EBIT/TA	strongly falling
EBIT/S	strongly falling, then weakly rising
	Leverage
OK/TA	strongly falling
CL/TA	strongly rising
TD/TA	rising
	Cost Structure
INTpaid/FD	flat, then falling, then flat
INT/D	flat, then rising
EBIT/INTpaid	weakly rising, then falling
EBIT/INT	weakly rising, then falling
	Liquidity
STD/D	weakly rising, then weakly falling, then flat
CASH/TA	strongly falling, then weakly rising
CASH/CL	strongly falling, then flat
CRL/D	weakly rising, then falling
CRL/CL	weakly rising, then falling
QA/CL	falling
CA/CL	falling
WC/TA	falling
CL/TL	flat, then rising
CL/TL2	weak
	Activity
TA/S	strongly rising, then flat
INV/S	weakly falling, then rising
AR/S	flat, then rising
AP/S	strongly rising
	Dynamics
SGM/SGV	weakly falling
SM/SV	weakly falling
SG05	strongly falling, then weakly rising
	Size
LogSM	weakly rising, then weakly falling
LogTA	rising, then weakly falling
LogS	flat, then weakly falling
	Non-accounting Characteristics
AGE	weakly falling
EN	weak
EMA	weak
EE	weak
EE2	weak
MS	falling
HW	weak
MEL	falling, then flat
	Diversification
NB	strongly rising
HER	strongly falling

with a higher dispersion. These transformations are routinely applied to the data prior to analysis.

Variable selection was performed using the forward selection procedure which starts with univariate models. At step one the first variable is selected that produces the most accurate univariate model as judged by its median AR. At step two, in addition to that variable the second variable from the remaining is chosen which has the highest meadian AR among all alternatives. At step three a trivariate model is selected, etc. The variables selected by Logit, DA and SVM are presented in table 3. After a certain step the accuracy of a model begins to drop since new included variables do not contain additional information but inject noise into the model.

The SVM was always applied with $R = 4\sqrt{d/2}$ and $C = (10000/n)\sqrt{d/2}$ which were chosen without numerical optimisation by plotting two-dimensional classification results and selecting the parameters which provided a good separation without using too complex curves. Here the transformations were used for C and R to keep SVM invariant of to the dimension of the data d and the number of observations in the training set n (cf. equation 2.5 of the optimisation problem).

All models considered – Logit, DA and SVM – include many identical variables, e.g. NB, LogS, INT/D, CRL/D and AP/S. It should be stressed that EBIT/INTpaid was selected by SVM, although not by Logit and DA. This results from the fact that default exhibits a non-monotonic relation with EBIT/INTpaid, as reported in the previous section. Additionally, SVM consider the leverage measure OK/TA, while Logit and DA consider CL/TA.

6 Probability of Default Estimates for Portuguese Firms

Rating models including Logit, DA and SVM produce the score as their output. In order to evaluate how high insolvency risk is score values must be converted into PD. For some models, e.g. Logit, the score is calibrated to lie between 0 and 1 and can be interpreted as PD. Logit, however, has a substantial drawback, namely the logistic distribution function for PD. This assumption is hardly plausible and does not reflect the dependence between PD and underlying predictors. We estimate PD following a much more flexible approach without prespecifying the form of the dependence and only imposing the monotonic relationship between the score and PD.

The estimation procedure consists of three steps. First, we derive univariate PDs as k nearest neighbour estimates (k-NN) with Gaussian weights in the same way as for the univariate analysis of variables by applying equation 4.1.

Score is, however, a special characteristic of a company which concentrates all information that the model extracts from the data. It is assumed and is by large confirmed by historical observations that PD depends monotonically from the score. Since the first step does not guarantee monotonicity, monotonicity is imposed at the second step by using the Pool Adjacent Violator (PAV) algorithm (Barlow, Bartholomew, Bremmer & Brunk, 1972). PAV could be applied as the single step for estimating PDs, however, it would produce a higher estimation bias, especially for the companies with the highest score.

Table 3: The variables selected at each step by the forward selection procedure for Logit, DA and SVM. Median AR is reported for each step as well as the confidence level (p-value) for the test with H_0: The model with the highest meadian AR (the "best" model) is not significantly different from the current one in terms of AR. It shows the number of cases out of 100 when the current model exceeded in accuracy the best model.

	Logit			DA			SVM		
Step	Var.	Med. AR	p	Var.	Med. AR	p	Var.	Med. AR	p
1	NI/TA	32.9	0	NI/TA	32.9	0	EBIT/INTpaid	33.9	0
2	NB	40.2	0	NB	40.2	0	NB	41.2	0
3	LogS	45.7	0	LogS	45.7	0	LogS	47.4	0
4	INT/D	50.1	0	INT/D	50.1	0	INT/D	51.8	0
5	CRL/D	52.6	0	CRL/D	52.6	0	CRL/D	54.0	2
6	AP/S	55.0	3	AP/S	55.0	4	AP/S	55.7	3
7	SM/SV	55.6	9	SM/SV	55.7	12	OK/TA	56.9	7
8	CL/TA	56.0	11	CL/TA	56.1	12	TA/S	57.7	18
9	STD/D	56.4	15	STD/D	56.5	16	MEL	58.2	18
10	TA/S	57.5	33	TA/S	57.4	32	MS	58.3	19
11	HER	57.7	26	EE	57.7	32	INTpaid/FD	58.4	24
12	INV/S	57.8	33	INV/S	57.9	29	TD/TA	58.7	26
13	EE	57.9	31	OI/S	58.0	30	CASH/TA	58.9	30
14	CL/TL	58.0	27	SGM/SGV	58.0	30	LogTA	58.9	34
15	TD/TA	58.0	30	HER	58.0	24	NI/S	59.1	40
16	MS	58.2	26	MEL	58.1	44	INV/S	59.0	32
17	MEL	58.1	36	OK/TA	58.2	45	CASH/CL	59.3	36
18	OK/TA	58.3	38	MS	58.2	–	SM/S	59.2	42
19	SM/S	58.3	48				LogSM	59.3	–
20	AGE	58.4	–						

For this reason additional smoothing was added at step 1 and 3.

The final third step is the smoothing of the PDs obtained with PAV using the k-NN method with Gaussian weigths:

$$PD(q) = \frac{\sum_{i=1}^{n} PD_i e^{-\frac{(q-q_i)^2}{2\sigma^2}}}{\sum_{i=1}^{n} e^{-\frac{(q-q_i)^2}{2\sigma^2}}},$$

where q is the percentile of a company to be evaluated, q_i is the percentile of the score for observation i of the training set. The smoothing parameters σ at step 1 and 3 are both set equal to 0.1.

Graphs 13 and 14 show estimated PDs coded with colour from green for the lowest to red for the highest level and a separating line corresponding to PD equal 50%. The estimations were done by Logit, DA and SVM on the same random subsample with 800 companies with 400 non-bankrupt companies marked with black triangles and 400 bankrupt ones marked with white circles.

The first observation valid for all predictors (figures 13 and 14) is that Logit and DA produce almost identical PDs. SVM, compared to Logit and DA, produces a non-linear separating line providing the highest accuracy of separation. This allows the detection and estimation of non-monotonic dependencies evident from section 4.

Figure 13 shows the PD and the 50% separating line for two most prominent predictors picked at the first variable selection step by Logit and DA (NI/TA) and SVM (EBIT/INTpaid). Although EBIT/INTpaid is one of the best predictors Logit and DA can not use it efficiently due to the way it is constructed. For positive values higher coverage ratios correspond to lower PD as a company finds it easier to pay interest on its debt. For very low and negative ratios this relation is broken. Logit and DA which expect the same relationship to hold for the whole range of predictors are unable to account for the break of the dependence around EBIT/INTpaid equal to 0. The SVM is able to separate two regions with a curve which draws a border between them.

In figure 14, profitabilty (NI/TA) vs. sales growth (SG05) the SVM produces non-linear separation more consistent with non-monotonous dependence revealed SG05 in section 4.

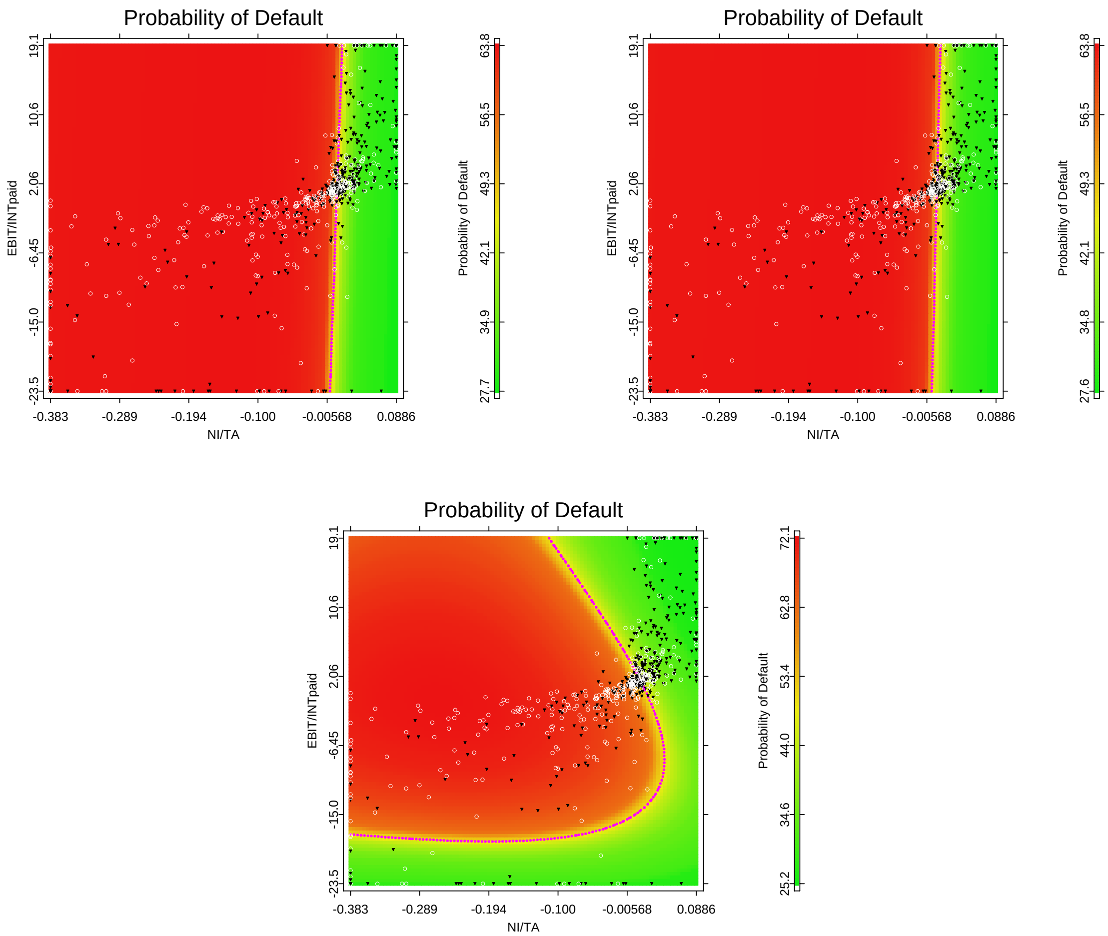

Figure 13: PDs estimated with Logit (left above), DA (right above) and SVM (below) for NI/TA and EBIT/INTpaid.

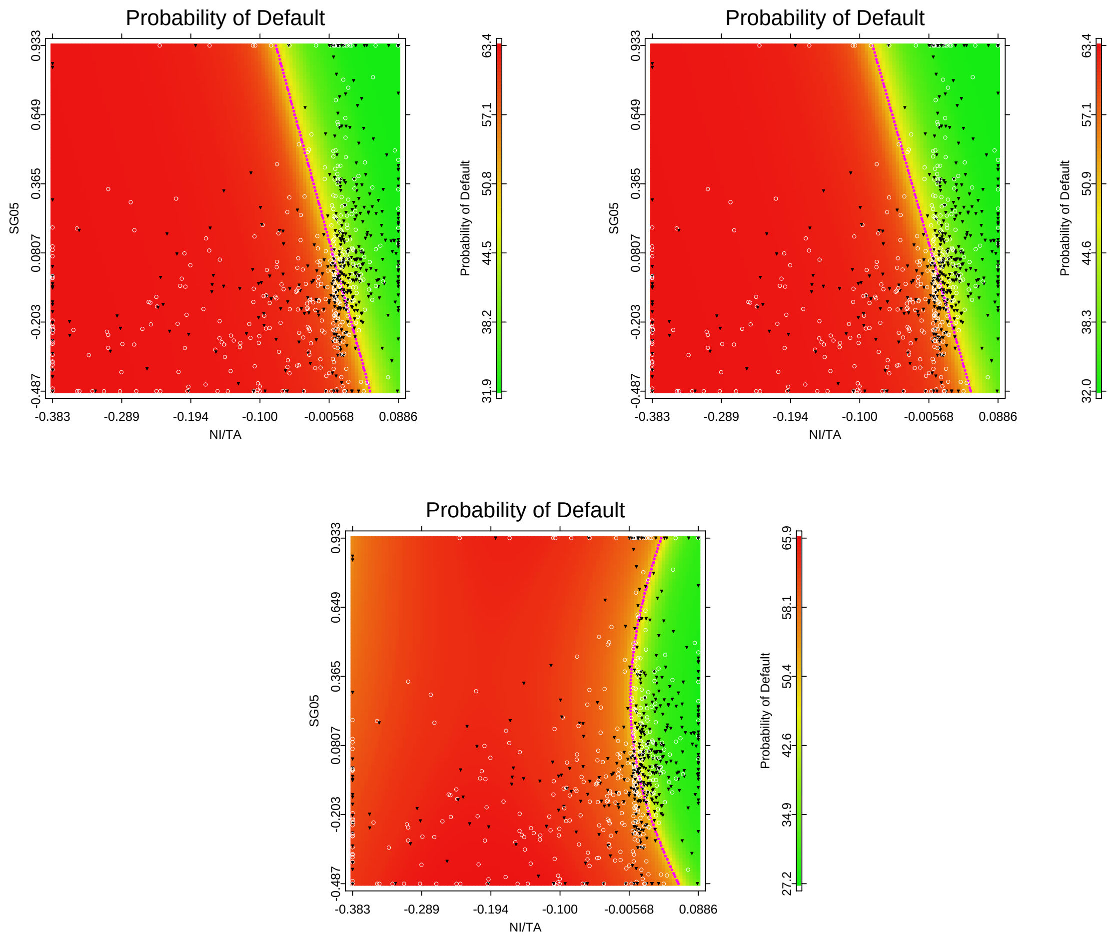

Figure 14: PDs estimated with Logit (left above), DA (right above) and SVM (below) for NI/TA and SG05.

7 Conclusion

In this work using three different methodologies an assessment was performed of the firm level information which proves to be valuable to the explanation of default. In total, 47 indicators were tested. These indicators contain information on the profitability, leverage, cost structure, liquidity, activity, time dynamics and size. Some other indicators, non-accounting characteristics (as information about firm age and employees), as well as credit diversification were tested, which is a novelty in credit risk analysis. This was made possible by combining different data sources. The SVM proves to have a superior performance since for the same number of explanatory variables it always achieves a higher accuracy ratio. However, besides the superiority demonstrated by the SVM, the predictors found by the three methods seem to play an important role.

The reported superiority of SVM is explained by the model technology as well as by its ability to use a broader range of firm indicators that prove to be relevant to predicting the default of Portuguese firms. The latter superiority is attributed to the fact that SMV is much more flexible, capturing non-monotonic dependence of the PD from the predictors. This feature of SVM, which is absent in Logit or DA, is of extreme importance since many strong predictors of default display a non-monotonic dependence with PD, i.e. the same change in a predictor can have either a positive or a negative effect on the probability depending on its magnitude.

Financial ratios characterising the cost structure, the liquidity, the activity as well as the leverage seem to be relevant for all the three methods. Among others, we should stress the interest over debt ratio, the credit lines over debt ratio and accounts payable over sales. The firm size also proves to be relevant for the explanation of default. Other variables, such as earnings before interest over interest paid, only seem to be selected if SVM is considered. This is caused by their non-monotonic dependence with PD, which can be captured by SVM but not by Logit or DA. The inclusion of these new variables explaines the superiority of SVM, which has higher reported accuracy ratios.

The paper is not limited to the analysis of financial ratios. Actually, one of the innovations of this work is the assessment of the employee profiles and debt concentration on PD, confirming their significance. On one hand, the increase of the number of banking relationships is a sign of a weaker financial position. This variable is a strong predictor. On the other hand, the median employment length also is valuable for explaining defaults of Portuguese firms.

Acknowledgements

R. A. Moro gratefully acknowledges the support of Banco de Portugal in writing this paper.

References

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

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

Antunes, A., N. Ribeiro, and P. Antão, 2005: Estimating probabilities of default under macroeconomic scenarios. Financial stability report, Banco de Portugal.

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.

Beaver, W., 1966: Financial ratios as predictors of failures. empirical research in accounting: Selected studies. *Journal of Accounting Research* 71–111. supplement to vol. 5.

Bharath, S. T. and T. Shumway, 2008: Forecasting default with the merton distance to default model. *Review of Financial Studies*, **21**(3), 1339–1369.

Bonfim, D., 2008: Credit risk drivers: Evaluating the contribution of firm level information and macroeconomic dynamics. Working paper no. 7, Banco de Portugal.

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

Falkenstein, E., A. Boral, and L. Carty, 2000: Riskcalc for private companies: Moody's default model.

Fernandes, J. E., 2005: Corporate credit risk modeling: Quantitative rating system and probability of default estimation. *http://pwp.netcabo.pt/jed_fernandes/JEF_CorporateCreditRisk.pdf*.

Fisher, R. A., 1936: The use of multiple measurements in taxonomic problems. *The Annals of Eugenics*, **7**, 179–188.

Fletcher, R., 1987: *Practical Methods of Optimization*. John Wiley and Sons, Inc., 2nd edition.

Friedman, C., 2002: Creditmodel technical white paper. discussion paper, Standard & Poor's.

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.

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, 317–329.

Härdle, W., R. A. Moro, and D. Schäfer, 2005: *Predicting Bankruptcy with Support Vector Machines*. Statistical Tools for Finance and Insurance. Springer Verlag, Berlin.

Härdle, W., R. A. Moro, and D. Schäfer, 2007: Estimating probabilities of default with support vector machines. Discussion paper 2007-035, SFB 649, Humboldt-Universität zu Berlin.

Manning, M. J., 2004: Exploring the relationship between credit spreads and default probabilities. *Working Paper No. 225, Bank of England*.

Martens, D., B. Baesens, T. van Gestel, and J. Vanthienen, 2006: Comprehensible credit scoring models using rule extraction from support vector machines. working paper 878283, Social Science Research Network.

Martin, D., 1977: Early warning of bank failure: A logit regression approach. *Journal of Banking and Finance*, **1**, 249–276.

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.

Merton, R. C., 1974: On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, **29**(2), 449–470.

Ohlson, J. A., 1980: Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, **18**(1), 109–131.

Platt, J., 1998: Sequential minimal optimization: A fast algorithm for training support vector machines. technical report msr-tr-98-14, Microsoft Research.

Soares, M. C., 2006: Modelling of an indicator for credit scoring of non-financial corporations – a preliminary research based on discriminant analysis. Financial stability report, Banco de Portugal.

Tam, K. and M. Kiang, 1992: Managerial application of neural networks: the case of bank failure prediction. *Management Science*, **38**(7), 926–947.

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

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

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

Wilcox, J. W., 1971: A simple theory of financial ratios as predictors of failure. *Journal of Accounting Research*, **9**(2), 389–395.

Winakor, A. and R. Smith, 1935: Changes in the financial structure of unsuccessful industrial corporations. bulletin no. 51. Bureau of Business Research, University of Illinois, Urbana, Ill.

WORKING PAPERS

2000

1/00 UNEMPLOYMENT DURATION: COMPETING AND DEFECTIVE RISKS
— *John T. Addison, Pedro Portugal*

2/00 THE ESTIMATION OF RISK PREMIUM IMPLICIT IN OIL PRICES
— *Jorge Barros Luís*

3/00 EVALUATING CORE INFLATION INDICATORS
— *Carlos Robalo Marques, Pedro Duarte Neves, Luís Morais Sarmento*

4/00 LABOR MARKETS AND KALEIDOSCOPIC COMPARATIVE ADVANTAGE
— *Daniel A. Traça*

5/00 WHY SHOULD CENTRAL BANKS AVOID THE USE OF THE UNDERLYING INFLATION INDICATOR?
— *Carlos Robalo Marques, Pedro Duarte Neves, Afonso Gonçalves da Silva*

6/00 USING THE ASYMMETRIC TRIMMED MEAN AS A CORE INFLATION INDICATOR
— *Carlos Robalo Marques, João Machado Mota*

2001

1/01 THE SURVIVAL OF NEW DOMESTIC AND FOREIGN OWNED FIRMS
— *José Mata, Pedro Portugal*

2/01 GAPS AND TRIANGLES
— *Bernardino Adão, Isabel Correia, Pedro Teles*

3/01 A NEW REPRESENTATION FOR THE FOREIGN CURRENCY RISK PREMIUM
— *Bernardino Adão, Fátima Silva*

4/01 ENTRY MISTAKES WITH STRATEGIC PRICING
— *Bernardino Adão*

5/01 FINANCING IN THE EUROSYSTEM: FIXED VERSUS VARIABLE RATE TENDERS
— *Margarida Catalão-Lopes*

6/01 AGGREGATION, PERSISTENCE AND VOLATILITY IN A MACROMODEL
— *Karim Abadir, Gabriel Talmain*

7/01 SOME FACTS ABOUT THE CYCLICAL CONVERGENCE IN THE EURO ZONE
— *Frederico Belo*

8/01 TENURE, BUSINESS CYCLE AND THE WAGE-SETTING PROCESS
— *Leandro Arozamena, Mário Centeno*

9/01 USING THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION INDICATOR
— *José Ferreira Machado, Carlos Robalo Marques, Pedro Duarte Neves, Afonso Gonçalves da Silva*

10/01 IDENTIFICATION WITH AVERAGED DATA AND IMPLICATIONS FOR HEDONIC REGRESSION STUDIES
— *José A.F. Machado, João M.C. Santos Silva*

2002

1/02 QUANTILE REGRESSION ANALYSIS OF TRANSITION DATA
— *José A.F. Machado, Pedro Portugal*

2/02 SHOULD WE DISTINGUISH BETWEEN STATIC AND DYNAMIC LONG RUN EQUILIBRIUM IN ERROR CORRECTION MODELS?
— *Susana Botas, Carlos Robalo Marques*

3/02 MODELLING TAYLOR RULE UNCERTAINTY
— *Fernando Martins, José A. F. Machado, Paulo Soares Esteves*

4/02 PATTERNS OF ENTRY, POST-ENTRY GROWTH AND SURVIVAL: A COMPARISON BETWEEN DOMESTIC AND FOREIGN OWNED FIRMS
— *José Mata, Pedro Portugal*

5/02 BUSINESS CYCLES: CYCLICAL COMOVEMENT WITHIN THE EUROPEAN UNION IN THE PERIOD 1960-1999. A FREQUENCY DOMAIN APPROACH
— *João Valle e Azevedo*

6/02 AN "ART", NOT A "SCIENCE"? CENTRAL BANK MANAGEMENT IN PORTUGAL UNDER THE GOLD STANDARD, 1854 -1891
— *Jaime Reis*

7/02 MERGE OR CONCENTRATE? SOME INSIGHTS FOR ANTITRUST POLICY
— *Margarida Catalão-Lopes*

8/02 DISENTANGLING THE MINIMUM WAGE PUZZLE: ANALYSIS OF WORKER ACCESSIONS AND SEPARATIONS FROM A LONGITUDINAL MATCHED EMPLOYER-EMPLOYEE DATA SET
— *Pedro Portugal, Ana Rute Cardoso*

9/02 THE MATCH QUALITY GAINS FROM UNEMPLOYMENT INSURANCE
— *Mário Centeno*

10/02 HEDONIC PRICES INDEXES FOR NEW PASSENGER CARS IN PORTUGAL (1997-2001)
— *Hugo J. Reis, J.M.C. Santos Silva*

11/02 THE ANALYSIS OF SEASONAL RETURN ANOMALIES IN THE PORTUGUESE STOCK MARKET
— *Miguel Balbina, Nuno C. Martins*

12/02 DOES MONEY GRANGER CAUSE INFLATION IN THE EURO AREA?
— *Carlos Robalo Marques, Joaquim Pina*

13/02 INSTITUTIONS AND ECONOMIC DEVELOPMENT: HOW STRONG IS THE RELATION?
— *Tiago V.de V. Cavalcanti, Álvaro A. Novo*

2003

1/03 FOUNDING CONDITIONS AND THE SURVIVAL OF NEW FIRMS
— *P.A. Geroski, José Mata, Pedro Portugal*

2/03 THE TIMING AND PROBABILITY OF FDI: AN APPLICATION TO THE UNITED STATES MULTINATIONAL ENTERPRISES
— *José Brandão de Brito, Felipa de Mello Sampayo*

3/03 OPTIMAL FISCAL AND MONETARY POLICY: EQUIVALENCE RESULTS
— *Isabel Correia, Juan Pablo Nicolini, Pedro Teles*

2004

1/04 HOW DOES THE UNEMPLOYMENT INSURANCE SYSTEM SHAPE THE TIME PROFILE OF JOBLESS DURATION?
— *John T. Addison, Pedro Portugal*

2/04 REAL EXCHANGE RATE AND HUMAN CAPITAL IN THE EMPIRICS OF ECONOMIC GROWTH
— *Delfim Gomes Neto*

3/04 ON THE USE OF THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION INDICATOR
— *José Ramos Maria*

4/04 OIL PRICES ASSUMPTIONS IN MACROECONOMIC FORECASTS: SHOULD WE FOLLOW FUTURES MARKET EXPECTATIONS?
— *Carlos Coimbra, Paulo Soares Esteves*

5/04 STYLISED FEATURES OF PRICE SETTING BEHAVIOUR IN PORTUGAL: 1992-2001
— *Mónica Dias, Daniel Dias, Pedro D. Neves*

6/04 A FLEXIBLE VIEW ON PRICES
— *Nuno Alves*

7/04 ON THE FISHER-KONIECZNY INDEX OF PRICE CHANGES SYNCHRONIZATION
— *D.A. Dias, C. Robalo Marques, P.D. Neves, J.M.C. Santos Silva*

8/04 INFLATION PERSISTENCE: FACTS OR ARTEFACTS?
— *Carlos Robalo Marques*

9/04 WORKERS' FLOWS AND REAL WAGE CYCLICALITY
— *Anabela Carneiro, Pedro Portugal*

10/04 MATCHING WORKERS TO JOBS IN THE FAST LANE: THE OPERATION OF FIXED-TERM CONTRACTS
— *José Varejão, Pedro Portugal*

11/04 THE LOCATIONAL DETERMINANTS OF THE U.S. MULTINATIONALS ACTIVITIES
— *José Brandão de Brito, Felipa Mello Sampayo*

12/04 KEY ELASTICITIES IN JOB SEARCH THEORY: INTERNATIONAL EVIDENCE
— *John T. Addison, Mário Centeno, Pedro Portugal*

13/04 RESERVATION WAGES, SEARCH DURATION AND ACCEPTED WAGES IN EUROPE
— *John T. Addison, Mário Centeno, Pedro Portugal*

14/04 THE MONETARY TRANSMISSION N THE US AND THE EURO AREA: COMMON FEATURES AND COMMON FRICTIONS
— *Nuno Alves*

15/04 NOMINAL WAGE INERTIA IN GENERAL EQUILIBRIUM MODELS
— *Nuno Alves*

16/04 MONETARY POLICY IN A CURRENCY UNION WITH NATIONAL PRICE ASYMMETRIES
— *Sandra Gomes*

17/04 NEOCLASSICAL INVESTMENT WITH MORAL HAZARD
— *João Ejarque*

18/04 MONETARY POLICY WITH STATE CONTINGENT INTEREST RATES
— *Bernardino Adão, Isabel Correia, Pedro Teles*

19/04 MONETARY POLICY WITH SINGLE INSTRUMENT FEEDBACK RULES
— *Bernardino Adão, Isabel Correia, Pedro Teles*

20/04 ACOUNTING FOR THE HIDDEN ECONOMY: BARRIERS TO LAGALITY AND LEGAL FAILURES
— *António R. Antunes, Tiago V. Cavalcanti*

2005

1/05 SEAM: A SMALL-SCALE EURO AREA MODEL WITH FORWARD-LOOKING ELEMENTS
— *José Brandão de Brito, Rita Duarte*

2/05 FORECASTING INFLATION THROUGH A BOTTOM-UP APPROACH: THE PORTUGUESE CASE
— *Cláudia Duarte, António Rua*

3/05 USING MEAN REVERSION AS A MEASURE OF PERSISTENCE
— *Daniel Dias, Carlos Robalo Marques*

4/05 HOUSEHOLD WEALTH IN PORTUGAL: 1980-2004
— *Fátima Cardoso, Vanda Geraldes da Cunha*

5/05 ANALYSIS OF DELINQUENT FIRMS USING MULTI-STATE TRANSITIONS
— *António Antunes*

6/05 PRICE SETTING IN THE AREA: SOME STYLIZED FACTS FROM INDIVIDUAL CONSUMER PRICE DATA
— *Emmanuel Dhyne, Luis J. Álvarez, Hervé Le Bihan, Giovanni Veronese, Daniel Dias, Johannes Hoffmann, Nicole Jonker, Patrick Lünnemann, Fabio Rumler, Jouko Vilmunen*

7/05 INTERMEDIATION COSTS, INVESTOR PROTECTION AND ECONOMIC DEVELOPMENT
— *António Antunes, Tiago Cavalcanti, Anne Villamil*

8/05 TIME OR STATE DEPENDENT PRICE SETTING RULES? EVIDENCE FROM PORTUGUESE MICRO DATA
— *Daniel Dias, Carlos Robalo Marques, João Santos Silva*

9/05 BUSINESS CYCLE AT A SECTORAL LEVEL: THE PORTUGUESE CASE
— *Hugo Reis*

10/05 THE PRICING BEHAVIOUR OF FIRMS IN THE EURO AREA: NEW SURVEY EVIDENCE
— *S. Fabiani, M. Druant, I. Hernando, C. Kwapil, B. Landau, C. Loupias, F. Martins, T. Mathä, R. Sabbatini, H. Stahl, A. Stokman*

11/05 CONSUMPTION TAXES AND REDISTRIBUTION
— *Isabel Correia*

12/05 UNIQUE EQUILIBRIUM WITH SINGLE MONETARY INSTRUMENT RULES
— *Bernardino Adão, Isabel Correia, Pedro Teles*

13/05 A MACROECONOMIC STRUCTURAL MODEL FOR THE PORTUGUESE ECONOMY
— *Ricardo Mourinho Félix*

14/05 THE EFFECTS OF A GOVERNMENT EXPENDITURES SHOCK
— *Bernardino Adão, José Brandão de Brito*

15/05 MARKET INTEGRATION IN THE GOLDEN PERIPHERY – THE LISBON/LONDON EXCHANGE, 1854-1891
— *Rui Pedro Esteves, Jaime Reis, Fabiano Ferramosca*

2006

1/06 THE EFFECTS OF A TECHNOLOGY SHOCK IN THE EURO AREA
— *Nuno Alves , José Brandão de Brito , Sandra Gomes, João Sousa*

2007

2/07 THE FORWARD PREMIUM OF EURO INTEREST RATES
— *Sónia Costa, Ana Beatriz Galvão*

3/07 ADJUSTING TO THE EURO
— *Gabriel Fagan, Vítor Gaspar*

4/07 SPATIAL AND TEMPORAL AGGREGATION IN THE ESTIMATION OF LABOR DEMAND FUNCTIONS
— *José Varejão, Pedro Portugal*

5/07 PRICE SETTING IN THE EURO AREA: SOME STYLISED FACTS FROM INDIVIDUAL PRODUCER PRICE DATA
— *Philip Vermeulen, Daniel Dias, Maarten Dossche, Erwan Gautier, Ignacio Hernando, Roberto Sabbatini, Harald Stahl*

6/07 A STOCHASTIC FRONTIER ANALYSIS OF SECONDARY EDUCATION OUTPUT IN PORTUGAL
— *Manuel Coutinho Pereira, Sara Moreira*

7/07 CREDIT RISK DRIVERS: EVALUATING THE CONTRIBUTION OF FIRM LEVEL INFORMATION AND OF MACROECONOMIC DYNAMICS
— *Diana Bonfim*

8/07 CHARACTERISTICS OF THE PORTUGUESE ECONOMIC GROWTH: WHAT HAS BEEN MISSING?
— *João Amador, Carlos Coimbra*

9/07 TOTAL FACTOR PRODUCTIVITY GROWTH IN THE G7 COUNTRIES: DIFFERENT OR ALIKE?
— *João Amador, Carlos Coimbra*

10/07 IDENTIFYING UNEMPLOYMENT INSURANCE INCOME EFFECTS WITH A QUASI-NATURAL EXPERIMENT
— *Mário Centeno, Alvaro A. Novo*

11/07 HOW DO DIFFERENT ENTITLEMENTS TO UNEMPLOYMENT BENEFITS AFFECT THE TRANSITIONS FROM UNEMPLOYMENT INTO EMPLOYMENT
— *John T. Addison, Pedro Portugal*

12/07 INTERPRETATION OF THE EFFECTS OF FILTERING INTEGRATED TIME SERIES
— *João Valle e Azevedo*

13/07 EXACT LIMIT OF THE EXPECTED PERIODOGRAM IN THE UNIT-ROOT CASE
— *João Valle e Azevedo*

14/07 INTERNATIONAL TRADE PATTERNS OVER THE LAST FOUR DECADES: HOW DOES PORTUGAL COMPARE WITH OTHER COHESION COUNTRIES?
— *João Amador, Sónia Cabral, José Ramos Maria*

15/07 INFLATION (MIS)PERCEPTIONS IN THE EURO AREA
— *Francisco Dias, Cláudia Duarte, António Rua*

16/07 LABOR ADJUSTMENT COSTS IN A PANEL OF ESTABLISHMENTS: A STRUCTURAL APPROACH
— *João Miguel Ejarque, Pedro Portugal*

17/07 A MULTIVARIATE BAND-PASS FILTER
— *João Valle e Azevedo*

18/07 AN OPEN ECONOMY MODEL OF THE EURO AREA AND THE US
— *Nuno Alves, Sandra Gomes, João Sousa*

19/07 IS TIME RIPE FOR PRICE LEVEL PATH STABILITY?
— *Vitor Gaspar, Frank Smets , David Vestin*

20/07 IS THE EURO AREA M3 ABANDONING US?
— *Nuno Alves, Carlos Robalo Marques, João Sousa*

21/07 DO LABOR MARKET POLICIES AFFECT EMPLOYMENT COMPOSITION? LESSONS FROM EUROPEAN COUNTRIES
— *António Antunes, Mário Centeno*

2008

1/08 THE DETERMINANTS OF PORTUGUESE BANKS' CAPITAL BUFFERS
— *Miguel Boucinha*

2/08 DO RESERVATION WAGES REALLY DECLINE? SOME INTERNATIONAL EVIDENCE ON THE DETERMINANTS OF RESERVATION WAGES
— *John T. Addison, Mário Centeno, Pedro Portugal*

3/08 UNEMPLOYMENT BENEFITS AND RESERVATION WAGES: KEY ELASTICITIES FROM A STRIPPED-DOWN JOB SEARCH APPROACH
— *John T. Addison, Mário Centeno, Pedro Portugal*

4/08 THE EFFECTS OF LOW-COST COUNTRIES ON PORTUGUESE MANUFACTURING IMPORT PRICES
— *Fátima Cardoso, Paulo Soares Esteves*

5/08 WHAT IS BEHIND THE RECENT EVOLUTION OF PORTUGUESE TERMS OF TRADE?
— *Fátima Cardoso, Paulo Soares Esteves*

6/08 EVALUATING JOB SEARCH PROGRAMS FOR OLD AND YOUNG INDIVIDUALS: HETEROGENEOUS IMPACT ON UNEMPLOYMENT DURATION
— *Luis Centeno, Mário Centeno, Álvaro A. Novo*

7/08 FORECASTING USING TARGETED DIFFUSION INDEXES
— *Francisco Dias, Maximiano Pinheiro, António Rua*

8/08 STATISTICAL ARBITRAGE WITH DEFAULT AND COLLATERAL
— *José Fajardo, Ana Lacerda*

9/08 DETERMINING THE NUMBER OF FACTORS IN APPROXIMATE FACTOR MODELS WITH GLOBAL AND GROUP-SPECIFIC FACTORS
— *Francisco Dias, Maximiano Pinheiro, António Rua*

10/08 VERTICAL SPECIALIZATION ACROSS THE WORLD: A RELATIVE MEASURE
— *João Amador, Sónia Cabral*

11/08 INTERNATIONAL FRAGMENTATION OF PRODUCTION IN THE PORTUGUESE ECONOMY: WHAT DO DIFFERENT MEASURES TELL US?
— *João Amador, Sónia Cabral*

12/08 IMPACT OF THE RECENT REFORM OF THE PORTUGUESE PUBLIC EMPLOYEES' PENSION SYSTEM
— *Maria Manuel Campos, Manuel Coutinho Pereira*

13/08 EMPIRICAL EVIDENCE ON THE BEHAVIOR AND STABILIZING ROLE OF FISCAL AND MONETARY POLICIES IN THE US
— *Manuel Coutinho Pereira*

14/08 IMPACT ON WELFARE OF COUNTRY HETEROGENEITY IN A CURRENCY UNION
— *Carla Soares*

15/08 WAGE AND PRICE DYNAMICS IN PORTUGAL
— *Carlos Robalo Marques*

16/08 IMPROVING COMPETITION IN THE NON-TRADABLE GOODS AND LABOUR MARKETS: THE PORTUGUESE CASE
— *Vanda Almeida, Gabriela Castro, Ricardo Mourinho Félix*

17/08 PRODUCT AND DESTINATION MIX IN EXPORT MARKETS
— *João Amador, Luca David Opromolla*

18/08 FORECASTING INVESTMENT: A FISHING CONTEST USING SURVEY DATA
— *José Ramos Maria, Sara Serra*

19/08 APPROXIMATING AND FORECASTING MACROECONOMIC SIGNALS IN REAL-TIME
— *João Valle e Azevedo*

20/08 A THEORY OF ENTRY AND EXIT INTO EXPORTS MARKETS
— *Alfonso A. Irarrazabal, Luca David Opromolla*

21/08 ON THE UNCERTAINTY AND RISKS OF MACROECONOMIC FORECASTS: COMBINING JUDGEMENTS WITH SAMPLE AND MODEL INFORMATION
— *Maximiano Pinheiro, Paulo Soares Esteves*

22/08 ANALYSIS OF THE PREDICTORS OF DEFAULT FOR PORTUGUESE FIRMS
— *Ana I. Lacerda, Russ A. Moro*