

# Forecasting Corporate Distress in the Asian and Pacific Region\*

Russ Moro<sup>†</sup>   Wolfgang Härdle<sup>‡</sup>  
Saeideh Aliakbari<sup>§</sup>   Linda Hoffmann<sup>¶</sup>

\*The authors are grateful for the financial support, data access and excellent research facilities provided by the Risk Management Institute (RMI) of the National University of Singapore (NUS) for the study of corporate distress in Asia and the Pacific region. We personally thank Prof. Duan Jin-Chuan and Dr. Oliver Chen for their inspiration and valuable suggestions. We also would like to express our thanks to the staff of RMI NUS for their patience and readiness to help in conducting our research. The work of R. A. Moro in development of the statistical algorithms was partially supported by the German Academic Exchange Service (DAAD). W. K. Härdle and L. Hoffmann were also assisted by the Deutsche Forschungsgemeinschaft through the SFB 649 ‘Economic Risk’.

<sup>†</sup>Brunel University, Department of Economics and Finance, School of Social Sciences, Uxbridge, Middx UB8 3PH, United Kingdom and DIW Econ, Mohrenstraße 58, 10117 Berlin, Germany; e-mail: Russ.Moro@brunel.ac.uk; phone: +44 1895 266058.

<sup>‡</sup>Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany and National Central University, Department of Finance, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan (R.O.C.); e-mail: stat@wiwi.hu-berlin.de.

<sup>§</sup>Brunel University, Department of Economics and Finance, School of Social Sciences, Uxbridge, Middx UB8 3PH, United Kingdom; e-mail: Saeideh.Aliakbari@brunel.ac.uk.

<sup>¶</sup>Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany; e-mail: Linda.Hoffmann@cms.hu-berlin.de.

This study analyses credit default risk for firms in the Asian and Pacific region by applying two methodologies: a Support Vector Machine (SVM) and a logistic regression (Logit). Among different financial ratios suggested as predictors of default, leverage ratios and the company size display a higher discriminating power compared to others. An analysis of the dependencies between PD and financial ratios is provided along with a comparison with Europe (Germany). With respect to forecasting accuracy the SVM has a lower model risk than the Logit on average and displays a more robust performance. This result holds true across different years.

KEYWORDS: Credit risk, Bankruptcy, Asian companies, SVM

JEL CLASSIFICATION: C14, G33, C45

# 1 Introduction

Although credit risk has always been a major concern for investors, in recent years high profile insolvencies have attracted widespread attention, first, after the dot-com bubble and then in connection with the subprime mortgage crisis. In the Asian and Pacific region, particularly the crisis of 1998, caused a wave of insolvencies. The announcement of the Basel III Capital Accord in 2010 after the adoption of Basel II in 2004 and Basel I in 1992 indicates both the concern of banks and regulators about providing protection against credit risk and, at the same time, inadequacy of the existing protection measures and methods for measuring risk.

As early as in the beginning of the XXth century [Winakor & Smith \(1935\)](#) proposed the use of financial ratios for separating firms into solid stable and potentially bankrupt ones. [Ramser & Foster \(1931\)](#) and [Fitzpatrick \(1932\)](#) also applied financial ratios for bankruptcy prediction. The systematic application of statistics to bankruptcy analysis began with the works of [Beaver \(1966\)](#) and [Altman \(1968\)](#). They introduced the univariate and multivariate discriminant analysis (DA), respectively. In 1968 Altman presented a formula for predicting bankruptcy known as the linear Z-score model ([Altman, Haldeman & Narayanan, 1977](#)). This formula remains popular for forecasting default rates even today due to its simplicity. The drawback of the Z-score model is the assumption of equal normal distributions for both failing and successful firms with the same covariance matrix.

Later the focus of research shifted towards the logit and probit models ([Ohlson \(1980\)](#), [Martin \(1977\)](#), [Wiginton \(1980\)](#), [Zavgren \(1983\)](#) and [Zmijewski \(1984\)](#)). Other statistical methods which were introduced at the same time, such as the gambler's ruin model ([Wilcox, 1971](#)) and option pricing theory ([Merton, 1974](#)), were based on time series data. Later hazard or survival models ([Glennon & Nigro, 2005](#)) and Forward Intensity Approach ([J. C. Duan & Wang, 2010](#)) used both time series and cross-sectional data. Another type of models such as recursive partitioning ([Frydman, Altman & Kao, 1985](#)), neural networks ([Tam & Kiang, 1992](#)), rough sets ([Dimitras, Slowinski, Susmaga & Zopounidis, 1999](#)) and Support Vector Machines (SVM) ([Martens, Baesens, van Gestel & Vanthienen, 2006](#)) were mostly applied to cross-sectional data.

One of the major shortcomings of many methodologies is the fact that they ignore non-monotonic dependence between some financial ratios and the PD such as the logistic regression or are badly suited for credit risk modelling such as neural networks due to their multiple local equilibria. For further information please refer to [Falkenstein, Boral & Carty \(2000\)](#), [Manning \(2004\)](#), [Fernandes \(2005\)](#) and [Härdle, Moro & Schäfer \(2010\)](#). For instance, the probability of default (PD) is non-

monotonically dependent from the net income (NI) growth. Negative or very slowly growing NI may create problems with paying company debt obligations. On the other hand, high NI is likely to be non-sustainable in the long run causing high volatility. Both situations can lead to a higher PD, what is in accordance with the existing literature (Merton (1974), Bharath & Shumway (2008)). The identification of the shape of the dependence, however, still remains a problem.

The non-monotonic and non-linear dependence between some financial ratios and the PD has been addressed by introducing non-linear models such as recursive partitioning, also known as classification and regression trees (Frydman, Altman & Kao (1985), Frydman, Altman & Kao (1985)), neural networks (Tam & Kiang (1992)), Proximal Support Vector Machines (PSVM) ((Friedman, 2002)) and Support Vector Machines (SVM) ((Martens, Baesens, van Gestel & Vanthienen, 2006), (Härdle, Moro & Schäfer, 2010)).

When classifying distressed vs. solvent companies, the SVM allows adjustment of its complexity. The complexity can be then optimised with respect to some accuracy measure, for example the Accuracy Ratio (AR), for the data and predicting variables at hand. Figure 1 illustrates the classical trade-off between the good in-sample performance and the generalisation ability. In this example by changing complexity of the classification method between it possible For more details on the SVM please see Appendix A.

In this study we use the Logit and SVM approaches, both in their cross-sectional and dynamic setting, to analyse credit risk of firms in the Asian and Pacific region and to establish the most important predictors of default selected from financial ratios.

## 2 Data Description

The data used in this study were collected and prepared by the Risk Management Institute (RMI) of the National University of Singapore (NUS). The data contain quarterly and annual company reports, default indicators and stock prices for 25,000 listed firms from the Asian and Pacific region as well as the macroeconomic and selected financial data for the countries in which the firms operate. The time coverage spans from 1980 to 2010. The database also indicates the relevant industry of operation for each firm. In our analysis we exclude companies in the financial sector, asset backed securities, funds and government owned enterprises since the nature of these businesses is different from non-governmental manufacturing firms and service providers.

At the first stage the financial data are converted into financial ratios. These ratios are

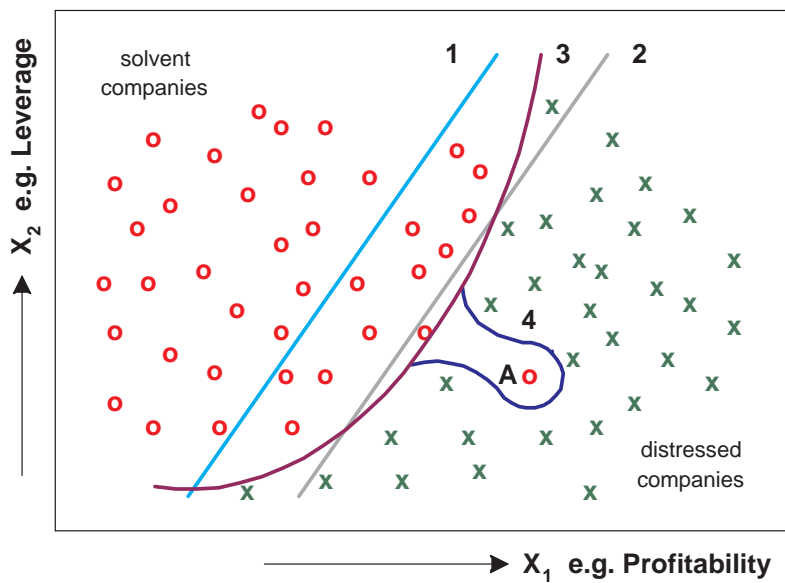


Figure 1: A classification example. The boundary between the classes of solvent and insolvent companies can be either linear (1 or 2) or non-linear (3 and 4). A model capable of producing non-linear boundaries can have low (linear cases 1 and 2), moderate (case 3) and high (case 4 where overfitting is evident) complexities. By optimising the complexity with respect to some accuracy criterion, the optimal boundary can be established (e.g. case 3).

grouped into seven categories: profitability, leverage, liquidity, activity, cost structure, dynamics and size, characterising company performance from different sides. A summary statistics of the indicators is presented in tables 2 – 4.

Financial reports in the database are released quarterly, semi annually and annually, however, the beginning of a financial year and, hence, reporting dates for companies are different and spread throughout the year. To reflect this situation we index each financial report by a unique time ID number according to the year and month of the report in order to have the financial information on regular monthly basis for all firms. Since the reporting date almost invariably falls on the last day of a month, this encoding gives us the precise time of a default event.

After assigning the report time ID number to each observation, distressed firms are defined based on the default information in the database. Each monthly report of a firm receives the default indicator  $y = 1$  if the firm files a credit event report within a period with a one year long period starting after one year after the date of the financial report (distressed observations). For the rest of the observations (solvent observations) the default indicator is  $y = -1$ . In this study we call this horizon specification *design 1*. This horizon is considered to analyse the effects of the default on the long term debt which has maturity of over one year.

Additionally, to see the effects of the short term debt on PD, we analyse distress for a different horizon, when the default indicator  $y = 1$  is assigned to those observations recording a credit event report filing within the two year period from the date of the financial report (distressed observations) and for the rest (solvent observations) the default indicator is  $y = -1$ . This horizon specification is called *design 2*.

A broad range of credit events is applied to identify distressed firms and assign the default indicator ( $y = 1$ ), including filings under Chapter 11, Chapter 15, Chapter 7, restructuring, liquidation, being sued by creditors and failing in coupon and principle payments. Overall, the bankruptcy events coded from 100 to 120 and 300 to 333 in the database are included to define distressed observations.

In the dataset with the horizon under the *design 2* specification, there are 311,682 observations from which 7,449 (2.39%) observations are indicated as distressed and 304,233 (97.61%) as solvent. The distribution of solvent and distressed observations among countries varies substantially. For instance, for Australia and Hong Kong, there are respectively only 6 (4.03%) and 19 (0.34%) of distressed observations out of 149 and 5,524 observations whereas for China there are 4,182 (7.22%) distressed observations out of 57,921 observations (see table 1).

Country	Horizon: <i>Design 1</i>		Horizon: <i>Design 2</i>	
	Distressed firms	Solvent firms	Distressed firms	Solvent firms
Australia	3 (2.01 %)	146	6 (4.03 %)	143
China	1088 (2.42 %)	43784	4182 (7.22 %)	53739
Hong Kong	10 (0.18 %)	5514	19 (0.34 %)	5505
India	48 (0.17 %)	28840	148 (0.51 %)	28775
Indonesia	26 (0.42 %)	6131	70 (1.12 %)	6186
Japan	124 (0.17 %)	71489	258 (0.36 %)	71380
Malaysia	385 (1.17 %)	32512	1100 (3.12 %)	34173
Philippines	113 (1.90 %)	5839	267 (4.16 %)	6154
Singapore	34 (0.48 %)	7009	77 (1.08 %)	7050
South Korea	99 (0.20 %)	49828	232 (0.46 %)	50153
Taiwan	260 (1.08 %)	23906	604 (2.47 %)	23809
Thailand	202 (1.19 %)	16702	486 (2.77 %)	17028

Table 1: Distribution of distressed and solvent firms across countries.

## 2.1 Variable Description

The components of the financial ratios which are estimated from data are explained below and the summary statistics for them for distressed and solvent firms are provided in tables 2 and 3.

### *Profitability Ratios*

1. NI/TA : return on assets; net income / total assets.
2. NI/S : net profit margin; net income / sales.
3. OI/TA: operating return on assets; operating income / total assets.
4. OI/S : operating profit margin; operating income / sales.
5. EBIT/TA: gross return on assets; earnings before interest and taxes / total assets.
6. EBIT/S : gross profit margin; earnings before interest and taxes / sales.

### *Leverage Ratios*

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

### *Liquidity Ratios*

1. STD/D : fraction of debt which is short term debt (liquidity).
2. CASH/TA : cash and cash equivalents / total assets.
3. CASH/CL : cash ratio; the ratio of cash and cash equivalents / current liabilities.

4. QA/CL : quick ratio; quick assets (current assets inventories) / current liabilities.
5. CA/CL : current ratio; the ratio of current assets / current liabilities.
6. WC/TA : working capital (current assets minus current liabilities) / total assets.
7. CL/TL : current liabilities / total liabilities.

#### *Activity Ratios*

1. TA/S : asset turnover; total assets / sales.
2. INV/S : inventory turnover; inventories / sales.
3. AR/S : account receivable turnover; account receivables / sales.
4. AP/CS : account payable turnover; account payables / cost of sales.

#### *Cost Structure Ratios*

1. INT/D : average cost of debt; the ratio of interest payments to debt.
2. EBIT/INT paid : interest coverage ratio; the ratio of earnings before interest and taxes to interest paid.

#### *Dynamic Ratios*

1. Sales-Growth : one year growth in sales.
2. NI-Growth : one year growth in income.

#### *Size*

1. log(TA) : company size; logarithm of total assets.
2. log(S) : logarithm of total sales.

## **2.2 Summary Statistics**

In this section summary statistics of the financial ratios for distressed and solvent companies are provided. They are reported for *Design 1* (table 2) and *Design 2* (table 3) horizon designs, pooled across countries. The first five columns in each table summarize the estimates for distressed companies and the next five columns report the estimates for solvent companies.  $q_{0.05}$  and  $q_{0.95}$  are 5% and 95% quantiles.  $N$  is the number of observations for which the ratio can be computed based on the available data and IQR represents the interquartile range for each ratio.

1986-2010	Distressed Firms					Solvent Firms				
Variable	N	$q_{0.05}$	Med	IQR	$q_{0.95}$	N	$q_{0.05}$	Med	IQR	$q_{0.95}$
<b>Profitability</b>										
NI/TA	1529	-0.10	-0.00	0.01	0.04	231456	-0.05	0.01	0.02	0.05
NI/S	2062	-2.90	-0.00	0.05	0.30	281329	-0.57	0.03	0.09	0.27
OI/TA	1704	-0.06	0.00	0.01	0.03	232939	-0.03	0.01	0.02	0.05
OI/S	1723	-1.17	0.02	0.09	0.30	249187	-0.37	0.05	0.11	0.29
EBIT/TA	1523	-0.06	0.00	0.01	0.04	231084	-0.03	0.01	0.02	0.05
EBIT/S	1542	-1.41	0.02	0.10	0.31	247091	-0.37	0.05	0.11	0.29
<b>Leverage</b>										
OK/TA	1716	-0.68	0.36	0.49	0.67	234206	0.16	0.54	0.70	0.88
CL/TA	1716	0.15	0.48	0.65	1.51	233974	0.07	0.32	0.45	0.68
TD/TA	1506	0.12	0.44	0.58	0.87	228469	0.00	0.20	0.36	0.58
<b>Liquidity</b>										
STD/D	1497	0.12	0.80	0.97	1.00	204949	0.08	0.69	0.97	1.00
CASH/TA	1685	0.00	0.04	0.10	0.27	233234	0.00	0.09	0.17	0.38
CASH/CL	1685	0.00	0.08	0.24	0.84	232945	0.01	0.27	0.65	2.70
QA/CL	1680	0.11	0.68	1.08	2.15	231103	0.30	1.11	1.91	5.58
CA/CL	1713	0.18	1.00	1.47	2.79	233920	0.50	1.52	2.46	6.60
WC/TA	1713	-0.96	-0.00	0.17	0.42	233948	-0.22	0.17	0.34	0.58
CL/TL	1716	0.28	0.81	0.95	1.00	233921	0.32	0.78	0.92	1.00
<b>Activity</b>										
TA/S	1697	2.29	8.67	16.61	90.13	232584	1.74	4.69	7.59	22.83
INV/S	1657	0.05	0.80	1.62	8.57	229603	0.01	0.47	0.84	2.26
AR/S	1653	0.17	0.93	1.66	4.82	230847	0.08	0.72	1.07	2.05
AP/CS	1085	0.09	0.65	1.14	4.52	174888	0.04	0.41	0.68	1.33
<b>Cost Structure</b>										
INT/D	712	0.01	0.03	0.08	0.70	130408	0.00	0.02	0.06	0.56
EBIT/INT	803	-20.08	0.45	2.68	20.53	172564	-28.08	4.25	19.86	326.33
<b>Dynamics</b>										
Sales-Growth	1617	-72.55	-2.81	27.42	119.62	226216	-48.84	5.21	23.89	97.17
NI-Growth	1744	-5.59	0.46	1.26	13.34	229415	-4.69	0.18	0.93	5.14
<b>Size</b>										
log(TA)	1721	4.73	7.57	9.12	12.21	234284	4.95	9.31	11.02	13.51
log(S)	2249	1.80	4.94	6.45	9.83	284275	2.32	7.09	9.24	11.85

Table 2: Summary statistics for distressed firms (the left five columns) and solvent firms (the right five columns) across countries. Horizon: *Design 1*. N indicates the number of observations which contain the variable.  $q_{0.05}$  and  $q_{0.95}$  are respectively 5% and 95% quantiles. IQR is the interquartile range.

1986-2010	Distressed Firms					Solvent Firms				
Variable	N	$q_{0.05}$	Med	IQR	$q_{0.95}$	N	$q_{0.05}$	Med	IQR	$q_{0.95}$
<b>Profitability</b>										
NI/TA	4989	-0.15	-0.00	0.00	0.03	240454	-0.06	0.01	0.02	0.05
NI/S	6657	-4.18	-0.03	0.04	0.26	292615	-0.69	0.03	0.09	0.28
OI/TA	5342	-0.09	0.00	0.01	0.03	241847	-0.04	0.01	0.02	0.05
OI/S	5397	-1.81	0.01	0.08	0.29	257567	-0.44	0.05	0.11	0.29
EBIT/TA	4967	-0.09	0.00	0.01	0.03	240011	-0.04	0.01	0.02	0.05
EBIT/S	5022	-2.12	0.00	0.09	0.30	255490	-0.43	0.05	0.11	0.29
<b>Leverage</b>										
OK/TA	5381	-1.25	0.34	0.48	0.63	243232	0.12	0.53	0.70	0.88
CL/TA	5381	0.18	0.54	0.72	2.02	242997	0.08	0.32	0.47	0.73
TD/TA	4913	0.15	0.44	0.58	1.06	237398	0.00	0.21	0.37	0.60
<b>Liquidity</b>										
STD/D	4890	0.17	0.89	1.00	1.00	213146	0.08	0.71	0.97	1.00
CASH/TA	5296	0.00	0.05	0.11	0.30	242186	0.00	0.08	0.17	0.38
CASH/CL	5296	0.00	0.08	0.24	0.71	241890	0.01	0.26	0.63	2.63
QA/CL	5289	0.10	0.64	1.00	1.91	239845	0.27	1.09	1.87	5.47
CA/CL	5375	0.15	0.90	1.37	2.40	242893	0.44	1.49	2.41	6.51
WC/TA	5375	-1.67	-0.04	0.16	0.37	242925	-0.29	0.16	0.33	0.58
CL/TL	5381	0.33	0.88	0.97	1.00	242941	0.32	0.79	0.93	1.00
<b>Activity</b>										
TA/S	5294	2.53	9.79	17.62	107.11	240836	1.75	4.77	7.89	25.78
INV/S	5193	0.09	0.90	1.78	9.48	237729	0.01	0.48	0.87	2.54
AR/S	5185	0.19	1.00	1.82	6.54	238940	0.08	0.72	1.09	2.18
AP/CS	3427	0.08	0.65	1.20	4.97	178398	0.04	0.42	0.69	1.38
<b>Cost Structure</b>										
INT/D	1802	0.01	0.04	0.12	3.02	131670	0.00	0.02	0.06	0.57
EBIT/INT	2095	-22.50	-0.38	1.68	12.90	174050	-28.00	4.18	19.60	322.00
<b>Dynamics</b>										
Sales-Growth	5389	-78.25	-7.82	20.61	117.90	235485	-51.28	5.12	24.24	100.80
NI-Growth	5792	-6.71	0.48	1.43	19.97	239939	-4.77	0.19	0.94	5.31
<b>Size</b>										
log(TA)	5393	4.70	7.31	8.32	11.49	243336	4.89	9.16	10.97	13.46
log(S)	7029	1.39	4.70	5.94	9.39	295454	2.18	6.93	9.17	11.80

Table 3: Summary statistics for distressed firms (the left five columns) and solvent firms (the right five columns) across countries. Horizon: *Design 2*. N indicates the number of observations which contain the variable.  $q_{0.05}$  and  $q_{0.95}$  are respectively 5% and 95% quantiles. IQR is the interquartile range.

Country	Horizon: <i>Design 1</i>		Horizon: <i>Design 2</i>	
	Distressed firms	Solvent firms	Distressed firms	Solvent firms
Australia	0 (0.00 %)	48	0 (0.00 %)	48
China	634 (1.83 %)	34048	2639 (6.10 %)	40627
Hong Kong	3 (0.19 %)	1588	7 (0.44 %)	1584
India	0 (0.00 %)	156	0 (0.00 %)	156
Indonesia	26 (0.45 %)	5758	70 (1.19 %)	5811
Japan	104 (0.16 %)	64735	227 (0.35 %)	64637
Malaysia	274 (1.14 %)	23693	813 (3.21 %)	24512
Philippines	39 (2.04 %)	1870	102 (4.84 %)	2007
Singapore	29 (0.53 %)	5431	70 (1.27 %)	5454
South Korea	77 (0.16 %)	48115	177 (0.37 %)	48385
Taiwan	77 (0.35 %)	22025	226 (1.02 %)	21947
Thailand	161 (1.07 %)	14856	352 (2.27 %)	15124

Table 4: Distribution of distressed and solvent firms across countries after removing 6 variables with most missing values. These variables are : INT/D, EBIT/INT, AP/CS, STD/D, Sales-Growth and NI-Growth.

As we can see from table 3, the lowest number of available observations belong to 6 variables: INT/D, EBIT/INT, AP/CS, STD/D, Sales-Growth and NI-Growth. Table 4 presents the distribution of distressed and solvent firms after removing these 6 variables with most missing values. After removing them and cleaning missing values the total number of distressed observations in the data set increases from 1,182 to 4,688.

### 3 Univariate Analysis of the Predictors of Default

The analysis of financial ratios and their individual power as predictors of default can be concisely done by estimating univariate dependence 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 2 – 8. They were obtained as  $k$  nearest neighbor estimates ( $k$ -NN) with Gaussian weights:

$$PD(q) = \frac{\sum_{i=1}^n \mathcal{I}(y_i = 1) e^{-\frac{(q-q_i)^2}{2\sigma^2}}}{\sum_{i=1}^n e^{-\frac{(q-q_i)^2}{2\sigma^2}}}, \quad (3.1)$$

where  $0 \leq q \leq 1$  is a percentile of a company for which PD is estimated,  $q_i$  is the percentile of company  $i$  of the data set and the smoothing parameter  $\sigma$  is set to 0.08.  $\mathcal{I}(y_i = 1)$  is the distress indicator which equals 1 if  $y_i = 1$  when company  $i$  is defined as distressed and 0 otherwise.

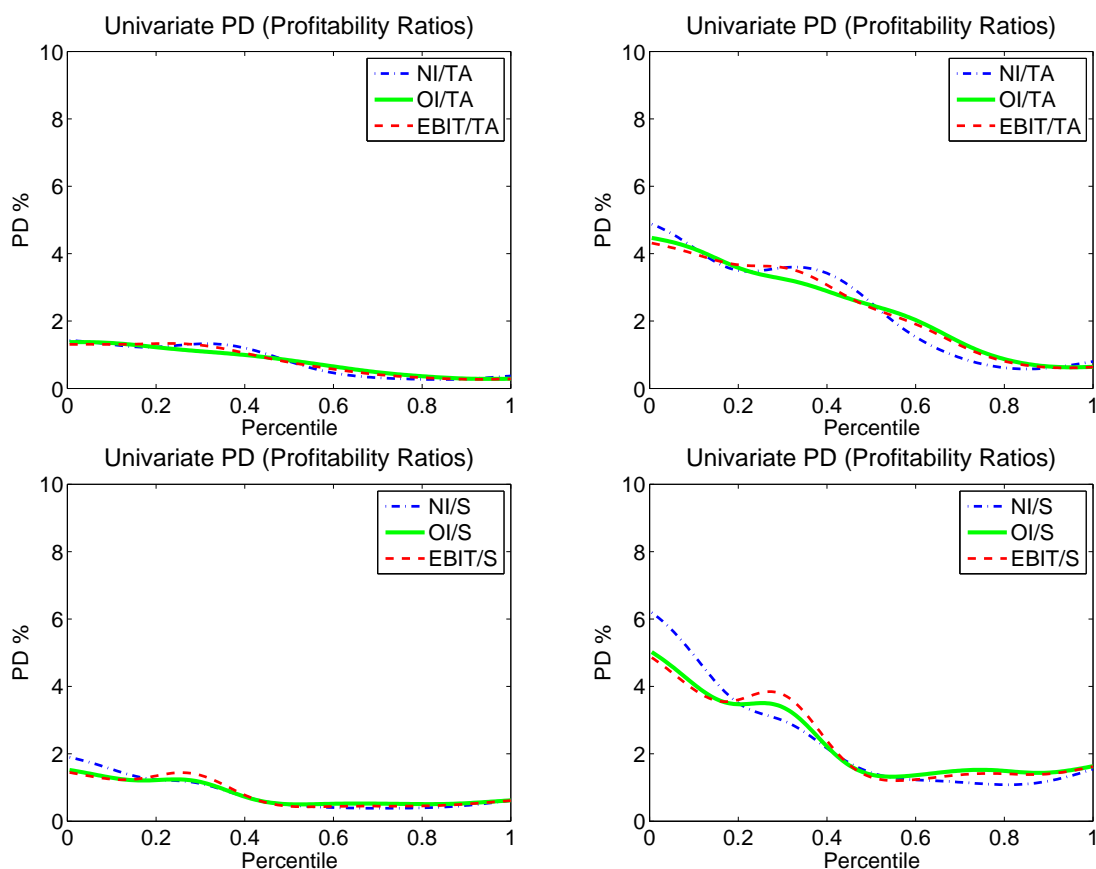


Figure 2: Univariate probabilities of default for *Profitability Ratios* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

The variables differ substantially in their predictive power. For instance, variables EBIT/TA, CL/TA and  $\log(S)$  indicate strong predictive power and also traditionally appear in the literature as strong indicators. In contrast some variables such as STD/D, AR/S and Sales-Growth show less discriminating power.

Another important observation from the plots is that some predictors, many of which with high discriminating power, such as CL/TA, OK/TA, CA/CL, EBIT/INT<sub>paid</sub>,  $\log(TA)$ , INT/D and CL/TL have a non-monotonic dependence with PD.

We analyse the relationship between each predictor of default with PD and their predictive power on data pooled over countries. The results are presented for the two horizon designs, *Design 1* and *Design 2*.

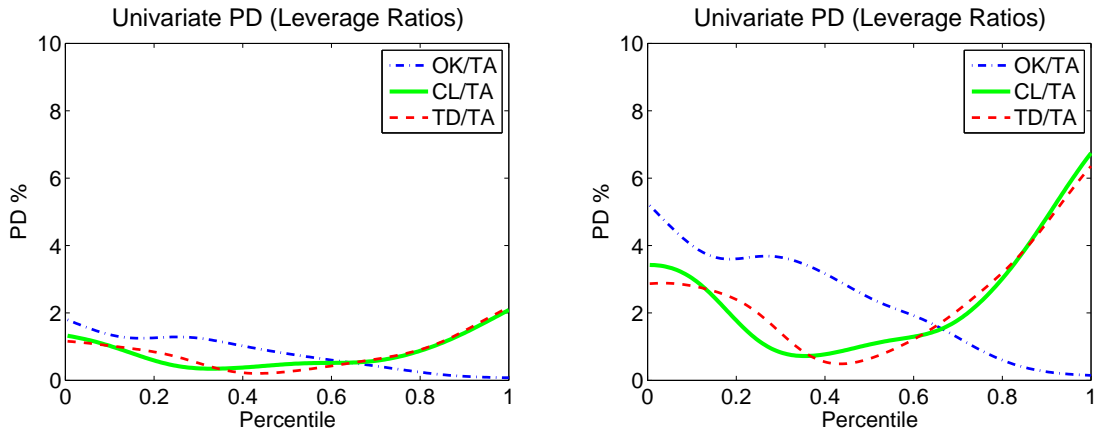


Figure 3: Univariate probabilities of default for *Leverage Ratios* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

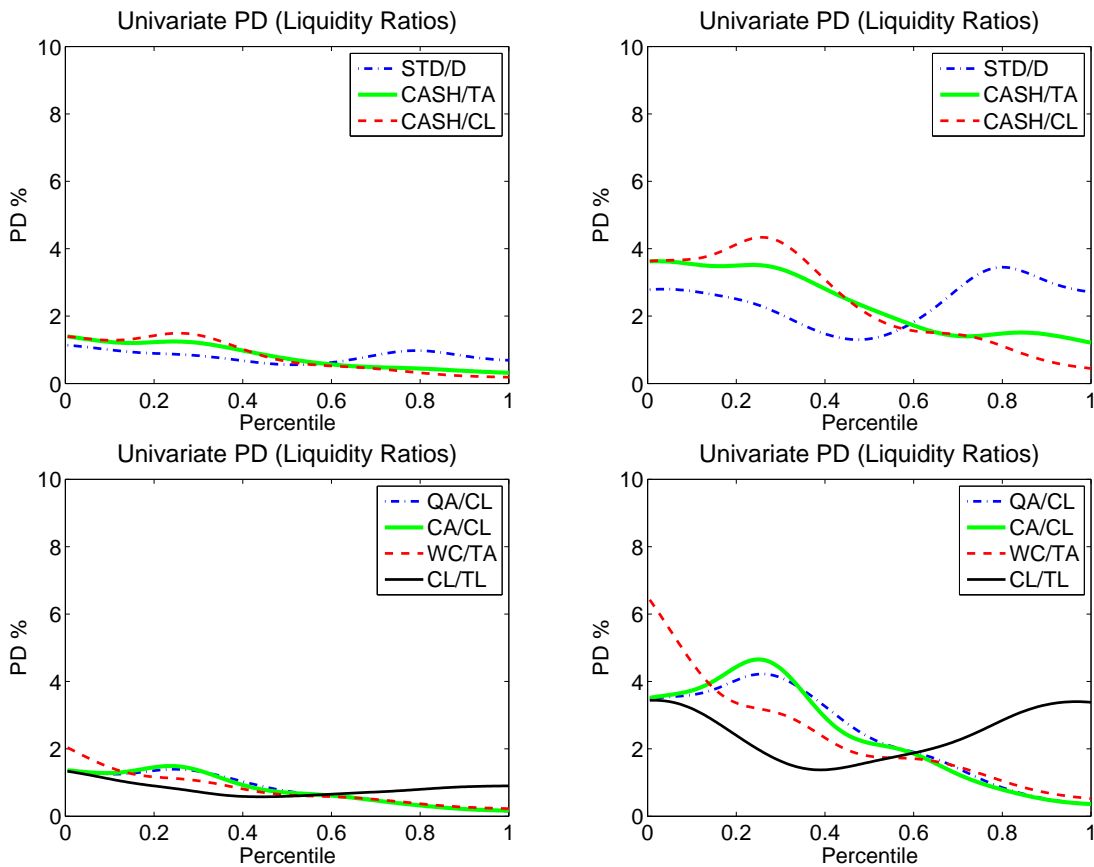


Figure 4: Univariate probabilities of default for *Liquidity Ratios* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

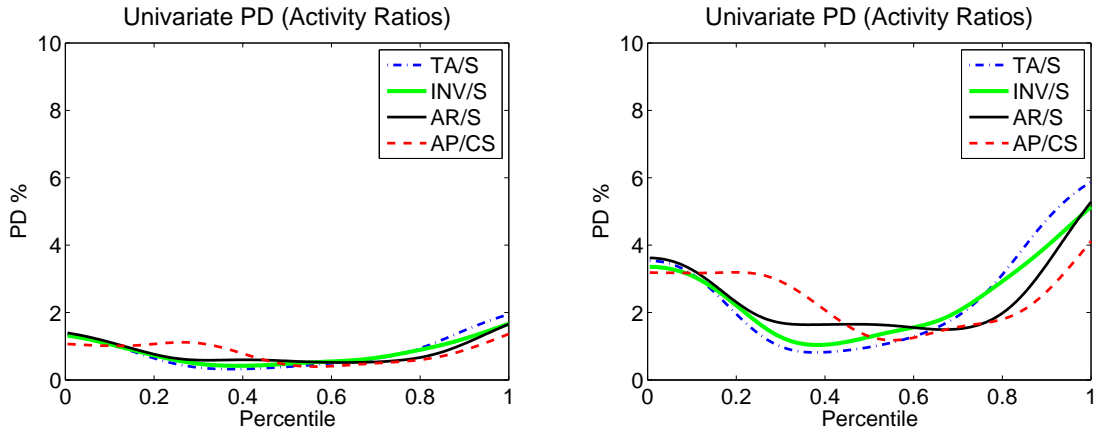


Figure 5: Univariate probabilities of default for *Activity Ratios* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

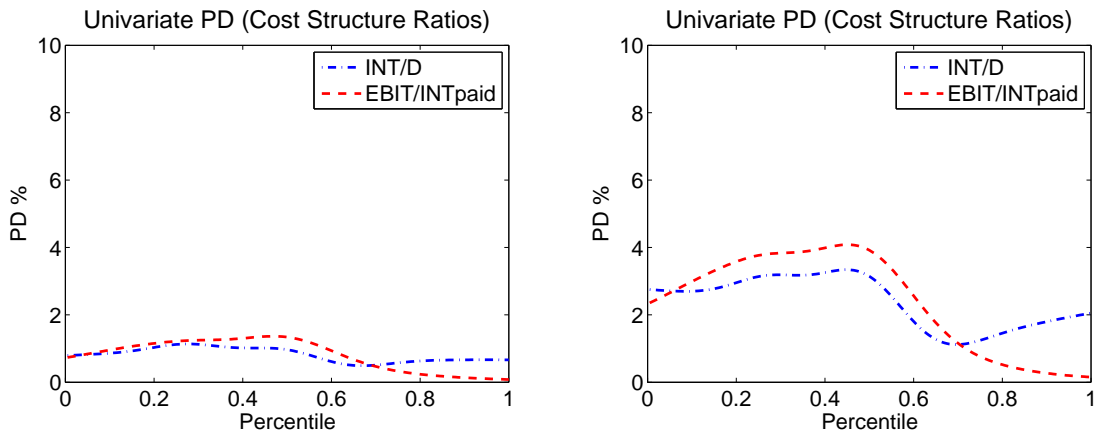


Figure 6: Univariate probabilities of default for *Cost Structure Ratios* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

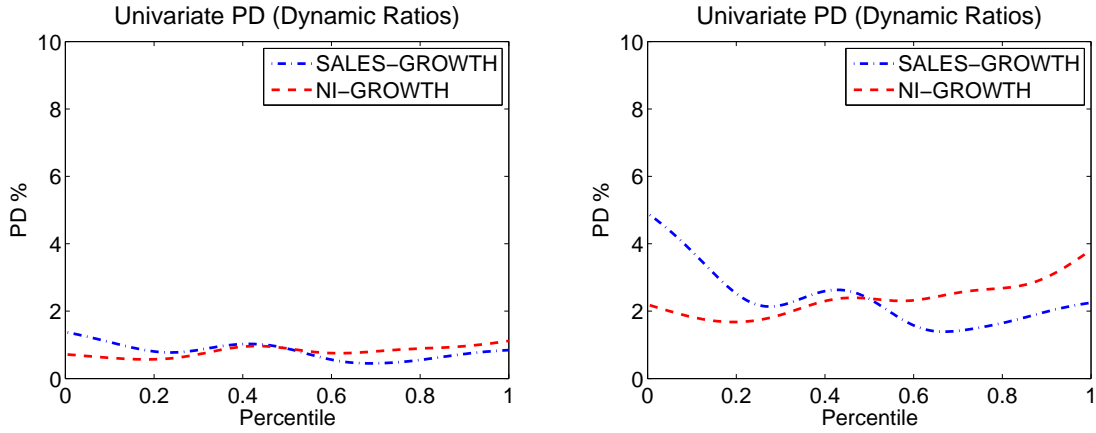


Figure 7: Univariate probabilities of default for *Dynamic Ratios* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

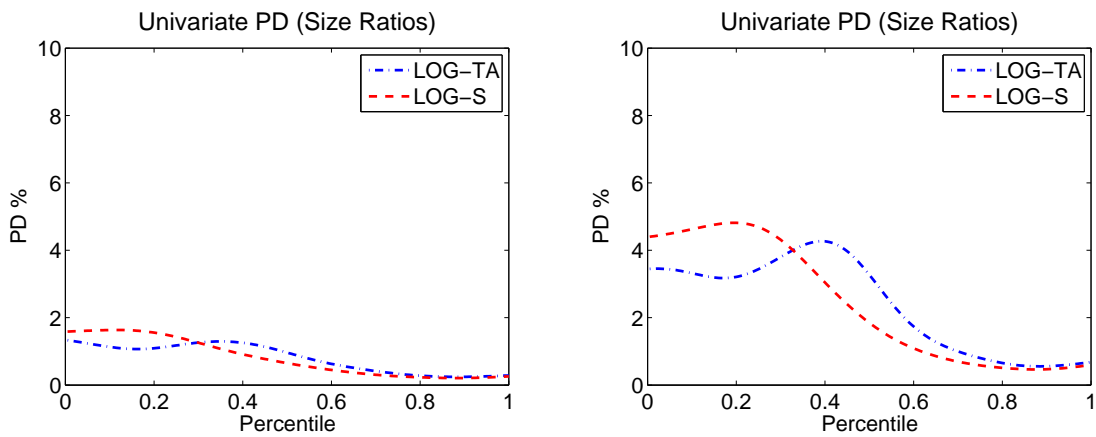


Figure 8: Univariate probabilities of default for *Company Size* pooled over countries and years. Horizon: *Design 1* (left panel), Horizon: *Design 2* (right panel).

## 4 Comparison of the PD between Asian and German Companies

In this section we compare the results of the PD univariate analysis for Asian firms (horizon: *Design 1*) pooled across all countries with the analysis of the same or very close financial ratios for German firms. The dataset for German companies was kindly provided for our analysis by the Deutsche Bundesbank and covers the years 1987 – 2005 containing around 500,000 balance sheets and income statements from which 8,000 belong to bankrupt firms. Some of the financial ratios that are used by the Deutsche Bundesbank for company rating are the same as constructed for Asian companies, while others are specific for Germany. We report the comparison of the common financial ratios in figures 9 – 15.

The graphs for Germany report the cumulative default rate with the horizon of default of three years and above, whereas the horizon of default for RMI data lies between one and two years. The possibility of registering default within a much broader range of horizons explains the higher PD for German data. For more information see (Härdle, Moro & Schäfer, 2010).

German companies in contrast to Asian ones are primarily private (non-traded) and are of a smaller size. Moreover, the sample of firms in the Bundesbank database is expected to be biased. These are the firms who voluntarily applied for rating in order to receive refinancing from commercial banks and are mostly self-selected solvent companies.

Despite many similarities, the dependence of PD from the individual financial ratios can display certain differences between German and Asian companies which can be attributed to a more homogeneous sample for Germany, disparity in company registration forms and sizes and the self-selected nature of the German sample. These differences are mostly proclaimed if the dependence for the companies from one of the regions has a U shape.

## 5 Variable Selection 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 the zero predictive power and (ii) between the cumulative default curves for the ideal model and the model with the zero predictive power (figure 16). AR is used

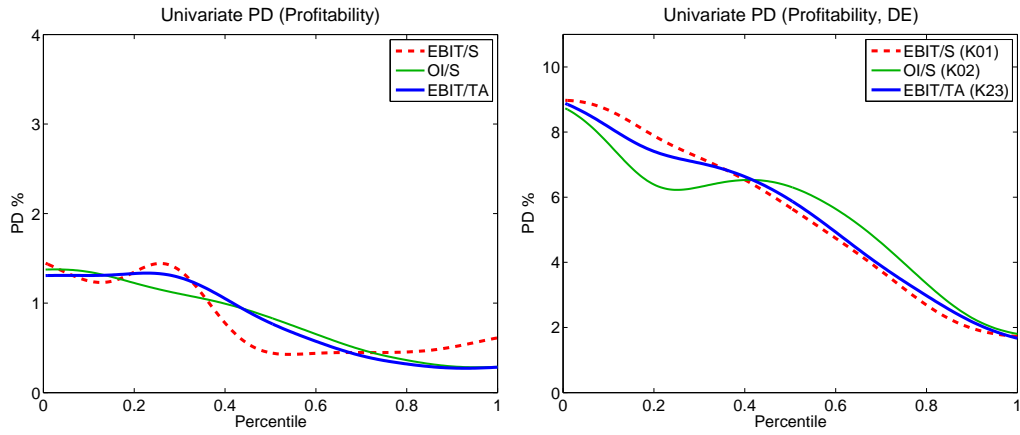


Figure 9: Univariate probabilities of default for *Profitability Ratios* pooled over countries and years for Asia (horizon: *Design 1*, left panel) and Germany (right panel).

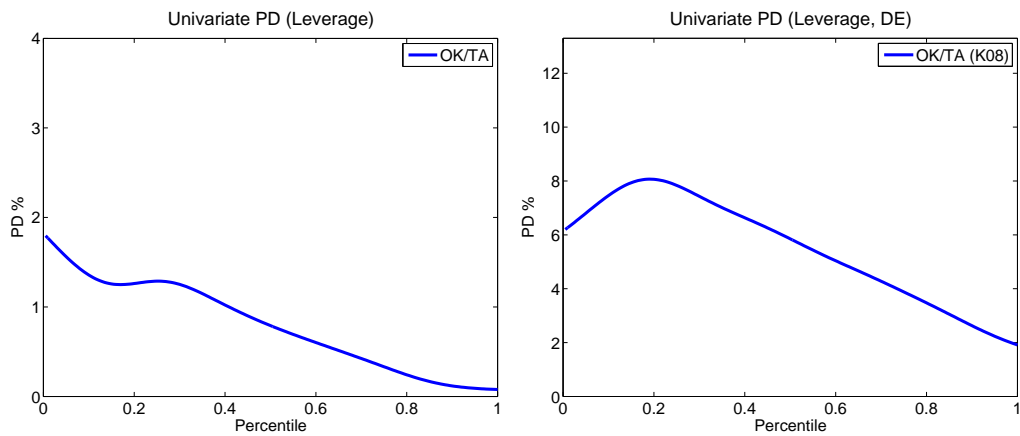


Figure 10: Univariate probabilities of default for *Leverage Ratios* pooled over countries and years for Asia (horizon: *Design 1*, left panel) and Germany (right panel).

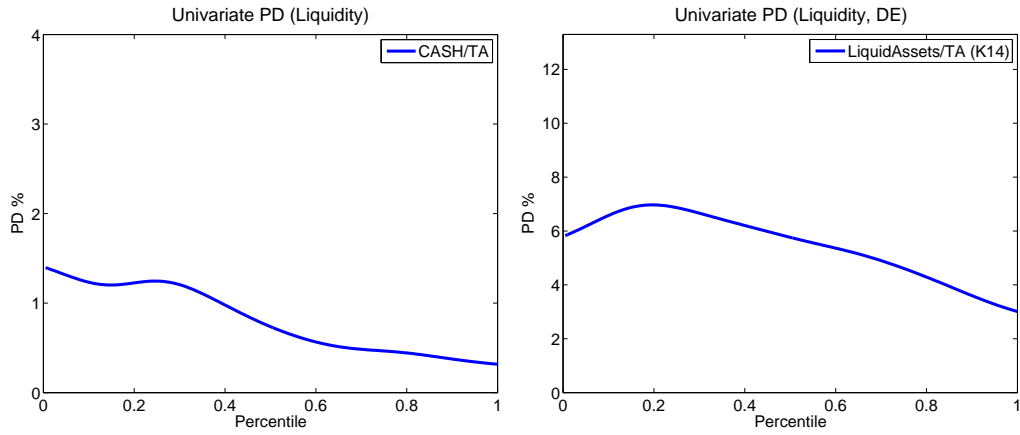


Figure 11: Univariate probabilities of default for *Liquidity Ratios* pooled over countries and years for Asia (horizon: *Design 1*, left panel) and Germany (right panel).

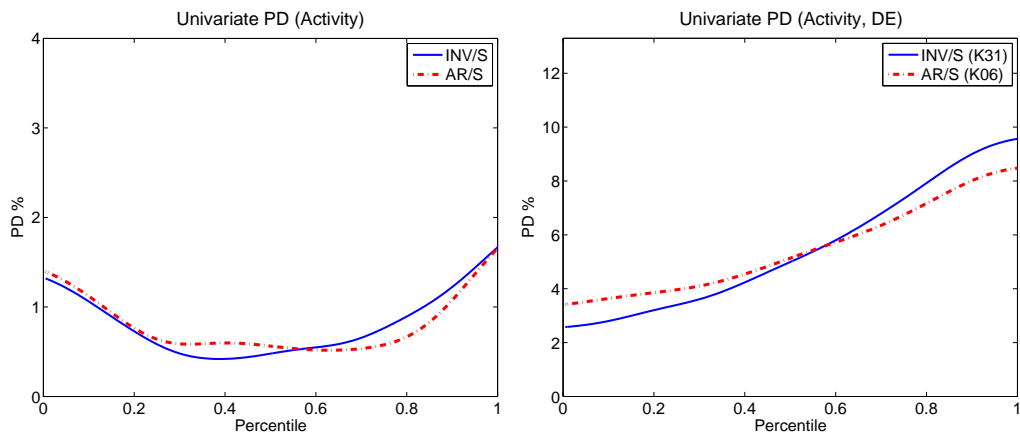


Figure 12: Univariate probabilities of default for *Activity Ratios* pooled over countries and years for Asia (horizon: *Design 1*, left panel) and Germany (right panel).

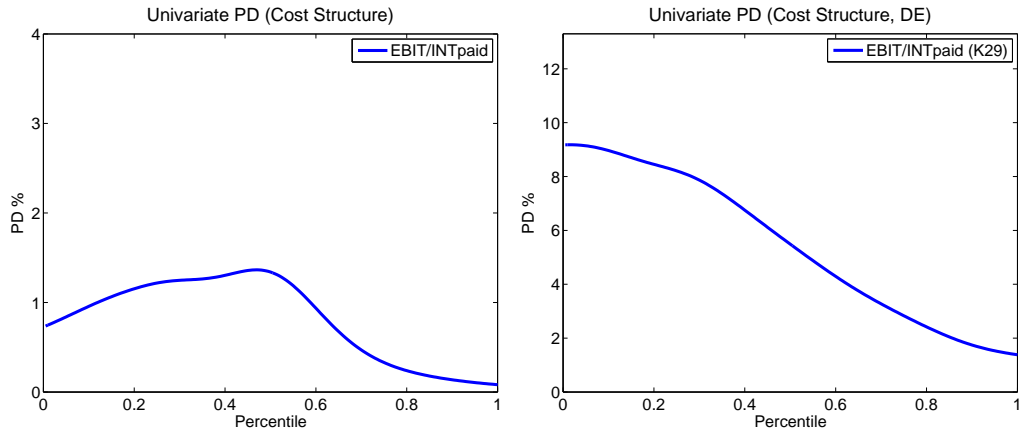


Figure 13: Univariate probabilities of default for *Cost Structure Ratios* pooled over countries and years for Asia (horizon: *Design 1*, left panel) and Germany (right panel).

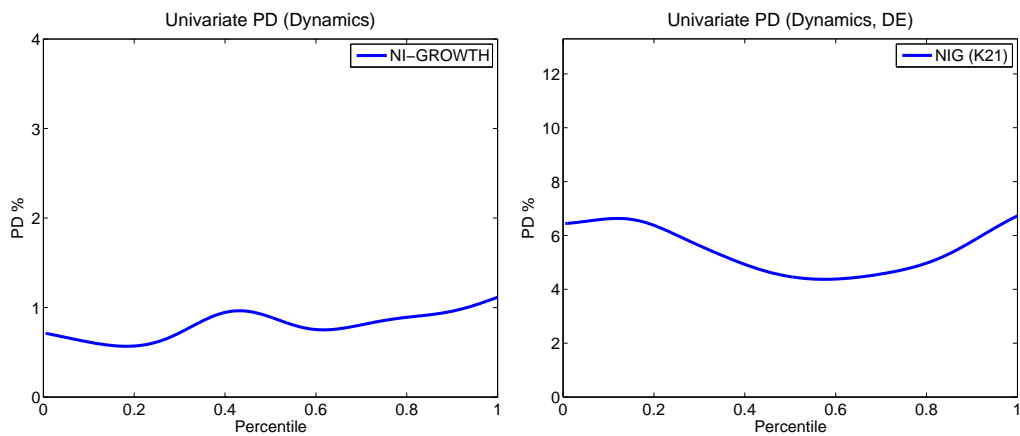


Figure 14: Univariate probabilities of default for *Dynamic Ratios* pooled over countries and years in Asia (horizon: *Design 1*, left panel) and Germany (right panel).

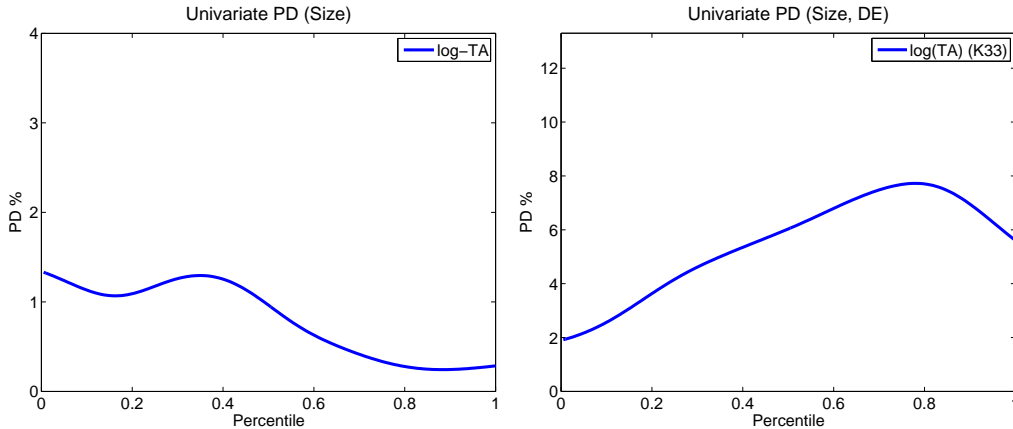


Figure 15: Univariate probabilities of default for *Company Size* pooled over countries and years in Asia (horizon: *Design 1*, left panel) and Germany (right panel).

since it is not sensitive to a monotonic transformation of a score in contrast to other accuracy measures such as hit rate or  $\alpha$  and  $\beta$  errors.

The bootstrap procedure (Efron & Tibshirani, 1993) for model comparison starts with the selection of two non-overlapping random subsamples of 1000 observations (500 non-defaulting and 500 defaulting firms) from the original data set. One of those subsamples is used as a training set and the other one as a testing set. A classification model (SVM or Logit) 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  $q_{sup}(x) = Median(x) + 1.5IQR(x)$ . Secondly, all data were standardised as  $x_{new} = (x - median(x))/\sigma(x)$ . This was done to avoid an excessive influence of the variables 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 estimated by bootstrapping. At step two, in addition to this variable, the second variable from the remaining is chosen which has the highest median AR among all alternatives. At step three a trivariate model is selected, etc. The variables selected by Logit and SVM for pooled data are presented in table 5. After a certain step four the accuracy of both the Logit and SVM

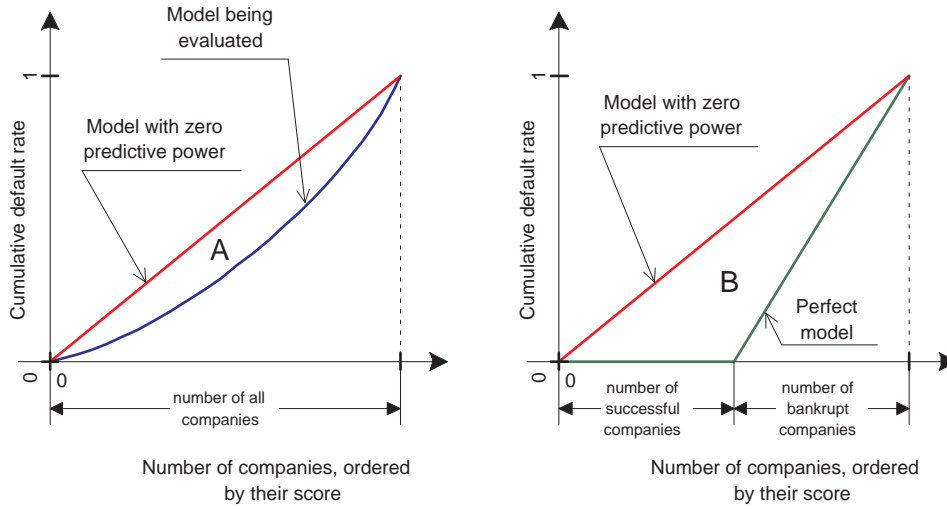


Figure 16: Accuracy Ratio (AR) is the ratio of two areas A and B.

models does not experience any significant improvements, what is evident from very high  $p$ -values.

The SVM was always applied with  $R = r\sqrt{d/2}$  and  $C = (c/n)(2/d)$ , where  $r$  and  $c$  were chosen based on the values reported as performing well for company rating (Lacerda & Moro (2008), Härdle, Moro & Schäfer (2010)). These two parameters of the SVM used in our study were  $r = 2.5$  and  $c = 1$  for a low complexity SVM with high generalisation ability, which is expected to perform well on a broad range of data sets. The performance of the SVM can be potentially further increased by optimising  $r$  and  $c$  for the studied data. The transformations for computing  $R$  and  $C$  figuring in the SVM formulation (see Appendix A) were applied to keep the SVM invariant of the data dimension  $d$  and the number of observations in the training set  $n$ .

As the table 5 indicates both models considered – Logit and SVM – have selected the first three variables identically: TD/TA,  $\log(S)$ , CL/TA. The fourth variable selected by the SVM is TA/S, while  $\log(TA)$  was selected by Logit. These variables form the basis for our model comparison.

## 6 Forecasting Accuracy

Figure 17 represents the time series of Accuracy Ratios (AR) for pooled data and a separate country with the highest number of distressed observations, China. The training set data are collected from the year indicated in the plots along the horizontal axis. The testing set data are collected for the year  $T + 2$ , where  $T$  is the training set year. The

Step	Logit				SVM ( $R = 2.5, C = 1$ )			
	Variable	Med. AR	$p_{max}$	$p$	Variable	Med. AR	$p_{max}$	$p$
1	TD/TA	57.5	0	–	TD/TA	57.5	0	–
2	log(S)	69.0	0	0	log(S)	69.7	0	0
3	CL/TA	71.1	0	0	CL/TA	71.7	3	5
4	log(TA)	<b>73.2</b>	–	<b>0</b>	TA/S	<b>73.5</b>	–	<b>3</b>
5	WC/TA	73.3	–	19	RV	73.4	–	75

Table 5: Variables selected at each step by the forward selection procedure for Logit and SVM for the pooled data. For computing the median AR for each combination of variables and the distributions of AR required for the tests, 100 bootstrapped subsamples were used. The confidence level  $p_{max}$  is reported for the test with  $H_0$ : the model is not significantly different from the four-variable model which was selected;  $p$  corresponds to the  $H_0$ : a model is not significantly different compared to a previous reduced model which has one variable less. Median AR,  $p_{max}$  and  $p$  are reported in percentage points.

used default horizon specification is *Design 2*. This arrangement guarantees that there are no overlapping observations in the data sets and forecasting is made out-of-sample. The parameters of the SVM are  $r = 2.5$  and  $c = 1$ . The variables are the same ones selected by variable selection procedure. For SVM these variables are: TD/TA, log(S), CL/TA and TA/S and for Logit: TD/TA, log(S), CL/TA and log(TA).

As it is evident from figure 17, SVM usually outperforms Logit in forecasting corporate distress. The difference in AR can be as high as 7.5%, as it is the case for China in 2005. On the other hand there are much fewer years when the SVM underperformed compared to the Logit. The maximum difference in this case is only 2.4%. For the pooled data in seven years out of eight the SVM has a higher performance than the Logit, although the differences in this case are more moderate than for China.

The similar conclusion about a higher predictive power of the SVM can be reached from analysing figure 18. It reports the distribution of the differences in AR between the SVM and Logit estimated on 100 bootstrapped subsamples of the data pooled across countries and years. Although the average improvement is moderate, around 0.5%, the SVM can achieve a much higher relative improvement for extreme scenarios. This is evident from a longer right tail of the probability density function. In other words, the SVM has a lower model misspecification risk compared with Logit, both on average and in the extreme cases.

To illustrate the performance of the SVM and Logit we will consider a two dimensional case (figures 19 and 20). The lines correspond to the isoquants with the PD equal to the average PD for the data. However small the differences may be, they are sufficient to

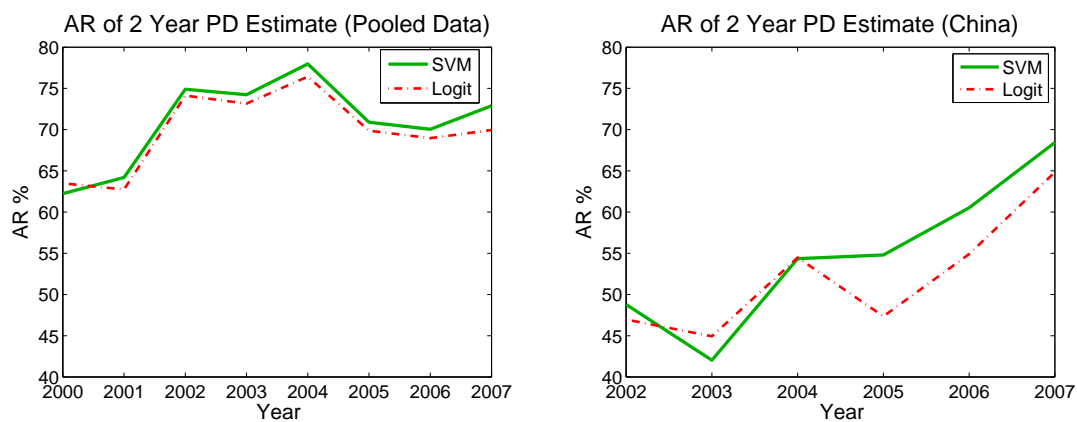


Figure 17: AR of 2 year probabilities of default estimated with SVM and Logit for pooled data (left panel) and China (right panel). Horizon: *Design 2*.

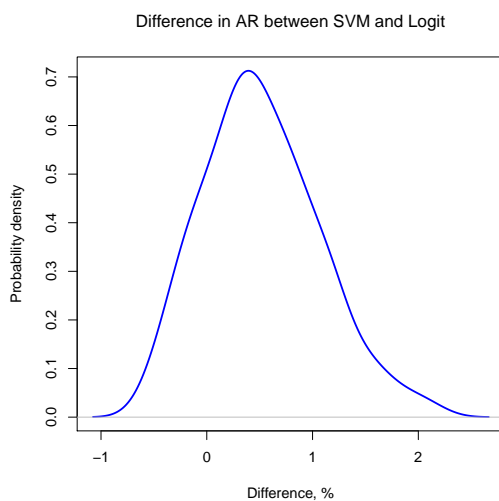


Figure 18: The distribution of differences in AR for the SVM vs. Logit estimated on 100 bootstrapped subsamples for a four-variable model on pooled data. Horizon: *Design 2*. A Gaussian kernel estimator was used with the bandwidth 0.191.

explain the higher AR achieved by the SVM. The background PDs coded with colours ranging from green to red are estimated for the SVM.

The PD range for the depicted area is from 0.11% to 8.87% for the pooled data and from 0.06% to 16.7% for China allowing for more than a 100 time difference between the highest and the lowest PD.

## 7 Conclusion

The focus of our study is the analysis of the ability of two models, SVM and Logit, to predict distress in the Asian and Pacific region in various settings.

Both models selected only four financial ratios as predictors of default, whereas three financial ratios are the same: TD/TA,  $\log(S)$  and CL/TA. They are leverage ratios and a company size. Surprisingly, no profitability ratios were selected.

A strong U-shaped dependence of PD from the leverage and activity ratios implies the existence of the optimal capital structure (TD/TA=15%, the figure being in accordance with the existing literature) and inventory stock (Inv/S=38%).

A comparison with the German data leads to a mixed conclusion. On one hand the dependence of PD from the profitability, leverage and liquidity ratios has a very similar shape. In all three cases for both countries it is falling. On the other hand, the dependencies for activity, cost structure and dynamic ratios and company size are dissimilar.

Comparison of forecasting accuracy reveals that the SVM has a lower model risk than the Logit. Firstly, on average SVM is more accurate than Logit. Secondly, in the extreme cases when discrepancies between the two models are the largest, the predictive power of the Logit can fall significantly below the SVM, while the probability that the SVM will significantly underperform relative to the Logit is much smaller.

Overall, an SVM with a high generalisation ability appears to be a promising method for distress forecasting in the Asian and Pacific region providing a reduction of model risk and a more robust performance compared to the Logit.

## Appendix A: Support Vector Machine (SVM)

The Support Vector Machine applied in this paper is a statistical method for binary classification that is a practical implementation of the Tikhonov regularisation principle (Tikhonov (1963), Tikhonov & Arsenin (1977)). It is based on linear classifiers that

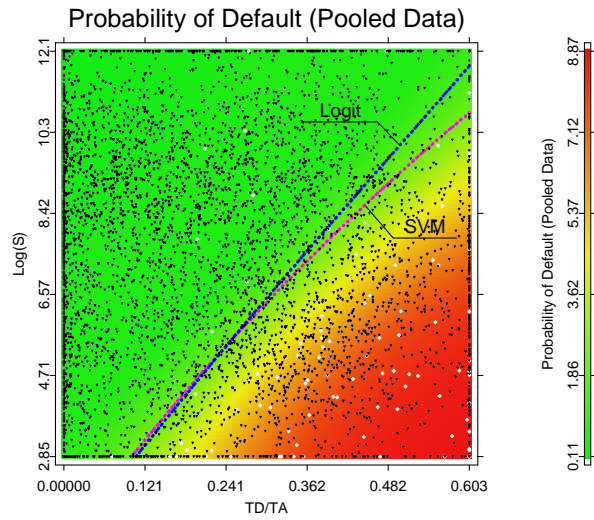


Figure 19: PD for  $\log(S)$  vs.  $TD/TA$  for pooled data are represented with a background colour.  $1/50^{th}$  of all data is represented ( $1/50^{th}$  of 228504 solvent and 4678 insolvent observations). Black dots denote solvent and white dots distressed companies. The separating curves are computed and plotted for the SVM and Logit for the average  $PD=2.01\%$ .

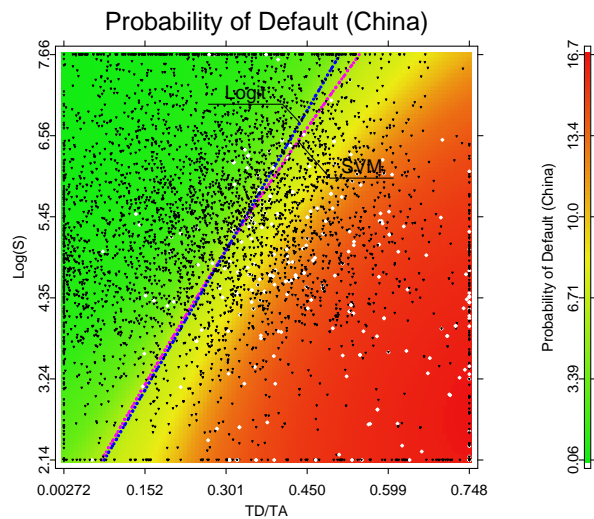


Figure 20: PD for  $\log(S)$  vs.  $TD/TA$  for pooled data are represented with a background colour.  $1/10^{th}$  of all data is represented ( $1/10^{th}$  of 40,627 solvent and 2,639 insolvent observations). Black dots denote solvent and white dots distressed companies. The separating curves are computed and plotted for the SVM and Logit for the average  $PD=6.10\%$ .

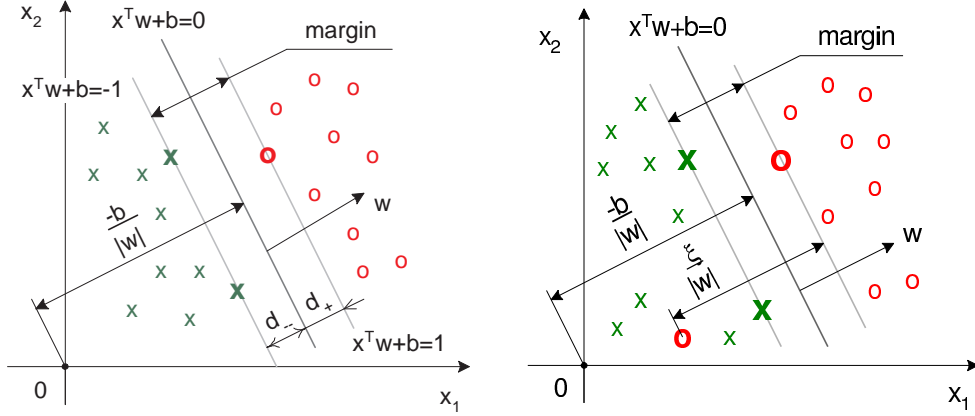


Figure 21: The separating hyperplane  $x^\top w + b = 0$  and the margin in a linearly separable (left) and non-separable (right) case. Crosses denote solvent companies, zeros are the insolvent ones. The hyperplanes bounding the margin zone equidistant from the separating hyperplane are represented as  $x^\top w + b = 1$  and  $x^\top w + b = -1$ . The misclassification penalty in the non-separable case is proportional to the distance  $\xi / \|w\|$ .

simultaneously maximise the margin or the distance between the classes and minimise empirical risk related to misclassifications on a given data set (Vapnik (1995)).

Figure 21 illustrates the maximum margin classification for linearly separable and non-separable data in a two-dimensional case. The separating function generated by a linear SVM is

$$x^\top w + b = 0. \quad (7.1)$$

Such a classification rule makes an SVM similar to Logit.  $x_i$  is a  $d \times 1$  vector of the characteristics of firm  $i$ , e.g. financial ratios described in Appendix B, whereas  $d$  is the number of characteristics or variables used.  $w$  is a  $d \times 1$  vector of weights which determine the slope of the separating function. The scalar  $b$  is a location parameter or a threshold.

The margin is the empirically estimated distance between the opposite classes of observations. In Figure 21 it is shown as the distance between the margin boundaries – the parallel lines located symmetrically on both sides of the separating function. In a perfectly separable case such as in Figure 21, left panel, no observations may lie in the

margin zone and all observations must satisfy the constraints:

$$x_i^\top w + b \geq 1 \quad \text{for } y_i = 1, \quad (7.2)$$

$$x_i^\top w + b \leq -1 \quad \text{for } y_i = -1. \quad (7.3)$$

The constraints ensure that the observations of the opposite classes lie on the opposite sides from the margin gap.

Misclassifications may occur if data are linearly non-separable as in Figure 21, right panel. Here the bold zero on the left-hand side of the separating line shows a solvent company that is classified as insolvent. SVM adjusts the weights  $w$  and the location parameter  $b$  in such a way that the margin is maximised and the sum of misclassification errors  $\xi_i$  is minimised.  $\xi_i \geq 0$  is also called a slack variable and is introduced to (7.2) and (7.3) to ensure that these constraints are satisfied. For any observation  $x_i$  the modified constraints must hold:

$$x_i^\top w + b \geq 1 - \xi_i \quad \text{for } y_i = 1, \quad (7.4)$$

$$x_i^\top w + b \leq -1 + \xi_i \quad \text{for } y_i = -1. \quad (7.5)$$

For the representation (7.2) – (7.5) when 1 appears on the right hand side the margin equals  $2/\|w\|$ . Here  $\|w\|$  is the Euclidean norm or the length of vector  $w$ .

Only the observations lying on the margin boundaries or on the wrong side of the margin determine the SVM solution. These observations are marked with bold crosses and zeros. They are called support vectors, hence the name of the method. This contrasts to Logit where all observations are used to derive the solution.

The primal minimisation problem to be solved is convex and has a unique solution:

$$\min_w \frac{1}{2} \|w\| + \sum_{i=1}^n C_i \frac{\xi_i}{\|w\|} \quad (7.6)$$

$$\text{s.t. } y_i(x_i^\top w + b) \geq 1 - \xi_i, \quad (7.7)$$

$$\xi_i \geq 0. \quad (7.8)$$

Here (7.4) and (7.5) are rewritten as one constraint. It is easier to see that the problem is convex if we rewrite the optimised functional in (7.6) as  $\frac{1}{2} \|w\|^2 + \sum_{i=1}^n C_i \xi_i$ . The first term is the inverse margin, which equals  $2/\|w\|$ . By minimising this term we maximise the margin. The second term is a sum of weighted errors that are measured as a distance to a misclassified observation  $i$  from the boundary of its class  $\xi_i/\|w\|$ . The parameters  $C_i$ 's which are called capacity represent the penalty weights of in-sample

false classifications for each company observation  $i$ . The SVM will give priority to the correct classification of the companies with higher  $C_i$ 's. Capacity is related to the width of the margin zone. Smaller  $C_i$ 's are associated with bigger margins. In our case  $C_i$  are set equal for the same class. In order to make SVMs comparable for a different number of observations and various ratios between solvent and insolvent companies we compute  $C_i$ 's as  $c/2n_+$  for insolvent and  $c/2n_-$  for solvent companies. Here  $n_+$ , and  $n_-$  are the numbers of insolvent and solvent companies in the training set,  $c$  is the coefficient that is used to control the capacity of the SVM. In contrast to  $C_i$  it is invariant of the number of observations in the training data set and provides a convenient basis for comparing SVMs. This formulation implies that in a sample with mostly solvent companies, misclassifications of insolvent companies are given a higher weight. If the number of solvent and insolvent companies is the same, then  $C_i = c/n$ .

The primal problem (7.6) – (7.8) rewritten in a Lagrangian formulation is

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

The Karush-Kuhn-Tucker Conditions or first order optimality conditions are:

$$\frac{\partial L_P}{\partial w_k} = 0 \Leftrightarrow w_k = \sum_{i=1}^n \alpha_i y_i x_{ik}, \quad k = 1, \dots, d, \quad (7.9)$$

$$\frac{\partial L_P}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^n \alpha_i y_i = 0, \quad (7.10)$$

$$\frac{\partial L_P}{\partial \xi_i} = 0 \Leftrightarrow C_i - \alpha_i - \mu_i = 0, \quad (7.11)$$

$$\alpha_i \{y_i(x_i^\top w + b) - 1 + \xi_i\} = 0,$$

$$\mu_i \xi_i = 0,$$

$$y_i(x_i^\top w + b) - 1 + \xi_i \geq 0,$$

$$\alpha_i \geq 0,$$

$$\mu_i \geq 0,$$

$$\xi_i \geq 0,$$

where  $x_{ik}$  is the  $k$ -th characteristic of company  $i$ . The dual problem is equivalent to the primal one since the minimised function is convex (Gale, Kuhn & Tucker (1951)). By substituting equations (7.9) – (7.11) into the primal Lagrangian we derive the dual

problem:

$$\begin{aligned} \max_{\alpha_i} \quad & \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^\top x_j, \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C_i, \\ & \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \tag{7.12}$$

The  $n$  Lagrange multipliers  $\alpha_i$  are the free parameters to be estimated. They represent the weights with which each observation influences the solution (see (7.14) and (7.18)). Those observations have higher weights which are harder to classify, i.e. which lie closer to the margin zone. On the contrary, the coefficients in the logistic regression are the weights assigned to each variable and can not be directly compared to Lagrange multipliers. Problem (7.12) can be equivalently expressed in a matrix notation:

$$\begin{aligned} \max_{\alpha} \quad & \iota^\top \alpha - \alpha^\top H \alpha, \\ \text{s.t.} \quad & 0 \leq \alpha \leq C, \\ & y^\top \alpha = 0. \end{aligned} \tag{7.13}$$

Here  $\alpha$  is a vector of Lagrange multipliers,  $\iota$  is a vector of 1's,  $y$  is a vector of company classes, +1 for solvent or -1 for insolvent ones and  $C$  here is a vector of the coefficients  $C_i$ ; all vectors are of the size  $n \times 1$ . The  $n$  components of the vector  $\alpha$  are obtained as the solution of the constrained maximisation problem (7.13). The  $i, j$ 'th element of the matrix  $H$  is

$$h_{ij} = y_i y_j x_i^\top x_j = y_i y_j \sum_{k=1}^d x_{ik} x_{jk}.$$

The reader who desires to construct an SVM independently may find the problem formulation in the matrix notation (7.13) more convenient. The SVM problem is a classical quadratic optimisation problem (Fletcher (1987)) that can be solved with numerous software packages such as Matlab (routines *minq* or *minqdef*) or using algorithms specifically developed for the SVM such as the Sequential Minimal Optimisation (SMO) (Platt (1998)).

Equation (7.9) of the KKT optimality conditions determines the weights  $w_k$ ,  $k = 1, \dots, d$  for the  $k$ -th characteristic of a company. By substituting (7.9) into (7.1) we

derive the classification rule:

$$f(x) = x^\top w + b = x^\top \sum_{i=1}^n \alpha_i y_i x_i + b = \sum_{i=1}^n \alpha_i y_i x_i^\top x + b \quad (7.14)$$

$$\begin{cases} f(x) < 0 & \Rightarrow x \text{ is solvent,} \\ f(x) \geq 0 & \Rightarrow x \text{ is insolvent.} \end{cases}$$

To derive the coefficient  $b$  we will use the fact that the separating hyperplane  $f(x) = 0$  (see Figure 21) lies equidistant from the hyperplanes bounding the classes:

$$x_+^\top w + b = 1 \quad \text{for } y_+ = 1, \quad (7.15)$$

$$x_-^\top w + b = -1 \quad \text{for } y_- = -1, \quad (7.16)$$

where  $x_+$  is any support vector that lies on or ‘supports’ the hyperplane for  $y = 1$  and  $x_-$  is any support vector that lies on the hyperplane for  $y = -1$ . Both  $x_+$  and  $x_-$  have dimensions  $d \times 1$ . By summing (7.15) and (7.16) we derive:

$$b = -\frac{1}{2} (x_+^\top + x_-^\top) w = -\frac{1}{2} \sum_{i=1}^n \alpha_i y_i (x_+^\top + x_-^\top) x_i. \quad (7.17)$$

To reduce numerical errors when training the SVM it is desirable to use averages over all  $x_+$  and  $x_-$  instead of two arbitrary chosen support vectors.

Note that the classification rule (7.14) depends only on the scalar product  $x^\top x_i$ , not on the original  $x$  and  $x_i$ . This makes possible a ‘kernel trick’, i.e. an implicit mapping of low dimensional data into a highly dimensional Hilbert feature space and performing a linear classification there, e.g. with an SVM. A kernel transformation corresponds to (i) performing a variable transformation and (ii) taking a scalar product of transformed variables.

In practice  $x^\top x_i$  in the SVM formulation (7.12) is replaced with a kernel function  $K(x, x_i)$  which represents a scalar product in a feature space (Weyl (1928)). Then the elements of the matrix  $H$  in (7.13) are  $h_{ij} = y_i y_j K(x_i, x_j)$ . A kernel function must satisfy the Mercer conditions (Mercer (1909)), i.e. be symmetric and semipositive definite as a scalar product. It can map data into infinitely dimensional spaces as in the case with Gaussian kernels. The number of Lagrange multipliers  $\alpha_i$  – parameters to be estimated – is  $n$  and can be large for large data sets. However, by selecting a small  $C_i$ ’s and, hence, a narrow interval  $[0, C_i]$  in which  $\alpha$  may vary we can avoid overfitting and extremely high complexities of the SVM classifier.

Figure 22 shows a simple mapping example. The quadratic kernel function

$$K(x, x_i) = (x^\top x_i)^2$$

maps two dimensional data into a three-dimensional space of features. The three features correspond to the three components of a quadratic form in two dimensions:  $\tilde{x}_1 = x_1^2$ ,  $\tilde{x}_2 = \sqrt{2}x_1x_2$  and  $\tilde{x}_3 = x_2^2$ . The transformation from a two dimensional data space into a three dimensional feature space is  $\Psi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^\top$ . However, we do not need to know the transformation  $\Psi$  explicitly and can equivalently apply the kernel  $K(x_1, x_2) = \Psi(x_1, x_2)^\top \Psi(x_1, x_2)$  to represent quadratic dependencies between input variables. The data separable in the data space of  $x_1$  and  $x_2$  only with a quadratic function will be separable in the feature space of  $\tilde{x}_1$ ,  $\tilde{x}_2$  and  $\tilde{x}_3$  with a linear function. Thus, a non-linear SVM in the data space is equivalent to a linear SVM in the feature space. The number of features is growing fast with the dimension of the data  $d$  and the degree of the polynomial kernel making a direct data transformation not feasible and the advantages of the data transformation via a kernel obvious.

By substituting the scalar product in (7.14) with a kernel function a non-linear score function  $f$  is derived:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b, \quad (7.18)$$

where, by analogy with (7.17):

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

The non-parametric score function (7.18) does not have a compact closed form representation.

In our study we applied an SVM with an anisotropic Gaussian or radial basis kernel

$$K(x, x_i) = \exp \left\{ -(x - x_i)^\top r^{-2} \Sigma^{-1} (x - x_i) / 2 \right\}, \quad (7.19)$$

where  $r$  is a coefficient and  $\Sigma$  is a scaling matrix, which in our case is a variance-covariance matrix of the training characteristics  $x$ . The  $k_1, k_2$ -th element of the matrix

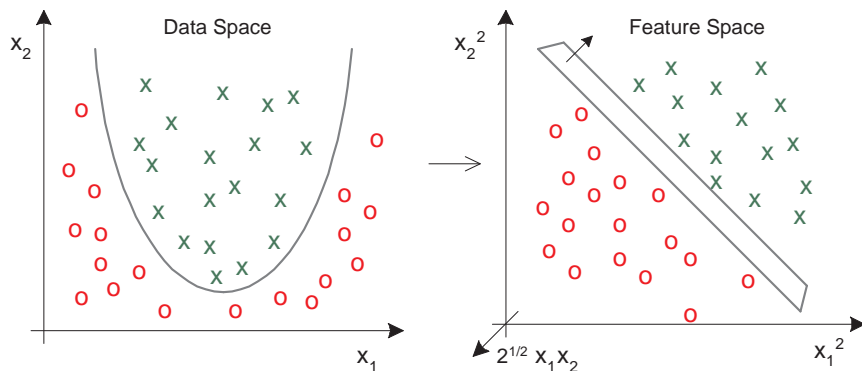


Figure 22: Mapping from a two-dimensional data space into a three-dimensional space of features  $\mathbb{R}^2 \mapsto \mathbb{R}^3$ .

is:

$$\sigma_{k_1, k_2} = \frac{1}{n} \sum_{i=1}^n \left( x_{i, k_1} - \frac{1}{n} \sum_{j=1}^n x_{j, k_1} \right) \left( x_{i, k_2} - \frac{1}{n} \sum_{j=1}^n x_{j, k_2} \right).$$

Here  $\sigma_{k_1, k_2}$  is the covariance between two financial ratios  $x_{k_1}$  and  $x_{k_2}$ , e.g. a profitability and a leverage financial ratio.  $\Sigma$  is used to bring all variables to the same scale and exclude the excessive influence of the variables with high variance. The ability to use differently scaled data explains the term ‘anisotropic’ in the kernel name. Before computing  $\Sigma$  and training an SVM the outliers should be processed, e.g. capped. The coefficient  $r$  is related to the complexity of classifying functions: the higher the  $r$  is, the lower is the complexity. If kernel functions allow for sufficiently rich feature spaces, the performance of SVMs with different kernels is comparable in terms of out-of-sample forecasting accuracy (Vapnik (1995)). Note that only the capacity  $C_i$  and the complexity coefficient  $r$  are to be set a priori. The Lagrange multipliers are the free parameters that are computed when training the SVM.

The SVM has a substantial advantage in comparison to the logistic regression with transformed variables, namely, it does not require to specify the transformation but estimates it from a broad range of possible ones defined implicitly by the kernel function type and the SVM capacity coefficient. This advantage of the SVM is fully revealed when data are new or the relevance of well known transformations must be tested.

## Appendix B: Conversion of Scores into PDs

The conversion of rating scores into PDs provides us with a link to the existing rating classes reported by rating agencies such as Moody's and S&P. In the Logistic model a sigmoid function is applied to estimate PD assuming the logistic distribution of the latent variable. However, such an assumption is often not compatible with reality. In general the company score, as it is computed by the SVM or Logit, defines the distance between companies in terms of PD: the lower the difference in scores, the closer are companies. If a company has a higher score, it lies farther from successful companies and, therefore, its PD should be higher. This means that the dependence between scores and PDs is assumed to be monotonic. No further assumptions about the form of this dependence will be made in contrast to the already mentioned Logit model with a prespecified functional transformation from the score to PD.

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

Step one is the estimation of PDs for the companies of the training set. This is done using standard smoothing techniques in order to preliminary evaluate PDs for all  $i = 1, 2, \dots, n$  observations of the training set:

$$\widetilde{PD}(z) = \frac{\sum_{i=1}^n w(z - z_i) I(y_i = 1)}{\sum_{i=1}^n w(z - z_i)}, \quad (7.20)$$

where  $w(z - z_i) = \exp\{(z - z_i)^2/2h^2\}$ . The rank of the  $i$ -th company  $z_i = Rank\{f(x_i)\}$  can be 1, 2, 3, ... up to  $n$  depending on its score  $f(x_i)$ ; the higher the score is, the higher is the rank.  $h$  is a bandwidth, in our case  $h = 0.09n$ . The smaller is the bandwidth, the smoother is  $\widetilde{PD}(z)$ . When  $h \rightarrow 0$  no smoothing is performed and all  $\widetilde{PD}(z_i)$ ,  $i = 1, 2, \dots, n$ , will be either 1 or 0; when  $h \rightarrow \infty$ , all  $\widetilde{PD}(z_i)$  will have the same value equal to the average probability of default for the training set.

Using the company rank  $z$  instead of the score  $f(x)$  we obtain a  $k$ -NN smoother with Gaussian weights  $\frac{w(z - z_i)}{\sum_{j=1}^n w(z - z_j)}$  which decay gradually as  $|z - z_i|$  grows. This differs from the most commonly used  $k$ -NN smoother that relies on the uniform weights  $\frac{1}{k} I(|z - z_i| < \frac{k}{2} + 1)$ .

The preliminary PDs evaluated at step one are not necessarily a monotonic function of the score. This is due to the fact that companies with close scores may have for different reasons a non-concordant binary survival indicator  $y$ . The monotonisation of  $\widetilde{PD}(z_i)$ ,  $i = 1, 2, \dots, n$  is achieved at step two using the Pool Adjacent Violator (PAV)

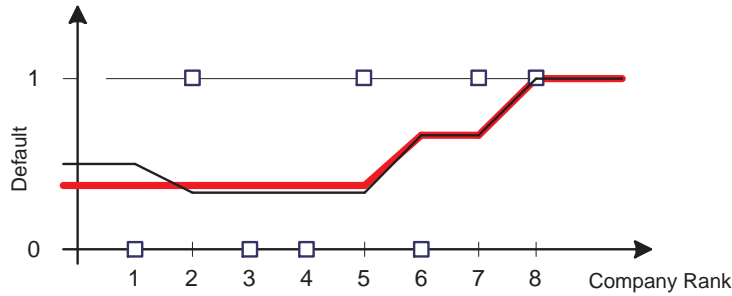


Figure 23: Monotonisation of PDs with the pool adjacent violator algorithm. The thin line denotes PDs estimated with the  $k$ -NN method with uniform weights and  $k = 3$  before monotonisation and the bold line after monotonisation. Here  $y = 1$  for insolvencies,  $y = 0$  for solvent companies.

algorithm (Barlow, Bartholomew, Bremner & Brunk, 1972). Figure 23 illustrates the workings of the algorithm. The companies are ordered according to their rank and have here the indicator  $y = 1$  for insolvent and  $y = 0$  for solvent companies. The thin line denotes the PDs estimated using the  $k$ -NN method with uniform weights and  $k = 3$ . At the interval between the observations with rank 1 and 2 monotonicity is violated and is corrected with the PAV algorithm. The bold line shows PDs after monotonisation.

The PAV algorithm solves the following optimisation problem: given data  $\{z_i, y_i\}_{i=1}^n$  with  $z_1 \leq z_2 \leq \dots \leq z_n$  find the monotonic increasing function  $m(z_i)$ , i.e.  $m(z_1) \leq m(z_2) \leq \dots \leq m(z_n)$  that minimises  $\sum_{i=1}^n \{y_i - m(z_i)\}^2$ . The solution to this problem is pooling (averaging) the adjacent observations that are violating monotonicity. The PAV acronym comes from this property. Mammen (1991) has shown that one can equivalently start with the PAV step and then smooth with a Nadaraya-Watson kernel estimator (Nadaraya (1964)).

As a result we obtain monotonised probabilities of default  $PD(x_i)$  for the observations of the training set. A PD for any observation  $x$  of the testing set is computed by interpolating PDs for two adjacent, in terms of the score, observations from the training set. If the score for  $x$  lies beyond the range of the scores of the training set, then  $PD(x)$  is set equal to the score of the first neighbouring observation of the training set.

## 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.
- Barlow, R. E., J. M. Bartholomew, J. M. Bremner, 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.
- Dimitras, A., R. Slowinski, R. Susmaga, and C. Zopounidis, 1999: Business failure prediction using rough sets. *European Journal of Operational Research*, **114**, 263–280.
- 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](http://pwp.netcabo.pt/jed_fernandes/JEF_CorporateCreditRisk.pdf).
- Fitzpatrick, P., 1932: A comparison of the ratios of successful industrial enterprises with those of failed companies.
- 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.

- Glennon, D. and P. Nigro, 2005: Measuring the default risk of small business loans: A survival analysis approach. *Journal of Money, Credit and Banking*, **37**, 923–947.
- Härdle, W. K., R. A. Moro, and D. Schäfer, 2010: Estimating probabilities of default with support vector machines. *Preprint Submitted to Elsevier*.
- J. C. Duan, J. S. and T. Wang, 2010: Multiperiod corporate default prediction – a forward intensity approach. working paper, Risk Management Institute, National University of Singapore.
- Lacerda, A. I. and R. A. Moro, 2008: Analysis of the predictors of default for portuguese firms. Working paper 22, Banco de Portugal.
- Mammen, E., 1991: Estimating a smooth monotone regression function. *Annals of Statistics*, **19**, 724–740.
- 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.
- Nadaraya, E. A., 1964: On estimating regression. *Theory of Probability and its Applications*, **10**, 186–190.
- 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.

- Ramser, J. and L. Foster, 1931: A demonstration of ratio analysis. bulletin no. 40. Bureau of Business Research, University of Illinois, Urbana, Ill.
- 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.
- Weyl, H., 1928: *Gruppentheorie und Quantenmechanik*. Hirzel, Leipzig.
- Wiginton, J., 1980: A note on the comparison of logit and discriminant models of consumer credit behaviour. *Journal of Financial and Quantitative Analysis*, **15**(3), 757–770.
- 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.
- Zavgren, C., 1983: The prediction of corporate failure: The state of the art. *Journal of Accounting Literature*, **2**, 1–38.
- Zmijewski, M., 1984: Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, **20**(0), 59–82.