CORPORATE CREDIT RISK MODELING: QUANTITATIVE RATING SYSTEM AND PROBABILITY OF DEFAULT ESTIMATION

João Eduardo Fernandes[1]

April 2005 (Revised October 2005)

ABSTRACT: Research on corporate credit risk modeling for privately-held firms is limited, although these firms represent a large fraction of the corporate sector worldwide. Research in this area has been limited because of the lack of public data. This study is an empirical application of credit scoring and rating techniques to a unique dataset on private firms bank loans of a Portuguese bank. Several alternative scoring methodologies are presented, validated and compared. Furthermore, two distinct strategies for grouping the individual scores into rating classes are developed. Finally, the regulatory capital requirements under the New Basel Capital Accord are calculated for a simulated portfolio, and compared to the capital requirements under the current regulation.

KEYWORDS: Credit Scoring, Credit Rating, Private Firms, Discriminatory Power, Basel Capital Accord, Capital Requirements

JEL CLASSIFICATION: C13, C14, G21, G28

[1] Correspondence Address: R. Prof. Francisco Gentil, E1 5E, 1600-625 Lisbon Portugal, email: joao.eduardo.fernandes@gmail.com

1 Introduction

The credit risk modeling literature has grown extensively since the seminal work by Altman (1968) and Merton (1974). Several factors contribute for an increased interest of market practitioners for a correct assessment of the credit risk of their portfolios: the European monetary union and the liberalization of the European capital markets combined with the adoption of a common currency, increased liquidity, and competition in the corporate bond market. Credit risk has thus become a key determinant of different prices in the European government bond markets. At a worldwide level, historically low nominal interest rates have made the investors seek the high yield bond market, forcing them to accept more credit risk. Furthermore, the announced revision of the Basel capital accord will set a new framework for banks to calculate regulatory capital[2]. As it is already the case for market risks, banks will be allowed to use internal credit risk models to determine their capital requirements. Finally, the surge in the credit derivatives market has also increased the demand for more sophisticated models.

There are three main approaches to credit risk modeling. For firms with traded equity and/or debt, Structural models or Reduced-Form models can be used. Structural Models are based on the work of Black and Scholes (1973) and Merton (1974). Under this approach, a credit facility is regarded as a contingent claim on the value of the firm's assets, and is valued according to option pricing theory. A diffusion process is assumed for the market value of the firm's assets and default is set to occur whenever the estimated value of the firm hits a pre-specified default barrier. Black and Cox (1976) and Longstaff and Schwartz (1993) have extended this framework relaxing assumptions on default barriers and interest rates.

For the second and more recent approach, the Reduced-Form or Intensity models, there is no attempt to model the market value of the firm. Time of default is modeled directly as the time of the first jump of a Poisson process with random intensity. These models were first developed by Jarrow and Turnbull (1995) and Duffie and Singleton (1997).

[2] For more information see Basel Committee on Banking Supervision (2003).

For privately held firms with no market data available, accounting-based credit scoring models are the most common approach. Since most of the credit portfolios of commercial banks consist of loans to borrowers that have no traded securities, these will be the type of models considered in this research[3]. Although credit scoring has well known disadvantages, it remains as the most effective and widely used methodology for the evaluation of privately-held firms' risk profiles[4].

The corporate credit scoring literature as grown extensively since Beaver (1966) and Altman (1968), who proposed the use of Linear Discriminant Analysis (LDA) to predict firm bankruptcy. On the last decades, discrete dependent variable econometric models, namely logit or probit models, have been the most popular tools for credit scoring. As Barniv and McDonald (1999) report, 178 articles in accounting and finance journals between 1989 and 1996 used the logit model. Ohlson (1980) and Platt and Platt (1990) present some early interesting studies using the logit model. More recently, Laitinen (1999) used automatic selection procedures to select the set of variables to be used in logistic and linear models which then are thoroughly tested out-of-sample.

The most popular commercial application using logistic approach for default estimation is the Moody's KMV RiskCalc Suite of models developed for several countries[5]. Murphy et al. (2002) presents the RiskCalc model for Portuguese private firms. In recent years, alternative approaches using non-parametric methods have been developed. These include classification trees, neural networks, fuzzy algorithms and k-nearest neighbor. Although some studies report better results for the non-parametric methods, such as in Galindo and Tamayo (2000) and Caiazza (2004), we will only consider logit/probit models since the estimated parameters are more intuitive, easily interpretable and the risk of over-fitting to the sample is lower. Altman, Marco and Varetto (1994) and Yang et al. (1999) present some evidence, using several types of neural network models, that these do not yield superior results than the classical models. Another potential relevant extension to traditional credit modeling is the inference on the often neglected rejected data. Boyes et al. (1989) and Jacobson and Roszbach (2003) have used bivariate probit models with sequential events to model a lender' decision problem. In the first equation, the decision to grant

---

[3] According to the Portuguese securities market commission (CMVM), at 31 December 2004 only 82 firms had listed equity or debt (CMVM 2005).
[4] See, for example, Allen (2002).
[5] See Dwyer et al. (2004).

the loan or not is modeled and, in the second equation, conditional on the loan having been provided, the borrowers' ability to pay it off or not. This is an attempt to overcome a potential bias that affects most credit scoring models: by considering only the behavior of accepted loans, and ignoring the rejected applications, a sample selection bias may occur. Kraft et al. (2004) derive lower and upper bounds for criteria used to evaluate rating systems assuming that the bank storages only data of the accepted credit applicants. Despite the findings in these studies, the empirical evidence on the potential benefits of considering rejected data is not clear, as shown in Crook and Banasik (2004).

The first main objective of this research is to develop an empirical application of credit risk modeling for privately-held corporate firms. This is achieved through a simple but powerful quantitative model built on real data randomly drawn from the database of one of the major Portuguese commercial banks. The output of this model will then be used to classify firms into rating classes, and to assign a probability of default for each one of these classes. Although a purely quantitative rating system is not fully compliant with the New Basel Capital Accord (NBCA), the methodology applied could be regarded as a building block for a fully compliant system[6].

The remainder of this study is structured as follows. Section 2 describes the data and explains how it is extracted from the bank's database. Section 3 presents the variables considered and their univariate relationship with the default event. These variables consist of financial ratios that measure Profitability, Liquidity, Leverage, Activity, Debt Coverage and Productivity of the firm. Factors that exhibit a weak or unintuitive relationship with the default frequency will be eliminated and factors with higher predictive power for the whole sample will be selected. Section 4 combines the most powerful factors selected on the previous stage in a multivariate model that provides a score for each firm. Two alternatives to a simple regression will be tested. First, a multiple equation model is presented that allows for alternative specifications across industries. Second, a weighted model is developed that balances the proportion of regular and default observations on the dataset, which could be helpful to improve

[6] For example, compliant rating systems must have two distinct dimensions, one that reflects the risk of borrower default and another reflecting the risk specific to each transaction (Basel Committee on Banking Supervision 2003, par. 358). The system developed in this study only addresses the first dimension. Another important drawback of the system presented is the absence of human judgment. Results from the credit scoring models should be complemented with human oversight in order to account for the array of relevant variables that are not quantifiable or not included in the model (Basel Committee on Banking Supervision 2003, par. 379).

the discriminatory power of the scoring model, and to better aggregate individual firms into rating classes. Results for both alternatives are compared and thoroughly validated. All considered models are screened for statistical significance, economic intuition, and efficiency (defined as a parsimonious specification with high discriminatory power). In Section 5 several applications of the scoring model are discussed. First, two alternative rating systems are developed, using the credit scores estimates from the previous section. One alternative will consist into grouping individual scores into clusters, and the other to indirectly derive rating classes through a mapping procedure between the resulting default frequencies and an external benchmark. Next, the capital requirements for an average portfolio under both the NBCA and the current capital accord are derived and compared. Section 6 concludes.

2 Data Considerations

A random sample of 11,000 annual, end-of-year corporate financial statements, is extracted from the financial institution's database. These yearly statements belong to 4,567 unique firms, from 1996 to 2000, of which 475 have had at least one defaulted loan in a given year. The default definition considered is compliant with the NBCA proposed definition, it classifies a loan as default if the client misses a principal and/or interest payment for more than 90 days.

Furthermore, a random sample of 301 observations for the year 2003 is extracted in order to perform out-of-sample testing. About half of the firms in this testing sample are included in the main sample, while the other half corresponds to new firms. In addition, the out-of-sample data contains 13 defaults, which results in a similar default ratio to that of the main sample (about 5%). Finally, the industry distribution is similar to the one in the main sample.

Firms belonging to the financial or real-estate industries are excluded, due to the specificity of their financial statements. Furthermore, firms owned by public institutions are also excluded, due to their non-profit nature.

The only criteria employed when selecting the main dataset is to obtain the best possible approximation to the industry distribution of the Portuguese economy. The objective is to produce a sample that could be, as best as possible, representative of the whole economy, and not of the bank's portfolio. If this is indeed the case, then the results of this study can be related to a typical, average credit institution operating in Portugal.

Figure 1 shows the industry distribution for both the Portuguese economy and for our dataset[7]. The two distributions are similar, although our sample has a higher concentration on industry *D – Manufacturing*, and lower on *H – Hotels & Restaurants* and *MNO – Education, Health & Other Social Services Activities*.

[7] Source: INE 2003.

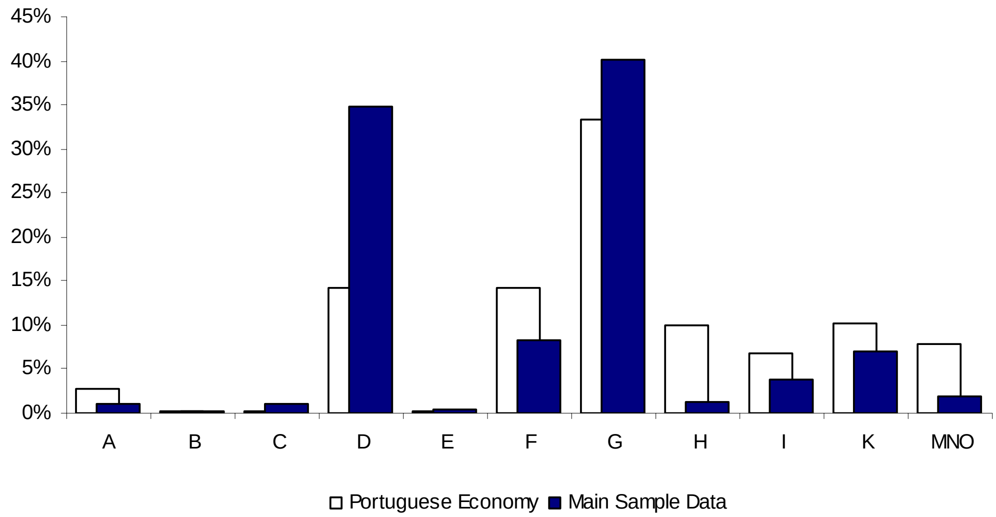

Figure 1 – Economy-Wide vs. Main Sample Industry Distribution

This figure displays the industry distribution for the firms in our dataset and for the Portuguese economy in 2003 (INE 2003). The two distributions are similar with a higher concentration in sector D for the study dataset and lower for sectors H and MNO. The industry types considered are: A – Agriculture, Hunting & Forestry; B – Fishing; C – Mining & Quarrying; D – Manufacturing; E – Electricity, Gas & Water Supply; F – Construction; G – Wholesale & Sale Trade; H – Hotels & Restaurants; I – Transport, Storage & Communications; K – Real Estate, Renting & Business Activities; MNO - Education/ Health & Social Work/ Other Personal Services Activities.

Figures 2, 3 and 4 display the industry, size (measured by annual turnover) and yearly distributions respectively, for both the default and non-default groups of observations of the dataset.

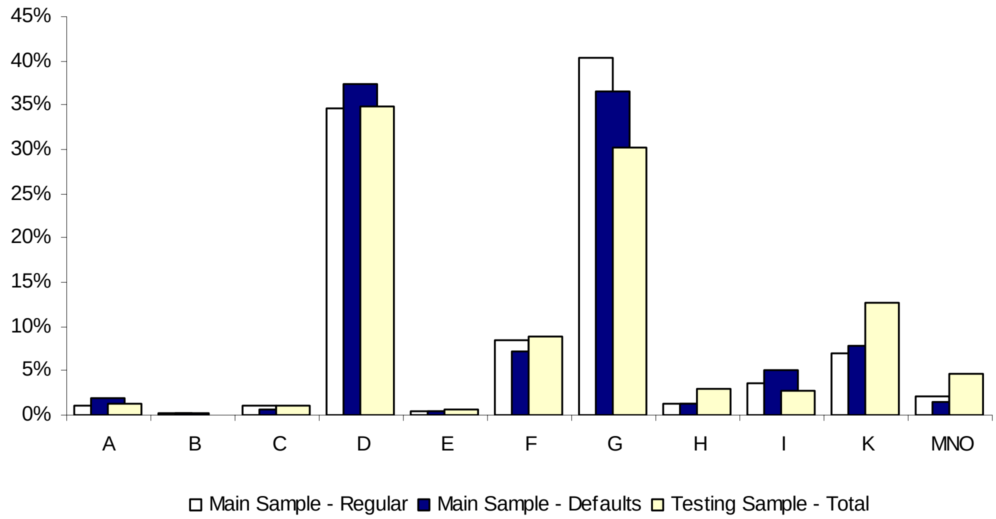

Figure 2 – Sample Industry Distribution

This figure shows the study sample industry distributions for the firms that defaulted, the firms that have not defaulted and for all firms. Both default and non-default distributions are similar, with high concentration on sectors D and G, accounting together to about 75% of the whole sample. The industry types considered are: A – Agriculture, Hunting & Forestry; B – Fishing; C – Mining & Quarrying; D – Manufacturing; E – Electricity, Gas & Water Supply; F – Construction; G – Wholesale & Sale Trade; H – Hotels & Restaurants; I – Transport, Storage & Communications; K – Real Estate, Renting & Business Activities; MNO - Education/ Health & Social Work/ Other Personal Services Activities.

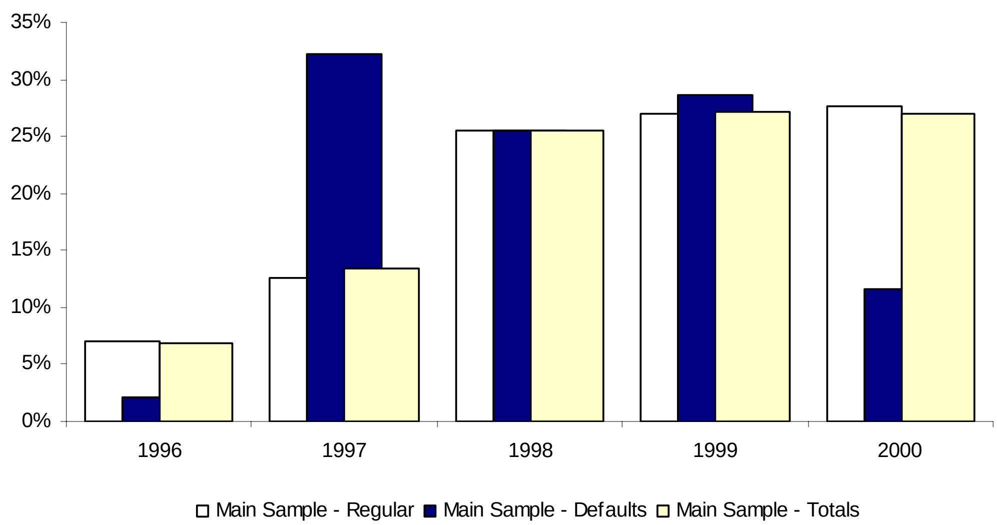

Figure 3 – Accounting Statement Yearly Distribution

The figure above displays the yearly distribution of the financial statements in the dataset for the default, non-default and total observations. The non-default yearly distribution is similar to that of the whole sample, with the number of observations rising until the third period and then remaining constant until the last period. The non-default distribution is concentrated in the 1997-1999 period, with fewer observations in the first and last periods.

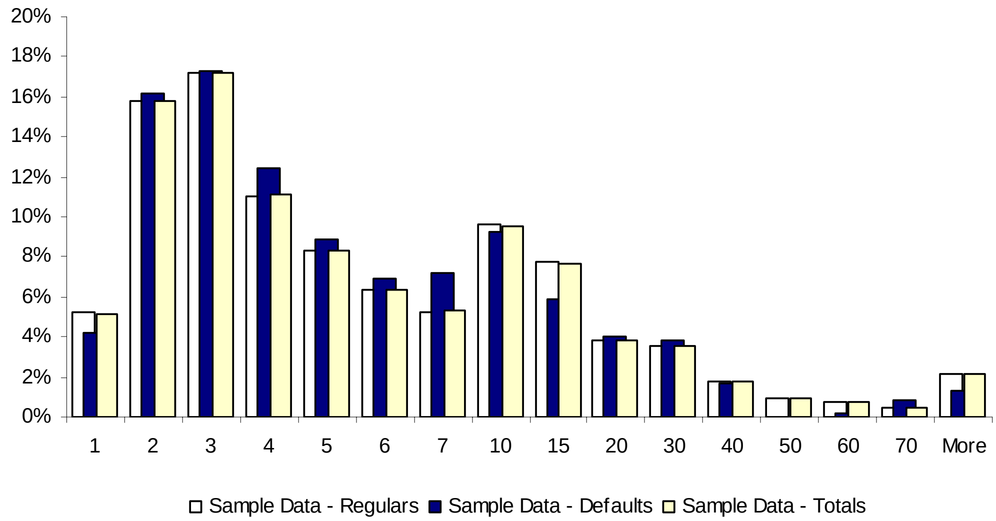

Figure 4 – Size (Turnover) Distribution, Millions of Eur

The figure shows the size distribution of the default, non-default and total observations. Size is measured by the firms' Turnover, defined has the sum of total sales plus services rendered. The tree distributions are similar, showing predominance of Small and Medium size Enterprises, with annual Turnover up to 50 Million EUR.

Analysis of industry distribution (Figure 2) suggests high concentration on industries *G – Trade* and *D – Manufacturing*, both accounting for about 75% of the whole sample. The industry distributions for both default and non-default observations are very similar.

Figure 3 shows that observations are uniformly distributed per year, for the last three periods, with about 3.000 observations per year. For the regular group of observations, the number of yearly observations rises steadily until the third period, and then remains constant until the last period. For the default group, the number of yearly observations has a great increase in the second period and clearly decreases in the last.

Figure 4 shows size distribution and indicates that most of the observations belong to the Small and Medium size Enterprises - SME segment, with annual turnover up to 50 million EUR (according to the NBCA SME classification). The SME segment accounts for about 95% of the whole sample. The distributions of both regular and default observations are very similar.

3 Financial Ratios and Univariate Analysis

A preliminary step before estimating the scoring model is to conduct an univariate analysis for each potential explanatory varaiable, in order to select the most intuitive and powerful ones. In this study, the scoring model considers exclusively financial ratios as explanatory variables. A list of twenty-three ratios representing six different dimensions – Profitability, Liquidity, Leverage, Debt Coverage, Activity and Productivity – is considered. The univariate analysis relates each of the twenty-three ratios and a default indicator, in order to assess the discriminatory power of each variable. Appendix 1 provides the list of the variables and how they are constructed. Figures 5 – 10 provide a graphical description, for some selected variables, of the relationship between each variable individually and the default frequency[8].

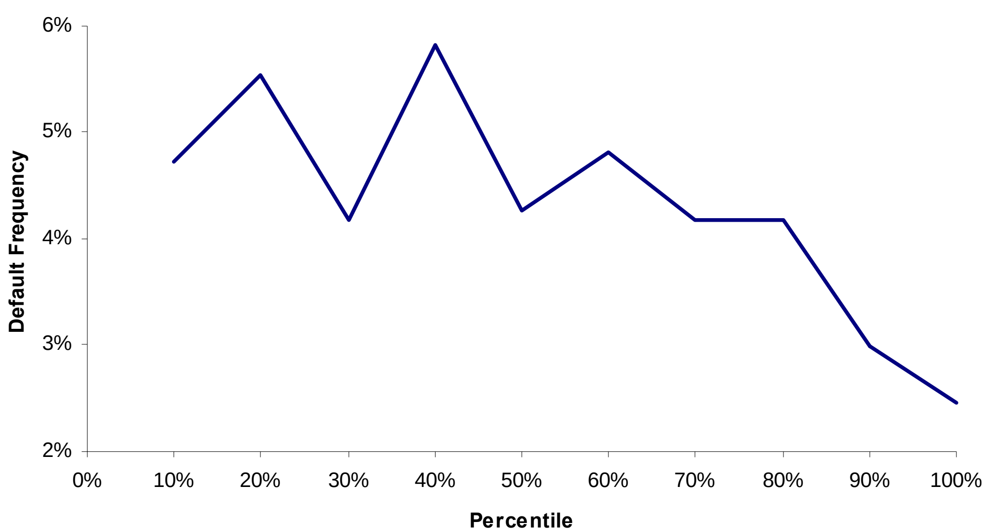

Figure 5 – Univariate Relationship Between *Liquidity / Current Liabilities* and Default Frequency

[8] The data is ordered ascendingly by the value of each ratio and, for each decile, the default frequency is calculated (number of defaults divided by the total number of observations in each decile).

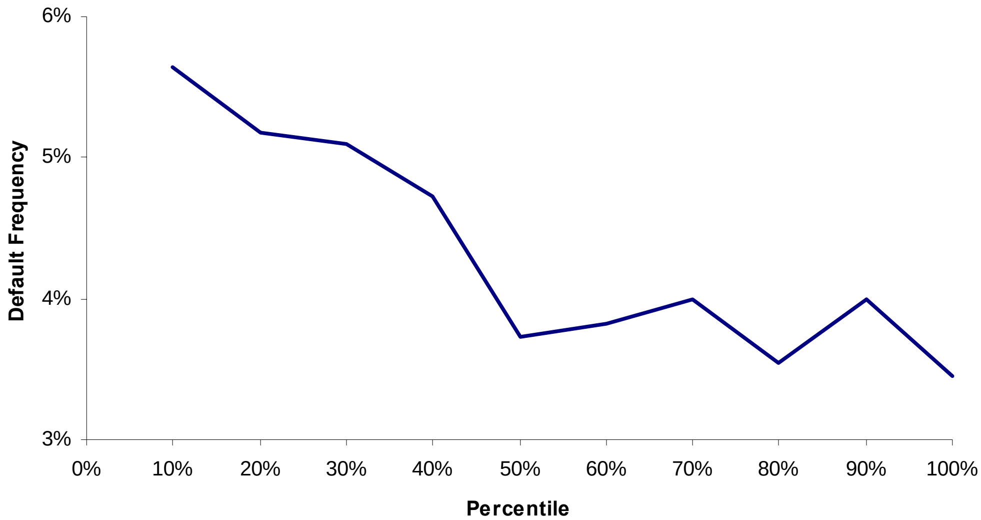

Figure 6 – Univariate Relationship Between *Current Ratio* and Default Frequency

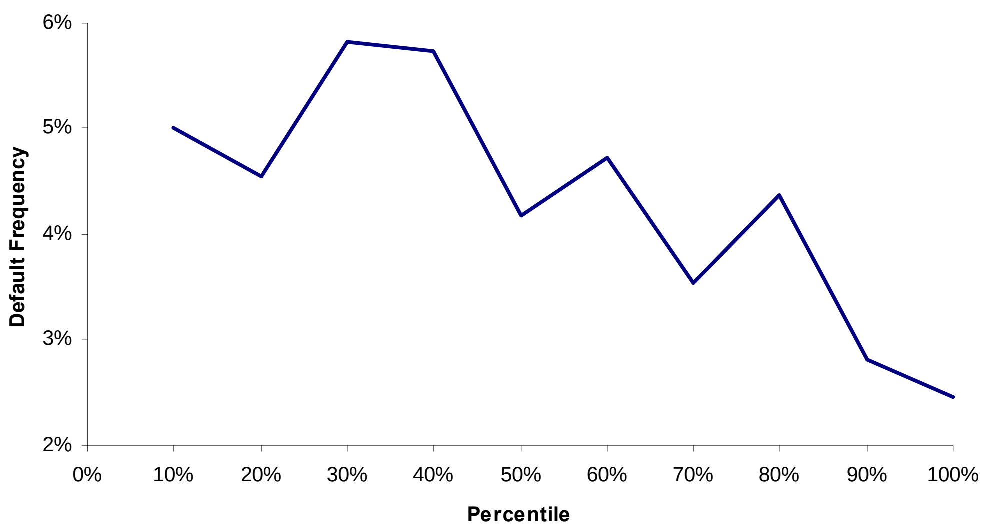

Figure 7 – Univariate Relationship Between *Liquidity / Assets* and Default Frequency

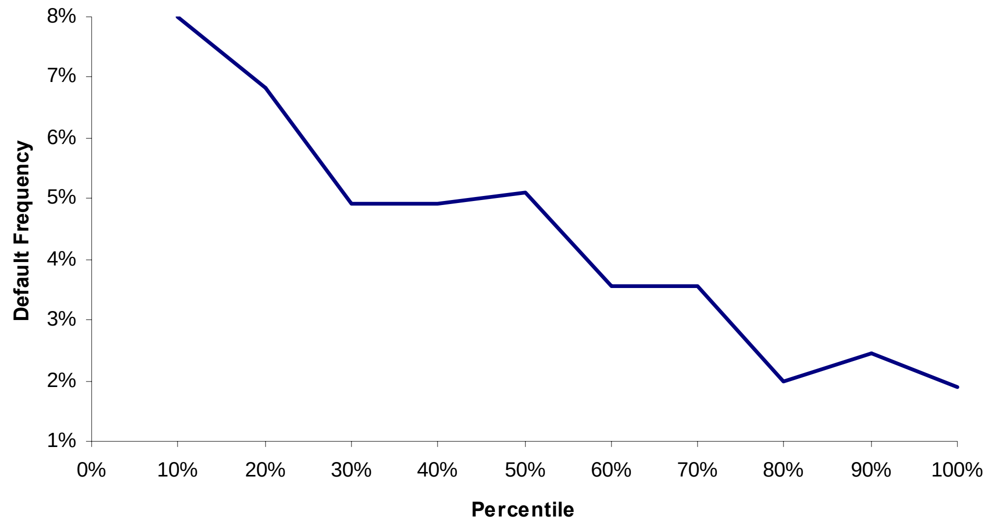

Figure 8 – Univariate Relationship Between *Debt Service Coverage* and Default Frequency

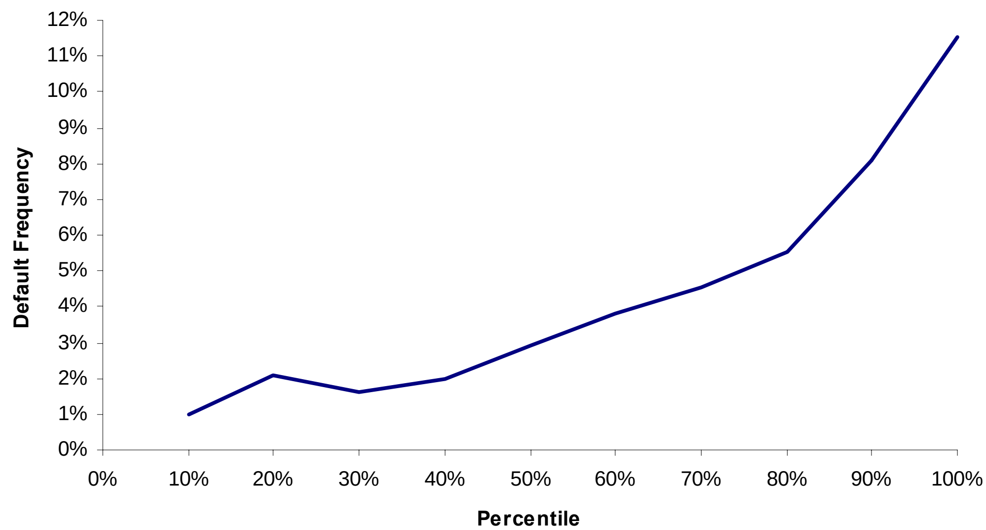

Figure 9 – Univariate Relationship Between *Interest Costs / Sales* and Default Frequency

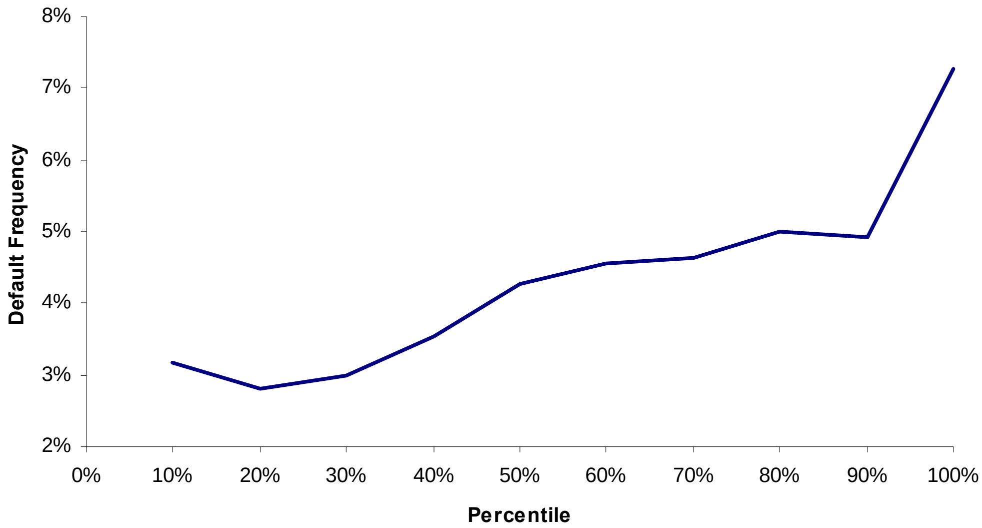

Figure 10 – Univariate Relationship Between *Productivity Ratio* and Default Frequency

In order to have a quantitative assessment of the discriminating power of each variable, the Accuracy Ratio is used[9]. The computed values of the Accuracy Ratios are reported in Appendix 1.

The selected variables for the multivariate analysis comply with the following criteria:

- They must have discriminating power, with an Accuracy Ratio higher than 5%;
- The relationship with the default frequency should be clear and economically intuitive. For example, ratio *Current Earnings and Depreciation / Turnover* should have a negative relationship with the default frequency, since firms with a high percentage of EBITDA over Turnover should default less frequently; analyzing Figure 11, there seems to be no clear relationship for this dataset;
- The number of observations lost due to lack of information on any of the components of a given ratio must be insignificant. Not all firms report the exact same items on their accounting reports, for example, ratios *Bank Debt / Accounts Payable* and *P&L / L-T Liabilities* have a significant amount of

[9] The Accuracy Ratio can be used as a measure of the discriminating power of a variable, comparing the ability of the variable to correctly classify the default and non-default obsevations against that of a random variable, unrelated to the default process. Section 4.1.1 provides a more detailed description.

missing data for the components *Debt to Credit Institutions* and *Long-Term Liabilities* respectively.

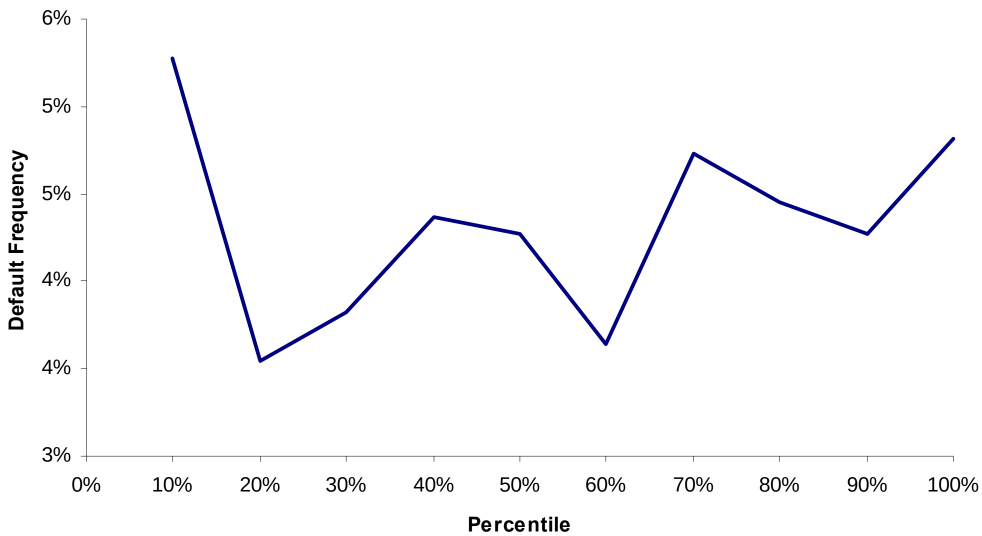

Figure 11 – Univariate Relationship Between *Current Earnings and Depreciation / Turnover* and Default Frequency

At this point, nine variables are eliminated and are not considered on the multivariate analysis. All the remaining variables are standardized in order to avoid scaling issues[10].

[10] Standardization consists on subtracting the value of the variable by its average on the sample and dividing the result by its sample standard deviation.

4 Scoring Model and Validation

The dependent variable Y_{it} is the binary discrete variable that indicates whether firm *i* has defaulted (one) or not (zero) in year *t*. The general representation of the model is:

$$Y_{it} = f\left(\beta_k, X_{it-1}^k\right) + e_{it}$$

where X_{it-1}^k represents the values of the *k* explanatory variables of firm *i*, one year before the evaluation of the dependent variable. The functional form selected for this study is the Logit model[11]. Alternative specifications can be considered, such as Probit, Linear Probability Model, or even Genetic Algorithms, although there is no evidence in the literature that any alternative specification can consistently outperform the Logit specification in credit default prediction (Altman, Marco and Varetto, 1994 and Yang et al., 1999).

Using both forward and backward procedures, the selected model is the one that complies with the validation criteria and has the higher discriminating power, measured by the Accuracy Ratio.

4.1 Model Validation

The variables selected on Section 3 are pooled together in order to obtain a model that is at the same time:

- Parsimonious but powerful: high discriminating power with few parameters to estimate;
- Statistically significant: all variables individually and the model as a whole must be significant, with low correlation between the variables;
- Intuitive: the sign of the estimated parameters should make economic sense and the selected variables should represent the various relevant risk factors.

[11] Refer to Appendix 3 for a description of the Logit model.

4.1.1 Efficiency

A model with high discriminatory power is a model that can clearly distinguish the default and non-default populations. In other words, it is a model that makes consistently "good" predictions relative to few "bad" predictions. For a given cut-off value[12], there are two types of "good" and "bad" predictions:

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="2">Estimated</th></tr>
<tr><th>Non-Default</th><th>Default</th></tr>
<tr><th rowspan="2">Observed</th><th>Non-Default</th><td>True</td><td>False Alarm (Type II Error)</td></tr>
<tr><th>Default</th><td>Miss (Type I Error)</td><td>Hit</td></tr>
</table>

- The "good" predictions occur if, for a given cut-off point, the model predicts a default and the firm does actually default (Hit), or, if the model predicts a non-default and the firm does not default in the subsequent period (True).
- The "bad" prediction occurs if, for a given cut-off point, the model predicts a default and the firm does not actually defaults (False-Alarm or Type II Error), or if the model predicts a non-default and the firm actually defaults (Miss or Type I Error).
- The Hit Ratio (HR) corresponds to the percentage of defaults from the total default population that are correctly predicted by the model, for a given cut-off point.
- The False Alarm Ratio (FAR) is the percentage of False Alarms or incorrect default predictions from the total non-defaulting population, for a given cut-off point.

Several alternatives could have been considered in order to analyze the discriminating power of the estimated models. In this study, both ROC/CAP analysis and Kolmogorov-Smirnov (KS) analysis are performed.

[12] The cut-off point is the value from which the observations are classified as "good" or "bad". For example, given a cut-off point of 50%, all observations with an estimated score between 0% and 50% will be classified as "good", and those between 50% and 100% will be considered "bad".

Receiver Operating Characteristics (ROC) and Cumulative Accuracy Profiles (CAP) curves are two closely related graphical representations of the discriminatory power of a scoring system. Using the notation from Sobehart and Keenan (2001), the ROC curve is a plot of the HR against the FAR, while the CAP curve is a plot of the HR against the percentage of the sample.

For the ROC curve, a perfect model would pass through the point (0,1) since it always makes "good" predictions, and never "bad" predictions (it has FAR = 0% and a HR = 100% for all possible cut-off points). A "naïve" model is not able to distinguish defaulting from non-defaulting firms, thus will do as many "good" as "bad" predictions, though for each cut-off point, the HR will be equal to the FAR. A better model would have a steeper curve, closer to the perfect model, thus a global measure of the discriminant power of the model would be the area under the ROC curve. This can be calculated has[13]:

$$AUROC = \int_0^1 HR(FAR)d(FAR)$$

For the CAP or Lorenz curve, a perfect model would attribute the lowest scores to all the defaulting firms, so if x% of the total population are defaults, then the CAP curve of a perfect model would pass through the point (x,1). A random model would make as many "good" as "bad" predictions, so for the y% lowest scored firms it would have a HR of y%. Then, a global measure of the discriminant power of the model, the Accuracy Ratio (AR), compares the area between the CAP curve of the model being tested and the CAP of the random model, against the area between the CAP curve of the perfect model and the CAP curve of the random model.

It can be shown that there is a linear relationship between the global measures resulting from the ROC and CAP curves[14]:

$$AR = 2\left(AUROC - 0.5\right)$$

The KS methodology considers the distance between the distributions of *1 – HR* (or Type I Errors) and *1 – FAR* (or True predictions) [15]. The higher the distance between the two distributions, the better the discriminating power of the model. The

[13] Refer to Appendix 2 for a technical description of the AUROC calculation.
[14] See, for example, Engelmann, Hayden and Tasche (2003).
[15] The Kolmogorov-Smirnov statistic is a non-parametric statistic used to test whether the density function of a variable is the same for two different groups (Conover, 1999).

KS statistic corresponds to the maximum difference for any cut-off point between the *1 – FAR* and *1 – HR* distributions.

4.1.2 Statistical Significance

All estimated regressions are subject to a variety of statistical tests, in order to ensure the quality of the results at several levels:

i. Residual Analysis is performed with the purpose of testing the distributional assumption of the errors of the regression. Although the logistic regression assumes that the errors follow a binomial distribution, for large samples (such as the one in this study), it approximates the normal distribution. The standardized residuals from the logistic regressions should then follow a standard normal distribution[16]. At this stage, severe outliers are identified and eliminated. These outliers are observations for which the model fits poorly (has an absolute studentized residual[17] greater than 2), and that can have a very large influence on the estimates of the model (a large DBeta[18]).

ii. The significance of each estimated coefficient is tested using the Wald test. This test compares the maximum likelihood value of the estimated coefficient to the estimate of its standard error. This test statistic follows a standard normal distribution under the hypothesis that the estimated coefficient is null. For the three models, all of the estimated coefficients are significant at a 90% significance level.

iii. In order to test the overall significance of each estimated model, the Hosmer-Lemeshow (H-L) test is used. This goodness-of-fit test compares the predicted outcomes of the logistic regression with the observed data by grouping observations into risk deciles.

[16] The standardized residuals correspond to the residuals adjusted by their standard errors. This adjustment is made in logistic regression because the error variance is a function of the conditional mean of the dependent variable.

[17] The studentized residual corresponds to the square root of the change in the -2 Log Likelihood of the model attributable to deleting the case from the analysis. It follows an asymptotical normal distribution and extreme values indicate a poor fit.

[18] DBeta is an indicator of the standardized change in the regression estimates obtained by deleting an individual observation.

iv. After selecting the best linear model, the assumption of linearity between each variable and the logit of the dependent variable is checked. This is performed in four stages:

1- The Box-Tidwell test (Box-Tidwell, 1962) is performed on all continuous variables, in order to confirm the linearity assumption;

2- For all variables that failed the linearity test in the previous step, a plot of the relationship between the covariate and the logit is presented, allowing to investigate the type of non-linear relationship;

3- For all continuous variables with significant non-linear relationships with the logit, the fractional polynomial methodology is implemented (Royston and Altman, 1994) in order to adequately capture the true relationship between the variables;

4- Check whether the selected transformation makes economic sense.

v. The last assumption to be checked is the independence between the explanatory variables. If multicolinearity is present, the estimated coefficients will be unbiased but their estimated standard errors will tend to be large. In order to test for the presence of high multicolinearity, a linear regression model using the same dependent and independent variables is estimated, and the tolerance statistic is calculated for each independent variable[19]. If any of the tolerance statistics are below 0.20 then it is assumed that we are in the presence of high multicolinearity, and the estimated regression is discarded.

4.1.3 Economic Intuition

All estimated coefficients follow economic intuition in the sense that the sign of the coefficients indicates the expected relationship between the selected variable and the default frequency. For example, if for a given model the estimated coefficient for variable *Productivity Ratio* is +0.123, this means that the higher the Personnel Costs relative to the Turnover, the higher the estimated credit score of the firm. In other words, firms with lower labor productivity have higher credit risk. For the non-linear relationships it is best to observe graphically the estimated relationship between the

[19] The tolerance statistic corresponds to the variance in each independent variable that is not explained by all of the other independent variables.

independent variable and the logit of the dependent. As for the linear case, this relationship should be monotonic, either always positive or negative. The difference is that the intensity of this relationship is not constant, it depends on the level of the independent variable.

During the model estimation two hypotheses are tested:

1. Whether a system of unrelated equations, by industry group yields better results than a single-equation model for all industries;
2. Whether a model where the observations are weighted in order to increase the proportion of defaults to non-defaults in the estimation sample, performs better than a model with unweighted observations.

4.2 Model A – Multiple Industry Equations

In order to test the hypothesis that a system of unrelated equations by industry group yields better results than a single-equation model for all industries, the dataset is broken into two sub-samples: the first one for *Manufacturing & Primary Activity* firms, with 5,046 observations of which 227 are defaults; and the second for *Trade & Services* firms, with 5,954 observations and 248 defaults. If the nature of these economic activities has a significant and consistent impact on the structure of the accounting reports, then it is likely that a model accommodating different variables for the different industry sectors performs better than a model which forces the same variables and parameters to all firms across industries[20]. The model is:

$$\hat{Y}_i = \frac{\exp(\hat{\mu}_i)}{1+\exp(\hat{\mu}_i)}$$

for the two-equation model,

$$\hat{\mu}_i = \begin{cases} X_i^a{}'\hat{\beta}^a & \text{if } i \text{ belongs to industry a} \\ X_i^b{}'\hat{\beta}^b & \text{if } i \text{ belongs to industry b} \end{cases}$$

for the single-equation model,

$$\hat{\mu}_i = X_i{}'\hat{\beta} \quad \forall i$$

[20] Model performance is measured by the ability to discriminate between default and regular populations, which can be summarized by the Accuracy Ratio.

For the final model, the selected variables and estimated coefficients are presented in the table below[21]:

Two-Equation Model (A)					
Industry a			**Industry b**		
Variable	**β^**	**Wald Test P-Value**	**Variable**	**β^**	**Wald Test P-Value**
Liquidity / Current Liabilities	-0.381	0.003	Current Ratio	-0.212	0.005
Debt Service Coverage	-0.225	0.021	Liquidity / Assets	-0.160	0.063
Interest Costs / Sales_1	2.011	0.002	Debt Service Coverage	-0.184	0.041
Interest Costs / Sales_2	-0.009	0.000	Interest Costs / Sales_1	1.792	0.000
Productivity Ratio	0.200	0.028	Interest Costs / Sales_2	-0.009	0.038
Constant	-3.259	0.000	Constant	-3.426	0.000
Number of Observations		5,044	**Number of Observations**		5,951
-2 LogLikelihood		1,682	**-2 LogLikelihood**		1,913
H-L Test P-Value		0.415	**H-L Test P-Value**		0.615

Table 1 – Estimated Model Variables and Parameters, Two-Equation Model (A)

The table above presents the estimation results for the two industry equation model. *Industry a* represents Manufacturing & Primary Activity firms, and *Industry b* Trade & Services firms. The sign of the estimated parameters for both regressions is in accordance with economic intuition. The significance of each parameter is demonstrated by the low p-values for the Wald test, while the overall significance of each regression is verified by the high p-values for the Hosmer-Lemeshow test.

The Hosmer-Lemeshow test is a measure of the overall significance of the logistic regression. Through the analysis of Figure 12 we can conclude that the estimated logistic regressions significantly fit the observed data.

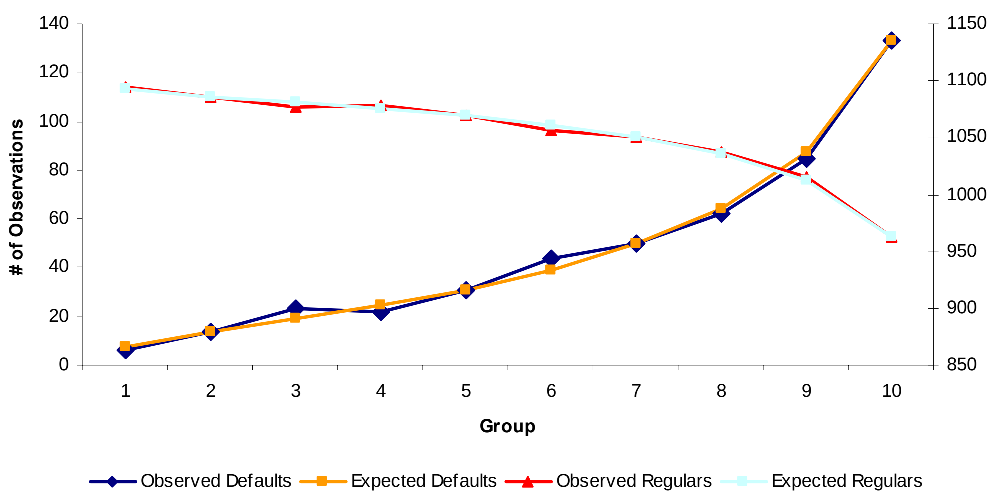

Figure 12 – Model A: Hosmer-Lemeshow Test

[21] Refer to Appendix 4 for full estimation results.

4.3 Model B – Single Equation, Unweighted Sample

In order to test our two hypotheses both the Two-Equation Model and the Weighted Sample Model will be evaluated against the standard setting of a single equation across all industries, using an unweighted sample. Table 2 summarizes the final results under this standard setting and Figure 13 provides a graphical description of the overall significance of the estimated model.

Single-Equation Model (B)		
Variable	**β^**	**Wald Test P-Value**
Current Ratio	-0.171	0.001
Liquidity / Assets	-0.211	0.002
Debt Service Coverage	-0.231	0.001
Interest Costs / Sales_1	1.843	0.007
Interest Costs / Sales_2	-0.009	0.000
Productivity Ratio	0.124	0.003
Constant	-3.250	0.000
Number of Observations		10,995
-2 LogLikelihood		3,600
H-L Test P-Value		0.973

Table 2 – Estimated Model Variables and Parameters, Single-Equation Model (B)
This table displays the estimation results for the single-equation model. The sign of the estimated parameters agrees with economic intuition. The significance of each parameter is demonstrated by the low p-values for the Wald test, while the overall significance of the regression is verified by the high p-values for the Hosmer-Lemeshow test.

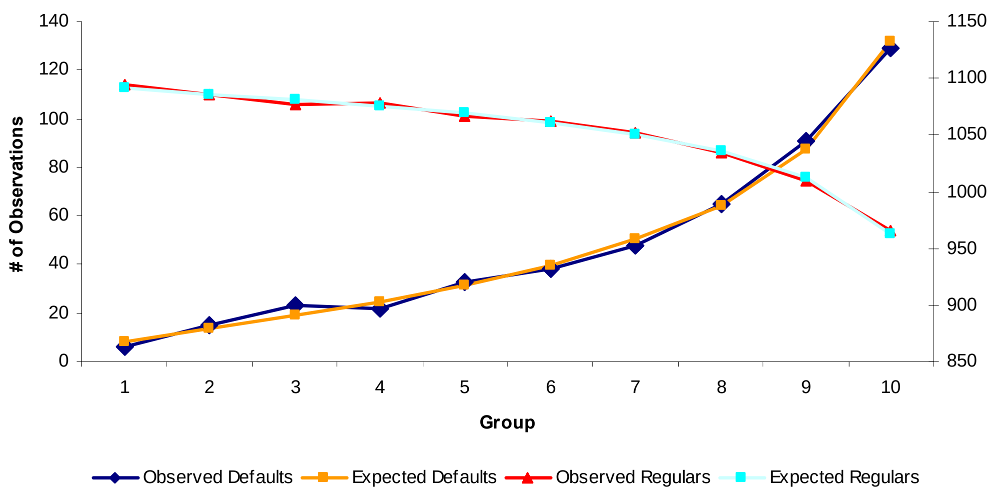

Figure 13 – Model B: Hosmer-Lemeshow Test

4.4 Model C – Weighted Sample

The proportion of the number of defaults (450) to the total number of observations in the sample (11,000) is artificially high. The real average annual default frequency of the bank's portfolio and the Portuguese economy is significantly lower than the 4.32% suggested by our sample for the corporate sector. However, in order to be able to correctly identify the risk profiles of "good" and "bad" firms, a significant number of observations for each population is required. For example, keeping the total number of observations constant, if the correct default rate is about 1%, extracting a random sample in accordance to this ratio would result in a proportion of 110 default observations to 11,000 observations.

A consequence of having an artificially high proportion of default observations is that the estimated scores cannot be directly interpreted as real probabilities of default. Therefore, these results have to be calibrated in order to obtain default probabilities estimates.

A further way to increase the proportion of the number of default observations is to attribute different weights to the default and regular observations. The weightening of observations could potentially have two types of positive impact in the analysis:

1. As mentioned above, a more balanced sample, with closer proportions of defaults and regular observations, could help the Logit regression to better discriminate between both populations;
2. The higher proportion of default observations results in higher estimated scores. As a consequence, the scores in the weighed model are more evenly spread throughout the]0,1[interval (see Figure 14). If, in turn, these scores are used to group the observations into classes, then it could be easier to identify coherent classes with the weighed model scores. Thus, even if weightening the observations does not yield a superior model in terms of discriminating power, it might still be helpful later in the analysis, when building the rating classes.

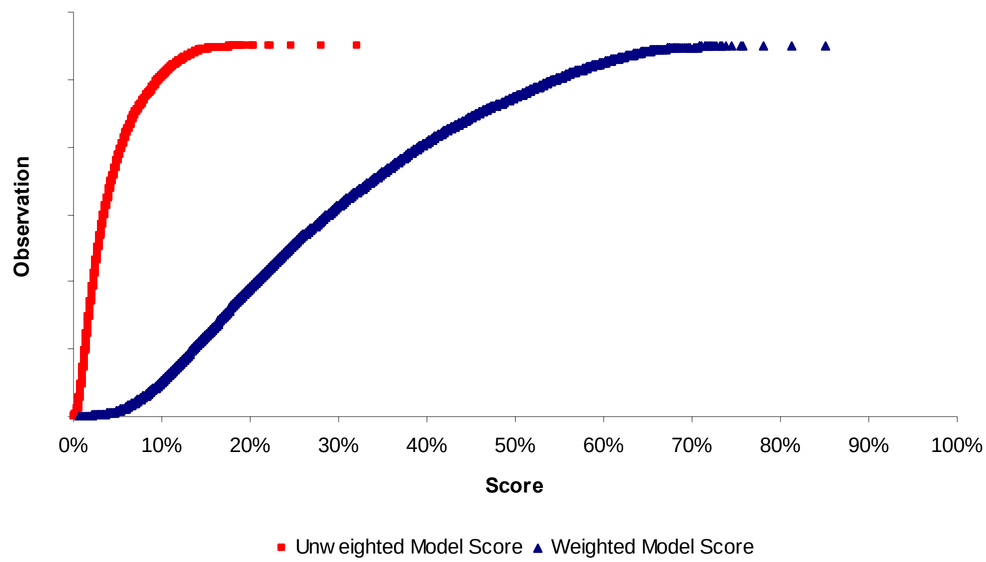

Figure 14 – Weighted vs. Unweighted Score

The weighted model estimated considers a proportion of one default observation for two regular observations. The weighed sample consists of 1,420 observations, of which 470 are defaults and the remaining 950 are regular observations[22]. The optimized model selects the same variables has the unweighted model though with different estimated coefficients.

Weighted Model (C)		
Variable	**β^**	**Wald Test P-Value**
Current Ratio	-0.197	0.003
Liquidity / Assets	-0.223	0.006
Debt Service Coverage	-0.203	0.013
Interest Costs / Sales_1	1.879	0.050
Interest Costs / Sales_2	-0.009	0.000
Productivity Ratio	0.123	0.023
Constant	-0.841	0.000
Number of Observations		1,420
-2 LogLikelihood		1,608
H-L Test P-Value		0.465

Table 3 – Estimated Model Variables and Parameters, Weighted Model (C)

This table shows the estimation results for the weighted sample model. The selected variables are the same as for the unweighted model (B). All estimated parameters are significant at 95% confidence level and the suggested relationship with the dependent variable concurs with economic intuition. The high p-values for the Hosmer-Lemeshow test attest the overall significance of the regression.

[22] Other proportions yield very similar results (namley the one default for one non-default and one default for three non-defaults proportions).

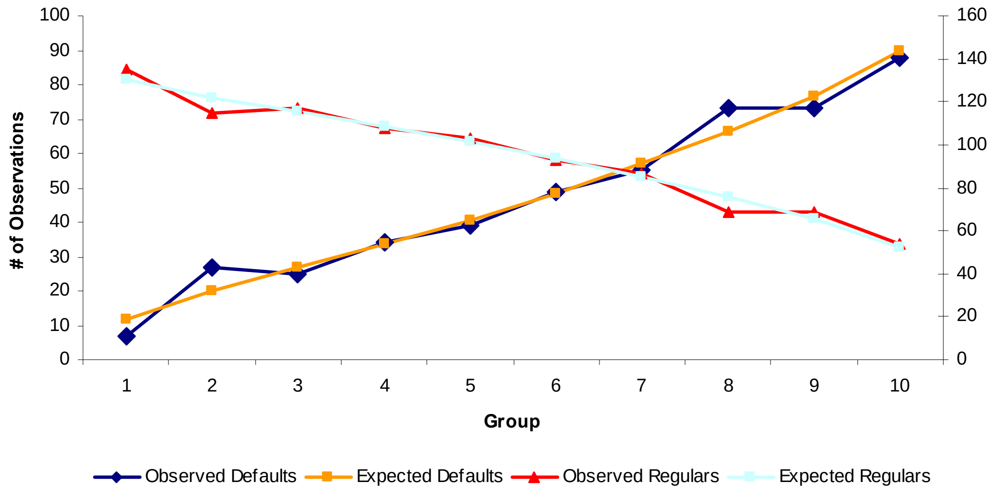

Figure 15 – Model C: Hosmer-Lemeshow Test

The following section analyses the estimation results in more detail and compares the different approaches in terms of efficiency.

4.5 Analysis of the Results

In Appendix 4, the final results of the estimations are presented for all three models: the two-equation model (Model A), the unweighted single-equation model (Model B) and the weighted single-equation model (Model C). The first step to obtain each model is to find the best linear combination through backward and forward selection procedures. The estimation equation that complies with both economic intuition and positive statistical diagnosis (described in steps i. to iii. of section 4.1.2), and had the higher discriminating power is considered the optimal linear model.

The second step is to check for non-linear relationships between the independent variables and the logit of the dependent. Results indicate that for all four selected linear regressions, there is a clear non-linear relationship between variable *Interest Costs / Sales* and the logit of the dependent variable. In order to account for this fact, the procedure described in step iv. of section 4.1.2 is implemented. The resulting non-linear relationship for the four regressions is illustrated in Figure 16 below:

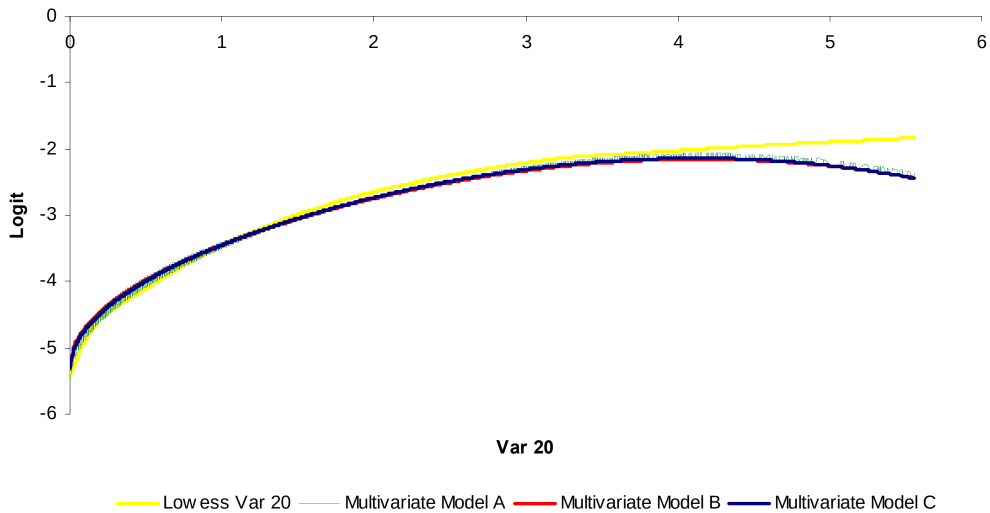

Figure 16 – Plot of the Univariate Smoothed Lowess Logit vs. Multivariate Fractional Polynomial Adjustment of Var. *Interest Costs / Sales*

The figure above demonstrates the non-linear nature of the relationship between the variable *Interest Costs / Sales* and the default event. In order to depict graphically the relationship between the covariate and the binary dependent variable, a Locally Weighted Scatterplot Smoothing, or Lowess (Cleveland 1979), was created. In addition, the quality of the fit of this relationship for the three estimated models can be accessed by comparing the multivariate adjustment for each model with the lowess curve. For all three models, the quality of the adjustment is high but deteriorates for very high values of the explanatory variable.

After the optimal non-linear regressions are selected, a final test for multicolinearity is implemented. Only the *Trade & Services* regression of the Two-Equation Model presented signs of severe multicolinearity. Since there is no practical method to correct this problem, the model is discarded and the second best model suggested by the fractional polynomial procedure is selected. This alternative specification does not suffer from multicolinearity, as it can be observed in the results presented in Appendix 4[23]. In short, the modeling procedure consisted on selecting the best discriminating regression from a pool of possible solutions that simultaneously complied with economic and statistical criteria.

In terms of efficiency, all three models have a small number of selected variables: model A has five variables for each equation, while models B and C have six variables each. Analyzing Figures 17 – 24, we can conclude that all three models have significant discriminating power and have similar performances. Results for

[23] In order to ensure stability of the final results, the whole modeling procedure is repeated with several random sub-samples of the main dataset. Across all sub-samples the variables selected for each model are the same, the values of the estimated coefficients are stable, and the estimated AR's are similar.

Altman's Z'-Score Model for Private Firms (Altman, 2000) are also reported as a benchmark (Model D):

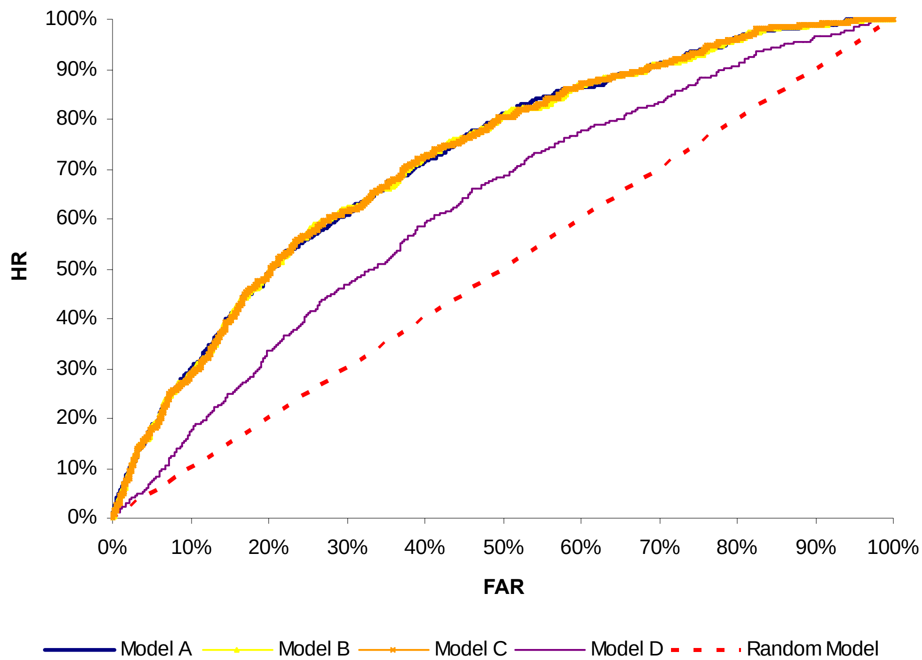

Figure 17 – Receiver Operating Characteristics Curves

The figure above displays the Receiver Operating Characteristics curves for the three estimated models and for the Z'-Score Model for Private Firms (Altman, 2000). The ROC curve provides for each possible cut-off value the proportion of observations incorrectly classified as default by the model against the proportion correctly classified as default. The three suggested models have similar ROC curves, clearly above the Random Model and Z'-Score Model curves.

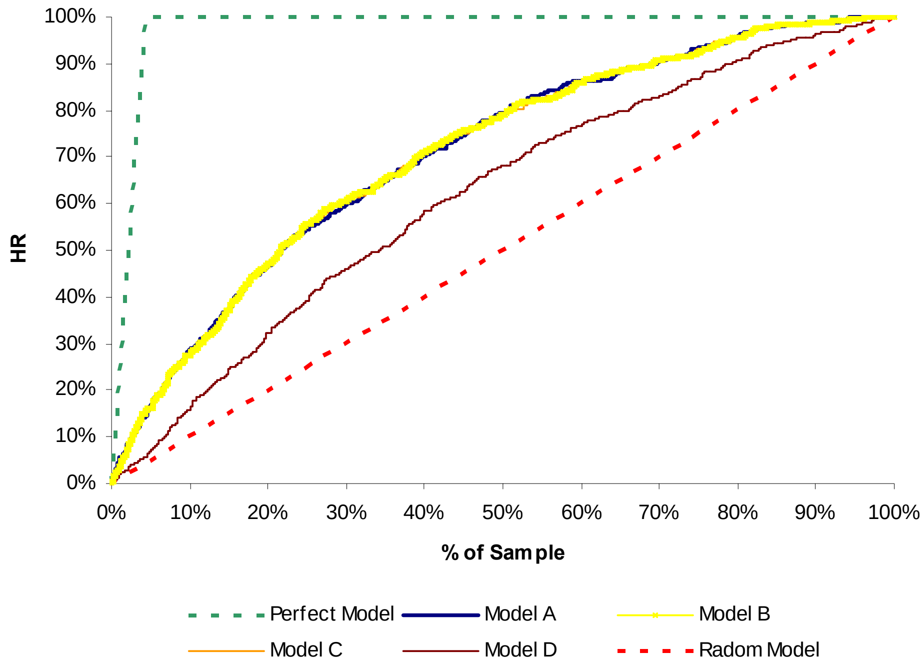

Figure 18 – Cumulative Accuracy Profiles Curves
This figure displays the Cumulative Accuracy Profiles curves for the three models estimated and for the Z'-Score model. The CAP curve provides, for a given proportion of observations with the highest estimated scores, the proportion of correctly classified default observations. As with the ROC analysis, the curves for the three selected models are similar and clearly above the Random Model and Z'-Score Model curves.

Figures 19-24 provide both the Kolmogorov-Smirnov (KS) analysis and Error Type curves. The KS analysis consists on evaluating for each possible cut-off point the distance between the Type I Error curve and the True Prediction curve. The higher the distance between the curves, the better the discriminating power of the model. The Error Type curves display for each cut-off point the percentages of Type I (incorrectly classifying an observation as non-default) and Type II (incorrectly classifying an observation as default) errors for each model.

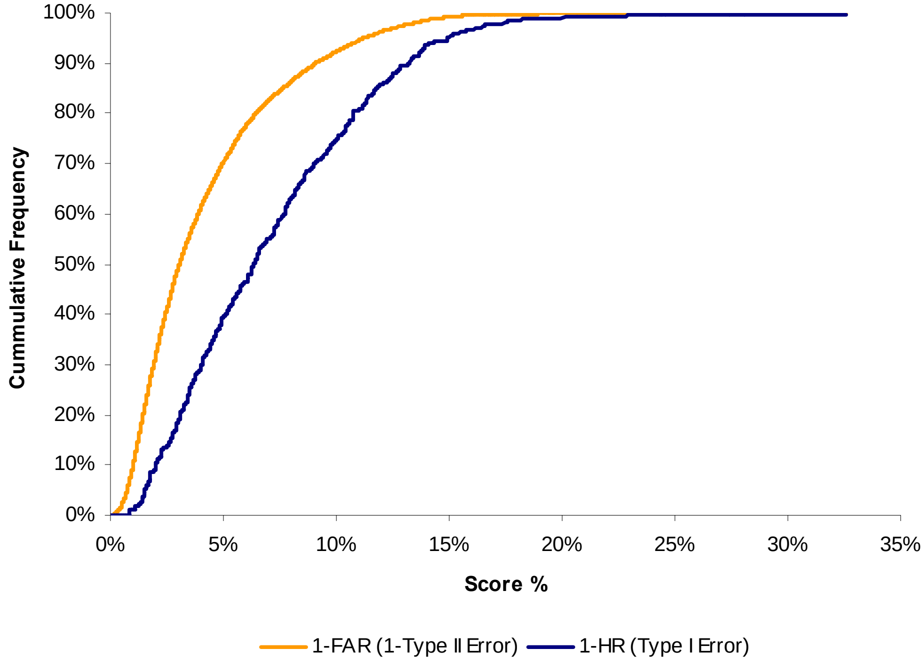

Figure 19 – Model A: Kolmogorov-Smirnov Analysis

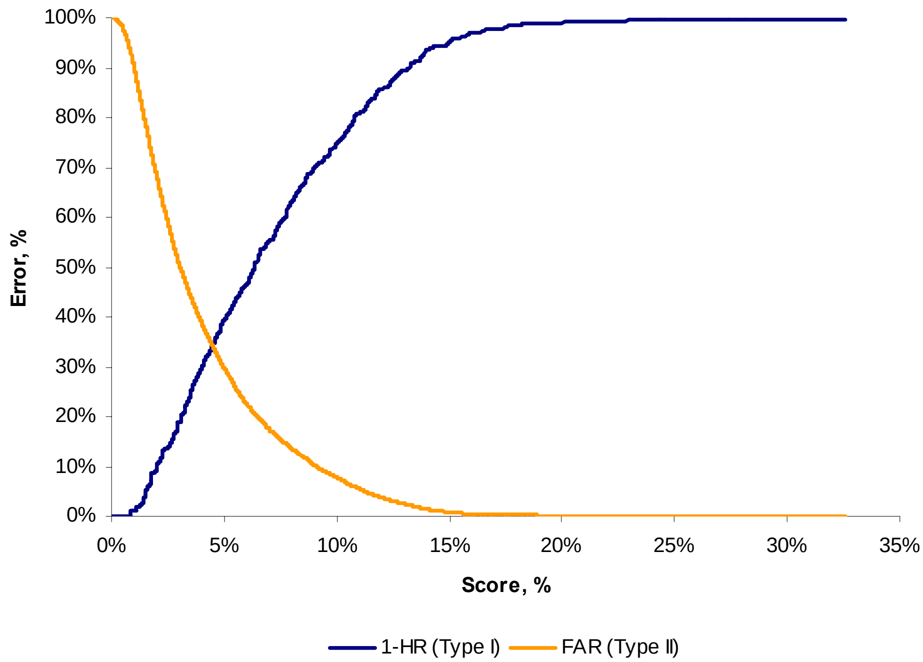

Figure 20 – Model A: Types I & II Errors

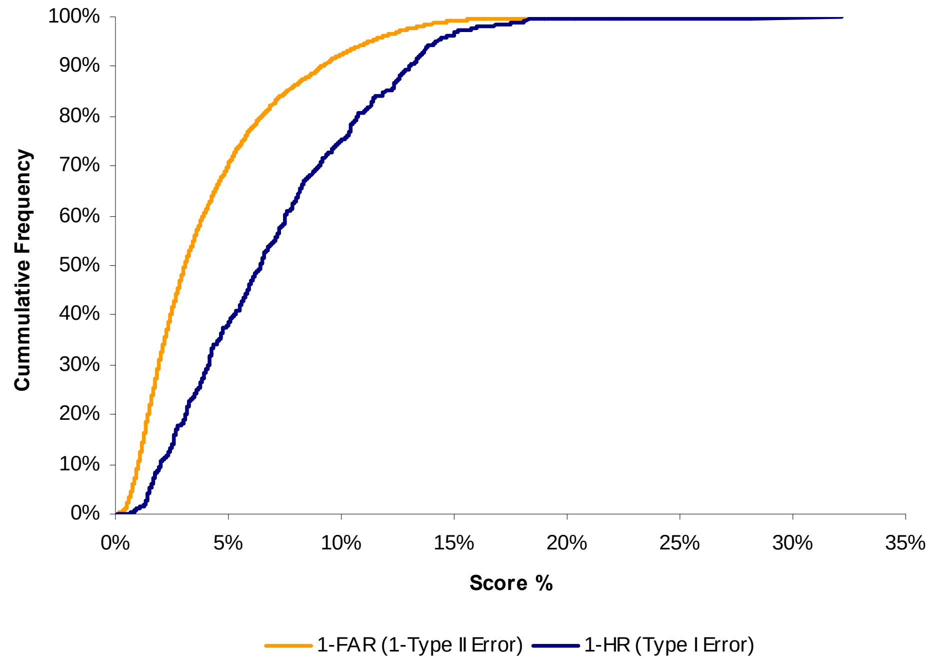

Figure 21 – Model B: Kolmogorov-Smirnov Analysis

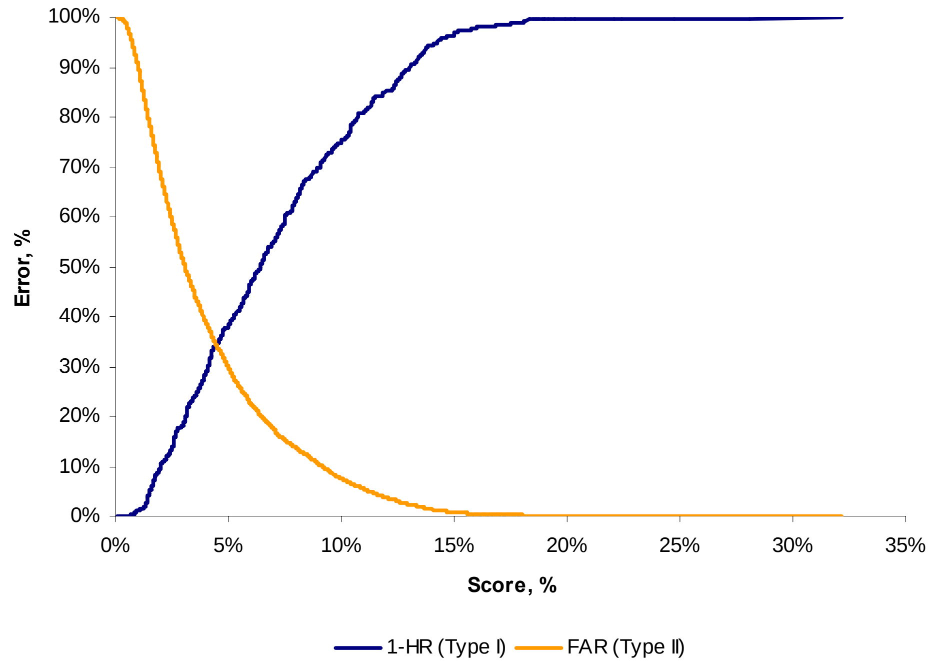

Figure 22 – Model B: Types I & II Errors

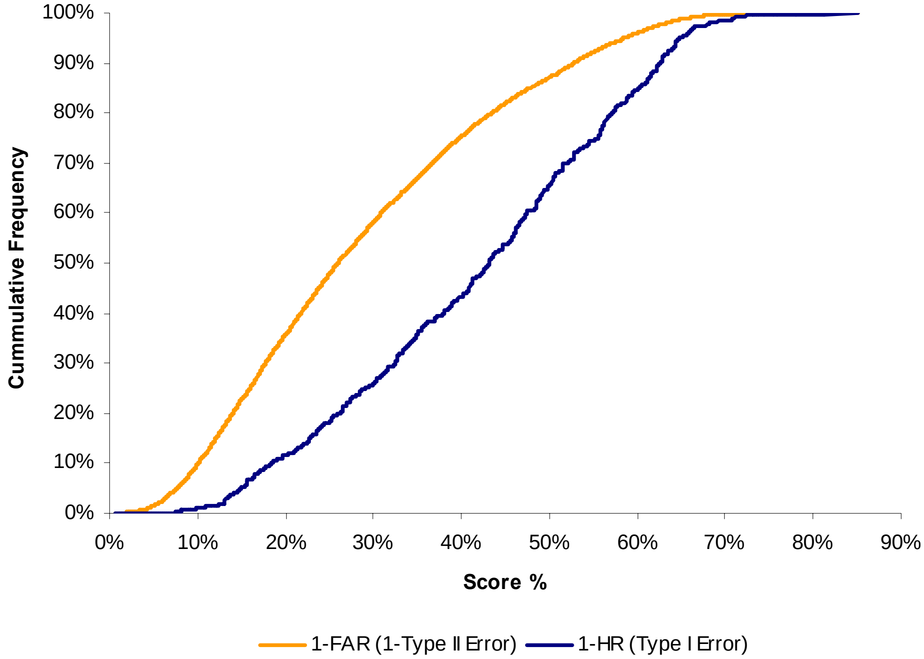

Figure 23 – Model C: Kolmogorov-Smirnov Analysis

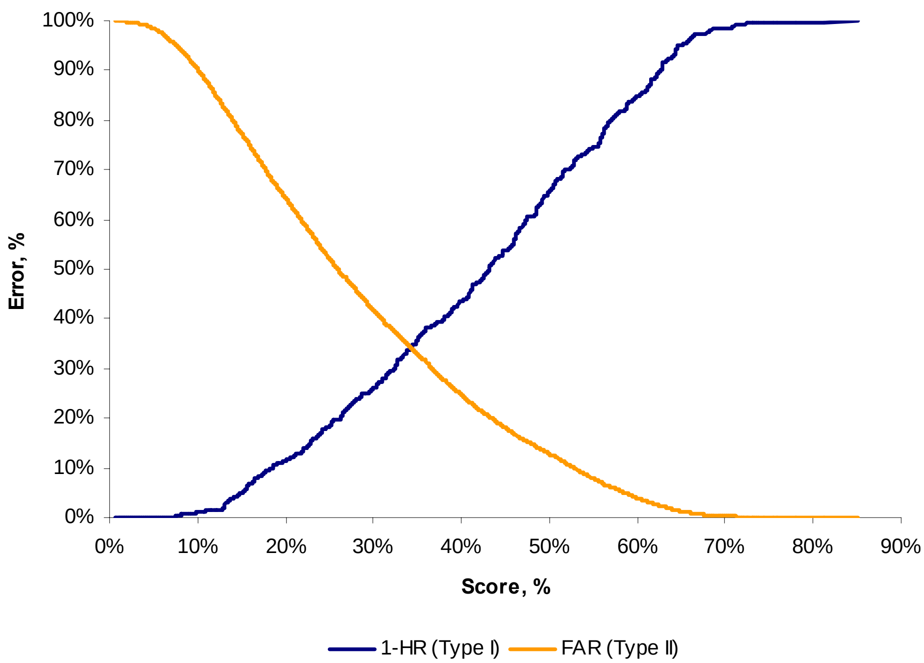

Figure 24 – Model C: Types I & II Errors

Table 4 reports the results for both ROC/CAP analysis and KS analysis (Model D is the Z'-Score):

	Main Sample				Out-of-Sample		
	AUROC	**σAUROC**	**AR**	**KS**	**AUROC**	**σAUROC**	**AR**
A	71.88%	1.15%	43.75%	32.15%	73.04%	7.53%	46.07%
B	71.88%	1.15%	43.77%	32.97%	75.29%	6.55%	50.59%
C	71.87%	1.15%	43.74%	32.94%	74.15%	6.88%	48.29%
D	62.53%	1.25%	25.07%	19.77%	61.11%	6.87%	22.22%

Table 4 – AUROC, AR and KS Statistics

This table reports the Area Under the ROC curves, Accuracy Ratios and Kolmogorov-Smirnov statistics estimated for the three suggested models and the Z'-Score, under both the estimation and testing samples. All three measures of discriminating power, under both samples, indicate similar and positive values for the three models estimated.

A more rigorous comparison of the discriminating power of the models can be obtained through a statistical test for the difference between the estimated AUROC's of the different models[24]. Table 5 presents the results of applying this test to the differences between all models for both samples:

Test	Main Sample			Out-of-Sample		
	θi - θj	**σ (θi - θj)**	**P-Value**	**θi - θj**	**σ (θi - θj)**	**P-Value**
A - B	-0.0089%	0.2225%	96.83%	-2.2571%	2.8844%	43.39%
A - C	0.0053%	0.2372%	98.23%	-1.1086%	2.7449%	68.63%
A - D	9.3425%	1.7807%	0.00%	11.9256%	7.7745%	12.50%
B - C	0.0141%	0.0476%	76.68%	1.1485%	0.5115%	2.47%
B - D	9.3514%	1.7788%	0.00%	14.1827%	6.7577%	3.58%
C - D	9.3372%	1.7751%	0.00%	13.0342%	7.0051%	6.28%

Table 5 – Testing the Differences between AUROC's

The table above provides a statistical test for comparing the estimated AUROC curves between the different models (DeLong et al., 1988). For both the estimation and testing samples there is no evidence to support the two alternative specifications to the standard model, in terms of discriminatory power. Considering multiple industry equations (A) or a weighted sample (C) does not yield a significant increase on the discriminatory power over the standard model (B), for both samples. All three suggested specifications have significant better performances than the Z'-Score model.

The results indicate that for both samples, Models A, B and C have similar discriminating power, and all three perform significantly better that the Z'-Score model.

[24] For a description of the test consult Appendix 2.

Regarding our first hypothesis that a setting with multiple equations could yield better results, both in-sample and out-of-sample results suggest there is no improvement from the standard approach. The estimated Accuracy Ratio for the two-equation model is 43.75%, which is slightly worse than the Accuracy Ratio of the single-equation model, 43.77%. The out-of-sample results confirm this tendency, the AR of the two-equation model is 46.07%, against 50.59% of the single-equation model, according to the test results presented in Table 5 above, none of these differences is statistically significant. Since the two-equation model involves more parameters to estimate and is not able to better discriminate to a significant extent the default and regular populations of the dataset, the single-equation specification is considered superior in terms of scoring methodology for this dataset. Regarding the hypothesis that balancing the default and non-default populations could help the logistic regression to better discriminate them, again both in-sample and out-of-sample results do not provide positive evidence. The estimated Accuracy Ratio for the weighed model is 43.74%, marginally worse than the 43.77% of the unweighted model. Again, the out-of-sample results confirm that the weighted model does not have a higher discriminating power (AR of 48.29%) than the unweighted model (AR of 50.59%).

As reference, the private-firm model developed by Moody's to the Portuguese market has an in-sample AR of 61.1% (unfortunately no out-of-sample AR is reported)[25]. The selected variables are: *Equity / Total Accounts Payable, Bank Debt / Total Liabilities, Net P&L / Assets, (Ordinary P&L+ Depreciation)/ Interest and similar Expenses, (Ordinary P&L + Depreciation + Provisions) / Total Liabilities, Current Assets / Accounts Payable (due within 1 year)* and *Interest and similar Expenses / Turnover*. The sample data comprised financial statements of 18,137 unique firms, of which 416 had defaulted (using the "90 days past due" definition), with a time span from 1993 to 2000. Hayden (2003) reports an in-sample AR of 50.3% and an out-of-sample AR of 48.8% for a logistic regression model applied to the Austrian market, with the "90 days past due" default definition. The variables selected are *Equity / Assets, Bank Debt / Assets, Current Liabilities / Assets, Accounts Payable / Mat. Costs, Ordinary Business Income / Assets* and *Legal Form*. The sample data included 16,797 observations, of which 1,604 were defaults, for a time

[25] See Murphy et al. (2002)

period ranging from 1992 to 1999. Due to differences in the dataset, such as different levels of data quality or the ratio of default to non-default observations, the reported AR's for both studies presented above cannot be directly comparable to the AR's reported in our study. Despite this fact, they can still be regarded as references that attest the quality of the model presented in terms of discriminatory power.

The following chapter discusses possible applications of the scoring model presented. We start by discussing the creation of a quantitative rating system, followed by the estimation of probabilities of default and rating transition matrixes. Finally the capital requirements for a simulated portfolio are calculated under both the NBCA and current regulations.

5 Applications

5.1 Quantitative Rating System and Probability of Default Estimation

The scoring output provides a quantitative assessment of the credit quality of each firm. Rating classes can be built through a partition of the scoring scale into *k* groups. A default frequency can, in turn, be estimated for each partition, dividing the number of default observations by the total number of observations for each rating class. Furthermore, these default frequencies can be leveled in order to allow for the global default rate of the dataset to be similar to the projected default rate of the universe. These adjusted default frequencies represent the Probability of Default (PD) estimates of the quantitative rating system for each rating class. In light of the NBCA, these can be interpreted as an approximation to the long-run averages of one-year realized default rates for the firms in each rating class[26].

The quantitative rating system presented in this section is not directly comparable to the traditional rating approaches adopted by the rating agencies. The two main differences between the systems are the scope of the analysis and the volatility of the rating classes. Regarding the scope of the analysis, the system developed in this study is concerned with only one risk dimension, the probability of default. Ratings issued by the agencies address not just obligor risk but the facility risk as well. The other major difference is related to the time horizon, the quantitative system has a specific one-year time horizon, with high volatility subject to economic cycle fluctuations. The agencies approach is to produce through-the-cycle ratings, with unspecific, long-term time horizon. Cantor and Packer (1994) provide a description of the rating methodologies for the major rating agencies, while Crouhy et al. (2001) present the major differences between the internal rating system of a bank and the rating systems of two major credit rating agencies.

[26] Basel Committee on Banking Supervision (2003), par. 409.

Regarding the quantitative rating system, two alternative methodologies are employed in order to obtain the optimal boundaries for each rating class. The goal is for the rating system to be simultaneously stable and discriminatory. A stable rating system is one with infrequent transitions, particularly with few ample transitions[27]. A discriminatory rating system is a granular system with representative and clear distinct classes, in terms of the frequency of default that should increase monotonically from high to low rating classes.

The first methodology employed consists on obtaining coherent rating classes through the use of cluster analysis on the scoring estimates. The second methodology is devised as an optimization problem that attempts to map the historical default frequencies of rating agency whole letter obligor ratings.

5.1.1 Cluster Methodology

Clustering can be described as a grouping procedure that searches for a "natural" structure within a dataset. It has been used thoroughly in a wide range of disciplines as a tool to develop classification schemes. The observations in the sample are reduced to *k* groups in a way that within each group, these observations are as close as possible to each other than to observations in any other group.

K-Means algorithm is implemented due to the large number of observations[28]. In order to determine the optimal number of clusters, the Calinski and Harabasz (1974) method is used. This index has been repeatedly reported in the literature as one of the best selecting procedures (Milligan and Cooper, 1985). The index is calculated as:

$$CL(k) = \frac{\sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{Y})^2 / (k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n} n_i(Y_{ij} - \overline{Y}_i)^2 / (n-k)} = \frac{BSS/(k-1)}{WSS/(k-n)}$$

where *BSS* is the Between Sum-of-Squares; *WSS* the Within Sum-of-Squares; *k* the number of clusters; *n* the number of observations; Y_{ij} estimated score for observation *j* in cluster *i*.

[27] An ample transition is a rating upgrade/downgrade involving several rating notches. For example, if a firms has a downgrade from Aaa to Caa in just one period.

[28] Refer to Appendix 5 for a description of the algorithm used.

The optimal *k* is the one that maximizes the value of *CL(k)*, since it will be at this point that the relative variance between groups respective to the variance within the groups will be higher.

The cluster analysis is performed on the scoring estimates of the three models estimated previously. Table 5 reports the *CL(k)* index for *k = 2* up to *k = 20.*

k	Model A	Model B	Model C
2	25,092	25,644	28,240
3	30,940	32,046	36,176
4	35,105	36,854	44,639
5	39,411	42,252	50,774
6	43,727	45,889	58,179
7	48,015	51,642	65,751
8	54,666	49,930	72,980
9	55,321	56,529	77,201
10	61,447	62,321	86,546
11	55,297	57,629	93,152
12	62,620	63,913	95,021
13	69,788	71,726	104,821
14	65,603	78,093	110,153
15	73,152	73,530	116,503
16	78,473	75,129	126,060
17	74,141	84,335	129,162
18	**79,710**	82,801	138,090
19	75,293	78,527	**138,461**
20	79,154	**87,544**	134,544

Table 6 – Calinski-Harabasz CL(k) index for k = 2 up to k = 20

The table above reports the Calinski and Harabasz (1974) index for the three alternative specifications using 2 to 20 clusters. The optimal number of clusters is the one that maximizes the index, where the relative variance between groups respective to the variance within the groups will be higher.

For Model A, the optimal number of clusters is 18, for Model B is 20, and for Model C is 19. In order to directly compare the resulting rating systems, classes are aggregated into *k = 7*[29]. This class aggregation is performed taking in consideration both stability and discriminatory criteria. Figures 25 and 26 present the distribution of the default frequency and of the number of observations by rating class, for each model.

[29] *K = 7* is the minimum number of classes recommended in the NBCA (Basel Committee on Banking Supervision 2003, par. 366) and it is also the number of whole letter rating classes of the major rating agencies.

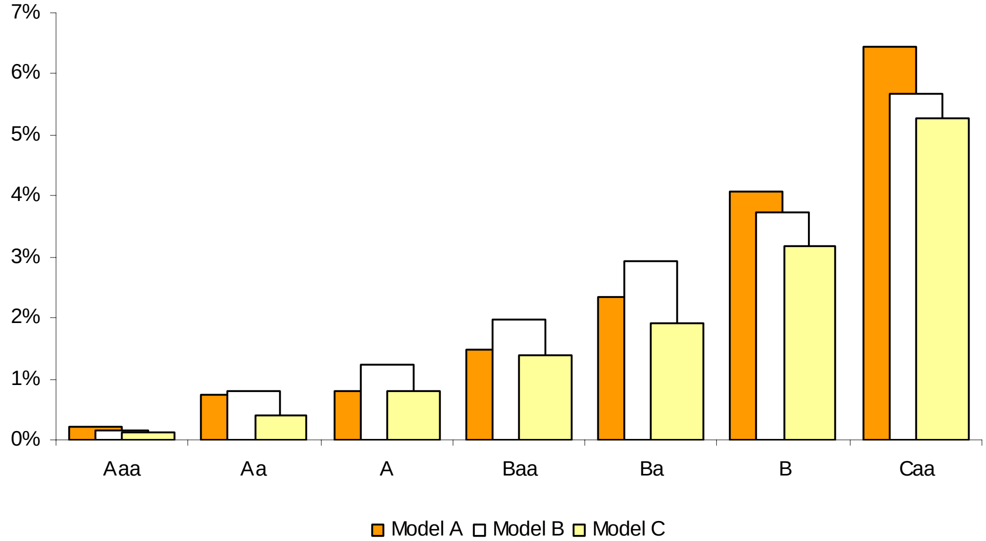

Figure 25 – Default Frequency by Rating Class (Cluster Method)

This figure provides the default frequency for each rating class determined by the cluster methodology. The default frequency corresponds to the proportion of default observations to the total number of observations in each rating class, adjusted by a calibration factor. As expected the defaulted frequencies are higher for the lower rating classes. The three alternative specifications yield similar results.

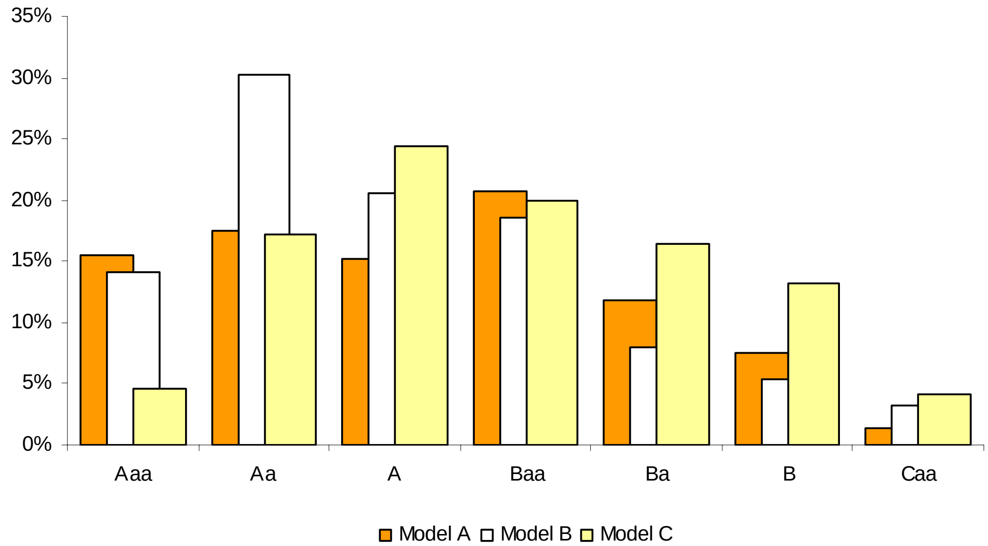

Figure 26 – Number of Observations Distribution by Rating Class (Cluster Method)

The figure above displays the distribution of the number of observations by rating class, with the rating classes determined by the cluster methodology. The distribution for the weighted sample specification (Model C) is clearly more balanced than the distribution for the other two specifications, avoiding excessive concentration on the higher rating classes.

Results in Figure 25 are similar across all three models, the default frequency rises from lower to higher risk ratings (only exception being the inflection point for Model A between classes Aa and A), although this rise is only moderate. The defaulted frequencies reported are calibrated frequencies that, as mentioned before, can be interpreted has the actual PD estimates for each rating class. Since the dataset is biased towards the default observations, the resulting default frequencies are leveled so that the overall default ratio would equal 1.5%[30].

Regarding the distribution of observations (Figure 26 above), it is interesting to observe that the three models that have in so far presented very similar results actually produce clearly distinct rating classes. Model A suggests a more uniformly distributed system, with only the lowest rating class having fewer observations. Model B presents a distribution more concentrated on the higher rating classes, while Model C presents a more orthodox distribution, with higher concentration on the middle ratings and lower weight on the extremes.

With the assumptions made, for the cluster methodology, Model B is the one that presents the less attractive rating system: it is not able to better discriminate between rating classes in terms of default frequency to a significant extent, and it assigns very high ratings too often. Models A and C rating systems have a similar discriminating power, although the rating distribution suggested by Model C is the one closer to what should be expected from a balanced portfolio. Thus, the empirical evidence seems to corroborate the hypothesis advanced in section 4.4, the weighting of the sample for the scoring model is helpful in order to identify coherent classes through a cluster methodology.

5.1.2 Historical / Mapping Methodology

The second methodology tested consists on defining the class boundaries in such a way that the resulting default frequencies for each class (after calibration) would approximate as best as possible a chosen benchmark. For this study, the benchmark is Moody's historical one-year default frequencies for corporate whole rating grades. Table 7 provides descriptive statistics for the Moody's ratings.

[30] The calibration value should be similar to the best estimate of the annual default ratio of the universe. For this study, it is estimated that this value should be equal to 1.5% for the non-financial private Portuguese firms.

Rating	Min	1st Quartile	Median	Mean	StDev	3rd Quartile	Max
Aaa	0	0	0	0	0	0	0
Aa	0	0	0	0.06	0.18	0	0.83
A	0	0	0	0.09	0.27	0	1.7
Baa	0	0	0	0.27	0.48	0.37	1.97
Ba	0	0	0.64	1.09	1.67	1.29	11.11
B	0	0.38	2.34	3.71	4.3	5.43	20.78
Caa-C	0	0	7.93	13.74	17.18	20.82	100
Investment-Grade	0	0	0	0.15	0.28	0.21	1.55
Speculative-Grade	0	0.59	1.75	2.7	3.04	3.52	15.39
All Corporate	0	0.18	0.67	1.1	1.38	1.32	8.4

Table 7 – Annual Global Issuer-Weighted Default Rate Descriptive Statistics, 1920-2003[31]

It is relevant to point out that this is not an attempt to create an alternative to Moody's ratings. The objective is to obtain a rating system whose default frequencies share some properties with an external reference. A downside of this mapping methodology is that implicitly we assume that our benchmark has the desired properties, and that the underlying structure of our population is similar to the one used to produce the benchmark statistics. The methodology is set up as an optimization problem that can be formalized as follows:

$$\min_{x_1,\ldots,x_{k-1}} \sum_{i=1}^{k} \left(y_i^b - y_i \right)^2$$

$$\text{subject to } \; y_i = \frac{d_i}{x_i} \;,\; x_i > 0 \;,\; d_i > 0 \;,\; \forall i$$

where y_i^b is the default frequency of the benchmark for class *i*, y_i is the default frequency of the model for class *i*, d_i is the number of default observations in class *i* and x_i is the number of observations in class *i*.

Figures 27 and 28 present the results of applying this methodology to the scoring models estimated previously.

[31] Source: Hamilton, *et al.* 2004.

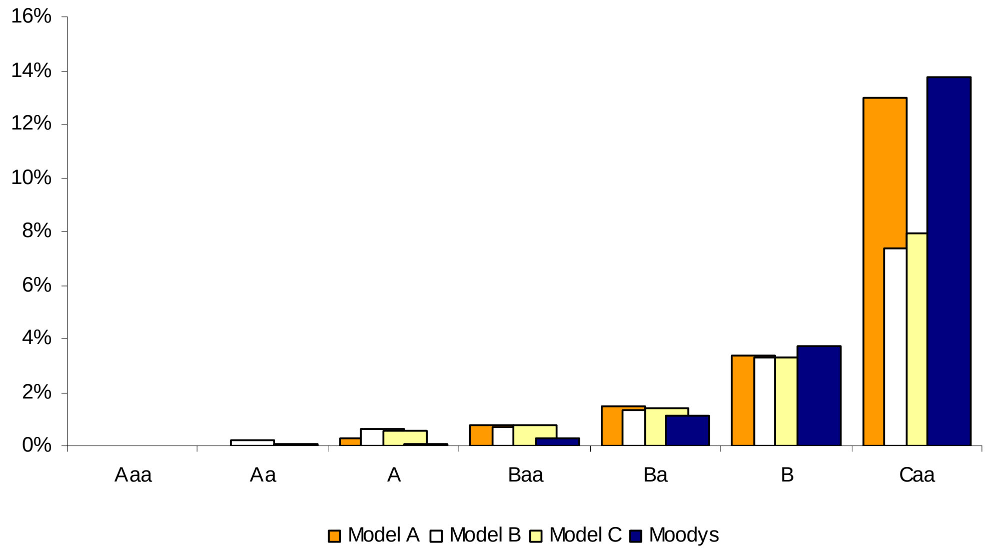

Figure 27 – Default Frequency by Rating Class (Historical Method)
This figure provides the default frequency for each rating class determined by the historical methodology. The default frequency corresponds to the proportion of default observations to the total number of observations in each rating class, adjusted by a calibration factor. The average default rate for Moody's whole letter ratings in the period 1920-2003 is also presented. The defaulted frequencies are higher for lower rating classes but the rate of increase is steeper than for the results reported under the cluster methodology. The three alternative specifications yield similar results.

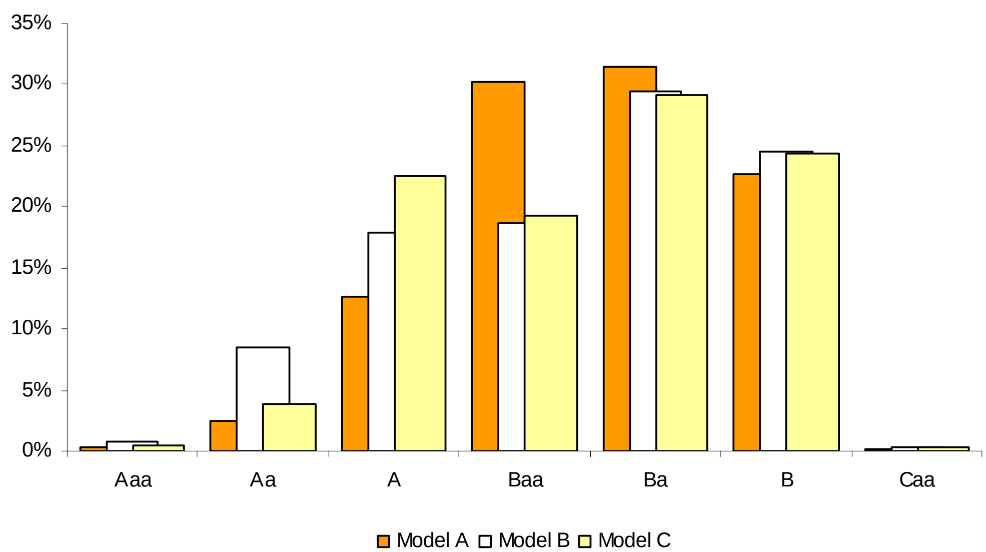

Figure 28 – Number of Observations Distribution by Rating Class (Historical Method)
The figure above displays the distribution of the number of observations by rating class, with the rating classes determined by the historical methodology. Comparing with the cluster methodology there is a much higher concentration on the middle classes and very few observations on the best and worst classes.

Figure 27 shows the default frequency by rating class for each model and selected benchmark. The default frequency presented are calibrated frequencies, the calibration is similar to the one described in the previous section. All three models can moderately approximate the benchmark, although only results for Model A provide a good fit for the default frequency in the lowest rating class. Even so, the results for the three models are clearly positive in terms of discriminatory power. When comparing the default frequencies between the two methodologies, it is clear that the historical methodology yields much steeper rating scales, starting at lower default rates for the higher rated classes, and ending at clearly higher default rates for the lower rated classes than the cluster methodology[32]. Consequently, the resulting distributions of observations for the rating systems based on the historical methodology (Figure 28, above) are less granular, with higher concentrations in the middle / lower classes. For all three models, only the very best firms belong to one of the two higher rating classes, and the worst class is reserved for the very worst performing firms. Comparing the distributions of observations by rating class based on the three scoring models, there are no clear differences between them.

5.1.3 Rating Matrixes and Stability

Once the optimal boundaries for each rating class are determined, a rating classification can be attributed for each observation of the dataset. Tracking the evolution of the yearly observations of each firm enables the construction of one-year transition matrixes. If, for example, a firm is classified as Baa in the fist period considered, in the next period it could either have an upgrade (to Aaa, Aa or A), a downgrade to (Ba, B, Caa), remain at Baa, default, or have no information in the dataset (Without Rating – WR).

The analysis of the transition matrix is helpful in order to study the stability of the rating system. The fewer transitions, i.e., low percentages in the off-diagonal elements of the matrix, the more stable is the rating system. Furthermore, transitions involving jumps of several notches (for example, a transition from Aaa to Caa) are

[32] The default rates for the higher rating class, resulting from the historical methodology, are 0% because historically there are no observed one-year defaults for the benchmark, in the period considered.

undesirable. Thus, a stable rating system is one whose rating transitions are concentrated in the vicinity of the main diagonal elements of the matrix.

Another relevant aspect of the transition matrix is the transition from each rating class to default. In terms of discriminatory power, a better rating system is one where the transitions to default rise at an exponential rate, from the higher rating to the lower rating classes.

Tables 8 – 10 present the transition matrices for the three models considered, with the class boundaries determined by the cluster methodology, while Tables 11 – 13 present the matrices based on the historical methodology:

	Aaa	**Aa**	**A**	**Baa**	**Ba**	**B**	**Caa**	**D**	**WR**
Aaa	41.15%	20.37%	4.71%	1.22%	1.06%	0.65%	0.00%	0.81%	30.03%
Aa	19.13%	29.82%	15.93%	6.11%	3.49%	1.16%	0.07%	2.98%	21.31%
A	7.15%	23.74%	25.84%	12.88%	8.25%	2.78%	0.08%	3.20%	16.08%
Baa	2.31%	14.08%	19.47%	19.25%	17.60%	6.93%	0.33%	5.39%	14.63%
Ba	1.21%	4.93%	10.85%	16.70%	29.69%	15.20%	0.43%	5.92%	15.06%
B	0.37%	1.76%	2.61%	4.27%	18.30%	42.96%	2.40%	11.85%	15.47%
Caa	0.00%	0.95%	0.95%	0.00%	3.81%	34.29%	9.52%	25.71%	24.76%

Table 8 – Model A - 1 Year Transition Matrix (Cluster Method)

	Aaa	**Aa**	**A**	**Baa**	**Ba**	**B**	**Caa**	**D**	**WR**
Aaa	42.43%	23.57%	1.72%	0.36%	0.27%	0.00%	0.09%	0.63%	30.92%
Aa	13.42%	46.57%	14.09%	3.32%	0.38%	0.21%	0.04%	3.20%	18.76%
A	1.58%	26.52%	32.79%	14.78%	2.19%	0.43%	0.43%	4.81%	16.48%
Baa	0.46%	8.27%	24.22%	33.40%	8.27%	2.99%	1.04%	7.55%	13.80%
Ba	0.14%	2.32%	9.71%	29.28%	21.16%	8.12%	3.77%	10.72%	14.78%
B	0.00%	1.89%	2.95%	14.74%	21.47%	21.05%	7.58%	13.05%	17.26%
Caa	0.00%	1.54%	0.00%	7.34%	11.97%	16.22%	18.15%	21.62%	23.17%

Table 9 – Model B - 1 Year Transition Matrix (Cluster Method)

	Aaa	**Aa**	**A**	**Baa**	**Ba**	**B**	**Caa**	**D**	**WR**
Aaa	26.67%	28.80%	7.20%	0.80%	0.00%	0.27%	0.00%	0.53%	35.73%
Aa	7.10%	41.27%	19.01%	3.85%	0.74%	0.44%	0.07%	1.63%	25.89%
A	1.15%	18.70%	40.28%	15.40%	3.72%	0.68%	0.10%	3.20%	16.76%
Baa	0.25%	3.84%	23.96%	32.14%	13.96%	2.89%	0.50%	5.41%	17.04%
Ba	0.07%	1.25%	9.42%	24.50%	29.29%	13.10%	1.62%	7.28%	13.47%
B	0.09%	0.35%	2.26%	9.03%	23.00%	32.38%	6.34%	11.37%	15.19%
Caa	0.00%	0.00%	1.15%	0.29%	5.48%	26.51%	23.34%	19.88%	23.34%

Table 10 – Model C - 1 Year Transition Matrix (Cluster Method)

	Aaa	Aa	A	Baa	Ba	B	Caa	D	WR
Aaa	13.33%	26.67%	16.67%	6.67%	0.00%	0.00%	0.00%	0.00%	36.67%
Aa	1.44%	14.83%	25.84%	15.31%	0.48%	0.96%	0.00%	0.00%	41.15%
A	0.40%	5.23%	34.97%	26.83%	3.52%	0.60%	0.00%	1.01%	27.44%
Baa	0.04%	0.80%	13.43%	45.59%	16.65%	1.69%	0.00%	2.96%	18.84%
Ba	0.04%	0.16%	1.65%	21.18%	44.64%	11.38%	0.04%	5.69%	15.22%
B	0.00%	0.10%	0.26%	3.64%	22.44%	44.83%	0.26%	12.47%	16.00%
Caa	0.00%	0.00%	0.00%	0.00%	0.00%	22.22%	0.00%	55.56%	22.22%

Table 11 – Model A - 1 Year Transition Matrix (Historical Method)

	Aaa	Aa	A	Baa	Ba	B	Caa	D	WR
Aaa	18.97%	36.21%	12.07%	0.00%	0.00%	0.00%	0.00%	0.00%	32.76%
Aa	2.85%	32.43%	24.32%	6.16%	1.05%	0.30%	0.00%	0.75%	32.13%
A	0.07%	12.79%	35.91%	16.80%	6.75%	0.98%	0.00%	2.46%	24.24%
Baa	0.00%	3.44%	22.66%	31.96%	21.35%	2.00%	0.00%	2.96%	15.63%
Ba	0.00%	0.42%	5.65%	17.58%	43.55%	11.80%	0.04%	5.31%	15.64%
B	0.00%	0.05%	0.76%	2.48%	20.88%	47.44%	0.43%	12.18%	15.77%
Caa	0.00%	0.00%	0.00%	0.00%	5.00%	25.00%	0.00%	30.00%	40.00%

Table 12 – Model B - 1 Year Transition Matrix (Historical Method)

	Aaa	Aa	A	Baa	Ba	B	Caa	D	WR
Aaa	18.92%	29.73%	16.22%	0.00%	0.00%	0.00%	0.00%	0.00%	35.14%
Aa	2.24%	21.41%	35.14%	3.83%	0.96%	0.32%	0.00%	0.00%	36.10%
A	0.22%	5.17%	46.79%	14.62%	5.29%	0.84%	0.00%	2.08%	24.97%
Baa	0.00%	1.13%	24.27%	33.36%	20.49%	2.06%	0.00%	3.18%	15.52%
Ba	0.00%	0.21%	5.42%	18.13%	43.13%	11.77%	0.04%	5.33%	15.96%
B	0.00%	0.05%	0.81%	2.49%	20.90%	47.51%	0.34%	12.13%	15.77%
Caa	0.00%	0.00%	0.00%	0.00%	0.00%	22.22%	0.00%	38.89%	38.89%

Table 13 – Model C - 1 Year Transition Matrix (Historical Method)

Results based on the historical methodology are more stable and display higher discriminatory power than the results based on the cluster methodology. In terms of stability, the historical based results have less high level transitions. For example, none of the three matrixes based on this methodology have transitions from the high classes Aa, A, Baa to lowest class Caa, while all of the three matrices based on the cluster methodology have such transitions.

In terms of discriminatory power, the matrixes based on the historical methodology also present better results, since the transitions to default start at lower percentages for the higher classes and increase continuously to considerable higher percentages than the transitions based on the cluster methodology.

Regarding the results for each model, within each methodology, none of them produces a clearly more attractive rating matrix.

5.2 Regulatory Capital Requirements

Under the New Basel Capital Accord (NBCA), financial institutions will be able to use their internal risk assessments in order to determine the regulatory capital requirements[33]. In the first pillar of the Accord – Minimum Capital Requirements – two broad methodologies for calculating capital requirements for credit risk are proposed. The first, the Standardized Approach, is similar to the current capital accord, where the regulatory capital requirements are independent of the internal assessment of the risk components of the financial institutions. Conversely, in the second methodology – the Internal Ratings-Based Approach – banks complying with certain minimum requirements can rely on internal estimates of risk components in order to determine the capital requirements for a given exposure. Under this methodology, two approaches are available: a Foundation and an Advanced approach. For the Foundation Approach, credit institutions will be able to use their own estimates of the PD but rely on supervisory estimates for the other risk components. For the Advanced Approach, banks will be able to use internal estimates for all risk components, namely the PD, Loss-Given-Defaults (LGD), Exposure-At-Default (EAD) and Maturity (M). These risk components are transformed into Risk Weighted Assets (RWA) through the use of risk weight functions[34].

Up to this point we have devised six alternative methodologies for determining one of the risk components, the PD. Assuming fixed estimates for the other risk components we are able to estimate capital requirements under the IRB Foundation approach, and compare them to the capital requirements under the current accord. The parameters assumed are LGD = 45%, M = 3 years, EAD for SME = 0.3 Million Eur and EAD for large firms = 1.5 Million Eur. The PD used for firm i corresponds to the maximum PD estimated for the rating class where i belongs. For the calculations under the current capital accord, it is considered that all exposures have the standard risk weight of 100%. Table 14 provides results for all six models.

[33] Basel Committee on Banking Supervision (2003).

[34] Appendix 6 provides a description of the formulas used to compute the RWA for corporate exposures.

	Model	Average RWA %	Capital Reqts (Eur)		Capital Difference
			Basel II - IRB Found.	Basel I / Basel II Standardized	
Historical	**Multiple Industry Equations (A)**	90.95%	272,396,629.19	299,496,000.00	27,099,370.81
	Single Equation, Unweighted Sample (B)	91.45%	273,885,689.68	299,496,000.00	25,610,310.32
	Weighted Sample (C)	91.59%	274,315,296.00	299,496,000.00	25,180,704.00
Cluster	**Multiple Industry Equations (A)**	96.24%	288,220,548.59	299,496,000.00	11,275,451.41
	Single Equation, Unweighted Sample (B)	91.11%	272,881,107.86	299,496,000.00	26,614,892.14
	Weighted Sample (C)	91.79%	274,906,654.20	299,496,000.00	24,589,345.80

Table 14 – Average RWA and Total Capital Requirements

The table above provides the average risk weighted assets for a standard portfolio using the rating classifications obtained through the three scoring model specifications and for each rating methodology. In addition, capital requirements under the current capital accord and the NBCA (IRB-Foundation) are also calculated and compared. For all six alternatives the capital requirements under the IRB-Foundation rules are lower than under Basel I legislation.

Results are similar for all models, the capital requirements under the IRB Foundation approach are lower than those that would be required under the current capital accord. For the Historical rating methodology, the two-equation scoring specification (Model A) is the one that provides the highest capital relief, but for the Cluster rating methodology it is the one that provides the lowest.

Figures 29 – 34 provide the distribution if the relative RWA for each rating class of all six methodologies, weighted by the number of observations attributed to each class by each rating methodology:

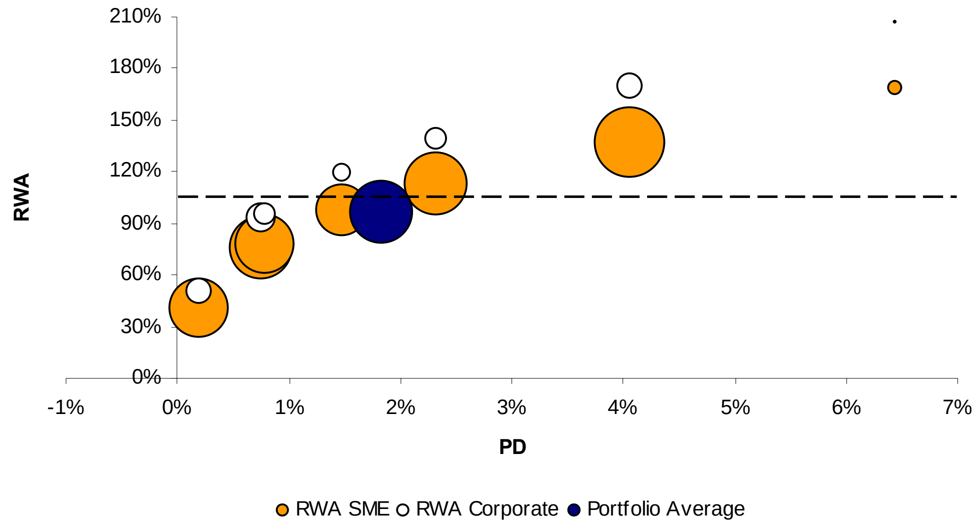

Figure 29 – Model A - IRB Capital Requirements (Cluster Method)

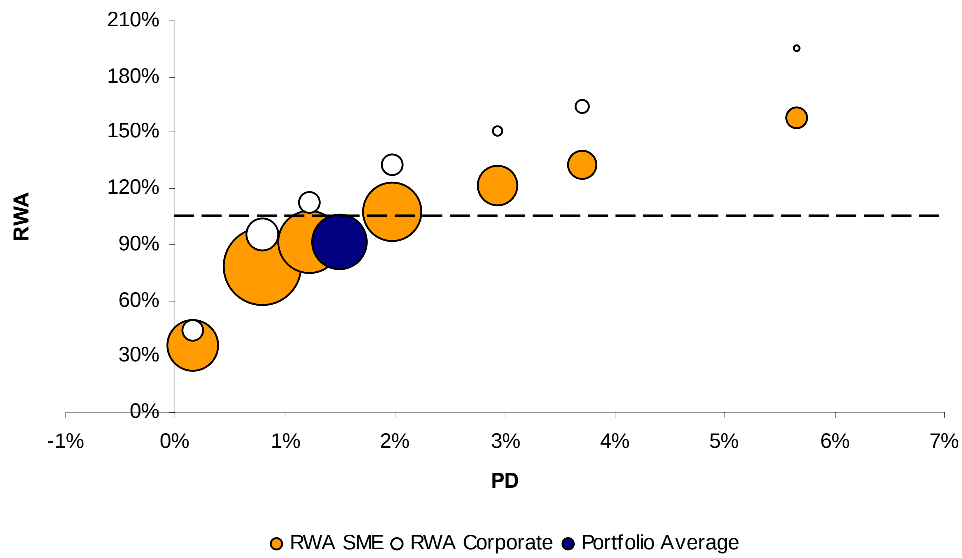

Figure 30 – Model B - IRB Capital Requirements (Cluster Method)

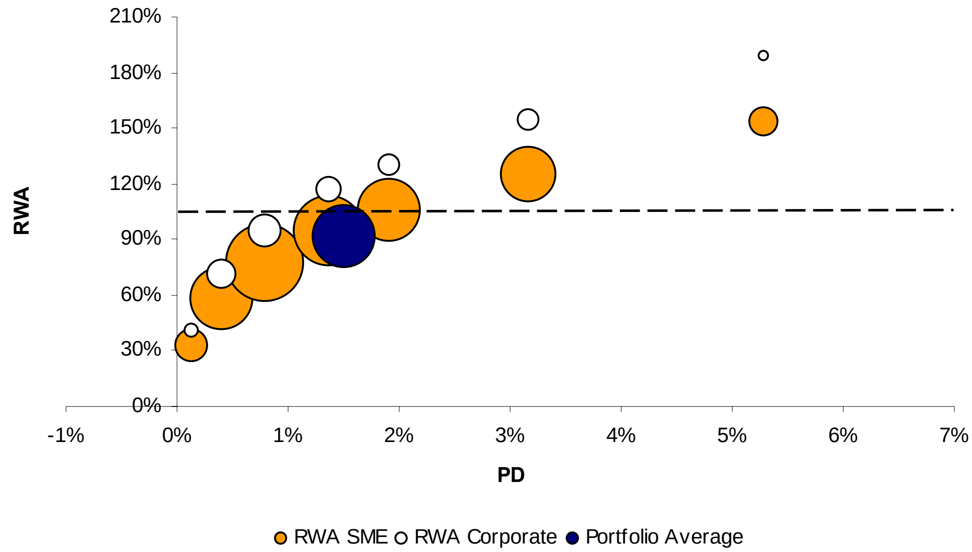

Figure 31 – Model C - IRB Capital Requirements (Cluster Method)

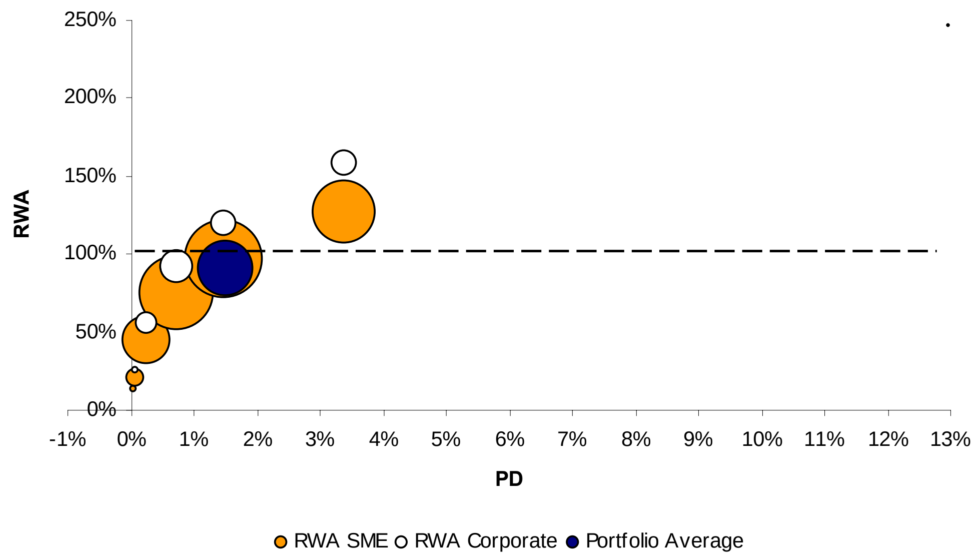

Figure 32 – Model A - IRB Capital Requirements (Historical Method)

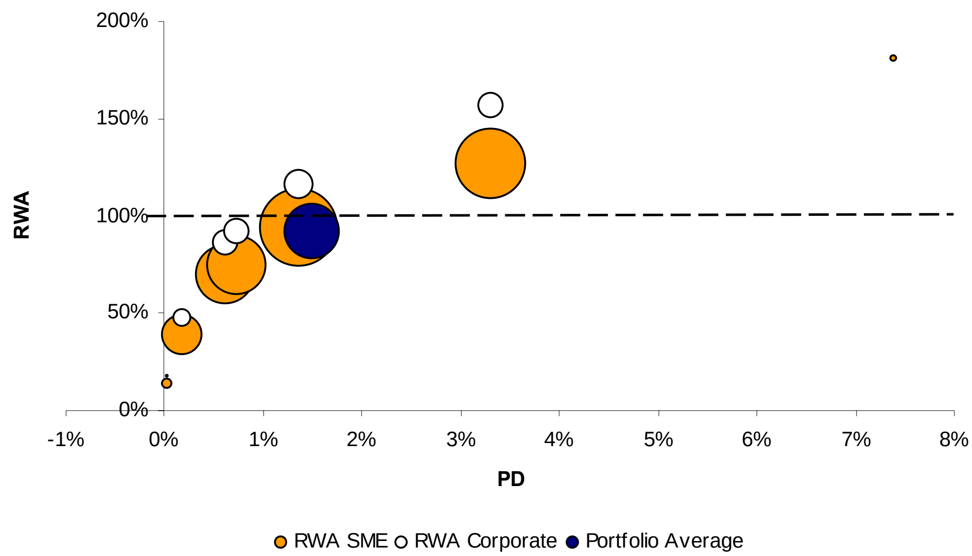

Figure 33 – Model B - IRB Capital Requirements (Historical Method)

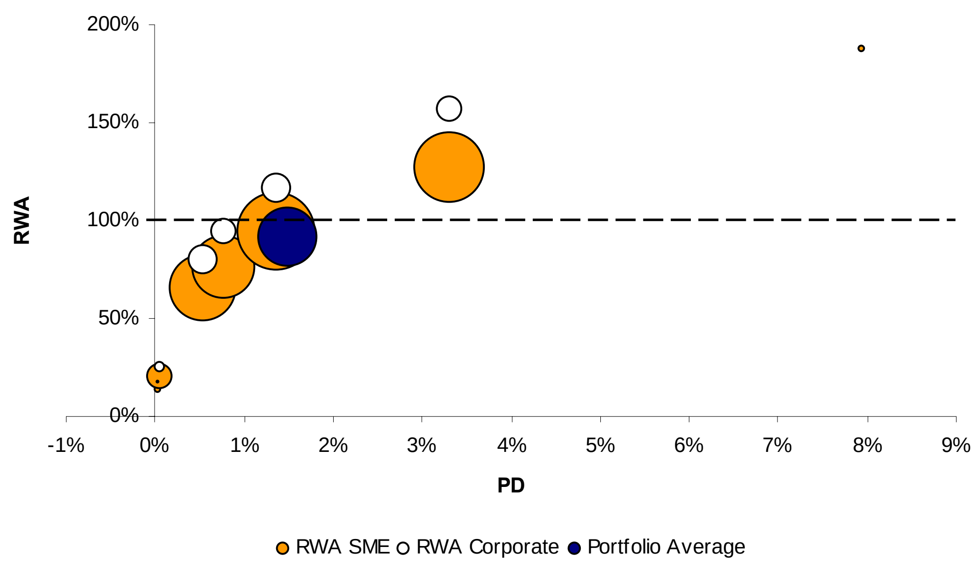

Figure 34 – Model C - IRB Capital Requirements (Historical Method)

The results based on the Historical Methodology are more concentrated on the middle classes, and typically only the two lowest rating classes have a risk weight above the standard Basel I weight. Results under the Cluster Methodology are more evenly spread out through the different classes, with the three up to five lowest rating classes having a risk weight above Basel I requirements.

6 Conclusion

The first and main result from this research is that it is possible to build a relatively simple but powerful and intuitive rating system for privately-held corporate firms, with few data requirements. In order to set up a similar system, it is only necessary to retrieve for a given time frame (at very least 4 years, better would be a full economic cycle) yearly default data and the accounting reports used to concede these loans. This purely quantitative system is enough to provide a scoring rule that, for this dataset, is able to discriminate to a very satisfactory extent the defaulting and non-defaulting populations, both in and out-of-sample. It is also capable of classifying the various firms into meaningful and coherent rating classes. Meaningful in the sense that firms belonging to a certain rating class have distinct probabilities of default from firms belonging to other classes, and to lower ratings correspond significantly higher probabilities of default. Coherent in the sense that rating transitions are stable: if a firm has a given rating for a given year, the probability that in the following period it would be either upgraded or downgraded several notches is very small. Furthermore, the probabilities of default associated to each rating class are calibrated to the estimated real average default frequency of the portfolio, and can therefore be used to access the potential impact of introducing the IRB – Foundation approach of the NBCA, for a given portfolio.

In terms of the scoring methodology, two alternatives to the classical regression are presented. The first alternative is a two-equation specification that allows for industry differentiation. The second is a weighted model that balances the proportion of defaulting and non-defaulting observations. In terms of the discriminating power of the scoring model, both in-sample and out-of-sample results indicate that neither of the two alternative specifications provide significant improvement to the classical regression. However, both alternatives have proven useful later when building the rating classes. The weighted model provides the best results when using a cluster methodology to group individual observations into rating classes, while the two-equation specification provides the most discriminating system when rating classes are built through a mapping methodology. Comparing the two ranting methodologies, the mapping methodology yields more discriminating systems but, on the other hand, the cluster methodology provides more granular rating distributions. Regarding the rating matrixes, the mapping methodology provides more discriminatory power with considerably less ample rating transitions.

There are, however, important extensions to the basic setup that should be considered. The first one derives from the fact that the scoring model only considers a subset of all the variables that can potentially help to discriminate the defaulting and non-defaulting populations. A more complete setup would then consider alternative explanatory variables (such as the reputation of management, the quality of the accounting reports or the relationship of the client to the bank), but more importantly, it should incorporate the subjective opinion or expertise of the credit analyst. A desirable feature of a rating system is giving the possibility for the credit analyst to override the rating decision provided by the mechanical score. This is particularly relevant in the corporate segment, since a wide array of idiosyncrasies (such as creative accounting) could distort the results of the quantitative assessment.

Another potentially useful extension would be to develop a system that provides ratings based not just on the most current available information, but also on the information available on the previous periods. This would result in a more stable system: for a firm to have a very good / bad classification, it would have to present very good / bad indicators for several periods. There is however, a trade-off between stability and discriminatory power: for example, if a firm has in the past produced consistently good indicators, but in the present is rapidly becoming on the verge of bankruptcy, such a system may not downgrade the rating classification of such a firm fast enough.

One final point worth mentioning is that the system developed only provides borrower ratings. In order to use such a system to concede loans, the variables specific to each loan (such as collateral) should be taken in consideration together with the borrower rating. In other words, the final rating assigned to a certain loan is a function of the borrower rating and the Loss-Given-Default (LGD).

Bibliography

Allen, L. (2002) "Credit Risk Modelling of Middle Markets." *presented at Conference on Credit Risk Modelling and Decisioning, Wharton FIC, University of Pennsylvania*.

Altman, E. (2000) "Predicting Financial Distress Of Companies: Revisiting The Z-Score And Zeta® Models." *Working Paper, Stern School of Business, New York University*.

Altman, E. (1968) "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance, Sep, 589–609*.

Altman, E., G. Marco, and F. Varetto (1994) "Corporate Distress Diagnosis: Comparison Using Linear Discriminant Analysis and Neural Networks (the Italian Experience)." *Journal of Banking and Finance, 18, 505-529*.

Bamber, D. (1975) "The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Operating Graph" *Journal of Mathematical Psychology 12, 387–415*.

Braga, A. (2000) "Curvas Roc: Aspectos Funcionais e Aplicações." *PhD Thesis, Universisdade do Minho, Portugal*.

BarNiv, R. and J. McDonald (1999) "Review of Categorical Models for Classification Issues in Accounting and Finance" *Review of Quantitative Finance and Accounting, Vol. 13, 39-62*.

Basel Committee on Banking Supervision (2003) "The New Basel Capital Accord." *Bank for International Settlements Consultative Document, April*.

Beaver, W. (1966) “Financial ratios as predictors of bankruptcy.” *Journal of Accounting Research (Supplement), 71–102*.

Black, F. and J. Cox (1976) “Valuing corporate securities: Some effects of bond indenture provisions.” *Journal of Finance, 351–367*.

Black, F. and M. Scholes (1973) “The pricing of options and corporate liabilities.” *Journal of Political Economy 81, 637–54*.

Box, G. and P. Tidwell (1962) "Transformation of independent variables", *Technometrics, 4, 531-550*.

Boyes, W., D. Hoffman and S. Low (1989) “An econometric analysis of the bank credit scoring problem.” *Journal of Econometrics 40, 3–14*.

Caiazza, S. (2004) “The comparative performance of credit scoring models: an empirical approach.” *Monetary Integration, Markets and Regulation, Research in Banking and Finance, 4, Elsevier Science Ltd, forthcoming*.

Calinski, T. and J. Harabasz (1974) “A dendrite method for cluster analysis.” *Communications in Statistics, 3, 1–27*.

Cantor, R. and F. Packer (1994) “The Credit Rating Industry.” *Federal Reserve Bank of New York Quarterly Review, Summer/Fall*.

Cleveland, W. (1979) “Robust locally weighted regression and smoothing scatterplots.” *Journal of the American Statistical Association 74, 829-836*.

CMVM (2005) *Relatórios de Auditoria às Contas das Sociedades com Valores Mobiliários Cotados*. Comissão do Mercado de Valores Mobiliários, Lisbon.

Conover, W. (1999) *Practical Nonparametric Statistics. 3rd Edition.* John Wiley & Sons, New-York.

Crook, J. and J. Banasik (2004) “Does reject inference really improve the performance of application scoring models?” *Journal of Banking & Finance 28, 857–874*.

Crouhy, M., D. Galai and R. Mark (2001) “Prototype risk rating system.” *Journal of Banking & Finance 25, 47-95*.

DeLong, E., D. DeLong, and D. Clarke-Pearson (1988) “Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach” *Biometrics 44, 837–845*.

Duffie, D. and K. Singleton (1997) “An econometric model of the term structure of interest-rate swap yields.” *The Journal of Finance 52(4), 1287–1322*.

Dwyer, D., A. Kocagil and R. Stein (2004) “Moody’s Kmv Riskcalc™ v3.1 Model.” Moody’s KMV Company.

Engelmann, B., E. Hayden and D. Tasche (2003) “Measuring the Discriminative Power of Rating Systems.” *Deutsche Bundesbank Discussion paper Series 2: Banking and Financial Supervision No 01/2003*.

Galindo, J. and P. Tamayo (2000) “Credit risk assesment using statistical and machine learning: basic methodology and risk modelling applications.” *Computational Economics, 15, 107-143*.

Hamilton, D., *et. al.* (2004), "Default and recovery rates of corporate bond issuers," *Moody's Special Comment, Jannuary*.

Hartigan, J. (1975) *Clustering Algorithms*. John Wiley and Sons, New-York.

Hayden, E. (2003) “Are Credit Scoring Models Sensitive With Respect to Default Definitions? Evidence from the Austrian Market” *Working Paper, Department of Business Administration, University of Vienna*.

Hosmer, D. and S. Lemeshow (2000) *Applied Logistic Regression. 2nd Edition.* John Wiley & Sons, New-York.

INE (2003) *Anuário Estatístico de Portugal 2003*. Instituto Nacional de Estatística, Lisbon.

Jacobson, T. and K. Roszbach (2003) "Bank lending policy, credit scoring and value-at-risk." *Journal of Banking & Finance 27, 615-633*.

Jarrow, R. and S. Turnbull (1995) "Pricing derivatives on financial securities subject to credit risk." *Journal of Finance 50, 53–85*.

Kraft, H., G. Kroisandt and M. Müller (2004) "Redesigning Ratings: Assessing the Discriminatory Power of Credit Scores under Censoring." *Working Paper, Fraunhofer Institut für Techno- und Wirtschaftsmathematik*.

Laitinen, E. (1999) "Predicting a corporate credit analyst's risk estimate by logistic and linear models" *International Review of Financial Analysis, 8, 97-121*.

Longstaff, F. and E. Schwartz (1993) "A simple approach to valuing risky fixed and floating rate debt." *Working Paper 22–93, Anderson Graduate School of Management, University of California*.

Mann, H. and D. Whitney (1947) "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other." *Annals of Mathematical Statistics 18, 50–60*.

Menard, S. (2002) *Applied logistic regression analysis, 2nd Edition.* Thousand Oaks, CA: Sage Publications. Series: Quantitative Applications in the Social Sciences, No. 106.

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

Metz, C. (1978) "Basic Principles of ROC Analysis." *Seminars in Nuclear Medicine, VIII (4), 283–298*.

Murphy, A., A. Kocagil, P. Escott and F. Glormann (2002) "Moody's RiskCalc™ For Private Companies: Portugal." *Moody's Investors Service Rating Methodology*.

Milligan, G. and M. Cooper (1985) "An examination of procedures for determining the number of clusters in a data set." *Psychometrika, 50, 159-179*.

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

Platt, H. and M. Platt (1990) "Development of a class of stable predictive variables: the case of bankruptcy prediction" *Journal of Business Finance and Accounting, 17, 31-51*.

Royston, P. and D. Altman (1994) "Regression using fractional polynomials of continuous covariates: Parsimonious parametric modeling." *Applied Statistics, 43, 429–467*.

Sobehart, J. and S. Keenan (2001) "Measuring default accurately." *Risk – Credit Risk Special Report, March, S31–S33*.

Yang, Z., M. Platt and H. Platt (1999) "Probabilistic Neural Networks in Bankruptcy Prediction." *Journal of Business Research, Feb., 67-74*.

Appendix 1 – Description of Financial Ratios and Accuracy Ratios

Type	Name	Definition	Expected Effect on PD	AR
Profitability	**Net P&L / Assets**	[Net Profit & Loss] / [Total Assets]	-	15.60%
	Current Earnings / Assets	[Current Earnings] / [Total Assets]	-	16.60%
	Current Earnings and Depreciation / Turnover	[Current Earnings + Depreciation] / [Turnover]	-	-1.00%
	P&L / Assets	[Net Profit & Loss + Depreciation + Provisions] / [Total Assets]	-	5.80%
	Gross Earnings / Production	[EBT + Depreciation + Provisions] / [Production]	-	-3.20%
	EBITDA / Production	[EBITDA] / [Production]	-	-15.80%
Liquidity	**Liquidity / Current Liabilities**	[Bank Deposits & Cash + Marketable Securities] / [Short-Term Liabilities]	-	10.60%
	Current Ratio	[Current Assets] / [Short-Term Liabilities]	-	9.20%
	Liquidity / Assets	[Bank Deposits & Cash + Marketable Securities] / [Total Assets]	-	12.00%
Leverage / Gearing	**Equity / Assets**	[Equity] / [Total Assets]	-	3.80%
	Equity / Accounts Payable	[Equity] / [Accounts Payable]	-	4.40%
	Bank Debt / Accounts Payable	[Bank Debt] / [Accounts Payable]	+	1.80%
	Accounts Payable / Assets	[Accounts Payable] / [Total Assets]	+	5.20%
	Liabilities / Assets	[Total Liabilities] / [Total Assets]	+	3.80%
	Net Current Accounts Payable / Assets	[Short-Term Accounts Payable - Bank Deposits & Cash] / [Total Assets]	+	-0.40%
Debt Coverage	**Gross Earnings / Liabilities**	[Current Earnings + Depreciation + Provisions] / [Total Liabilities]	-	9.80%
	Debt Service Coverage	[Current Earnings + Depreciation] / [Interest & Similar Costs]	-	25.20%
	P&L / L-T Liabilities	[Net Profit & Loss + Depreciation + Provisions] / [Long-Term Liabilities]	-	0.00%
	Operating Earnings / Debt Service	[Operating Earnings] / [Interest & Similar Costs]	-	16.60%
Activity	**Interest Costs / Sales**	[Interest & Similar Costs] / [Turnover]	+	40.20%
	Inventories / Turnover	[Inventories] / [Turnover]	+	12.00%
	Turnover / Assets	[Turnover] / [Total Assets]	-	20.40%
Productivity	**Productivity Ratio**	[Personnel Costs] / [Turnover]	+	15.80%

Note: Turnover = Total Sales + Services Rendered

Appendix 2 – Estimating and Comparing the Area Under the ROC curves

The estimated ROC curve and, consequently, the AUROC are outcomes of random variables, since we only have one sample of the scoring of the borrowers and their realized defaults. Following DeLong et al. (1988), the area under the population ROC curve can be defined has the probability that, when the estimated scoring is observed for a randomly selected borrower from the default population and a randomly selected borrower from the non-default population, the resulting scores will be in the correct order (the scoring of the default observation is higher than the scoring of the regular observation). For a given sample, the AUROC can be estimated either trough parametric or nonparametric methods. A parametric approach would involve distributional assumptions on the observed variable, although these distributions cannot be uniquely determined from the ROC curve (see, for example, the binormal model used in Metz 1978). The nonparametric approach used in this study relates the estimation of the AUROC to the Mann-Whitney (1947) U-statistic[35]. Let d_i ($i = 1,\dots, m$) be the estimated scores for the default observations and r_j ($j = 1,\dots, n$) be the estimated scores for the regular observations. An unbiased estimator of the probability of correctly classifying two randomly chosen subjects from the default and regular populations is given by the average over a kernel ψ:

$$\widehat{AUROC} = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}\psi\left(d_i, r_j\right)$$

where,

$$\psi\left(d,r\right) = \begin{cases} 1 & d > r \\ \tfrac{1}{2} & d = r \\ 0 & d < r \end{cases}$$

The variance of this estimator can be computed through the use of placement values. Let $V(d_i)$ be the placement of the estimated score d_i in the distribution of r scores (i.e.,

[35] Bamber (1975) originally developed this relationship. For more details see, for example, Braga (2000).

the fraction of r scores that it exceeds). In addition, let $V(r_j)$ be the placement of the estimated score r_j in the distribution of d scores:

$$V(d_i) = \frac{\sum_{j=1}^{n} \psi(d_i, r_j)}{n} \quad \text{and} \quad V(r_j) = \frac{\sum_{i=1}^{m} \psi(d_i, r_j)}{m}$$

The variance of the estimator for large samples can then be computed as the sum of the scaled variances for the placement values of d and r:

$$\operatorname{var}\left(\widehat{AUROC}\right) = \frac{m\sum_{j=1}^{m} V(r_j)^2 - \left[\sum_{j=1}^{m} V(r_j)\right]^2}{m^2(m-1)} + \frac{n\sum_{i=1}^{n} V(d_i)^2 - \left[\sum_{i=1}^{n} V(d_i)\right]^2}{n^2(n-1)}$$

If we wish to build a test to compare the AUROC estimates for two alternative models, A and B based on the same dataset it is also relevant to compute the covariance of the estimates:

$$\operatorname{cov}\left(\widehat{AUROC_A}, \widehat{AUROC_B}\right) = \frac{\sum_{j=1}^{m}\left[V(r_{Aj}) - \overline{V}(r_A)\right]\left[V(r_{Bj}) - \overline{V}(r_B)\right]}{m(m-1)} + \\ + \frac{\sum_{i=1}^{n}\left[V(d_{Ai}) - \overline{V}(d_A)\right]\left[V(d_{Bi}) - \overline{V}(d_B)\right]}{n(n-1)}$$

The test statistic for testing $H_0 : \widehat{AUROC_A} = \widehat{AUROC_B}$ is given by:

$$T = \frac{\left(\widehat{AUROC_A} - \widehat{AUROC_B}\right)^2}{\operatorname{var}\left(\widehat{AUROC_A} - \widehat{AUROC_B}\right)}$$

where,

$$\operatorname{var}\left(\widehat{AUROC_A} - \widehat{AUROC_B}\right) = \operatorname{var}\left(\widehat{AUROC_A}\right) + \operatorname{var}\left(\widehat{AUROC_B}\right) - 2\operatorname{cov}\left(\widehat{AUROC_A}, \widehat{AUROC_B}\right)$$

The test statistic T is asymptotically χ^2-distributed with one degree of freedom.

Appendix 3 – Binomial Logistic Regression Estimation and Diagnostics[36]

a) Binomial Logistic Regression

Binomial (or binary) logistic regression is a type of regression useful to model relationships where the dependent variable is dichotomous (only assumes two values) and the independent variables are of any type. Logistic regression estimates the probability of a certain event occurring, since it applies maximum likelihood estimation after transforming the dependent variable into a logit variable (the natural log of the odds of the dependent occurring or not). Unlike OLS regression, it estimates changes in the log odds of the dependent variable, not changes in the dependent itself.

Let y_i be a binary discrete variable that indicates whether firm i has defaulted or not in a given period of time, and let x_i^k represent the values of the k explanatory variables for the firm i. The conditional probability that firm i defaults is given by $P\left(y_i = 1 \mid x_i^k\right) = \pi\left(x_i^k\right)$, while the conditional probability that the firm does not default is given by $P\left(y_i = 0 \mid x_i^k\right) = 1 - \pi\left(x_i^k\right)$. Thus, the odds that this firm defaults is simply: $odds_i = \pi\left(x_i^k\right) \big/ 1 - \pi\left(x_i^k\right)$. The estimated regression relates a combination of the independent variables to the natural log of the odds of the dependent outcome occurring:

$$g\left(x, \beta\right) = \ln\left[\frac{\pi\left(x\right)}{1 - \pi\left(x\right)}\right] = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

or,

$$\pi\left(x\right) = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)}$$

[36] This Appendix is based on Menard (2002) and Hosmer and Lemeshow (2000).

<u>Assumptions</u>

i. Each y_i follows a Bernoulli distribution with parameter $\pi\left(x_i^k\right)$. Which is equivalent to saying that each y_i follows a Binomial distribution with 1 trial and parameter $\pi\left(x_i^k\right)$;

ii. The error terms are independent;

iii. No relevant variables are omitted, no irrelevant variables are included, and the functional form is correct;

iv. There is a linear relationship between the logit of the independent variables and the dependent;

v. There is no significant correlation between the independent variables (no multicolinearity).

<u>Estimation</u>

Estimation of the binomial logistic regression is made through the maximum likelihood methodology. The expression of the likelihood function of a single observation is given by:

$$l_i = \pi\left(x_i\right)^{y_i}\left[1-\pi\left(x_i\right)\right]^{1-y_i}$$

Since independence between the observations is assumed, the likelihood function will be the product of all individual likelihoods:

$$l(\beta) = \prod_{i=1}^{n} \pi\left(x_i\right)^{y_i}\left[1-\pi\left(x_i\right)\right]^{1-y_i}$$

The log-likelihood function to be maximized will be:

$$L(\beta) = \ln\left[l(\beta)\right] = \sum_{i=1}^{n}\left\{y_i \ln\left[\pi\left(x_i\right)\right] + \left(1-y_i\right)\ln\left[1-\pi\left(x_i\right)\right]\right\}$$

The ML estimators correspond to the values of β that maximize the previous expression.

b) Residual Analysis

For the logistic regression, the residuals in terms of probabilities are given by the difference between the observed and predicted probabilities that default occurs:

$$e_i = P(y_i = 1) - \hat{P}(y_i = 1) = \pi(x_i) - \hat{\pi}(x_i)$$

Since these errors are not independent of the conditional mean of *y*, it is useful to adjust them by their standard errors, obtaining the Pearson or Standardized residuals:

$$r_i = \frac{\pi(x_i) - \hat{\pi}(x_i)}{\sqrt{\hat{\pi}(x_i)\left[1 - \hat{\pi}(x_i)\right]}}$$

These standardized residuals follow an asymptotically standard normal distribution. Cases that have a very high absolute value are cases for which the model fits poorly and should be inspected.

In order to detect cases that may have a large influence on the estimated parameters of the regression, both the Studentized residuals and the Dbeta statistic are used. The studentized residual corresponds to the square root of the change in the -2 Log-Likelihood of the model attributable to deleting the case from the analysis:

$$s_i = \sqrt{d_i^2 - \frac{r_i^2 h_i}{1 - h_i}}$$

The dbeta is an indicator of the standardized change in the regression estimates obtained by deleting an individual observation:

$$dbeta_i = \frac{r_i^2 h_i}{(1 - h_i)^2}$$

In the previous two expressions, h_i corresponds to the leverage statistic and d_i to the deviance residual. The leverage statistic is derived from the regression that expresses the predicted value of the dependent variable for case *i* as a function of the observed values of the dependent for all cases (for more information see Hosmer and Lemeshow (2000), 168-171). The deviance residual corresponds to the contribution of each case to the -2 Log-Likelihood function (the deviance of the regression).

c) Testing Coefficient Significance: the Wald Chi-Square Test

For the purpose of testing the statistical significance of the individual coefficients, the Wald Chi-Square test is implemented. Under the hypothesis that $\beta_i = 0$, the test statistic bellow follows a chi-square distribution with one degree of freedom:

$$W_i = \frac{\hat{\beta}_i^2}{\widehat{SE}\left(\hat{\beta}_i\right)^2}$$

d) Testing Regression Significance: the Hosmer & Lemeshow Test

In order to evaluate how effectively the estimated model describes the dependent variable the Hosmer & Lemeshow goodness-of-fit test is applied. The test consists in dividing the ranked predicted probabilities into deciles (*g=10* groups) and then computing a Pearson chi-square statistic that compares the predicted to the observed frequencies in a 2x10 contingency table. Let o_i^0 be the observed count of non-defaults for group *i* and p_i^0 be the predicted count. Similarly, let o_i^1 be the observed count of defaults for group *i* and p_i^1 be the predicted count. Then the *HL* test statistic following a chi-square distribution with *g-2* degrees of freedom is:

$$HL = \sum_{i=1}^{g} \left[\frac{\left(o_i^0 - p_i^0\right)^2}{p_i^0} + \frac{\left(o_i^1 - p_i^1\right)^2}{p_i^1} \right]$$

Lower values of *HL*, and non-significance indicate a good fit to the data and, therefore, good overall model fit.

e) Testing for Non-Linear Relationships: the Box-Tidwell Test

If the assumption of linearity in the logit is violated, then logistic regression will underestimate the degree of relationship of the independents to the dependent and will lack power, thus generating Type II errors (assuming no relationship when there actually is). A simple method to investigate significant non-linear relationships is the Box-Tidwell (1962) Transformation Test. It consists on adding to the logistic model interaction terms corresponding to the cross-product of each independent variable with its natural logarithm *(x)ln(x)*. If any of these terms are significant, then there is evidence of nonlinearity in the logit. This procedure does not provide the type of nonlinearity, thus if present further investigation is necessary.

f) Fitting Non-Linear Logistic Regressions: the Fractional Polynomial Methodology

Whenever evidence of significant non-linear relationship between a given independent variable and the logit of the dependent is detected, the Fractional Polynomial methodology (Royston and Altman, 1994) is implemented, in order to detect the best non-linear functional form that describes the relationship. Instead of trying to directly estimate a general model, where the power parameters of the non-linear relationship is estimated simultaneously with the coefficients of the independents, this methodology searches for the best functional form from a given set of possible solutions.

As presented before, our logistic regression expression is given by:

$$g(x,\beta)=\ln\left[\frac{\pi(x)}{1-\pi(x)}\right]=\beta_0+\beta_1x_1+\ldots+\beta_kx_k$$

For this study, only one of the independent variables had a potentially non-linear relationship with the logit, let this variable be represented by x_k. In order to accommodate the non-linear relationship, the logistic regression expression could be generalized to:

$$g(x,\beta)=\beta_0+\beta_1x_1+\ldots+\beta_{k-1}x_{k-1}+\sum_{j=1}^{J}\beta_{j+k-1}H_j(x_k)$$

where, for $j = 1,\ldots,J$:

$$H_j(x_k)=\begin{cases}x_k^{p_j} & \text{if } p_j\neq p_{j-1}\\ H_{j-1}(x_k)\ln(x_k) & \text{if } p_j=p_{j-1}\end{cases}$$

Under this setting, p represents the power and j the number of polynomial functions. For example, a quadratic relationship would have $J=2$, $p_1=1$ and $p_2=2$:

$$g(x,\beta)=\beta_0+\beta_1x_1+\ldots+\beta_{k-1}x_{k-1}+\beta_kx_k+\beta_{k+1}x_k^2$$

In practice, as suggested by Royston and Altman (1994), it is sufficient to restrict *J* to 2 and *p* to the set $\Omega = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, where *p=0* denotes the natural log of the variable. The methodology is implemented through the following steps:

i. Estimate the linear model;

ii. Estimate the general model with *J=1* and $p \in \Omega$, and select the best *J=1* model (the one with lower deviance);

iii. Estimate the general model with *J=2* and $p \in \Omega$, and select the best *J=2* model;

iv. Compare the linear model with the best *J=1* and the best *J=2* models. This comparison is made through a likelihood ratio test, asymptotically chi-square distributed. The degrees of freedom in the test increases by 2 for each additional term in the fractional polynomial, one degree for the power, and another for the extra coefficient. The selected model is the one that represents a significant better fit than that of next lower degree, but not a significant worse fit than that of next higher degree;

v. Graphically examine the fit estimated by the model selected in the previous stage, in order to validate the economic intuition of the non-linear relationship suggested by the model. This is achieved by comparing the lowess[37] function of the relationship between the dependent and the independent variable in question, and the multivariable adjusted function that results from the model selected in the previous stage.

g) Testing for Multicolinearity: the Tolerance Statistic

As for linear regression, high colinearity between the independent variables in a logistic regression results in loss of efficiency, with unreasonably high estimated coefficients and large associated standard errors. Detection of multicolinearity can be made through the use of the Tolerance statistic, defined as the variance of each independent variable that is not explained by all of the other independent variables. For the independent variable X_i, the tolerance statistic equals $1 - R^2_{X_i}$, where $R^2_{X_i}$ is the

[37] The Lowess is the Locally Weighted Scatterplot Smoothing (Cleveland 1979) between two variables. Since the dependent is a binary variable, it is convenient to use this smoothed function to be able to graphically access the relationship in question.

R^2 of a linear regression using variable X_i as the dependent variable and all the remaining independents as predictors.

If the value of the statistic for a given independent is close to 0, it indicates that the information the variable provides can be expressed as a linear combination of the other independent variables. As a rule of thumb, only tolerance values lower than 0.2 are cause for concern.

Appendix 4 – Estimation Results

Linear Regressions General Results								
Regression	Nº Obs	Obs Y=0	Obs Y=1	Deviance	Hosmer & Lemeshow			AUROC
					χ^2	df	P-Value	
A - 2 Eq. Model / Sectors 1 & 2	5,044	4,819	225	1,696	8.74	8	36.51%	71.30%
A - 2 Eq. Model / Sector 3	5,951	5,706	245	1,928	6.79	8	55.89%	
B - Unweighted Model	10,995	10,525	470	3,626	7.07	8	52.94%	71.28%
C - Weighted Model	1,420	950	470	1,623	11.73	8	16.38%	71.44%

Linear Regressions Estimated Coefficients																
Variable	A - 2 Eq. Model / Sectors 1 & 2				A - 2 Eq. Model / Sector 3				B - Unweighted Model				C - Weighted Model			
	β^	σ^	Wald	P-Value	β^	σ^	Wald	P-Value	β^	σ^	Wald	P-Value	β^	σ^	Wald	P-Value
R7	-0.39246	0.12878	9.29	0.2307%	-	-	-	-	-	-	-	-	-	-	-	-
R8	-	-	-	-	-0.19705	0.07590	6.74	0.9427%	-0.16455	0.05230	9.90	0.1653%	-0.18762	0.06564	8.17	0.4258%
R9	-	-	-	-	-0.18184	0.08514	4.56	3.2691%	-0.22849	0.06887	11.01	0.0907%	-0.23442	0.08127	8.32	0.3923%
R17	-0.28779	0.09241	9.70	0.1843%	-0.24115	0.08659	7.76	0.5356%	-0.28909	0.06361	20.66	0.0005%	-0.26327	0.07845	11.26	0.0791%
R20	0.46940	0.06164	58.00	0.0000%	0.45161	0.05664	63.57	0.0000%	0.44002	0.04283	105.55	0.0000%	0.50697	0.06564	59.66	0.0000%
R23	0.23328	0.06380	13.37	0.0255%	-	-	-	-	0.15280	0.04436	11.86	0.0572%	0.15948	0.06234	6.54	1.0520%
K	-3.35998	0.08676	1,499.67	0.0000%	-3.33521	0.07658	1,896.73	0.0000%	-3.33613	0.05688	3,440.16	0.0000%	-0.94820	0.06586	207.30	0.0000%

Box-Tidwell Final Backward Stepwise Regression Coefficients																
Variable	A - 2 Eq. Model / Sectors 1 & 2				A - 2 Eq. Model / Sector 3				B - Unweighted Model				C - Weighted Model			
	β^	σ^	Wald	P-Value	β^	σ^	Wald	P-Value	β^	σ^	Wald	P-Value	β^	σ^	Wald	P-Value
R7	-0.38011	0.12830	8.78	0.3049%	-	-	-	-	-	-	-	-	-	-	-	-
R8	-	-	-	-	-0.21276	0.07622	7.79	0.5247%	-0.17143	0.05241	10.70	0.1073%	-0.19597	0.06552	8.95	0.2782%
R9	-	-	-	-	-0.15921	0.08632	3.40	6.5114%	-0.21020	0.06940	9.17	0.2454%	-0.22388	0.08196	7.46	0.6301%
R17	-0.22552	0.09719	5.38	2.0317%	-0.18249	0.09026	4.09	4.3184%	-0.23063	0.06677	11.93	0.0552%	-0.20282	0.08150	6.19	1.2824%
R20	1.68533	0.36083	21.82	0.0003%	1.58508	0.31588	25.18	0.0001%	1.57265	0.23829	43.56	0.0000%	1.57769	0.29246	29.10	0.0000%
R23	0.19889	0.06597	9.09	0.2570%	-	-	-	-	0.12254	0.04590	7.13	0.7586%	0.12243	0.06302	3.77	5.2037%
BT20*	-0.66208	0.19297	11.77	0.0601%	-0.63780	0.17459	13.34	0.0259%	-0.62538	0.12917	23.44	0.0001%	-0.62506	0.16384	14.55	0.0136%
K	-2.96198	0.13987	448.47	0.0000%	-2.91367	0.13336	477.33	0.0000%	-2.93971	0.09630	931.87	0.0000%	-0.53335	0.12495	18.22	0.0020%
***BT20 = R20*LN(R20)**																

Fractional Polynomial Model Comparisons (Best J=1,2,3 Models)

R20	df	A - 2 Eq. Model / Sectors 1 & 2				A - 2 Eq. Model / Sector 3				B - Unweighted Model				C - Weighted Model			
		Deviance	Gain	P-Value	Powers	Deviance	Gain	P-Value	Powers	Deviance	Gain	P-Value	Powers	Deviance	Gain	P-Value	Powers
Not in model	0	1750.177	-	-	-	1986.467	-	-	-	3724.482	-	-	-	1687.181	-	-	-
Linear	1	1696.043	0.000	0.000	1	1927.857	0.000	0.000	1	3626.025	0.000	0.000	1	1623.173	0	0.000	1
J = 1	2	1684.842	11.201	0.001	0	1915.064	12.793	0.000	0	3603.782	22.243	0.000	0	1610.633	12.54	0.000	0
J = 2	4	1682.437	13.605	0.301	.5 3	1913.080	14.778	0.371	1 1	3599.921	26.105	0.145	.5 3	1608.129	15.044	0.286	.5 3
J = 3	6	1681.540	14.503	0.639	-1 2 2	1911.768	16.089	0.519	2 3 3	3599.042	26.983	0.644	-1 1 2	1607.349	15.824	0.677	-1 1 2

Reported Deviances for Fractional Polynomial Search

Model #	Power 1	Power 2	Deviance			
			Model A1	Model A2	Model B	Model C
1	-2	-	1750.175	1986.353	3724.480	1687.179
2	-1	-	1699.910	1937.404	3636.893	1633.960
3	-0.5	-	1689.693	1922.565	3614.541	1618.907
4	0	-	1684.842	1915.064	3603.782	1610.633
5	0.5	-	1687.719	1918.091	3609.449	1613.104
6	1	-	1696.043	1927.857	3626.025	1623.173
7	2	-	1715.213	1949.074	3662.488	1646.956
8	3	-	1728.820	1962.952	3686.848	1663.446
9	-2	-2	1750.175	1986.353	3724.480	1687.179
10	-1	-2	1699.911	1937.404	3724.480	1687.179
11	-0.5	-2	1689.694	1922.565	3614.542	1618.908
12	0	-2	1684.842	1915.066	3603.784	1610.634
13	0.5	-2	1687.718	1918.071	3609.449	1613.103
14	1	-2	1696.040	1927.808	3626.023	1623.170
15	2	-2	1715.210	1948.992	3662.485	1646.953
16	3	-2	1728.817	1962.857	3686.846	1663.444
17	-1	-1	1750.175	1935.171	3724.480	1687.179
18	-0.5	-1	1689.685	1922.555	3614.528	1618.898
19	0	-1	1684.842	1915.064	3603.782	1610.633
20	0.5	-1	1685.583	1916.271	3605.742	1611.041
21	1	-1	1687.582	1919.556	3610.443	1613.928
22	2	-1	1692.009	1925.960	3620.050	1620.917
23	3	-1	1695.230	1930.067	3626.581	1626.092

Reported Deviances for Fractional Polynomial Search (Cont.)						
Model #	Power 1	Power 2	Deviance			
			Model A1	Model A2	Model B	Model C
24	-0.5	-0.5	1688.517	1920.858	3612.048	1617.124
25	0	-0.5	1684.839	1915.060	3603.776	1610.627
26	0.5	-0.5	1685.272	1915.696	3604.853	1610.790
27	1	-0.5	1686.189	1917.169	3606.940	1612.193
28	2	-0.5	1687.903	1919.588	3610.549	1615.131
29	3	-0.5	1688.928	1920.884	3612.589	1617.001
30	0	0	1684.776	1914.838	3603.591	1610.240
31	0.5	0	1684.827	1914.977	3603.738	1610.376
32	1	0	1684.838	1915.058	3603.778	1610.541
33	2	0	1684.661	1914.992	3603.492	1610.619
34	3	0	1684.353	1914.867	3603.072	1610.458
35	0.5	0.5	1684.297	1914.262	3602.552	1609.827
36	1	0.5	1683.755	1913.681	3601.482	1609.245
37	2	0.5	1682.890	1913.159	3600.148	1608.379
38	3	0.5	1682.437	1913.385	3599.921	1608.129
39	1	1	1683.014	1913.080	3600.178	1608.436
40	2	1	1682.485	1913.609	3600.057	1608.280
41	3	1	1682.816	1915.364	3601.938	1609.432
42	2	2	1684.157	1917.984	3605.421	1611.925
43	3	2	1687.085	1923.590	3613.244	1617.255
44	3	3	1692.467	1932.259	3626.168	1626.117

Non-Linear Regressions General Results								
Regression	Nº Obs	Obs Y=0	Obs Y=1	Deviance	Hosmer & Lemeshow			AUROC
					χ^2	df	P-Value	
A - 2 Eq. Model / Sectors 1 & 2	5,044	4,819	225	1,682	8.20	8	41.46%	71.88%
A - 2 Eq. Model / Sector 3	5,951	5,706	245	1,913	6.29	8	61.49%	
B - Unweighted Model	10,995	10,525	470	3,600	2.23	8	97.32%	71.88%
C - Weighted Model	1,420	950	470	1,608	7.68	8	46.53%	71.87%

Non-Linear Regressions Estimated Coefficients																
Variable	**A - 2 Eq. Model / Sectors 1 & 2**				**A - 2 Eq. Model / Sector 3**				**B - Unweighted Model**				**C - Weighted Model**			
	β^	**σ^**	**Wald**	**P-Value**	**β^**	**σ^**	**Wald**	**P-Value**	**β^**	**σ^**	**Wald**	**P-Value**	**β^**	**σ^**	**Wald**	**P-Value**
R7	-0.38053	0.12831	8.80	0.3020%	-	-	-	-	-	-	-	-	-	-	-	-
R8	-	-	-	-	-0.21229	0.07617	7.77	0.5321%	-0.17136	0.05241	10.69	0.1078%	-0.19728	0.06560	9.04	0.2637%
R9	-	-	-	-	-0.16045	0.08631	3.46	6.3017%	-0.21111	0.06940	9.25	0.2353%	-0.22341	0.08196	7.43	0.6414%
R17	-0.22465	0.09710	5.35	2.0686%	-0.18418	0.09013	4.18	4.1003%	-0.23136	0.06668	12.04	0.0521%	-0.20304	0.08142	6.22	1.2638%
R23	0.20007	0.06590	9.22	0.2398%	-	-	-	-	0.12378	0.04587	7.28	0.6964%	0.12343	0.06299	3.84	5.0039%
R20_1	2.01146	0.31598	40.52	0.0000%	1.79215	0.27152	43.56	0.0000%	1.84306	0.21015	76.92	0.0000%	1.87907	0.26051	52.03	0.0000%
R20_2	-0.00933	0.00424	4.83	2.7966%	-0.00873	0.00421	4.30	3.8206%	-0.00876	0.00297	8.72	0.3145%	-0.00907	0.00400	5.13	2.3451%
K	-3.25891	0.08887	1,344.58	0.0000%	-3.42640	0.08329	1,692.28	0.0000%	-3.24970	0.05921	3,012.06	0.0000%	-0.84100	0.07034	142.94	0.0000%

Multicolinearity Test									
Unweighted Reg.		**Weighted Reg.**		**Sectors 1&2 Reg.**		**Sector 3 Reg.**		**Sector 3 Reg.**	
Variable	**Tolerance**	**Variable**	**Tolerance**	**Variable**	**Tolerance**	**Variable**	**Tolerance**	**Variable**	**Tolerance**
R8	0.989	R8	0.988	R7	0.989	R8	0.9880	R8	0.9878
R9	0.964	R9	0.963	R17	0.763	R9	0.9700	R9	0.9685
R17	0.762	R17	0.722	R23	0.868	R17	0.8130	R17	0.8128
R23	0.854	R23	0.853	R20_1	0.379	R20_1	0.4200	R20_1	0.0646
R20_1	0.375	R20_1	0.336	R20_2	0.477	R20_2	0.4890	R20_2	0.0685
R20_2	0.477	R20_2	0.440	-	-	-	-	-	-
Model # 38		Model # 38		Model # 38		Model # 38		Model # 39	

Appendix 5 – K-Means Clustering

K-Means Clustering[38] is an optimization technique that produces a single cluster solution that optimizes a given criteria or objective function. In the case of the methodology applied in this study, the criteria chosen is the Euclidean Distance between each case, c_i and the closest cluster centre C_k:

$$d\left(c_i, C_k\right) = \sqrt{\left(c_i - C_k\right)^2}$$

Cluster membership is determined through an iterative procedure involving two steps:

i. The first step consists on selecting the initial cluster centers. Two conditions are checked for all cases: first, if the distance between a given case c_i and its closest cluster mean C_k is greater than the distance between the two closest means, C_n and C_m, then that case will replace either C_n or C_m, whichever is closer to it. If case c_i does not replace any cluster mean, a second condition is applied: if c_i is further from the second closest cluster's centre than the closest centre if from any other cluster's centre, then that case will replace the closest cluster centre. The initial k cluster centers are set after both conditions are checked for all cases;

ii. The second step consists of assigning each case to the nearest cluster, where the distance is the Euclidean Distance between each case and the cluster centers determined in the previous step. The final cluster means are then computed as the average values of the cases assigned to each cluster. The algorithm stops when the maximum change of cluster centers in two successive iterations is smaller than the minimum distance between initial cluster centers times a convergence criterion.

---

[38] For more information see, for example, Hartigan (1975).

Appendix 6 – IRB RWA and Capital Requirements for Corporate Exposures

The formulas for calculating the RWA for corporate exposures under the IRB approach are:

$$RWA = k * 12,5 * EAD$$

where *k* is the Capital Requirement, computed as:

$$k = LGD * \Phi\left[\frac{\Phi^{-1}(PD)}{\sqrt{1-R}} + \sqrt{\frac{R}{1-R}} * \Phi^{-1}(0,999)\right] * \frac{1+(M-2,5)*b(PD)}{1-1,5*b(PD)}$$

b(PD) is the Maturity Adjustment:

$$b = \left(0,08451 - 0,05898 * \log(PD)\right)^2$$

and *R* is the Default Correlation:

$$R = 0,12 * \frac{1-\exp(-50*PD)}{1-\exp(-50)} + 0,24 * \left[1 - \frac{1-\exp(-50*PD)}{1-\exp(-50)}\right]$$

PD and LGD are measured as decimals[39], Exposure-At-Default (EAD) is measured as currency, Maturity (M) is measured in years, and Φ denotes the cumulative distribution function for a standard normal random variable.

The Default Correlation (R) formula has a firm-size adjustment of $\left[0,04 * 1 - \left(\frac{S-5}{4}\right)\right]$ for SME borrowers, where *S* is the total annual sales in Millions of Eur, and $5 \le S \le 50$. SME borrowers are defined as "Corporate exposures where the reported sales for the consolidated group of which the firm is a part is less than 50 Millions of Eur" (Basel Committee on Banking Supervision 2003, par. 242). It is possible for loans to small business to be treated as retail exposures, provided that the borrower, on a consolidated basis, has a total exposure to the bank of less than one Million Eur, and the bank has consistently treated these exposures as retail. For the

[39] The PD for corporate exposures has a minimum of 0,03%.

purpose of this study it is assumed that all exposures are treated as corporate exposures.

Thus, ignoring both Market and Operational risks, we have:

$$\text{Capital Ratio} = \frac{\text{Regulatory Capital}}{\text{Total RWA}}$$

If the minimum value for the capital ratio (8%) is assumed, then:

$$\text{Regulatory Capital} = 8\% * \text{Total RWA}.$$