universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Statistical Credit Risk Models“

Verfasser

Racko Peter

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften

(Mag. rer. soc. oec.)

Wien, im November 2007

Studienkennzahl lt. Studienblatt	A157
Studienrichtung lt. Studienblatt	Internationale Betriebswirtschaftslehre
Betreuer	O. Univ.-Prof. Dipl.-Ing. Dr. Engelbert Dockner

Table of Contents

List of tables and figures

1. Introduction

With the implementation of Basel II, banks are faced with more demands on quantifying market, credit and operational risks in their portfolios. Especially the new rules for credit risk quantification, known as Internal Ratings Based Approach, constitute a large part of the entire Basel II Accord. In order to take full advantage of these new rules, banks can, unlike under the old 1988 Basel Accord, implement sophisticated models that capture the individual elements of credit risk in more detail. There are major challenges in the implementation of the IRB Approach, one of them being the internal assessment of the credit risk profile of a bank's counterparty – a task which has to be accomplished by assigning a default probability to the particular client[1]. If the counterparty has an assigned rating by an external rating agency, such as Standard & Poor's or Moody's, the default probability can be directly obtained from the respective rating as ratings are provided with a default probability percentage for each rating class. However, the majority of companies do not have a rating. For example, on its homepage, Moody's indicates to have ratings that cover over 11.000 company issuers which is a fairly small figure when compared to the number of companies in the US economy alone – nearly six million. As a result, for those counterparties that do not have a rating, a bank will have to be able to calculate the PD on its own. The vast majority of these firms will be small and middle sized enterprises (SMEs)[2]. While the exposures of SMEs may not be very large, their importance for every economy is undisputed. There are many arguments that underscore this fact: many new, innovative companies start as small start-ups, SMEs employ a large part of the population and, for example, between 2001 and 2003 in the EU they were a more significant force in driving job growth than their large counterparts[3]. Therefore, considering how vital these are for any economy it will not only be in the best interest of the bank, as the PD is a major component in the calculation of capital requirements. It will also be in the best interest of each country to have models in place that quantify the default probability as accurately as possible in order to make sure that loans are correctly priced. The following table gives an overview of the size structure of companies in the EU and US as of 2003:

[1] Please see Chapter 2 for a more detailed look into the Basel II Accord.
[2] In the EU, SMEs are defined as having less than 250 employees and a turnover of less than €50 Million or a balance sheet total of less than €43 Million.
[3] Schmiemann (2006).

European Union:

Firm size group by number of employees	Percentage of firms	Percentage of total employment
1 to 9	91,24%	28,77%
10 to 49	7,42%	20,52%
50 to 249	1,12%	16,91%
Less than 250	***99,78%***	***66,19%***
250 and more	***0,22%***	***33,81%***

USA:

Firm size group by number of employees	Percentage of firms	Percentage of total employment
1 to 9	78,55%	11,03%
10 to 99	19,69%	25,32%
100 to 499	1,47%	14,56%
Less than 500	***99,71%***	***50,92%***
500 and more	***0,29%***	***49,08%***

Table 1.1: Size structure of firms in the EU and USA. Sources: Eurostat, United States Small Business Association.

In both the EU and US, SMEs account for well over 90 percent of all companies and employ at least 50 percent of the workforce and thus it is very likely that the credit portfolio composition of most commercial banks will be dominated by SME exposures.

The problem with these firms is that in addition to the fact that they do not possess a credit rating, they also cannot supply any market data that would aid the assessment of their creditworthiness expressed in terms of PD. Instead, they will only be able to provide accounting data. It is an acknowledged fact that market data on firms is much more accurate and reliable than accounting data and in recent years a number of market data based models for credit risk appeared[4]. The challenge for banks is to be able to quantify credit risk without the knowledge of market data. As a result, the aim of this thesis will be to show the currently prevalent practice of quantifying credit risk in the absence of market data and external ratings where only accounting data can be utilized in the context of Basel II.

Throughout the credit risk literature, William H. Beaver and Edward I. Altman are considered to be the pioneers in the analysis of accounting data with respect to company defaults. Beaver (1966) looked at the effect of individual financial ratios on default and concluded that ratio analysis is useful in bankruptcy prediction. In 1968 E. Altman published his seminal work *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy* and it is widely considered to be the birth of the so-called statistical credit risk models. It was the first time that multiple ratios were analyzed at the same time in a bankruptcy prediction context as part of the classic Z-Score model. And even though the modelling

[4] See Chapter 7 for an overview of market data based credit risk models.

technique applied by Altman (1968) has since become outdated, as will be shown in Chapter 3, the framework of current statistical models shares the same idea as the original Z-Score model: a large amount of balance sheet data for many firms that both defaulted and survived over a given period of time is needed to analyze the influence of various factors, mainly financial ratios, on the creditworthiness of borrowers. Afterwards, a selected set of factors that are assumed to have the greatest influence on whether or not a firm defaults, when combined and weighted, will produce a score that will be used to assess the creditworthiness of a particular borrower.

Throughout this thesis, the major aspects and issues of currently used statistical models, most notably logistic models, as a means of quantifying the default probability of exposures will be discussed. These will also be complemented by discussions of already created credit risk models, namely the Z-Score model by Altman (1968), the model for US SMEs by Altman and Sabato (2006), the model proposed by Fernandes (2005), the proprietary RiskCalc model by Moody's KMV (2000), the model by Bank Austria Creditanstalt AG (2005), Austria's largest bank, and also a model by Halling and Hayden (2005) which includes a time component in addition to accounting data.

The aforementioned Z-Score by Altman (1968) was the first model to analyze multiple financial ratios in order to assess the creditworthiness of a firm. This model was based on observations of 33 defaulted and 33 non-defaulted manufacturing firms during the period between 1946 and 1965. The asset sizes of these firms range from \$0.7 million to \$25.9 million, thus omitting large and very small companies. Despite its age and the rather unsuitable functional form in terms of Basel II requirements, this model continues to be somewhat of a benchmark, with authors of other models using it to demonstrate their model's performance in terms of discriminative ability.

E. Altman and G. Sabato (2006) published a logistic model specifically designed for US SMEs. Using a sample that includes 2010 firms, of which 120 were defaulted, and that spans the time from 1994 until 2002, they mainly attempt to show the importance of having a separate model for assessing the rating of an

SME instead of using a generic model that is also used for large corporations. According to Atlman & Sabato (2006), by taking into account the differences between SMEs and large firms via separate models, banks can lower the required capital according to Basel II.

Fernandes (2005) uses a dataset of 11000 financial statements from Portuguese firms to firstly create two statistical, logistic models, one that takes different industries into account and one without regard for industries. He then uses these models to demonstrate how rating classes can be created from the output of the models.

Moody's KMV (2000), a provider of credit risk management solutions, also developed a model for PD estimation of unrated private companies, RiskCalc. While some of the details of the model are not disclosed as it is proprietary, Moody's KMV does provide extensive documentation for RiskCalc. It is not only a good example of a commercially available credit risk model but with its many regional adaptations (there is a RiskCalc version for Australian, Austrian, Belgian, Dutch, French, German, Italian, Japanese, Mexican, Scandinavian, Spanish, UK, and US firms) it also serves the purpose of showing differences in credit risk relevant attributes between countries.

The model by Bank Austria Creditanstalt AG (2005) is a demonstration of an Austrian commercial bank's internal rating model. While the exact design details are confidential and only available to the bank and its supervisor, it is a good example of how to include qualitative data into a rating model.

Finally, Halling and Hayden (2004) create a model, based on a sample spanning the years between 1994 and 1999 with 2283 firms of which 171 defaulted. This model is unique in that it represents a rare case in which the time component is explicitly taken into account in corporate default prediction. This so-called hazard model is perhaps a glimpse of the future in terms of statistical credit risk models and as such deserves to be mentioned.

The thesis is structured as follows: In Chapter 2, the Basel Accords and their implications on the credit risk assessment of banks' counterparties are outlined. Chapter 3 deals with the statistical framework that is the basis for models of credit risk and shows which functional form dominates the present landscape, Chapter 4 looks into the issue of identifying the most useful variables for a model, Chapter 5 will outline the process of calibration of models, an area of extreme importance in terms of Basel II compliance, Chapter 6 will detail the validation issues for statistical default prediction models and Chapter 7 will briefly outline alternatives to statistical default prediction. Chapter 8 will conclude.

2. Implications of the Basel Accords

This chapter will outline the main objectives of the Basel Accords, give an overview of the Basel II framework and focus on the implications on the assessment of counterparty credit risk.

The first Basel Accord, drafted in 1988, represented a huge step in banking regulation in that it required banks to hold specified amounts of equity capital. This capital is called regulatory capital and serves the purpose of helping banks absorb losses stemming various risks, credit risk included. For credit risk, the requirement was for a bank to hold 8 percent of risk weighted assets. More precisely, the regulatory capital to be set aside for an exposure would be given by multiplying the exposure amount with a risk weight and then with the 8 percent. The risk weights were set by the Basel Committee and were supposed to reflect the riskiness of certain customer types. For example, the risk weight for an exposure to an OECD country was set to zero percent while an exposure to a corporate received 100 percent. While this regulation certainly helped increase the capital held at banks, weaknesses were soon discovered, most notably the lacking risk sensitivity. This was especially true in the corporate segment: irrespective of whether a borrower was a rated AAA firm or a barely surviving SME, the risk weight remained the same for both borrowers. As a result, in 1999 the Committee began working on the Basel II Accord.

The new Accord indeed incorporates risk sensitivity in both of its main broad methodologies for the calculation of capital requirements for credit risk. The simpler Standardized Approach extends the old Basel I methodology by enabling a differentiation of borrowers according to their external credit rating assessment. As a result, an exposure to a AAA rated customer will result in a 1,6 percent capital requirement, much lower than the 8 percent under the old Accord. However, herein also lays the problem: in order to incorporate this risk sensitivity, external ratings have to be available for a particular borrower. However, as Chapter 1 demonstrated, this will rarely be the case. As a result, the true innovation in terms of adding risk sensitivity into the calculation of capital requirements and achieving a lower capital charge is the Internal Ratings Based (IRB) approach.

The innovation lies in the concept of the bank providing their own estimates of key parameters which are supposed to give a good indication about the overall credit situation of a borrower and his/her exposure:

- Probability of default (PD): the quantification of the borrower's creditworthiness, these estimates have to be provided for a one year horizon. As opposed to the other parameters, the PD is not facility-specific and has to be assigned to every borrower.
- Loss given default (LGD): the facility-specific quantification of the amount that a bank will lose once the borrower defaults on a particular exposure.
- Exposure at default (EAD): the amount of the exposure at the moment of default of the borrower.
- Maturity (M) of an exposure

These key parameters are then used to calculate the capital requirement for each exposure. This is accomplished via the so-called risk weight functions, for which the risk parameters serve as inputs. There are different risk weight functions for different asset classes[5], however, since the focus lies on corporate borrowers, only the corporate risk weight functions are presented:

[5] Basel II differentiates between the Sovereign, Institutions, Corporate, Retail, Equity, Securitization and Other Assets asset classes in the IRB Approach.

Correlation (R):

$$R = 0{,}12 \times \left(\frac{1 - e^{-50 \times PD}}{1 - e^{-50}} \right) + 0{,}24 \times \left(1 - \frac{1 - e^{-50 \times PD}}{1 - e^{-50}} \right)$$

Maturity Adjustment (b):

$$b = \left(0{,}11852 - 0{,}05478 \times \ln(PD)\right)^2$$

Capital requirement (K):

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

Risk-weighted assets (RWA):

$$RWA = K \times 12{,}5 \times EAD$$

The main equation is the capital requirement equation (K): its output is the percentage of the exposure at default that has to be set aside for each facility. Besides considering the PD, LGD and M, it also incorporates the effects of asset correlation[6] and maturity sensitivity[7], through the equations R and b respectively. In addition, the risk-weighted assets (RWA) of an exposure can be calculated by multiplying the capital requirement (K) with the EAD and the inverse of 8 percent, 12,5.

The risk weight functions use the Value at Risk[8] concept and were calibrated in order to obtain a reasonable capital requirement given the estimated input parameters. Figure 2.1 shows the capital requirements, expressed in percentage points, as a function of the default probability for an LGD of 50 percent and a maturity of three years:

[6] The main goal is to capture the effect of the overall economic environment on a borrower.
[7] The aim is to capture maturity effects. First of all, higher maturities imply higher risk and secondly, low-PD borrowers have a higher downside potential for downgrades than high-PD borrowers. As a result, an increase in the maturity will have a bigger impact on capital requirements for better rated firms than for worse rated ones.
[8] Value at Risk (VaR) is a measure of loss: VaR provides the maximum loss for a given probability for a given holding period. For example, if the 99 percent, 1 day VaR is € 1.000, it means that with a probability of 99 percent the loss on one day will not exceed € 1.000. The risk weight functions operate with a 99,9 percent confidence interval, i.e. the intention is for regulatory capital to be sufficient in 99,9 percent of all loss occurrences.

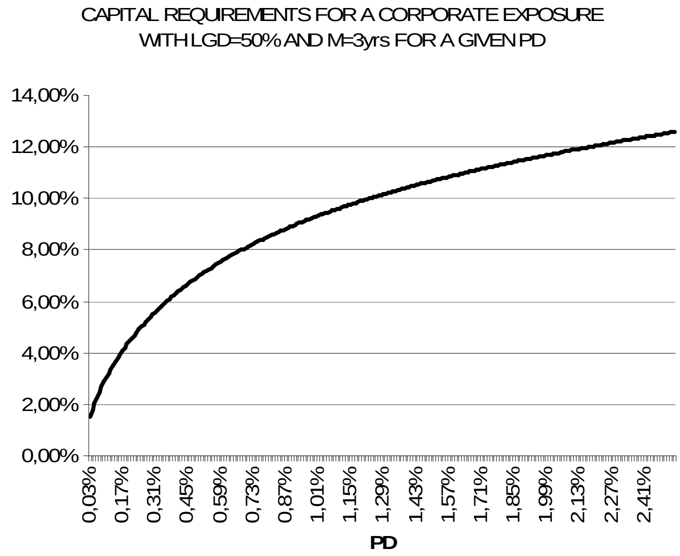

Figure 2.1: Capital requirements for a corporate exposure as a function of the PD

It becomes clear that the estimation of these risk parameters will pose serious challenges to banks applying for the IRB Approach. The Basel Committee acknowledges this fact by enabling the implementation of two variants of the IRB Approach. Under the Foundation IRB, the bank will only have to provide estimates for the PD while under the Advanced IRB, all inputs will have to be estimated by the bank itself. Nonetheless, irrespective of which IRB Approach a bank chooses, the quantification of counterparty credit risk, expressed as the borrower's PD, will be the bank's task. This fact also underscores the central importance of borrower PD estimation.

There are several implications for the evaluation of credit risk associated with borrowers using default prediction models arising from Basel II. First of all, the required final output of the credit risk assessment process was clearly specified – the final results obtained from default prediction models will have to be default probabilities. Secondly, the PD will have to be estimated for every corporate borrower in the portfolio and as a result, banks will have to develop models that accomplish this task with the available information at hand. As a result, when no market data is available, models that can work with accounting data as input will have to be implemented. Third, besides providing the risk weight functions, Basel II also specifies an extensive framework under which such models have to be operated and requirements that they have to meet. For example, the models have to be able to classify borrowers into seven rating classes consistently[9], the PDs that are estimated are subject to various rules, the PDs can only reflect borrower characteristics and models have to be validated in regular intervals. Finally, Basel II provided a huge boost for research in the area of credit risk measurement, which at the time was not yet well established and definitely lagging behind that of market risk. This research is mainly the result of the fact that while Basel II clearly defines the parameters and the framework of their use, it does not define how these parameters are to be estimated and which methodologies are to be applied. The subsequent research was aimed at closing this gap.

As a consequence of the entire Basel II regulation, which has had a substantial effect on the way borrower default risk is estimated, the focus throughout this thesis will not only lie in the design and statistical background of the models but also on areas which the Basel Committee emphasizes, such as the calibration of model outputs to PDs, the validation methods which should ensure a correct evaluation of borrower credit risk profiles over longer periods of time as well as the inclusion of qualitative data into models.

3. Modelling framework

This chapter will highlight the methods that can be used in construction of a default prediction model. It will show why in today's modelling practice logit/probit models are the clear favourites and linear discriminant analysis, as introduced by Altman (1968), is more or less only regarded as an important evolutionary step that nowadays does not find application in banking practice. Moreover, it will be shown why more advanced applications of logit/probit models are not being considered.

3.1 Discriminant analysis

Linear discriminant analysis is used to find a linear combination of various features and characteristics of a dataset that differentiates best between two or more classes within this dataset. In the context of credit risk, this means that in

[9] Each rating class has to be tied to a PD.

order to perform a discriminant analysis, one first needs data from both non-defaulted and defaulted firms as these will be the two groups to be differentiated. This data should contain financial/accounting variables. They are the independent variables and discriminant analysis attempts to find a linear equation that includes these variables such that, based on the outcome/score of the equation, the clearest and best differentiation between the two groups can be made. More formally, the equation will have the following form:

$$Z = a + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + e$$

Here, X_n is the value of the independent variable n, b_n is the coefficient for the independent variable n (i.e. the main result of the analysis), a is the constant, e is the error term and Z is the dependent variable based on which the classification into the two groups is done. As can be seen, LDA closely resembles linear regression. The main difference is that while in linear regression the dependent variables are of quantitative nature, LDA's dependent variable is of categorical nature, such as default or non-default. In order to provide the most fitting equation, discriminant analysis searches the optimal values for b_n, such that following term, as described in Backhaus et al. (2003) is maximized:

$$\frac{\sum_{g=1}^{2} N_g \left(\overline{Z}_g - \overline{Z}\right)^2}{\sum_{g=1}^{2} \sum_{i=1}^{n} \left(Z_{gi} - \overline{Z}_g\right)^2}$$

Variables:

g	the observed class
N_g	number of observations in a given class
$\overline{Z}_g$	centroid of a given class
$\overline{Z}$	mean score of both classes combined
Z_{gi}	score of firm i in class g

In both the numerator and denominator, the centroids, $\overline{Z}_g$, of the two classes can be found. The centroid of a given class is the mean score that is achieved in a particular class when the discriminant equation is applied to all observations within that class[10]. The centroids of each group are a basic way of describing the groups in a discrimination problem. Broadly speaking, the further they are apart, the better the discrimination function works. However, maximizing the distance

[10] More formally, $\overline{Z}_g = \frac{1}{N_g} \sum_{i=1}^{N_g} Z_{gi}$

$\left|\overline{Z}_{default} - \overline{Z}_{non-default}\right|$ need not necessarily yield an optimal result. The reason is the variance of scores within each group. As can be seen in Figure 3.1, the graph to the left shows a distribution of scores of two groups and it becomes apparent that there is a large area where the two distributions overlap due to the high variance within each group. As a result, in order to optimize the discriminant function, the variance has to be included in the calculation. The term above achieves this. The numerator measures the variance between the default and non-default groups, also referred to as sum of squares between. By maximizing it, the centroids of the respective groups will drift further apart. The denominator, on the other hand, measures the variance of scores within each class and it is to be minimized. This is done by calculating the squared sum of deviations of individual scores Z_{ig} from the group's respective centroid $\overline{Z}_g$, also referred to as sum of squares within. Figure 3.1 below illustrates this process. The two distributions on the left are distributions of scores that are based on a random discriminant model. As long as the model operates with some intuition, for example higher profitability leads to a higher score, the model will probably be able to discriminate between good and bad loans to some extent. However, there will be an interval of scores that will include good and bad borrowers, i.e. the part where the distributions overlap. LDA lets the distributions drift further apart and makes them more skewed – the overlapping part is being minimized (the two distributions to the right).

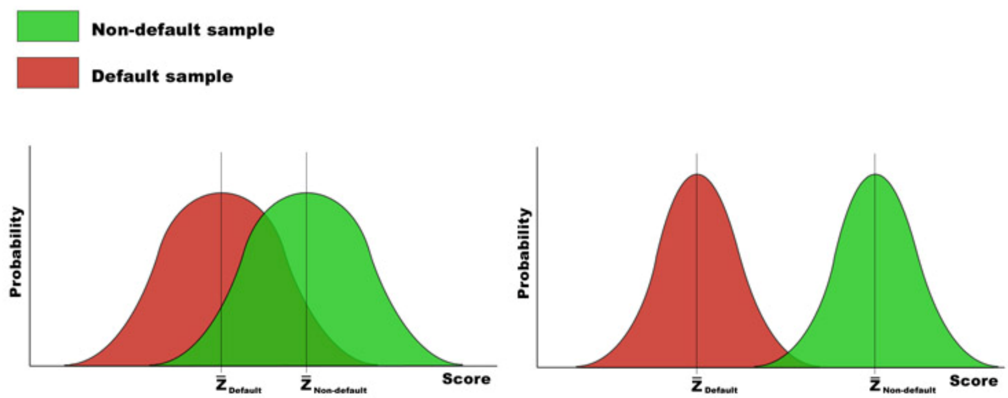

Figure 3.1: The process of the linear discriminant analysis: distributions of the default and non-default sample based on a random discriminant model (left), distributions of the two samples after linear discriminant analysis has been applied (right).

Upon maximizing the above term, firms can be classified into one of the two groups of interest. This can be done by comparing the squared distances between the score for an observation and the centroids the two groups[11]:

$$D_{ig}^2 = \left(Z_i - \overline{Z}_g\right)^2$$

When this squared distance to the default group centroid is smaller than the squared distance to the non-default centroid, the firm will be classified as defaulted and vice-versa. Derived from these squared distances can be a critical value of Z^*, for which if $Z_i < Z^*$, the firm will be assigned to the defaulted group. For $Z_i > Z^*$ it will be considered healthy.

It has to be pointed out, however, that it is almost impossible to obtain a discrimination function that perfectly discriminates between the two groups. A consequence of this is that there will always be score intervals where the distributions of the two groups overlap[12]. As a result, where the squared distances between a score and the two centroids are very similar, there is a high probability of committing either a type 1 or a type 2 error. In the context of default prediction, a type 1 error is made when a bad customer is classified as good based on model output. A type 2 error occurs when a good customer is classified as bad by the model. A type 1 error will only occur in the case of $Z_i > Z^*$, while a type 2 error only occurs when $Z_i < Z^*$.

Linear discriminant analysis has serious shortcomings. First of all, since the model is linear in nature, it assumes that there is a linear relationship between the independent variables and the eventual score. For example a 10% increase in the leverage ratio will always have the same effect on the score, irrespective of whether the leverage ratio is 5% or 75%. Intuitively, it is obvious that in the case of 5% leverage there will be a negative change in credit quality. However, it will not be as significant as in the latter case where a company loses virtually all equity. Another major problem is the weak interpretability of the score results. A score of, e.g. 2,04 does not indicate anything about the default probability. Hence,

[11] See Backhaus et al. (2003).
[12] Altman (1968) refers to this interval as a *„zone of ignorance“*.

while in terms of pure default and non-default classification discriminant analysis may be sufficient, in today's banking world where the Basel II Accords require an estimation of default probabilities and the creation of several rating classes, discriminant analysis is of limited use. As a result, experts have turned their attention to Logit/Probit models, rendering the LDA obsolete for default prediction purposes.

3.2 Logit and Probit Models

Logit and Probit models are based on regressions and have one very useful property: the dependent variable can only take on values between 0 and 1. This is achieved through the use of the logit and probit functions. The logit function is the inverse of the logistic function. The logit of a probability is the logarithm of the odds of an event[13], also referred to as the LogOdds. It has the following form:

$$logit(p) = log\left(\frac{p}{1\text{ - }p}\right)$$

The probit function, on the other hand, is the inverse cumulative distribution function associated with the standard normal distribution:

$$probit(p) = \Phi^{-1}(p)$$

The advantage of the logit and probit is the fact that they transform the probability *p* into values ranging outside the zero and one boundary. This is a useful property since both Logit and Probit models are generalized linear models: the dependent variable is estimated from a linear combination of independent variables. Through the use of the logit and probit as link functions, the dependent variable is connected to the aforementioned linear combination of predictors in a manner that is desirable in a default prediction context, i.e. the output, *p*, is constrained between zero and one. Thus the results of default prediction models built on the basis of Logit and Probit models can be interpreted as default probabilities.

---

[13] Odds are defined as $\frac{p}{1-p}$

At the heart of Logit models lies the logistic regression, which has the following form:

$$\ln\left(\frac{p}{1-p}\right) = a + b_1 X_1 + b_2 X_2 \ldots + b_n X_n + e$$

Again, a represents the constant, b_n are the coefficients, X_n are the independent variables and e is the error term. The left-hand term is the aforementioned logit. Through its use, the dependent variable, p, is constrained to values between zero and one. If Z is substituted for the entire right-hand side of the regression (i.e. Z is the score of the linear combination of dependent variables) then by reshuffling of the regression equation the default probability can be expressed as:

$$p = \frac{1}{1 + e^{-Z}}$$

The relationship between the independent variables (Z) and the eventual score (p) corresponds more with reality and intuition, as illustrated with a small example. Suppose there is a simple logistic model which consists of only one input variable – the leverage ratio – and there are three firms (A, B and C). Firm A has the lowest leverage ratio while firm C has the highest leverage ratio. If all three firms increased their leverage ratio by the same amount of percentage points, their scores from the linear combination inputs would increase by the same amount. This can be seen in Figure 3.2 on the x-axis. When it comes to default probability, this increase will most likely have different effects, depending on the firm. For firms A and C, the increased leverage ratio will not result in very dramatic increases of PD. Firm A had a very small leverage ratio to start with and even after the increase the ratio would still be considered a healthy figure. Firm C already had a very high leverage ratio initially and hence a very high PD and thus the increased leverage only marginally affects the PD. On the other hand, firm B will be hurt most in terms of PD and subsequently the rating: while not too high at the start, the leverage ratio increased to levels where the firm will be considered substantially more risky. As a consequence, the PD has to increase more dramatically than in cases of firms A and C.

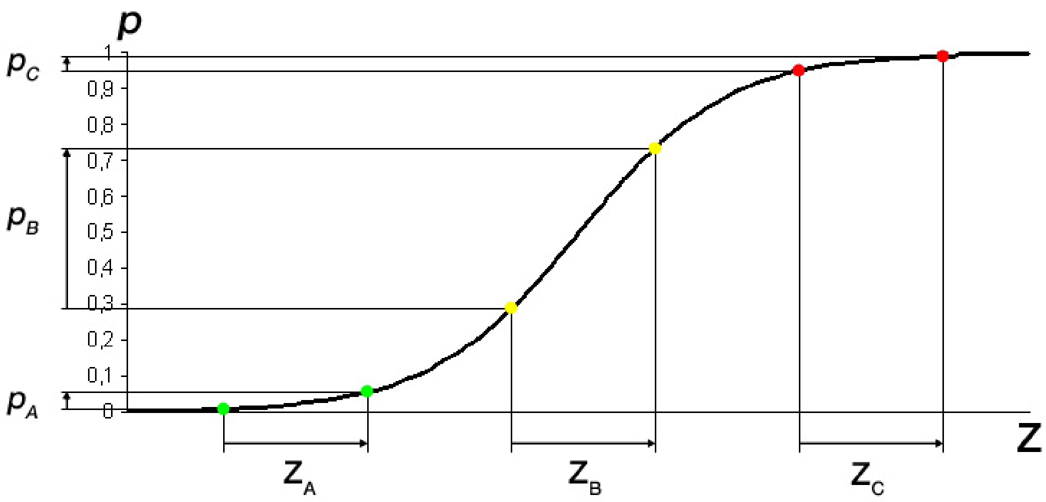

Figure 3.2: Relationship between independent variables (Z) and the eventual score (p) in a logistic regression.

The coefficients of the logit regression are obtained by means of maximum likelihood[14] estimation. This estimation method is a statistical procedure that attempts to find the optimal combination of coefficients by maximizing the following likelihood function:

$$LF = \prod_{i=1}^{n} p(x_i)^{y_i} \left[1 - p(x_i)\right]^{1-y_i}$$

Here, $p(x_i)$ denotes the default probability of firm i as given by the logistic regression equation and y_i is a binary variable that takes on the value 1 if firm i defaulted and 0 otherwise. The entire equation is essentially a product of the default probabilities of those companies in the dataset that defaulted and the survival probabilities (i.e. $1-p(x_i)$) of the dataset companies that did not default. The maximized likelihood function is the one that maximizes the likelihood that the default prediction based on the logit regression is correct.

Probit models, on the other hand, utilize the probit link function in the generalized linear model and implicitly assume that the underlying variable is normally distributed (i.e. the score of the linear combination of factors is normally distributed). Formally, the probit regression function has the following form:

[14] Likelihood is not to be confused with probability: while probability allows the inference of probabilities of unknown outcomes based on known parameters, the concept of likelihood allows the inference of likelihoods of known outcomes for different values of the parameters.

$$\Phi^{-1}(p) = a + b_1 X_1 + b_2 X_2 \ldots + b_n X_n + e$$

The variables on the right hand side of the term are the same as in the logit model, while $\Phi^{-1}(p)$ is the above mentioned probit. The relationship between the independent variables and the final score, which can also be interpreted as a default probability, is very similar to the relationship in the logit model depicted above.

The utilization of the probit link function is less common than that of logit in default prediction. However, the choice of one of the two link functions does not have a substantial effect on the entire model development process – the steps in variable selection remain the same and so does the calibration. Moody's KMV (2000) point out that in the past, logit was preferred due to computationally easier maximum likelihood estimation, an issue which has become irrelevant today. Even in professional literature on default prediction, there is no research into the matter of whether logit or probit delivers better results.

There is one assumption underlying these two models that should be kept in mind whenever one is designing a model using logit/probit: since they are both generalized linear models, both assume that the relationship between independents and the Logit/Probit (i.e. ln(p/(1-p) and $\Phi^{-1}(p)$ respectively) is linear. However, with some financial variables, this is not always the case. A part of the chapter on variable selection will discuss this problem and possible solutions to this problem will be presented.

3.3 Hazard models as an extension of traditional logit/probit models

In the recent past, some researchers have voiced their criticism of one particular characteristic of models based on either the discriminant analysis or logit/probit regressions: they do not specifically account for time. In order to overcome this deficiency, a small number of logit/probit models extended by inclusion of a hazard component which describes the default risk over time have been proposed. These models are referred to as hazard models and this section will briefly outline their basic principles. Hazard models are rooted in survival analysis[15], which is used in medicine to analyze the occurrence of death and time to death, such as the death rate of a population or the death rate past a certain age. Translated into credit risk modelling, it studies and analyzes the effect of time on the default probability of an exposure – in hazard models, the risk of bankruptcy changes with the passage of time.

At the heart of any hazard rate model in credit risk will be the modelling of the hazard rate, denoted in Cox & Oakes (1984) as:

$$\lambda(t)dt = P(t < T < t + dt \mid T > t)$$

$\lambda(t)dt$ is the hazard rate[16], t is a point in time, dt is one unit of time and T denotes the time of failure. In default prediction, T is the time of default. The hazard rate can be interpreted as the probability of default in the period between t and $t + dt$ (i.e. per unit time) of a company given that it has survived until time t. Halling and Hayden (2004) present an example of modelling the hazard rate through the use of logistic regression. One starts with a normal logistic regression with the typical financial ratios as explanatory variables. This model serves as the basis for classifying firms into risky firms – firms that are, based on the data, in a critical situation and especially close to default – and non-risky firms. Risky and non-risky firms are then separated by a cut-off point in the estimated score – firms above a certain score will be included in a separate model.

The new model is basically also a logit model, but in this case it takes the time dimension into account. This is the hazard component, namely the number of periods that elapse between being assigned to the risky category and the possible default (default, of course, does not necessarily have to occur, even in the risky group). The model then takes on the following form:

$$y_i = a + bX_i + cD_{it} + e$$

[15] For more details on survival analysis, please see Cox & Oakes (1984) or Hosmer & Lemeshow (1999).

[16] The hazard rate is sometimes also referred to as instantaneous failure probability, conditional failure rate or hazard function.

The inclusion of the time component is done via the vector D. D_{it} is a dummy variable that takes on the value of one for the particular firm-year observation that is t periods after the firm has been classified as risky. For example, in the firm-year observation of the firm i's second period after being classified as risky, D_{i1} will be zero and D_{i2} will be one. This way the dependent variable is also determined by the time spent in the group of risky firms and not just by accounting ratios. The independent variables of the vector X are chosen in the same fashion as those chosen for standard logit models.

3.4 Current state of modelling practice

The logistic functional form is currently dominating the field of default prediction in the absence of market data. The majority of internal rating models at banks use logistic regression to evaluate their corporate customers. Its main advantages lie in the fact that the output of such models is bound between zero and one. In addition, the relationship between the output and the independent variables is not linear and quite intuitive. Where linear discriminant analysis attempts to find a black and white solution by definition (i.e. classification into default and non-default), logit or probit models are more open towards producing a more diverse classification. In the past, where it would have been enough to find a tool that can make the decision whether to grant a loan or not, an LDA model would have been sufficient. However, with the emergence of the Basel II requirements and also with more risk-sensitive decision making tools such as RAROC[17], logit/probit models are simply the more intuitive and better choice. Models that incorporate the hazard rate suffer from shortcomings which can be attributed to their non-existence in the IRB landscape. They are more sophisticated and as a result are more difficult to implement. Moreover, the use of such models in the corporate default prediction segment is a very new topic with very little research. Hence, there is no indication as to their track record in default prediction. As a consequence, the focus throughout the next chapters will be on logit/probit models since they represent the present state of the art.

[17] Risk Adjusted Return on Capital (RAROC) is a commonly used risk-adjusted performance measure in banks.

4. Input variables

For a statistical default prediction model to work properly, i.e. to discriminate well between borrowers of different quality, the right combination of independent variables has to be found. Of course, one can estimate a regression with an arbitrarily chosen set of input variables and the significance of individual inputs varies with different datasets. As a result, the real challenge is to identify the first broad pool of “candidate” variables and then to find a procedure that narrows down the number of useful ratios to a few that will provide the best discrimination results.

Generally, input variables can be classified into two broad categories: quantitative data and qualitative data. While quantitative data utilizes financial figures taken from balance sheets and profit and loss statements, qualitative data attempts to capture information not reflected by hard numbers, such as management quality or market position. As opposed to quantitative data, the evaluation of qualitative data requires more input from loan specialists who also know the borrower. This process leans heavily on expert knowledge and is very similar to the development of scorecards for the evaluation of private individuals. Since the emphasis of this thesis is on statistical models which process financial data, only a broad overview of the issue of qualitative inputs will be given. In practice banks will have two separate models for the evaluation of quantitative data and qualitative data.

This chapter will deal with the topic of variable selection by first dealing with the choice of appropriate quantitative inputs and their combination into a model and presenting the quantitative variables selected into actual models from both research as well as practice. Finally, the topic of the inclusion of qualitative variables will be briefly discussed and an illustrative example of how quantitative and qualitative data are combined is given.

4.1 Quantitative inputs

In the absence of market data, the most viable input remains balance sheet as well as profit and loss statement data. As a result, most quantitative data will be in the form of financial ratios. The analysis of financial ratios is not new to economics as, for example, financial statement analysis is considered a basic tool to assess the performance, profitability and general health of a company. The main advantage of financial ratios is the fact that they represent a fraction of two figures from the balance sheet. A single figure from the balance sheet is not very useful as there are vast differences between individual companies: an EBIT of €100.000 is good for a small firm but for a large corporation it practically equals zero profit. Setting this figure in relation to a size measure, such as total assets, enables the comparison of the small and large firm.

The amount of useful financial figures goes into the hundreds. As a result, analyzing each one is an inefficient and daunting task and providing an exhaustive list of all possible ratios does not make sense. For example, a simple profitability measure can be expressed in several variations depending on which figure one takes as the actual profit number – possibilities include Net Income, Ordinary Profit, EBIT and EBITDA. Nevertheless, in order to give an overview, most accounting data that is being used as an input into default prediction models, can be broadly categorized as follows:

1. *Profitability ratios:* these represent the most obvious group of ratios, as profitability is the most crucial indicator of a firm's success. Higher profits are also reflected in higher equity and a profitable firm is much less likely to default on its obligations.
2. *Leverage ratios:* these ratios reflect the capital structure of a firm and indicate the degree to which it relies on external funding. Higher leverage translates into higher default probability as the buffer for losses – equity – is smaller and the threat of liquidation rises with a higher proportion of lenders.
3. *Liquidity ratios:* they measure the amount of a firm's cash as well as resources that can be converted to cash very easily. The higher the liquidity, the higher a firm's ability to repay obligations as they become due and to withstand unexpected shocks, and the lower the bankruptcy probability.
4. *Coverage ratios:* these ratios measure specifically the relation of performance and the amount of liabilities in the form of debt and interest payments. They are a combination of profitability and leverage.

5. *Activity ratios:* such ratios indicate the efficiency of the firm's operations, often in terms of the turnover of a particular figure and as a result, these ratios are sometimes referred to as efficiency ratios.
6. *Size:* although size is not a ratio, it is a very important factor that has a substantial effect on the default probability. Presumably, larger firms have a bigger and more experienced and capable management team, they are more diversified and more diversification implies less volatility.

4.1.1 Identifying the relevant variables

The identification of the most strongly discriminating variables is the main step towards creating a default prediction model. In fact, Moody's KMV (2000) state that "the selection of variables and their transformation are often the most important part of modelling default risk". There is a certain amount of truth in that statement: while the regression itself can be accomplished fairly easily with today's statistical software packages, a single set of clear, accurate rules and instructions to ensure that the modeller ends up with the best combination of factors and eventually an optimal model is non-existent. Instead, the variable selection is an iterative process that poses many challenges and demands some degree of creativity from the modeller.

Obviously, one has to start with a set of factors. However, as mentioned above, there are hundreds of potentially useful inputs. As a result, the analysis of all of them would be very time consuming and inefficient. Hence, one simply has to choose a smaller subset for evaluation. Altman (1968) described this process as a choice based on popularity in the literature, potential relevancy to the study as well as the inclusion of new ratios created by the study. This way, he ended up with 22 variables for the analysis. A closer look at other models also reveals that the first selection of variables is rather arbitrary and based on personal preference, as long as variables representing the categories described above (i.e. profitability, leverage, liquidity etc.) are considered. Fernandes (2005) starts the closer examination of the variables with 23 ratios, Altman (1977) uses a subset of 27 ratios, Altman (2005) starts with 17 variables. On the other hand, Moody's KMV does not disclose how many factors were closely analyzed for the Riskcalc models.

Once an initial set of variables has been identified, the variables contained in that initial set have to be analyzed individually in the univariate analysis. In essence, the univariate analysis is a thorough screening of a variable in terms of its suitability for default prediction and consists of three steps:

a) Evaluation of the economic intuitiveness of the variable
b) Evaluation of the discriminating power of the variable
c) Removal of outliers

The first step is very straightforward – one has to have a clear idea of the expected relationship between the ratio in question and the default probability and the relationship should follow basic economic intuition. For instance, the ratio Equity/Total Assets is expected to have a negative relationship with the default probability. In other words, one expects the default probability to rise with a declining equity ratio – this is the general assumption and is very much in line with common sense. While the equity ratio will rarely run the risk of not having an intuitive relation to bankruptcy, there may be ratios that are more ambiguous. If a variable doesn't show a clear and intuitive relationship to default, it should be dismissed from any further modeling considerations. Default frequency graphs are most widely used to graphically depict the relationship between an independent variable and bankruptcy. The following figure shows an example of a default frequency graph.

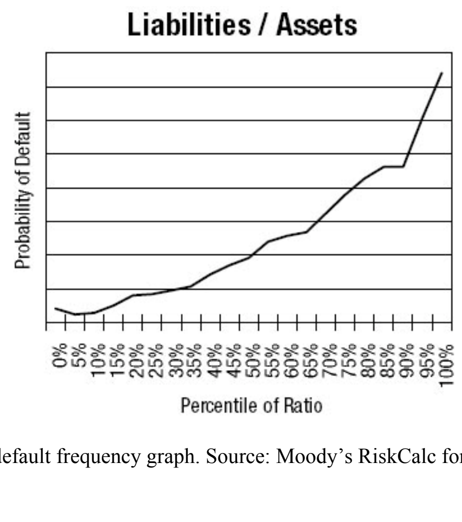

Figure 4.1: A sample default frequency graph. Source: Moody's RiskCalc for Private Companies: UK.

To create a default frequency graph for a particular ratio (i.e. Liabilities/Assets in this case), all firm observations have to be sorted according to the size of the ratio and then divided into equally sized groups in terms of numbers of observations so that each group corresponds to a different percentile range of the ratio in question. These groups are then plotted on the x-axis. In the figure above, each group contains five percent of the observations, however, the more observations are available the more groups can be created for this purpose. Finally, for each group the default frequency has to be calculated: the amount of defaults in that group divided by the total amount of firms in that group. This frequency is then plotted on the y-axis. The result is an easily observable relationship between the variable and the default probability.

In order to test the discriminative power of individual variables, power curves, also referred to as Cumulative Accuracy Profile (CAP) plots, are usually constructed and from those the accuracy ratio can be calculated. The power curve is a graphical depiction of the ability of any variable or even model – power curves are also used to test the discriminative power of entire models[18] – to predict or influence the outcome of a binary event, i.e. default/non-default. Figure 4.2 shows an example of a typical power curve graph with three power curves.

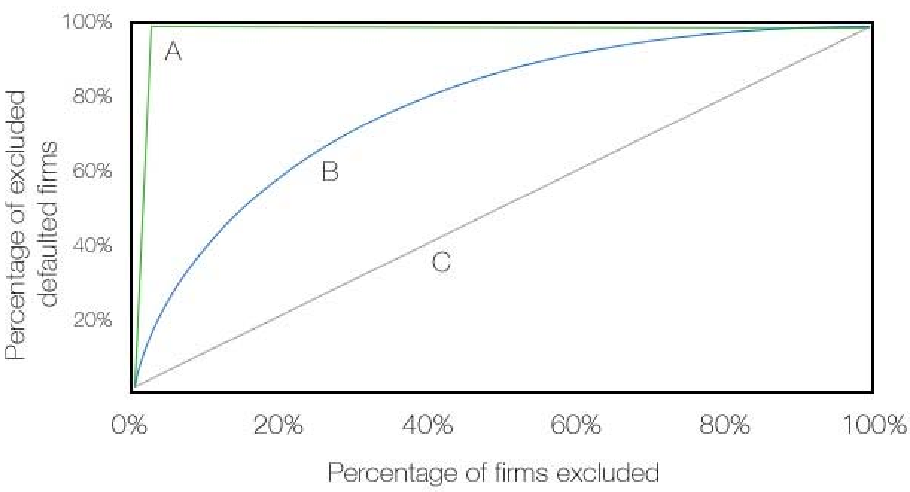

Figure 4.2: Three sample power curves: perfectly discriminating variable (A), reasonably well discriminating variable (B), non-discriminating variable (C).

Usually a power curve has the shape of the B curve: if the B curve above showed the power curve of the Liabilities/Total Assets ratio, then the x-axis would contain the percentages of firms whose Liabilities ratio is worse than a given cut-off point.

[18] See Chapter 6 for more details on model validation

In other words, as one moves from the worst value of the ratio to the best and at the same time from left to right on the x-axis, the percentage of firms will increase – the higher the cut-off ratio, the more firms will have a worse leverage. The y-axis plots, for any given cut-off point of the variable, the percentage of firms that actually defaulted and had a lower leverage ratio than that specified by the cut-off point. For example, curve B indicates that if the 20% of the worst performing firms in terms of leverage ratio are excluded from the dataset (x-axis), then approximately 60% of all defaulted firms from the dataset would be excluded. Curves A and C are extreme cases: curve A is a curve of a perfectly discriminating ratio – the kink is the point that separates the defaults from the non-defaults. As a result, if the population default rate is 3% and one excludes the worst 3% of the firms from the sample, based on that perfectly discriminating variable, all defaulted firms are identified. Curve C, on the other hand, is a power curve of a variable that has no discriminating power at all – for every cut-off, there is always the same proportion of defaults to non-defaults. Given these three exemplary curves, it is clear that the more northwesterly a curve lays, the higher its discriminating power.

The discriminating power can also be expressed in terms of a single number: the accuracy ratio[19]. In the example above, the accuracy ratio of the leverage ratio (curve B) would be calculated as:

$$\frac{\textit{(Area below curve B – Area below curve C)}}{\textit{(Area below curve A – Area below curve C)}}$$

The accuracy ratio sets the discriminatory power of the studied variable in relation to the power of a perfectly discriminating variable.

Alternatively, the discriminating power of individual variables can be judged by estimating a regression for each variable individually. Based on these regressions, power curves are constructed in a similar way, the only difference being that instead of using the value of the ratio itself, one utilizes the result of the regression, i.e. the estimated univariate default probability associated with a particular variable.

[19] The accuracy ratio is sometimes also referred to as *Gini Coefficien.t*

In either case, the accuracy ratio is a popular tool to choose which variables will be considered in the regression model estimation. In terms of accuracy ratio, Fernandes (2005) only allows variables to enter multivariate analysis if they have an accuracy ratio of over 5%. Altman and Sabato (2006) use the two ratios from each ratio category with the highest accuracy ratio for the regression.

It has to be pointed out, that power curves and accuracy ratios for a single variable will be different across models and they do not represent a universal truth. Instead, the results one obtains from these discriminative power measures will depend on the sample being used in model development. For example, in samples containing firms where the leverage ratios are somewhat constant across the sample, the power curves for that ratio will have little information value and the accuracy ratios will be low. On the other hand in other samples, the ratio may very well have an effect on default and generate high accuracy ratios.

Finally, outliers, i.e. extremely high or low values pose a major threat and can bias the eventual result. Eliminating these values would lead to a reduction of data and thus it is advisable to restrict variables that are below the 1st percentile and above the 99th percentile to the 1st percentile value and 99th percentile value, respectively[20]. Altman and Sabato (2006) deal with the high variability of values in their SME model by logging all input variables and thus limiting the high range of variable values: estimating two logistic models using non-transformed ratios and logarithmically transformed ratios, the model with the logged inputs has an accuracy ratio of 87% compared to 75% of the other model.

4.1.2 Dealing with the non-linearity assumption in logit/probit models

While the process of identifying suitable financial data inputs will remain the same regardless of the functional form of the model, the process of transforming variables in order to deal with the non-linearity assumption between the independent variable and the logit/probit is an issue that only concerns models of that functional form. However, since the majority of current default prediction models are logistic, the possible solutions to this challenge will be discussed.

[20] RiskCalc (2000) truncates input ratios at the 2nd and 98th percentile.

Altman and Sabato (2006) do not address this issue while Moody's KMV (2000), OeNB (2004) and Fernandes (2005) do take this problem into account with different methods.

Non-linearity is dangerous, because in its presence the regression will underestimate the relationship between the independent variable and the dependent. A good example to illustrate this fact is sales growth[21]. In one interval, from negative sales growth to certain positive values of sales growth, the relationship to default probability (and thus to the Logit/Probit) is negative: the higher the sales growth, the lower the PD. Nevertheless from a certain point, when sales begin to grow excessively, the PD starts to increase. There are many reasons for this phenomenon: sometimes the firm's expectations, based on the high growth, are too optimistic and it starts taking up too much debt financing. In addition, rapid growth presents numerous internal challenges to a company, such as expansion of the staff, new sales locations or a bigger management team. All these uncertainties cause the relationship to default become positive and higher sales growth will result in higher PDs. Eventually, the plot of the sales growth to the logit/probit will be a U-shaped curve, which is definitely not linear. A regression of this relationship will try to fit a straight line to depict the relation between the growth and the logit/probit. Since the actual relationship has the shape of a U, the fitted line will be rather flat with almost no slope. This is an indication of no or very weak relationship even though in reality there is a strong relationship.

Initially, the variables have to be checked for non-linearity. This can be accomplished by constructing graphs similar to the default frequency graphs where the x-axis remains the same and instead of plotting the empirical default probabilities, the empirical LogOdds (ln[p/(1-p)]) are plotted. The graph will indicate a non-linear relationship. It is also suggested to use a smoothing filter to smooth the data and reduce noise. The most widely used filter in credit risk literature is the Hodrick and Prescott[22] filter which was developed to eliminate short-term fluctuations in macroeconomic time series. Using the filter, one obtains

[21] Both Moody's KMV (2000) and Halling & Hayden (2004) cite sales growth as a good example of non-linearity between dependent and independent variable.
[22] See Hodrick & Prescott (1997) for further details.

an even clearer visual representation of the relationship between a variable and the logit. A more statistical method to evaluate the linearity in logistic models is using the Box-Tidwell[23] methodology: terms that multiply each independent variable with its natural logarithm are added to the regression, i.e. for each x_i, there is also an $x_i ln(x_i)$ term in the regression. If subsequent analysis shows that these terms are significant, nonlinearity is present.

The Moody's KMV (2000) solution to the nonlinearity problem is based on the univariate default frequency relationships. Instead of using non-transformed inputs into their probit model, they use the corresponding univariate default frequency for that particular variable. In essence, the inputs are derived directly from the smoothed default frequency graphs for each input variable. So for example, if firms with a leverage ratio of 20 percent had an empirical default rate of 0,5 percent, then 0,005 would be used as an input instead of 0,2. Moody's KMV (2000) uses the univariate relationship between Net Income/Assets and default to demonstrate how the nonlinearity is captured – above a certain point this relationship is flat, i.e. no matter how high the profitability is, the default probability stays flat and does not decrease. As a result, when using the transformed inputs which reflect the univariate level of the default probability, the effect of a very high Net Income/Assets ratio will stay the same above a certain point. In a similar fashion, the OeNB (2004) bank analysis model[24], which is based on a logistic regression, uses the smoothed (utilizing the Hodrick-Prescott Filter) univariate relationship between a variable and the empirical LogOdd as basis for the transformation. Afterwards the empirical LogOdds are used as inputs for variables that do not display a linear relationship with the LogOdd.

Another possibility to capture nonlinearity is by the use of polynomial expansion. For his logistic model, Fernandes (2005) applies the Fractional Polynomial Methodology developed by Royston and Altman (1994). In essence, this methodology attempts to find the best curve for a particular relationship between an independent and a dependent variable using polynomials that are limited to

[23] See Box & Tidwell (1962) for further details.
[24] This model is not further discussed in this thesis as it is aimed at default prediction of banks – a segment which differs substantially from the corporate segment. Most IRB banks acknowledge this fact by creating separate models for banks and corporate customers.

only a small set of values, which should be sufficient to model the most common curve shapes. In a default prediction model with *k* explanatory variables, where variable *X* is suspected to be non-linear, it can be transformed into a fractional polynomial of the following form:

$$\phi_m(X;\beta,p) = \sum_{j=1}^{m} \beta_{j+k-1} H_j(X)$$

where

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

Here, *m* denotes the number of polynomial functions and *p* the power of function *j*. To illustrate, a model with three explanatory variables (i.e. *k* equals 3) – *X, Y, Z* – where only *X* is non-linear and one would use two polynomial functions (i.e. *j*=2) with p_1=0.5 and p_2=1 would have following form:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 Y + \beta_2 Z + \beta_3 \sqrt{X} + \beta_4 X$$

Based on their experience, Royston and Altman suggest restricting the possible powers to *p* = {-2, -1, -0.5, 0, 0.5, 1, 2, …, max(3, *m*)} and *m*=2 at most. The optimal values for *p* and *m* are found using an iterative process, described by Fernandes (2005). Firstly, a normal logit model is estimated with the non-linear variable treated as if it were linear. Afterwards, the first set of models with fractional polynomials is estimated: *m* is set to one and for each *p* from the above set a model is estimated. The model with the lowest deviance[25] is then selected. Then the second set is estimated where *m* is set to two and for each combination of *p* a separate model is again estimated. The model with the lowest deviance is again selected. In the next step, the three estimated models (the regular logit model, the optimal *m*=1 and *m*=2 models) are compared via the likelihood ratio test. The optimal model is the one which has a significantly better fit than the

[25] See Chapter 6.1 for a description of deviance and likelihood ratio

model with a lower degree of *m* and not a significantly worse fit than the model with a higher degree of *m*.

4.1.3. Final variable selection

Once the univariate analysis helps identify the most appropriate variables, the optimal model with the optimal combination of these pre-selected variables has to be found. As Moody's KMV (2000) point out, from 20 independent variables one could create over one million possible models[26]. It is also not advisable to put all possible explanatory variables into the model, as it would then suffer from overfitting. An overfitted model produces very good prediction results when applied to the observations from within the sample used to estimate the model. However, it will have poor out-of-sample prediction results, i.e. when applied to completely new data it will have very low prediction reliability. The reason for this is multicollinearity: the presence of high correlation between some explanatory variables – for example, while the inclusion of all possible profitability measures in a model may reflect all aspects of profitability, the high correlation of these measures will cause serious instability when applied to out-of-sample data. A good way to eliminate this problem is to create a correlation matrix and check for pairs with very high correlation. After a model has been estimated, counter-intuitive signs of coefficients are hints of multicollinearity.

Hence, a major goal lies in the achievement of parsimony for the model. In fact, the models discussed in this thesis do not have more than ten input ratios. Moody's KMV (2000) mention many research papers and conclude that the optimal number of independent factors should be approximately seven. As a result, since it is suboptimal to use all variables from the univariate analysis and the combinations of possible models are numerous, automatic variable selection procedures, which are described below, have been devised.

For the commonly applied functional form of logit/probit[27] models the forward selection, backward selection as well as stepwise regression selection procedures exist. Forward selection starts with the estimation of univariate models for each

---

[26] If k denotes the number of candidate variables for possible inclusion, $2^k - 1$ models can be fitted.
[27] For their Hazard model, Halling and Hayden (2004) use the stepwise regression selection procedure.

variable that was selected to be a potential candidate for inclusion. Of those models, the one with the most significant variable, found via the likelihood ratio test, is selected as the starting point for further testing. In the next step, one estimates all possible two-variable models – the most significant variable found previously is kept and another variable from the set of the remaining candidates is added. These two-variable models are then compared to the one-variable model selected in the first step based on their likelihood ratios. The two-variable model that has the highest likelihood ratio and also is significantly different[28] from its one-variable counterpart then replaces the one-variable model. Further variables are added in the same way until the *n+1* variable model is no longer significantly different from the *n* variable model identified one iteration before.

Backward elimination is the opposite: at first, the largest possible model that includes all candidate ratios is estimated. Subsequently, models where one variable is dropped are estimated. Using likelihood ratio tests, the least significant variable, i.e. the one with the lowest significance and where there is also not a significant difference between the full and the reduced model, can be eliminated. This procedure is continued until no more variables can be dropped because the *n* variable model is significantly different from the *n-1* model.

Finally, stepwise regression is a combination of both forward selection and backward elimination. Stepwise regression starts with forward selection. At each iteration, once a variable has been added, all variables already in the model are checked for the possibility of elimination by means of the backward elimination. This process is continued until no additional variables can be added or removed.

Ultimately, for a predefined significance level, these procedures should help the modeler find a final form for the logit/probit prediction model. However, even after the model has been estimated, the coefficients should be checked for plausibility, i.e. ratios with a negative relation to default (such as profitability ratios) should also have negative coefficients.

[28] If it is not significantly different from the one-variable model, despite a higher likelihood ratio, this implies that the additional variable does not have a significant influence on the dependent variable.

With respect to linear discriminant analysis, similar procedures to the ones described above for logit/probit models exist, the main criterion being the common statistical F-Test.

4.1.4 Variables selected into models

This subchapter will highlight the different statistical models and the inputs that make up those models. As already mentioned previously, they cover the main factors that predict default, however, the actual ratios and weights differ substantially.

i) Variables selected into logit/probit models

Altman and Sabato's SME model

In their model, Altman and Sabato (2006) specifically tackle the default prediction problem for small and middle-sized firms in the United States. They make the case that banks should have different models for large corporates and SMEs: smaller firms have lower asset correlation but are in turn much riskier than their larger counterparts. As a result, the logistic model is based on a sample of firms whose asset size does not exceed $65 million. Eventually they present a set of five ratios that should be the most representative for SME default prediction in the United States. As already mentioned, the inputs were logged in order to constrain the large variability of the inputs.

	Factor	Ratio	Coefficient
SME Model	*Profitability*	-LN(1 - EBITDA / Total Assets)	4,09
		-LN(1 - Retained Earnings / Total Assets)	4,32
	Leverage	LN(Short Term Debt / Equity)	-1,13
	Liquidity	LN(Cash / Total Assets)	1,84
	Debt Coverage	LN(EBITDA / Interest Expenses)	1,97
	Constant		53,48

Table 4.1: input variables in Altman and Sabato's SME model for the US.

Fernandes' default prediction model

While not specifically focusing on small and medium-sized Portuguese firms, the dataset is dominated by SMEs, as they make up 95% of the dataset. Fernandes (2005) notes that the sample was chosen in a way as to approximate the overall structure of the Portuguese economy. An interesting aspect of his logistic default prediction model is the fact that the possibility of differentiation between different industries is studied. To achieve this, the dataset is split based on whether firms belong to the manufacturing and primary activity industry group or trade and services group and for each of the two subsamples a separate model is estimated and then compared to a model which was estimated on the entire dataset. While the three estimated models differ in the input variable composition, the results show only a marginally and statistically insignificant improved discriminating power of the models specifically aimed at the two industries. Fernandes (2005) thereby concludes that due to the increased demands on modeling effort by the industry-differentiating models and the low additional discriminating benefits, the non-differentiating model is superior. Hence only the inputs of this model are presented. The factor representing activity – Interest and similar expenses / Sales – was found to have a non-linear relationship with the logit and was thus transformed using the fractional polynomial method as described above.

	Factor	Ratio	Coefficient
Fernandes Model	*Liquidity*	Current Assets / Short Term Liabilities	-0,171
		(Bank deposits & Cash + Marketable Securities) / Total Assets	-0,211
	Activity	(Interest and Similar Expenses / Sales)^3	1,843
		(Interest and Similar Expenses / Sales)^0,5	-0,009
	Debt Coverage	(Current Earnings + Depreciation) / Interest and Similar Expenses	-0,231
	Productivity	Personnel Expenses / Sales	0,124
	Constant		-3,250

Table 4.2: input variables in Fernandes' Model for Portuguese Companies.

Moody's KMV RiskCalc for private firms

The various versions of RiskCalc are numerous and Moody's KMV base their analysis on large amounts of balance sheet data from the countries their cover. They also use the probit model instead of the logit model in two country-specific adaptations, namely in the Australian and US versions. In terms of the weights of individual factors and the exact form of the financial ratios used, these models are indeed very different. However, there are common threads in all country-specific versions of the model: they are not intended for analysis of firms with annual turnover of €500.000 or less and total assets of €100.000 or less. In case of very small firms the credit quality often not only depends on the state of the firm itself but also on the finances of the individual running the firm. Consequently, the lines between the business side and the personal side blur. Too large firms are also not targeted as the really large ones are usually listed in stock exchanges and for those, market data allows for a more accurate default prediction. Finally, financial institutions, real estate development firms, public sector institutions and holding companies are also not the focus of RiskCalc as their balance sheet peculiarities make them unsuitable for analysis together with private companies. As RiskCalc is proprietary, the exact coefficients are not made public. Despite this fact, Moody's KMV does publish the relative contributions of each factor in a model. The relative contribution is a measure of how a particular factor influences the eventual result. It is important to note however, that one should be careful when comparing the relative contributions of the same factor between countries: as some of the models have been designed by different individuals, sometimes the same financial ratios are assigned to different factors. For example, the ratio Gross Profit & Loss / Interest Expense and the very similar EBIT / Interest Expense is classified under the "Other" factor in the Japanese model, under "Debt Coverage" in the Mexican model and finally under "Profitability" in the US version. Reclassifying the ratios to achieve consistency would be impossible without losing the relative contributions of the various factors. After all, one factor may include more than one ratio but the contributions are only available for the factors alone. Hence, one cannot assign the relative contribution to a factor after a ratio has been moved from one factor classification to another. As a result, the ratios and their assignments to factors have been taken over from the original documentation.

	Factor	Ratio	Rel. Contribution
Australia	*Profitability*	Net Income / Total Assets EBIT / Interest	15%
	Leverage	Total Liabilities / Total Assets Retained Earnings / Total Assets	35%
	Liquidity	(Current Assets – Inventories) / Current Liabilities Cash / Total Assets	33%
	Activity	Inventories / Sales	7%
	Size	Assets	5%
	Debt Coverage		0%
	Growth	Sales Growth	5%
		Net Income Growth	
	Other		0%
Austria	*Profitability*	(Ordinary P&L + Depreciation) / Net sales	17%
	Leverage	Equity / Total Liabilities (Trade Liabilities + Bank Liabilities + Notes payable) / (Total Liabilities + Provisions for risks and charges)	29%
	Liquidity		0%
	Activity	(Trade Liabilities + Notes payable) / Net sales	16%
	Size		0%
	Debt Coverage	Ordinary P&L / Interest and similar expenses (Net P&L + Depreciation) / Accounts payable	24%
	Growth	Sales Growth	8%
	Other	Cash at bank and in hand / Current Assets	6%
Belgium	*Profitability*	Ordinary P&L / Total Assets	18%
	Leverage	Equity / Total Liabilities Retained earnings / Total Assets	25%
	Liquidity	Cash&Equivalents / Current Liabilities (Current Liabilities - Cash - Investments) / Total Assets	20%
	Activity		0%
	Size		0%
	Debt Coverage	(Ordinary P&L + Depreciation) / Financial Expenses (Net P&L + Depreciation) / Current Liabilities	27%
	Growth		0%
	Other	Cash / Current Assets	10%
France	*Profitability*	EBITDA / Sales (Net P&L + Tax) / Total Assets	21%
	Leverage	Equity / Assets (Total Liabilities - Cash&Equivalents - Advances) / Total Assets	30%
	Liquidity	Cash & Equivalents / Total Assets	6%
	Activity		0%
	Size		0%
	Debt Coverage	(Net P&L + Depreciation) / Total Liabilities	10%
	Growth	Asset Growth Sales Growth	20%
	Other	Financial Expenses / Sales	13%
Germany	*Profitability*	EBITD / Total Assets Ordinary P&L / Sales	25%
	Leverage	(Trade Liabilities + Notes Payable + Bank Liabilities) / (Total Liabilities - Advances) (Current Liabilities + Cash & Equivalents) / Total Assets (Equity - Intangible Assets) / (Assets - Intangible Assets - Cash & Equivalents - Land & Buildings)	38%
	Liquidity		0%
	Activity	[(Notes Payable + Trade Liabilities)*360] / Sales	10%
	Size		0%
	Debt Coverage	Cash Flow / (Total Liabilities - Advances)	9%
	Growth	Sales Growth	7%
	Other	Personnel Expenses / Sales	11%

	Factor	Ratio	Rel. Contribution
Italy	*Profitability*	Ordinary P&L / Total Assets	17%
	Leverage	(Equity - Intangible Fixed Assets) / (Total Assets - Intangible Fixed Assets) (Total Liabilities - Liquid Funds) / Total Assets	25%
	Liquidity		0%
	Activity	Interest and Other Financial Expenses / Sales	15%
	Size		0%
	Debt Coverage	(Ordinary P&L + Depreciation) / Interest and Other Financial Expenses (Net P&L + Depreciation) / Total Liabilities	23%
	Growth	Sales Growth	9%
	Other	Liquid Funds / Current Assets	11%
Japan	*Profitability*	Ordinary P&L / Total Assets	15%
	Leverage	Total Liabilities / Total Assets	29%
	Liquidity	Cash / Current Assets	13%
	Activity	Total Inventories / Sales	8%
	Size	Sales	8%
	Debt Coverage	Retained Earnings / Current Liabilities	7%
	Growth		0%
	Other	Gross P&L / Total Interest Expense	20%
Mexico	*Profitability*	(Gross P&L/Avg. CPI) / (Total Assets/CPI)	5%
	Leverage	[(Total Liabilities/CPI) + (Retained Earnings/Average CPI)] / [(Total Assets - Fixed Assets)/CPI]	8%
	Liquidity	Cash / Total Assets	5%
	Activity	[(Total Inventories + Total Accounts Receivable)/CPI + (Pre-paid Expenses/Average CPI)] / (Cost of Goods Sold / Average CPI)	14%
	Size	(Total Assets - Fixed Assets) / CPI	14%
	Debt Coverage	Gross P&L / Total Interest Expense	17%
	Growth	Sales Growth	9%
	Other	Short Term Notes / Cash	28%
Netherlands	*Profitability*	Net P&L / Total Assets	16%
	Leverage	Equity / Total Assets	19%
	Liquidity	(Current Liabilities - Liquid Funds) / Total Assets	15%
	Activity		0%
	Size		0%
	Debt Coverage	Ordinary P&L / Total Liabilities Operating P&L / Current Liabilities	34%
	Growth		0%
	Other	Liquid Funds / Current Assets	16%
Portugal	*Profitability*	Net P&L / Total Assets	17%
	Leverage	Equity / Total Accounts Payable Bank Debt / Total Liabilities	21%
	Liquidity	Current Assets / Accounts Payable	11%
	Activity	Interest and Similar Expenses / Sales	17%
	Size		0%
	Debt Coverage	(Ordinary P&L + Depreciation) / Interest and Similar Expenses (Ordinary P&L + Depreciation + Provisions) / Total Liabilities	34%
	Growth		0%
	Other		0%
Scandinavia	*Profitability*	Pre-Tax P&L / Total Assets	20%
	Leverage	Total Liabilities / Total Assets (Current Liabilities - Cash at Bank and in Hand) / Total Assets	34%
	Liquidity	(Current Assets - Current Liabilities) / Total Liabilities	6%
	Activity		0%
	Size		0%
	Debt Coverage	Ordinary P&L / Financial Expenses EBITDA / Current Liabilities	25%
	Growth		0%
	Other	Cash at Bank an in Hand / Current Assets	15%

	Factor	Ratio	Rel. Contribution
Singapore	*Profitability*	Operating P&L / Total Assets Net Worth / Total Interest Expense	26%
	Leverage	(Total Liabilities - Cash & Marketable Securities) / Assets Retained Earnings / Current Liabilities	24%
	Liquidity	Cash&Marketable Securities / Assets	13%
	Activity	Current Liabilities / Sales	10%
	Size	Assets	14%
	Debt Coverage		0%
	Growth	(Total Liabilities / Net Worth) Growth	13%
	Other		0%
Spain	*Profitability*	Ordinary P&L / Total Assets	5%
	Leverage	Retained Earnings / Total Liabilities (Total Liabilities - Cash) / Total Assets	28%
	Liquidity	Cash / Current Liabilities	4%
	Activity	Sales / Total Assets	25%
	Size		0%
	Debt Coverage	Cash Flow / Total Liabilities Cash Flow / Current Liabilities	34%
	Growth	Financial Expenses / Sales	4%
	Other		0%
UK	*Profitability*	Net P&L / Total Assets	18%
	Leverage	Total Liabilities / Total Assets (Current Liabilities - Cash) / Total Assets	29%
	Liquidity	Cash / Total Assets	14%
	Activity	Trade Creditors / Sales	9%
	Size		0%
	Debt Coverage	Ordinary P&L / Total Liabilities (Ordinary P&L + Depreciation) / Interest Expense	25%
	Growth	Sales Growth	5%
	Other		0%
USA	*Profitability*	Net Income / Total Assets Net Income Growth EBIT / Interest Expense	23%
	Leverage	Total Liabilities / Total Assets Retained Earnings / Total Assets	21%
	Liquidity	(Current Assets - Inventories) / Current Liabilities Cash&Equivalents / Total Assets	19%
	Activity	Inventories / Cost of Goods Sold	12%
	Size	Total Assets	14%
	Debt Coverage		0%
	Growth	Sales Growth	12%
	Other		0%

Table 4.3: Input variables for different versions of Moody's KMV RiskCalc suite for private companies.

ii) Variables in a hazard model: Halling and Hayden (2004)

As already discussed, Halling and Hayden (2004) estimate a logistic model enhanced by the hazard component which takes time explicitly into account. The first model, which is purely logistic and serves the purpose of classifying firms into healthy and risky firms, utilizes the following inputs:

	Factor	Ratio	Coefficient
Pure Logit	*Leverage*	Equity / Total Assets	-0,85
		Bank Liabilities / Total Assets	1,21
		Short-Term Debt / Total Assets	1,56
		Liabilities / Sales	1,53
	Activity	(Sales - Material Costs) / Personnel Costs	-0,23
	Constant		-0,95

Table 4.4: Variables selected into the pure logit model by Halling and Hayden (2004).

Interestingly, leverage plays an important role in classifying firms into healthy and risky categories. The second model is clearly different in that time is considered a factor and it is more balanced in terms of factors as both leverage and profitability are considered:

	Factor	Ratio	Coefficient
Hazard	*Time*	Dummy: Period 2	0,88
		Dummy: Period 3	1,47
	Profitability	Ordinary P&L / Total Assets	-2,15
	Leverage	Liabilities / Total Sales	1,57
	Activity	(Sales - Material Costs) / Personnel Costs	-0,34
	Constant		-2,32

Table 4.5: Variables selected into the hazard logit model by Halling and Hayden (2004).

iii) Variables in a linear discriminant model: Altman's Z-Score (1968)

The venerable Z-Score was the first model to simultaneously include multiple predictor variables. Today it still serves as a benchmark for other models in terms of their discriminating power. Altman (1968) intended it to be a default prediction tool for manufacturing companies that neither had total assets below $1 million nor above $25 million. The original version required the input of the market value of equity; however, Altman (2000) also proposed a variant that substitutes the market value of equity for its book value. Since the majority of unrated firms are private, the book value variant will be presented here. For the cut-off score that differentiates between good and bad firms, Altman (2000) proposes 1.23 for the revised model. It is mainly its historical importance that is responsible for the inclusion in the thesis as due to today's requirements on default prediction models, the Z-Score can only serve limited purposes. The main reason for this is, of course, the very nature of the discriminant analysis as described in Chapter 3. Apart from that, it was created in the late 1960s and its age also significantly reduces the application as the overall economic environment was different compared to today.

	Factor	Ratio	Coefficient
Z-Score	*Profitability*	Retained Earnings / Total Assets	0,717
		EBIT / Total Assets	0,847
	Leverage	Book Value of Equity / Book Value Total Debt	3,107
	Liquidity	Working Capital / Total Assets	0,420
	Activity	Sales / Total Assets	0,998

Table 4.6: Altman's Z-Score model input variables.

iv) Variables in the model used by a bank in Austria: the Bank Austria-Creditanstalt internal ratings model

The specifics of the model design, including the functional form, and parameters are confidential, but nonetheless, BA-CA (2005) does reveal the input variables that form the basis for the quantitative evaluation part of their ratings system for corporate customers. Only after the evaluation of the qualitative factors the firms are assigned a rating grade and a default probability.

	Factor	Ratio	Coefficient
BA-CA	*Profitability*	Net Income / Total Assets (ROI)	N/A
	Leverage	Equity / Total Assets	N/A
		Accounts payable to credit institutions / Total Liabilities	
	Activity	Cash Flow / Sales	N/A
	Debt Coverage	EBITDA / Accounts payable to credit institutions	N/A
		Cash Flow / (Total Liabilities - Liquid funds)	N/A
	Size	Sales	N/A

Table 4.7: Input variables for the quantitative evaluation of a firm in the BA-CA internal rating system for corporate customers.

v) Concluding observations

Taking a look at the composition of various models, it becomes very clear that a clear-cut one-size-fits-all solution in terms of quantitative variable selection for the statistical default prediction problem does not exist. With the exception of the Scandinavian RiskCalc version, models are estimated for individual countries. There seem to be large enough differences between the individual economies that would make a broader model less well performing. For example, the distributions of ratio values vary from one country to another. In addition, accounting standards are not the same for every country with sometimes different interpretations of various balance sheet items. However, there is no differentiation based on industries within an economy, even though in this case, too, the ratio values can exhibit varying distributions. The model by Fernandes (2005) shows only marginally better performance when industry differences are taken into account, hence, subsequently disregards industries. Moody's KMV acknowledge in some country models that there are sector-based differences, e.g. higher volatility of the construction industry. Still, only one model per country is developed and no sector dummy variables are included as validation results prove robustness across industries. Instead, as part of calibration, firms belonging to certain industries receive a bonus (penalty) which results in a lower (higher) default probability than the standard model output.

For three countries, there is more than one model presented (Austria: BA-CA, the Halling & Hayden Hazard model and RiskCalc; Portugal: Fernandes and RiskCalc, USA: Z-Score, SME model and RiskCalc) and these models are anything but identical. It becomes clear that the scope of the model as well as the underlying dataset have a major influence on the final form of a model. The three US models are very clearly aimed at different firm size segments, with the SME model having a limit of $65 million on total assets and RiskCalc also including firms that can be classified as large, but are not publicly traded. In addition, the datasets each span different years. Even the two Portuguese models whose scope is comparable only have two ratios in common. The Fernandes model does not include any distinct profitability or leverage measure. In this case, the datasets again span different time intervals and the size distribution within the datasets is different for the two models.

In terms of functional form influencing the variable selection, no statements can be made as there are very few models that utilize discriminant analysis or work with a hazard component in order to compare them with the numerous logistic models. However, from the models available, there does not seem to be a tendency of alternative functional forms to prefer certain types of factor, with the exception of time in the case of hazard models as a result of their design.

Finally, the modeler also has a large impact on the eventual model. After all, it is him/her who defines the first set of candidate ratios and who is responsible for the various modeling techniques and approaches that are applied.

Observing the relative contributions of factors as given by Moody's KMV, one pattern becomes apparent: the most important factors are leverage, profitability and debt coverage, which by itself is essentially a combination of the first two factors. The table below illustrates this fact: with the exception of Mexico (which is the only developing country on this list) and the USA, the three factors together account for at least 50% of the influence on default. The influence is even more pronounced in Europe, where the relative contribution ranges from 61 percent in France to 79 percent in Scandinavia. And while there are no relative contributions given for the other models that are also examined, this pattern seems to continue as well, since the majority of the input variables fall into the three aforementioned categories, the only exception being the Fernandes model.

		Profitability + Leverage + Debt Coverage	Liquidity	Activity	Size	Growth	Other
Europe	*Austria*	***70%***	0%	16%	0%	8%	6%
	Belgium	***70%***	20%	0%	0%	0%	10%
	France	***61%***	6%	0%	0%	20%	13%
	Germany	***72%***	0%	10%	0%	7%	11%
	Italy	***65%***	0%	15%	0%	9%	11%
	Netherlands	***69%***	15%	0%	0%	0%	16%
	Portugal	***72%***	11%	17%	0%	0%	0%
	Scandinavia	***79%***	6%	0%	0%	0%	15%
	Spain	***67%***	4%	25%	0%	4%	0%
	UK	***72%***	14%	9%	0%	5%	0%
Rest of world	*Australia*	***50%***	33%	7%	5%	5%	0%
	Japan	***51%***	13%	8%	8%	0%	20%
	Mexico	***30%***	5%	14%	14%	9%	28%
	Singapore	***50%***	13%	10%	14%	13%	0%
	USA	***44%***	19%	12%	14%	12%	0%

Table 4.8: Relative contributions of profitability, leverage and debt coverage factors combined compared to the other factors, sorted by two distinct regions.

4.2 Inclusion of Qualitative Inputs in Default Prediction

While highly objective, the exclusive use of accounting data has its drawbacks. It is strictly backward looking and it can only take into account information that is expressed in numerical terms. As a result, banks also turn to their own employees' credit expertise to judge the soft facts of a firm when evaluating it. In addition, Basel II also emphasizes the use of human judgment in the internal rating process. This subchapter will only give a broad overview on the subject since its origins are mainly rooted in the development of scorecards for private individuals[29]. In addition, due to the lack of data and its subjective nature, the use of statistical methods is only of limited use.

[29] Private individuals cannot provide accounting data. As a result, mostly qualitative factors are used as inputs into the evaluation of their creditworthiness.

The literature on the effectiveness of qualitative data is very sparse due to the fact that expert judgment data on soft facts is not disclosed by banks and historical qualitative data is usually not even stored at all while balance sheet data is available more readily. Lehmann (2003) analyzed whether the inclusion of soft facts considerably improved the default prediction ability of banks' internal rating models. Her analysis was based on a sample of 20.000 observations, which included both quantitative and qualitative data, of German small and middle-sized enterprises. The author constructs two models from the data. One model uses quantitative data exclusively while the other combines a rating based on the quantitative, hard facts with a rating based on the evaluation of soft facts. Her findings confirm that soft facts should not be disregarded. Firstly, when rated using the partial qualitative scoring, the mean of the non-default sample differs significantly from the mean of default sample. When the qualitative rating is compared to the quantitative rating, there is no clear dominance of either one, as their power curves intersect: for higher quality, low PD borrowers, the qualitative model performs better. On the other hand, for lower quality, high PD borrowers, the quantitative model performs better. However, the combined model dominates the purely quantitative model – the coefficient of concordance[30] for the combined model is significantly greater than that of the purely quantitative model.

In order to gain insightful information from soft facts, usually a bank has a questionnaire that covers the most important areas of a company and predefined answer possibilities that ensure that the information filled out by the various credit experts is standardized and comparable. For example, for its qualitative rating, BA-CA (2005) uses school grades (one to five) to evaluate individual factors. However, it is extremely important that an extensive documentation regarding the qualitative evaluation is provided to the credit experts to guarantee that the rating remains as objective as possible. There have to be clear and detailed rules in place that make sure that specific grades are only given when certain criteria are met. After all, the more vagueness there is behind qualitative grades, the more room there is for the subjective opinion of the expert.

[30] A measure similar to the accuracy ratio.

It has to be pointed out, that a structured selection process such as the one described for quantitative inputs is not possible. The main reason for this is that the subjective nature of the soft-fact information and evaluation makes it very hard to create a historical analysis with respect to past default experience. Usually, historical data on soft-fact scores will rarely be available (which is also the reason why there are no models in academic research that utilize qualitative data). As a result, the selection process of the various questions and topics that are part of the qualitative evaluation is especially driven by the expert knowledge and past experience of credit experts.

Nonetheless, in practice several key areas have been identified to have an impact on the creditworthiness of a borrower. The following list is based on rating guidebooks of three important credit institutions in Austria – Bank Austria Creditanstalt (2005), Erste Bank (2003), Raiffeisen (2005) – and summarizes the most common factors in qualitative evaluation.

1. Management:
    - Education
    - Experience, both overall professional as well as industry experience
    - The existence of a detailed long-term strategy that outlines the goals of the company and the plans how to reach them
    - The existence of a succession plan, which is an important issue especially in smaller companies where the management team is smaller and the departure of a single manager can already cause a serious loss of leadership, experience and know-how.
2. Accounting, reporting and controlling systems:
    - Presence of audited balance sheets
    - Quality of balance sheets, especially in terms of detailedness
    - Existence of efficient planning within the company, thorough and continuous comparisons of target and actual performance as well as defined consequences for future actions
3. Organization of the company's activities:

- Existence of an organizational charts, descriptions of employees' tasks
- Clearly defined processes that ensure the smooth running of day-to-day operations

4. Marketing, sales:
   - Evaluation of the company's distribution channels
   - Evaluation of advertising activities
   - Existence of market studies
5. Products/services:
   - Evaluation of the modernity of the firm's equipment
   - Evaluation of the quality of the company's products or services
   - Does the company invest in innovation/research?
   - Diversification and structure of the product/service portfolio
   - Evaluation of current order intake as well as the utilization of the production capacity
6. Market and market development:
   - Is the company affected by changes in economic cycles?
   - Positioning within the market (i.e. is the firm a market leader?)
   - The competitive situation in the market
   - The degree of dependence on individual customers or suppliers
7. Relationship with the bank
   - The willingness to provide information has a major influence on the credit quality as it directly influences the relationship between the bank and the firm
   - To what extent does the bank make use of other products of the bank?
   - The past behavior of the company when dealing with the bank, such as behavior when the client was temporarily unable to meet credit obligations

In order to combine the aforementioned qualitative factors into a qualitative score, a logit or probit regression can be performed as these can cope with categorical variables.

4.3 Combining quantitative and qualitative scores – an illustrative example

The process of combining qualitative and quantitative scores is a complex issue and highly dependent on the functional forms of the models that provide the respective scores. As such, it will not be presented in detail. Nevertheless, an illustrative example will be given to outline the most commonly applied process of arriving at a single score for one borrower.

The extent to which the qualitative rating score influences the overall rating and thus the default probability usually depends on the size of the company being rated. Both BA-CA (2005) and Erste Bank (2003) put a higher weight on the qualitative rating when the company is small and this weight gradually becomes smaller as the size of the company increases. In addition, there are also specifically predefined warning signs or negative information whose presence triggers an automatic up- or downgrade of the calculated rating. Such an event or information is for example the absence of a new balance sheet in past 18 months, which will result in the company being automatically downgraded to a specified score, regardless of the calculated, actual rating. At BA-CA, subsidiaries receive the rating of their parent companies by default. BA-CA also enables overrulings at different stages of the rating process by the credit analyst if he/she deems the calculated rating unsuitable. However, the reasons for overrulings have to be significant and well documented. Figure 4.3 shows the rating process at Bank Austria Creditanstalt.

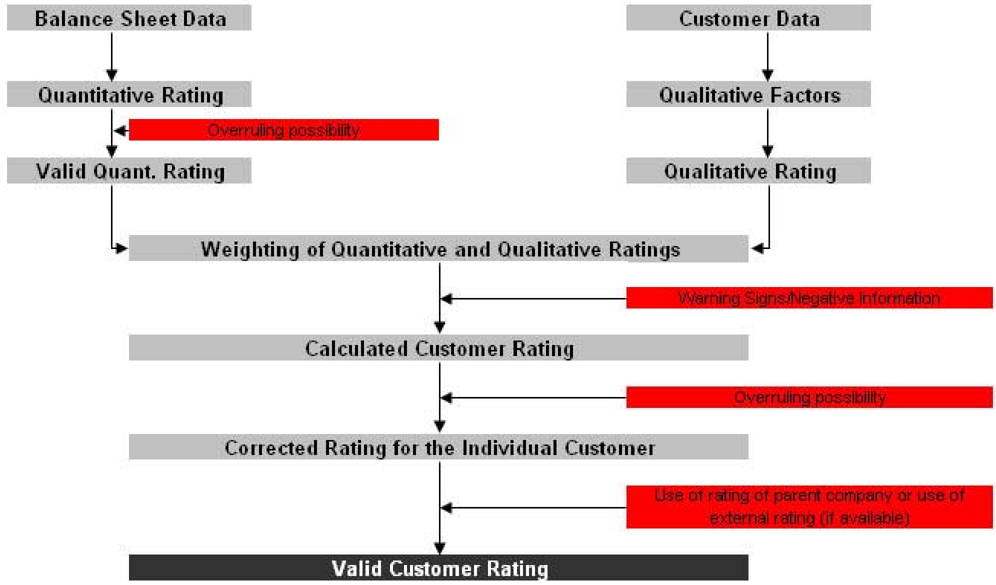

Figure 4.3: The rating process at Bank Austria Creditanstalt. Source: Bank Austria Creditanstalt: Unternehmensfinanzierung im Wandel – Rating als neuer Marktfaktor (2005).

5. Calibration

The previous chapter dealt with the selection of the proper input variables in order to estimate a model that is capable of the highest possible degree of discrimination between good and defaulting firms. Without a doubt, strong discrimination ability is the single most important attribute of a good model. Nevertheless, under Basel's IRB Approach PDs have to be estimated for each borrower. As a result, the main goal of calibration is to assign a default probability to any given model score[31]. Calibration establishes the link between model score and the output as stipulated by Basel II and should ensure that the final model output, the PD, will as accurately as possible reflect the realized default rates.

Obviously, logit and probit models provide percentage outputs. However, it is very often the case that such models have been estimated using development samples where the proportion of defaulted and non-defaulted firms does not reflect the default rate for the particular segment. The mean estimated score for all observations of the development sample will equal the default rate of the development sample. When applied to the actual, existing portfolio, the output of the model will produce either too high or too low default probabilities. In most

[31] See OeNB/FMA (2004)

models, it is the case that the development sample contains a higher ratio of defaulted firms than in reality in order to better capture and incorporate the information from the balance sheets of the defaulted firms. Table 5.1 shows the one-year aggregate default probabilities for private, middle-market firms as surveyed by Moody's KMV at the time of creating the RiskCalc suite:

		1 year PD
Europe	*Austria*	2,0%
	Belgium	1,5%
	France	2,2%
	Germany	1,6%
	Italy	2,1%
	Netherlands	1,7%
	Portugal	1,5%
	Scandinavia (Denmark)	1,7%
	Scandinavia (Finland)	1,9%
	Scandinavia (Norway)	2,1%
	Scandinavia (Sweden)	2,2%
	Spain	1,6%
	UK	2,0%
Rest of World	*Australia*	1,7%
	Japan	1,2%
	Mexico	5,1%
	Singapore	1,8%
	US	1,7%

Table 5.1: One-year aggregate default probabilities for private companies. Source: Moody's KMV RiskCalc models.

The table clearly illustrates the need to use disproportionately higher numbers of defaulted firms in model development: with average default probabilities hovering around 2%, a sample that contains 1000 non-defaulted firms would only contain approximately 20 observations of bad firms. While this would ensure that the final model's output would reflect real-world probabilities, it would probably not be able to discriminate well between good and bad firms due to the relative lack of information about unhealthy firms.

However, as already mentioned, in cases where relatively more defaults have been included in a model, the output will result in too high estimates of default probabilities. From a cautiousness perspective it is not a serious mistake to assign firms higher default probabilities. However, this would result in higher regulatory capital requirements and it would also not be in line with Basel II, which states that estimated default probabilities have to be accurate. As a result, the next step in statistical credit risk modeling is calibration which establishes a link between the model's output and the default probability so that it corresponds to the aggregate default probability.

The actual process of calibration is started by determining the accurate long-run aggregate default probability for the particular firm segment. The simplest solution is to take the entire segment population over at least two years[32] and divide the number of defaults by the total number of observations and finally divide this fraction by the number of years. However, there are several issues that have to be taken into account. The aggregate default probability will be the future reference point for the model. As a result, it is vital that it is calculated from a population of firms that is similar in composition and structure to the portfolio that will be eventually subjected to the model. Otherwise, the simple method described above will lead to wrong PD estimates. The time span is also important. For example, if the default rate is computed based on a limited period, one runs the risk of having a reference PD that is too high if the underlying period was in recessionary years[33] – this is especially true when using a data history of only two years. As a consequence, banks that have collected data on defaults only recently will also turn to external sources for aggregate default estimates[34]. However, the use of external data again raises the issue of portfolio applicability. As can be seen, a lot of considerations and adjustments have to be made before an aggregate PD is set for a portfolio. At any rate, the bank will have to prove to its regulator that the established aggregate PD is applicable to the portfolio it will be applied to.

In a next step, a sample of firms that reflects the composition of the bank's portfolio for which the model has been developed has to be specified first in order to ensure the most accurate calibration results. This calibration sample then has to be rated using the developed model and firms have to be sorted based on their score. This ranked sample is to be divided into several groups for each of which the default frequency has to be determined by dividing the number of defaults with the total number firms within a particular bucket. Usually, the grouping is

[32] In the initial stages of IRB implementation, PDs estimated on data history of at least two years have to be used. In successive years after the IRB implementation, this PD has to reflect at least five years worth of default experience.
[33] For example, Moody's KMV use country-specific PDs that are estimated over long time spans that are through the cycle and as a result, they capture the entire business cycle, with both recessions and expansions.
[34] A popular source of external default data in Austria is the Kreditschutzverein (KSV)

accomplished by creating buckets which each contain the same number firms[35]. The number of groups can vary and it also depends on the sample size. However, the fewer groups, the more ineffective the calibration will be. On the other hand, if too many groups are chosen, one runs the risk of not having enough observations within each group, which results in very inaccurate default frequencies. This step is important, since it ties the model output to actual, experienced default rates. Every group, characterized by a score interval, will have a default rate. However, if the mean default rate of the calibration sample is different from the aggregate default rate, which was set a step earlier, the default rates for each group have to be scaled.

According to OeNB/FMA (2004), in order to accomplish this scaling, the default frequencies (both for the individual groups as well as the calibration sample as a whole) and the aggregate default probabilities have to be transformed into odds; in this case they are called relative default frequencies:

$$RDF = \frac{PD}{1 - PD}$$

After this transformation, the modeler will have following relative default frequencies:

- $RDF_{bucket\ i}$ – the relative default frequency of group i within the calibration sample. This is the unscaled RDF for a particular group.
- RDF_{CS} – the overall, unscaled, relative default frequency of the entire calibration sample.
- $RDF_{aggregate}$ – the relative default frequency of the aggregate default probability to which the model is being calibrated.

Using these three relative default frequencies, the adjusted, scaled relative default frequency for every group of scores, $RDF_{adj,.\ bucket\ i}$, can be calculated:

[35] An alternative in terms of creating groups is to define equally spaced score intervals. However, this alternative is only suitable if the number of defaults in the intervals with higher scores is not significantly lower than in the worse score intervals – this is not the case in default prediction models.

$$RDF_{adj.,\ bucket\ i} = RDF_{bucket\ i} \frac{RDF_{aggregate}}{RDF_{CS}}$$

This equation, which is an adaptation of Bayes' Theorem, scales every default frequency of each group within the calibration sample in order to achieve the long-run aggregate default probability that was established prior to the calibration for the particular portfolio. Of course these, RDFs have to be retransformed into PDs.

However, the equation does not yet establish a link between any model output and the calibrated default probability. It is also possible that some groups which should have a lower PD based on the scores, will be assigned a higher PD than a group with worse scores[36]. As a result, in the next step, a regression is performed which ensures that every model score can be assigned a default probability that is in line with the aggregate default probability and that this relationship is smoothed. This way, better scores will always lead to lower PDs.

First, the average scores of the outputs from every group/bucket have to be calculated. Each average score can then be linked with the corresponding adjusted probability of default in the following way, using an exponential function:

$$\ln(PD_{adj.,\ bucket\ i}) = \alpha + \beta * Score_{avg.,bucket\ i}$$

Once the regression has determined the optimal α and β, for every score an appropriate default probability can be calculated:

$$PD = e^{(\alpha + \beta * Score)}$$

While the focus so far has mainly been on logit/probit regression models, where the output can already be interpreted as a default probability, it is important to note that this calibration method works for every output any model on default prediction produces.

[36] This is caused by inaccuracies in the model.

In practice, banks have universal rating scales, resembling those of Standard & Poor's or Moody's, onto which the calculated default probabilities are mapped. As a result, each counterparty receives a rating grade. Rating scales not only comply with the Basel II rules of assigning PDs to borrower grades. They also simplify internal communication of default analyses due to the aggregation of exposures into a manageable number of rating classes. In addition, all classes also have a verbal description of the credit risk tied to them. This simplification is also necessary in order to prevent classification of exposures based purely on the estimated default probability. This is the consequence of default prediction models not being completely accurate. Thus, it is more reasonable to assign a middle PD of a particular rating class to several exposures whose estimated PD falls into the PD interval of that rating class instead of assuming 100 percent accuracy of the PD estimation. Table 5.1 shows the rating scales of two major Austrian banks and how their rating classes relate to the Standard & Poor's rating classes:

BA-CA		S&P
1	*1+*	AAA/AA+
	1	AA
	1-	AA-
2	*2+*	A+
	2	A
	2-	A-
3	*3+*	
	3	BBB+
	3-	BBB
4	*4+*	BBB-
	4	
	4-	BB-
5	*5+*	
	5	BB
	5-	
6	*6+*	BB-
	6	
	6-	B+
7	*7+*	
	7	
	7-	B
8	*8+*	B-
	8	CCC
	8-	
9	*9*	
10	*10*	

Erste Bank	S&P	1 year PD
1	AAA	0,00%
2	AA+	
	AA	0,01%
	AA-	
3	A+	
	A	0,05%
	A-	
4a	BBB+	
4b	BBB	0,35%
4c	BBB-	
5a	BB+	0,52%
5b	BB	1,16%
5c	BB-	2,07%
6a	B+	3,29%
6b	B	9,31%
7	B-	13,15%
8	CCC	27,87%
R (Default)	D	

Table 5.1: Master Rating Scales used at Bank Austria Creditanstalt and Erste Bank. The corresponding 1 year default probabilities for BA-CA were not available. Sources: Bank Austria Creditanstalt (2005); Erste Bank (2003).

Table 5.1 also shows that banks cannot just take over the external rating grades of the large rating agencies as this would imply that banks' portfolio population is similar to the population of all rated companies. Moody's KMV (2000) point out that the average bank loan is comparable to Ba2[37]. As a result, the rating grades have to be adapted to accommodate this fact. For example, BA-CA uses a finer and more detailed grading between the BBB- and CCC external grading interval. Erste Bank, which utilizes a less detailed master scale, subsumes the AAA to A-interval in just three buckets.

6. Model validation

Besides model development and calibration, the validation process constitutes an equally important step in the construction and implementation of default prediction models. In the context of default prediction, validation encompasses all means, methods and techniques that test whether a satisfactory and stable categorization of exposures based on their default risk profile has been achieved by the model. The intuition is clear: even the most sophisticated and theoretically sound model is useless until it has shown that it can discriminate between good and bad credits. It is not just in the own interest of banks to have efficient validation procedures in place – even regulators stress the importance of validation as paragraph 500 of the Basel II document[38] clearly demonstrates:

"Banks must have a robust system in place to validate the accuracy and consistency of rating systems, processes and the estimation of all relevant risk components. A bank must demonstrate to its supervisor that the internal validation process enables it to assess the performance of internal rating and risk estimation systems consistently and meaningfully."

The actual validation consists of three dimensions, each of which will be outlined in more detail in the next subchapters:

1. *Validation of the fundamental statistical soundness of the model:* these are general statistical tests that indicate the significance of the entire

[37] Ba2 is equivalent to BB in Standard & Poor's rating grades.
[38] Paragraphs 500 – 505 of the Basel II document contain minimum requirements for validation processes for banks under the internal ratings based approach.

logit/probit regression and individual coefficients of the independent variables. These tests are not specifically aimed at ensuring that default is predicted correctly and they will not provide explicit results with respect to differentiation between good and bad companies.

2. *Validation of the discriminating power of the model:* in this process the ability of the model to discriminate between high risk and low risk companies is evaluated.
3. *Validation of the calibration of the model:* calibration is also subjected to testing as the estimated default probabilities have to be constantly compared to actual default frequencies to ensure an accurate PD estimation. This is also the least developed area in model testing with no truly established testing procedure.

It is important to note that the various validation techniques are not solely intended to be applied after the model is constructed. Many of the methods are already used in model development itself, such as the use of power curves in variable selection or the likelihood ratio in forward selection, backward elimination or stepwise regression.

6.1 Validation of the fundamental statistical soundness

The most commonly applied tests for logit/probit models revolve around the deviance and the likelihood ratio test. The major statistical software programs provide these figures when a regression is performed. The deviance of a logit/probit is calculated by multiplying the log-likelihood of the maximum likelihood estimation with -2. This value, also denoted as -2LL, approximately follows a Chi-Square distribution with (K-J-1) degrees of freedom, whereby K represents the number of observations and J the number of factors. Using the deviance, one can test the null-hypothesis that a model has perfect fit, which is the case when the likelihood equals one and thus when -2LL equals zero, given a chosen significance level α: if the χ^2 value is lower than that of the significance level of α, it can be assumed that the model has good fit. However, the likelihood ratio test is more frequently used in the context of logit and probit regressions. The reason lies in one major drawback of the deviance: regressions of data with highly skewed distributions of the dependent variable[39] tend to have higher deviances than regressions with more balanced distributions of the dependent variable[40].

The likelihood ratio test, which can be compared to the F-Test used in linear regression analysis, can be used to evaluate the entire model as well as individual independent coefficients. If used to judge a whole model, the likelihood ratio compares the -2LL of the model in question (also known as the full model) with the -2LL of a model with all the independents having coefficients of zero, i.e. this model consists only of the constant (the null model). The likelihood ratio test tests whether the factors have a sufficiently large influence on the dependent variable. The null hypothesis proposes that the coefficients equal zero and the number of independent variables also equals the degrees of freedom. If the absolute value of the difference between the -2LL of the full and the null models is higher than the χ^2 value for a given significance level, then the independents are assumed to have a significant impact on the dependent variable.

This test can also be used to assess the significance of individual independents. In this case, the -2LL of the full model is compared to the -2LL of a model where the coefficient for the particular independent is assumed to be zero. An insignificant result means that there is no substantial difference between the two models and thus the independent in question can be removed from the regression.

Logit/probit regression analysis lacks a measure that describes the goodness of fit in a way that the R^2 does in linear regression[41]. Several alternatives, also known as Pseudo-R-Square measures, have been proposed and the majority are reinterpretations of the relationship between the log-likelihood of the full (LL_F) and zero models (LL_0). One example[42] is McFadden's R^2:

[39] Default prediction data is a good example of a highly skewed distribution of the dependent variable: the vast majority of credits are good while usually only a single-digit percentage of the entire dataset are firm defaults.
[40] For more details, see Backhaus et al. (2003).
[41] More specifically, R^2 describes the fraction of variation that is accounted for by the model, often expressed in percentage terms.
[42] For other Pseudo-R-Square measures, see Backhaus et al. (2003).

$$McFaddenR^2 = 1 - \frac{LL_V}{LL_o}$$

Clearly, if the two log-likelihoods are very similar, McFadden's R^2 will be close to zero, indicating low fit. According to Backhaus et al. (2003), a McFadden's R^2 in excess of 0.2 indicates good fit.

6.2 Validation of the discriminatory power

One possible means of validation of the discriminatory power of models was already described in Chapter 4 – the power curve in combination with the accuracy ratio. Obviously, this method is also applied to whole models and not just on individual variables. The power curve is a very popular method to assess the differentiating abilities of various models.

One point that has to be stressed, however, is that a higher (equal) accuracy ratio, regardless of whether it is utilized in the assessment of individual variables or whole models, does by no means suggest absolute superiority (equality) of one variable or model over another. The accuracy ratio only summarizes the entire area under a respective power curve. Nonetheless it is possible that two power curves of two separate models intersect, as the following figure illustrates:

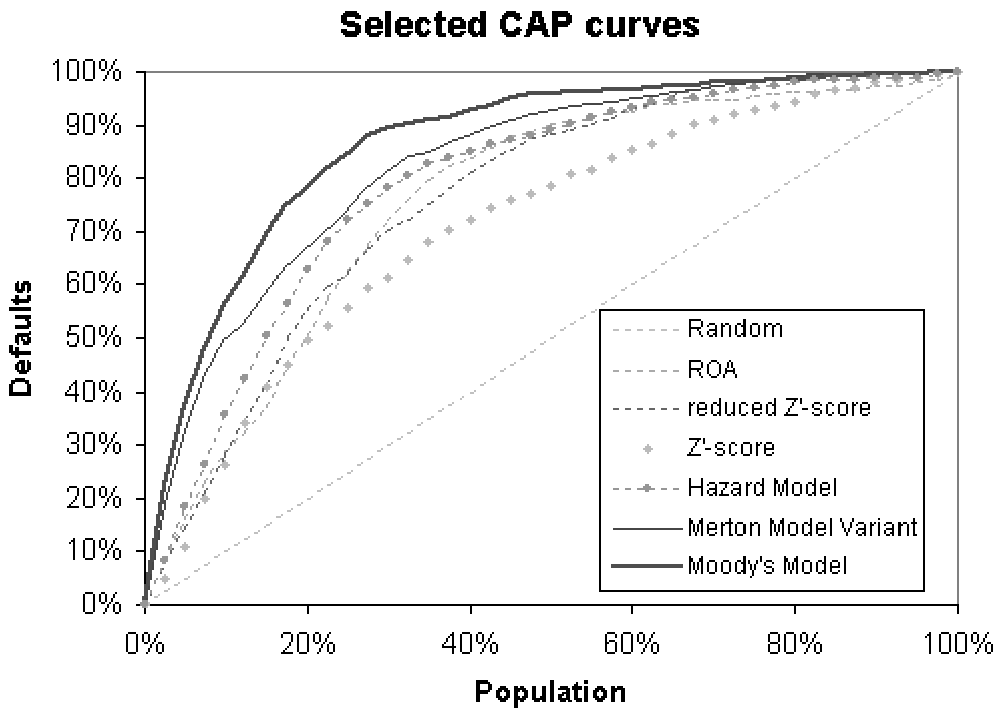

Figure 6.1: Power curves for various models, as tested by Moody's KMV. Source: Sobehart et al., *Benchmarking Quantitative Default Risk Models: A Validation Methodology*, Moody's Investor Services, 2000.

In the figure above, the reduced Z-Score model power curve intersects the ROA model power curve at a population of approximately 30 percent: up until that point the reduced Z-Score power curve lies above the ROA curve indicating a better discrimination ability of the reduced Z-Score. For the rest of the better rated sample, the ROA model has the better power. Thus, even though the Accuracy Ratios are similar for both models, their performance varies for different segments of the sample.

In addition to the power curve, the Receiver Operating Characteristic (ROC) curve is also frequently used in the validation of models. While the graphic representations of both the CAP plot and the ROC curve are similar in their shape, i.e. concave curves, the ROC curve is derived differently.

ROC curves take advantage of the idea of α and β errors (type I and type II errors) at different cut-offs of scores of a given default prediction model. A cut-off in this context is a score based on which loans are classified as good or bad, i.e. a higher score than the cut-off will suggest a good credit. In any model, the distributions of good and defaulted firms will overlap[43]. As a result, with the exception of the most extreme scores (best and worst), for every cut-off there will be defaulted firms classified as good (i.e. α error) and good firms classified as defaulted and thus rejected (i.e. β error). ROC curves use the concept of "Hit Rate" and "False Alarm Rate". The Hit Rate is the amount of correctly predicted defaults for a given cut-off divided by the total amount of defaults in the sample. The False Alarm Rate, on the other hand, is the number of predicted defaults that are healthy firms in reality for a given cut-off divided by the total amount of defaults in the sample. As a result, for every possible cut-off, there is a pair Hit and False Alarm Rates. Finally, the ROC Curve is constructed by plotting all False Alarm Rates on the x-axis and all Hit Rates on the y-axis. Figure 6.2 depicts three sample ROC Curves:

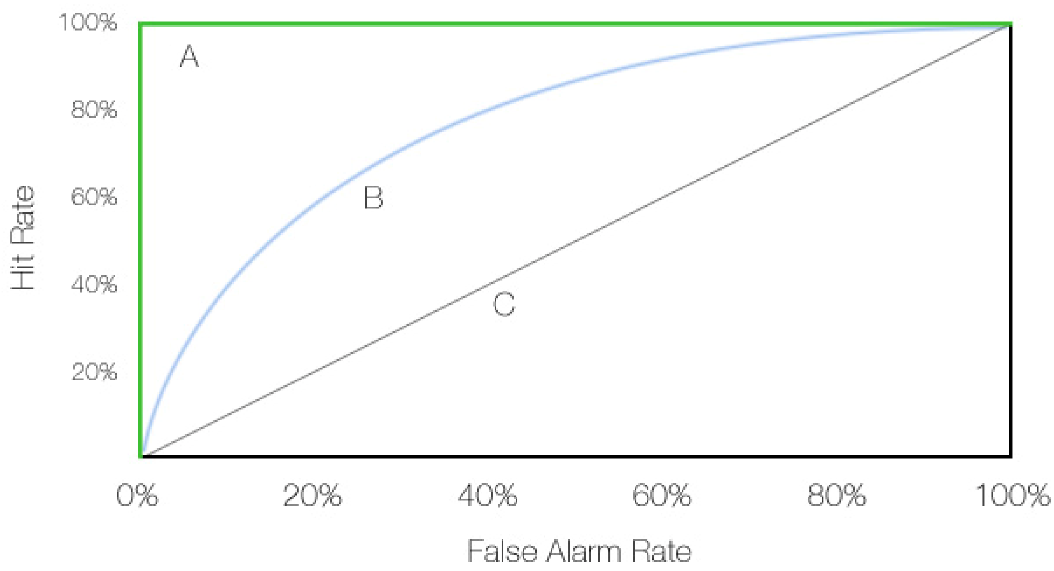

Figure 6.2: Three sample ROC Curves: perfectly discriminating model (A), reasonably well discriminating model (B), random model (C).

The steeper the ROC curve is at its left end and the nearer it is to the point (0% false alarm rate, 100% hit rate) – i.e. the more northwesterly it lies – the better a rating model discriminates between defaulters and healthy firms.

A similar measure to the power curve's accuracy ratio is the "area under curve" (AUC). The area under curve measures the area under a particular ROC curve. Since the ROC curve of the random model has an AUC equal to 0.5 and the AUC of the perfect model is one, a good model should have AUC values in the interval

[43] See Figure 3.1.

between 0.5 and 1, with higher AUC values preferable to lower ones. Nevertheless, models with different or equal AUC can have different differentiation power over different intervals, just as is the case in power curves and their accuracy ratios.

There is even a relationship between the area under the curve and the accuracy ratio (AR), as Engelmann et al. (2003) show:

$$AR = 2AUC - 1$$

When comparing the accuracy ratios or areas under curve of different models, it is important to remember that these comparisons are only meaningful if these figures are derived from evaluations of models based on the same sample. If model A has a higher AUC than model B, but model A was used to rate firms from sample C and model B evaluated firms from sample D, then the AUC has no explanatory power with respect to which model has the higher discriminating ability. Only if models A and B are used to rate the same sample, say sample C, the AUCs can be directly compared.

There are other methods to test the power of default prediction models, such as the Bayesian Error Rate or the Conditional Informational Entropy Ratio[44]. However, due to their easy and clear graphical interpretation as well as the possibility to express discriminatory power in one, highly understandable number, the power curve as well as the ROC curve are the preferred and predominantly used ways to evaluate the discriminatory power of models. The Basel Committee on Banking Supervision also advises to use these two methods[45].

In terms of actual validation of models using either the power curve or the ROC curve, it should take place on at least two levels. First of all, the model has to be applied to the data that was used in the models development, this data is referred to as the development sample. Secondly, data that was not used in any phase during the development of the model, also known as the holdout sample, should

[44] For a more detailed discussion of alternative discriminatory power measures, see Basel Committee on Banking Supervision (2005).
[45] See Basel Committee on Banking Supervision (2005).

be scored with this model. The main reason for this approach is to test the power of the model in the face of completely new data which the model has not been fitted for.

In an ideal world, the holdout sample consists of data that spans a period different than that of the development sample and data obtained from another source than the one which provided the development sample[46]. However, this is a luxury most modelers do not have. As a result, usually the initial sample is divided into the development sample and a holdout sample which is to be used exclusively in validation. This division can be made across time so that if, for example, a sample that spans the years 2000 until 2005, the years until 2004 are used in the development and the rest is used for validation. Alternatively, a certain percentage of the initial sample is randomly selected and thus the test sample spans the same period as the development sample. The SME model by Altman and Sabato (2006) uses the former approach. As previously noted, the development sample contains 2010 firms over a period from 1994 until 2002. Their holdout sample contains 432 firms, 26 of which were defaults, over the two years following 2002. Fernandes (2005) also follows this pattern: his model is estimated on 11.000 financial statements between 1996 and 2000 while it is tested on 301 observations from the year 2003. This observed strategy obviously has one main advantage: the predictive ability of the model over time can be tested: in the case of the Fernandes Model, it was estimated on data from a period of economic upturn while it was tested on data from the post September 11th period with much lower economic growth. Moody's KMV does not disclose the composition of development and validation samples of their RiskCalc models.

Table 6.1 presents the validation results, expressed in accuracy ratios, of the Altman and Sabato SME Model, the Fernandes Model as well as the various RiskCalc editions. These accuracy ratios are not directly comparable as the validation samples are different for each model, however, each model presented the accuracy ratio of the old Z-Score when applied to the respective testing sample

[46] Of course, there are limitations to the extent to which the holdout sample differs from the development sample. A holdout sample from a period in which economic conditions were dramatically different than in the current period or where the composition of firms in the holdout sample reflects a completely different firm segment clearly does not make sense and would lead to bad validation results.

model as a benchmark, another testament to the importance of the Z-Score in corporate default prediction.

		Model	Z-Score
Altman & Sabato SME Model		75,4%	68,8%
Fernandes Model		50,6%	22,2%
European RiskCalc Models	*Austria*	54,7%	34,5%
	Belgium	67,2%	46,7%
	France	76,3%	50,1%
	Germany	59,7%	30,2%
	Italy	67,6%	47,6%
	Netherlands	64,1%	49,6%
	Portugal	61,1%	22,2%
	Scandinavia (Denmark)	71,0%	60,3%
	Scandinavia (Finland)	75,1%	65,7%
	Scandinavia (Norway)	75,8%	64,9%
	Scandinavia (Sweden)	60,3%	43,2%
	Spain	64,3%	42,0%
	UK	58,5%	52,1%
Other RiskCalc Models	*Australia*	39,7%	27,7%
	Japan	69,4%	41,8%
	Mexico	37,2%	31,3%
	Singapore	60,8%	49,4%
	US	54,1%	45,5%

Table 6.1: Available Accuracy Ratios for models discussed throughout the thesis.

Clearly, each model beats the benchmark. However, there is a visible variability in the accuracy ratios between countries/models. There may be several reasons for a high variability in the values of accuracy ratios. One is data availability: for example, the Austrian RiskCalc model is based on a sample of 19.524 firms while the French RiskCalc is based on a sample of 253.268 firms while covering roughly the same time span. Intuitively, the more data there is available to estimate a model the more information can be included and more specific effects can be captured in the model. Another possible reason might be data quality as not each balance sheet from the data sample will be audited and, necessarily, there will be incomplete statements. This quality may vary from dataset to dataset. Additionally, another possibility that may explain the high variability is the differing scope of the models. While the SME model by Altman and Sabato uses a relatively small development dataset, it is aimed exclusively at the SME segment and excludes all firms with assets larger than €50 million. On the other hand, the Fernandes Model or the RiskCalc models do not impose this restriction and also feature firms with assets higher than €100 million. It is more than reasonable to assume that a more narrowly defined dataset will feature firms with more homogenous characteristics. As a result, it will be easier to fit an accurate model.

Finally, differences in the corporate landscape between regions and countries may also play a role.

Nonetheless, only two of the twenty presented models have an accuracy ratio below fifty percent while thirteen are above sixty percent. According to OeNB/FMA (2004), multivariate logit/probit or discriminant models should achieve an accuracy ratio between sixty and seventy percent.

6.3 Validation of the calibration of default prediction models

The need to validate the PD calibration, sometimes referred to as backtesting, arises from two main factors. First of all, loan pricing decisions are tied to internal ratings which are in turn tied to assigned default probabilities. A systematic under- or overestimation of PDs leads to misclassifications and thus mispricings. Secondly and equally as important, for banks using one of the two internal ratings based approaches, the PD is the main component for determination of the regulatory capital requirements. If the estimated PD constantly underestimates the true PD, the bank will hold less capital than required by law. The other case where estimated PDs are higher than true PDs, while from a conservative standpoint irrelevant, will cause inefficiencies by tying up too much equity capital.

It is highly unlikely that the actual PDs for a given year will match exactly those estimated by a model. As a result, the calibration tests have to offer a measure that provides information about the extent to which the realized PDs can deviate from the estimated ones.

A significant issue in PD backtesting is the assumption of correlation between defaults. Correlation between defaults tends to increase the variability of default probabilities and thus widens the interval for an acceptable variation in PD estimates. There are tests that do take this correlation into account and there are tests that don't. The correlation assumption increases tolerance for variability in PDs. As a consequence, those tests that operate under the assumption of uncorrelated defaults will be more conservative in that they will report a significant deviation from estimated PDs even if in fact there is still some room for actual/estimated PD variation because defaults are correlated to some extent.

Hence, these tests can be applied, even if their underlying assumptions do not reflect reality, because they are stricter in judging calibration errors.

Nevertheless, it is important to point out that even the Basel Committee concedes that the validation of calibration is more difficult than the validation of discrimination power and that no really powerful tests of adequate calibration are currently available. The Basel Committee stresses that the best use of specific models depends on a given circumstance and that a combined use of multiple tests will be most appropriate. The following paragraphs describe the most commonly used calibration backtesting methods.

6.3.1 Calibration validation without default correlation assumption

The most basic method to measure calibration accuracy is by means of the Brier Score[47], which is the average squared difference between estimated default probability and the realized default/non-default state for every exposure in a credit portfolio:

$$BS = \frac{1}{N}\sum_{n=1}^{N}(p_n^{estimate} - y_n)^2$$

N is total number of exposures in a portfolio, $p_n^{estimate}$ is the model's estimated default probability for exposure n and y_n is a binary variable taking on a value of zero if exposure n did not default and the value of one if it defaulted in a given observation period. The Brier-Score will always lie in an interval between zero and one and a lower Brier-Score indicates more accurate PD estimates. However, especially in default prediction, where the percentage of defaults is very low, the Brier-Score will be a low number. The Basel Committee suggests overcoming this disadvantage by setting the Brier-Score in relation to a trivial model for the same portfolio: a trivial model is a model that assigns the overall default frequency of the portfolio to each exposure. The Brier-Score of a trivial model is computed as follows:

$$BS_{trivial} = p(1-p)$$

[47] See Basel Committee on Banking Supervision (2005).

In this case, p denotes the overall default frequency of the portfolio. Nevertheless, the major drawback of the Brier-Score is its inability to judge significance or acceptability of the variance between the estimated and realized default probabilities.

As a result, significance tests can be used to assess the extent of PD estimation deviance. These significance tests test the Null hypothesis that the PD estimation is correct against the alternative hypothesis that the aforementioned estimation is not correct. It has been suggested that the confidence levels used in these tests can be tied traffic lights for easier interpretability[48]. Under this so-called "traffic lights approach", deviations significant at the 95% level or lower, fall under the green light as these deviations are not assumed to be serious and do not warrant special attention to the calibration of a model. On the other hand, deviations significant at a level of 99,9% or higher are assumed to be very serious and fall under the red light resulting in a definite need to adjust the calibration or even completely recalibrate a model. Deviations falling between the above mentioned significance levels trigger a yellow light. While not as serious as in the case of red light deviations, there should be at least some attention devoted to the accuracy of a model.

Under the assumption that defaults are uncorrelated, the Basel Committee suggests using the Binomial test which is applied when testing the significance of variations within a population whose observation fall into two categories. Using this test, one can determine whether the realized default rate variation is significant for given significance levels q. In the context of traffic lights, the two significance levels of interest will be 95% and 99,9%. The number of defaults in a portfolio N^D out of total population of N firms will represent a significant deviation from the estimated default rate, denoted as $PD_{est.}$ for the predefined confidence level q if following condition is satisfied:

$$\sum_{n=0}^{N^D} \binom{N}{n} (PD_{est.})^n (1 - PD_{est.})^{N-n} > q$$

[48] See Tasche (2003).

As a result, if the term on the left-hand side is smaller or equals 0,95, the deviation of PD estimates can still be considered as acceptable and classified as "green". If it is higher than 0,999, the deviations will have to be considered as severe and classified as "red". Values higher than 0,95 and lower or equal 0,999 will trigger a yellow light.

When the number of exposures is sufficiently high, the binomial distribution can be approximated by the normal distribution[49]. In this case, a significant deviation in terms of underestimating the realized default probabilities is given when following condition is met:

$$(DF_{real.} - PD_{est.}) > \Phi^{-1}(q)\sqrt{\frac{PD_{est.}(1 - PD_{est.})}{N}}$$

$DF_{real.}$ denotes the realized default frequency and Φ^{-1} is the inverse of the standard normal distribution, the rest of the notation is the same as in the binomial test formula.

6.3.2 Calibration validation under the assumption of correlated defaults

As mentioned above, the binomial test and its approximation using the normal distribution do not take default correlation into account. This is certainly not the case in reality and even the Basel II risk weight functions take correlation into account. For example, OeNB/FMA (2004) state that typically this correlation lies between 0,5% and 3% while Tasche (2003) suggests 5% being appropriate for Germany. The following significance test that takes correlation into account and is described in OeNB/FMA (2004) is more complicated than its correlation disregarding counterparts. Here, ρ denotes correlation:

[49] More specifically, when $np(1\text{-}p) \geq 9$, where n is the number of observations and p the probability, then an approximation by means of the normal distribution is acceptable. See Brannath & Futschik (2001), p. 124.

$$DF_{real.} > Q + \frac{1}{2N}\left(2Q - 1 - \frac{Q(1-Q)}{\Phi\left(\sqrt{\frac{\rho\Phi^{-1}(1-q)-t}{\sqrt{1-\rho}}}\right)}\left(\frac{(1-2\rho)\Phi^{-1}(1-q)-t\sqrt{\rho}}{\sqrt{\rho(1-\rho)}}\right)\right)$$

where

$$t = \Phi^{-1}(PD_{est.});\ Q = \Phi^{-1}\left(\frac{\sqrt{\rho}\Phi^{-1}(q)+t}{\sqrt{1-\rho}}\right)$$

When the realized default frequency is larger than the right-hand term, the estimated PD is a significant underestimation of reality.

7. Alternative approaches to default prediction

This chapter will provide a brief overview of the available alternatives for default prediction. The decision between different models is mainly driven by the presence of market data. As already pointed out in Chapter 1, market data is preferred to accounting data. Accounting data is essentially a snapshot of the state of firm at a given time and reflects past performance. Market data, on the other hand is forward-looking, in the sense that market prices reflect market participants' expectations of the future performance of the firm. Thus, they implicitly incorporate the risk of defaulting. Another factor in favour of market data is their availability: while accounting data is updated yearly, at best quarterly, market data, provided the market in question is liquid enough, is available daily. Moody's KMV (2000) state that as soon as sufficient market data on a firm is available, the RiskCalc suite is no longer the most effective tool in assessing credit risk. In situations where only accounting information is available, neural networks, which can also incorporate market data, are viable alternatives. The most commonly used methods used in the presence of market data are structural and reduced form models.

7.1 Artificial Neural Networks

Artificial neural networks attempt to recreate the mechanics and the learning process of the human brain where received information is processed simultaneously by interconnected nerve cells, also called neurons. Typically, an artificial neural network consists of three main components, as figure 7.1 shows. The first component is the input layer where input data is acquired by the network. In a neural network designed to predict bankruptcy, this input data will be a vector of relevant information about the company, which can be of quantitative (market and accounting data) as well as qualitative nature. From the input layer this vector is passed on to the second component which consists of multiple inner layers of neurons. Each of these neurons weighs and transforms the incoming vector into one single value and then transmits this value to other subsequent neurons within the inner layer. Again, these subsequent neurons receive the outputs of the preceding neurons and weigh and transform them. Finally the output layer, consisting of just one neuron combines the received data into one output value, for example a default probability.

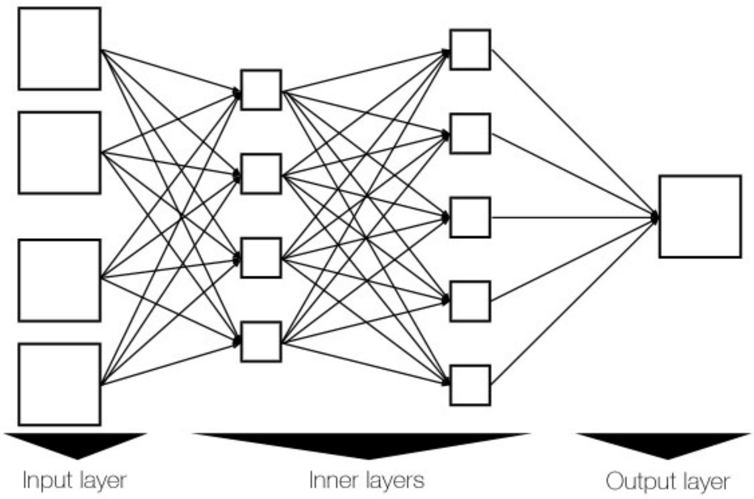

Figure 7.1: The structure of an artificial neural network.

The neural network has to be "trained" – i.e. it has to receive input data for which the output is already available. This way the optimal weights within the inner layer can be determined. The most popular training algorithm is the back-propagation algorithm[50]: the inputs are passed through the network and the resulting outputs are compared with the desired outputs – any difference is then used to adjust the weights within the inner layers. This process is repeated until the network can appropriately depict the relationship between inputs and the output.

[50] See Balcaen and Ooghe (2004).

The main advantage of artificial neural networks is their ability to analyse complex patterns in data[51] and as a consequence they can deliver good results. According to OeNB/FMA (2004), accuracy ratios of up to 80 percent are possible with neural networks. However, neural networks run the risk of being custom made for the training sample. As a result, they may be overfitted so that they incorporate development sample specific patterns in their analysis logic which are completely irrelevant out of sample. Another significant disadvantage is the "black-box" approach of these networks – they process the inputs into outputs, but in a way which is difficult to trace and understand. This leads to acceptance problems as people like to understand how a process works and are not just satisfied that the process works.

7.2 Structural Models

Structural models take advantage of a link between loans and optionality. First observed by Robert Merton in 1974, equity can be viewed as a call option on a firm's assets where total liabilities represent the strike price. The intuition is as follows: if a firm borrows D in debt and if at maturity the total asset value of the firm is larger than D, the equity holders can keep the residual, which, depending on the success of the firm, can be very high. On the other hand, if at maturity the asset value is below the debt borrowed, the equity holders don't earn anything since the assets were used to, at least partly, repay the loan. However, due to limited liability, the potential loss to the owners is limited. Just like in the case of an owner of a call option – limited downside risk with unlimited upside potential.

Formally and in a similar fashion as for european call options, the payoff to equityholders at maturity T can be denoted as:

$$E_T = \max(0, A_T - D)$$

A_T is the asset value at maturity and D is the amount of debt. The asset value is assumed to follow a Geometric Brownian Motion, where the change in Asset value, dA, is modelled in the following manner:

[51] See Balcaen & Ooghe (2004).

$$dA = \mu A dt + \sigma_A A dz$$

Here, μ is the asset drift, *dt* is one unit of time, σ_A is the volatility of assets and *dz* is a standard Wiener process. The term containing the drift specifies the general trend of the future development of the asset value while the term containing the asset volatility specifies the amount of variation from the trend and introduces a random component (the Wiener process).

Due to the link between debt and optionality, the default probability can be obtained via the Black-Scholes option pricing framework. According to Saunders & Allen (2002), in Merton's model the value of equity is a function of the firm's asset value *(A)*, its debt *(D)*, the volatility of assets *(*σ_A*)*, the maturity of the debt *(T)* and the risk-free rate *(r)* as can be seen from the following equations. The value of equity can be calculated using the equation below:

$$E = A\Phi(d_1) - De^{-rT}\Phi(d_2)$$

with

$$d_1 = \frac{\ln\left(\frac{A}{D}\right) + \left(r + \frac{\sigma_A^2}{2}\right)T}{\sigma_A \sqrt{T}}$$

$$d_2 = d_1 - \sigma_A \sqrt{T}$$

d_2 is also referred to as the distance to default, a measure of the riskiness of the firm. More specifically, it is the number of standard deviations the firm is removed from default – i.e. by how many standard deviations of the asset value can the asset value decrease before default is reached. $\Phi(-d_2)$ is the corresponding default probability estimate, often referred to as the expected default frequency.

The difficulty in the application of such models lies in the fact that the firm's asset value as well as the corresponding volatility of the assets are not directly observable. As a result, the calculation of the default probability is not possible without further assumptions[52]. In many applications of the model, a relationship between the volatility of equity – which is observable for publicly listed companies – and the volatility of assets is assumed. This relationship takes the form of another equation. As a result, one obtains two equations and two unknowns. For example, a method based on Ronn & Verma (1986) proposes the following relationship, with E and σ_E being the value of equity and the volatility of equity respectively:

$$\sigma_E = \frac{A}{E}\Phi(d_1)\sigma_A$$

From this brief introduction into structural models, it becomes apparent that they are not suitable for the evaluation of firms that are only able to provide accounting data. The main reason for this is the fact that two crucial inputs that by themselves are not directly observable – A and σ_A – are indirectly inferred from available market information. Structural models describe through their mechanics the process of default and why a firm defaults – this is their biggest advantage as they rely on a sound theoretical foundation. In addition, as Saunders & Allen (2002) point out, deteriorations in credit quality of high-profile defaults such as Enron were captured faster by a structural model than by agency ratings.

7.3 Reduced Form Models

Contrary to structural models, reduced form models, introduced by Robert A. Jarrow and Stuart Turnbull[53], do not model the process that determines default and they make no statements as to why it occurs. Instead, they derive the probability of default from market prices of debt. This process occurs by decomposition of the yields on risky bonds into the risk-free rate and a risk-premium. The PDs are obtained using risk-neutral valuation: in a risk-neutral world, all market participants are willing to accept an expected return on risky assets that equals the return on the risk-free asset. As a result, the current price of a risky asset, such as a bond, can be obtained by discounting the expected future

---

[52] The calculation is impossible because there is one equation (the value of equity) but two unknowns (asset value and its volatility).

[53] See Jarrow & Turnbull (1995).

pay-offs with the risk-free rate. To illustrate how the inference of PDs works, following simplified scenario is considered: credit risk is the only risk in a bond market, a risky zero-coupon bond with a maturity of 1 year has a face value of 100, is currently traded at a price of 92 and has a loss given default rate of 50% and the risk-free rate is 5%. The expected value for this bond will be 100*(1-PD*LGD) and by discounting this term with the risk-free rate (i.e. applying the risk-neutral valuation) it has to equal the current price, arriving at following relationship:

$$\frac{100(1-PD\times LGD)}{1{,}05}=92 \rightarrow PD=6{,}8\%$$

This is the basic intuition behind reduced form models: the expected value reflects the probability of default. In reality, this calculation is more complicated than illustrated above as the risk premium is a premium for various types of risk such as liquidity or political risks and not just credit risk. In addition the loss given default is in most cases unknown and the PD profile of a firm is hardly constant over time so additional modelling complexities have to be incorporated into reduced form models[54].

The main advantage of reduced form models, as opposed to structural models is their being fitted to existing bond price data – as Saunders and Allen (2002) point out *"they are data driven and should provide results that conform to the data better than structural models"*. Their drawbacks lie in their limited applicability as their scope can only cover the traded bond universe and the fact that bond spreads may arise not only due to credit risk but also due to other risks, such as the aforementioned liquidity risk, which are difficult to separate from each other.

9. Conclusions

The main focus of this thesis was to highlight the current practice of statistical credit risk models as a means for quantifying credit risk in banks' exposures. The

[54] For an excellent and more detailed introduction to reduced form models, please refer to Saunders & Allen (2002).

theoretical basis upon which these models build was detailed, the major challenges in model design were outlined and the possible solutions presented, the conversion of model output into Basel II relevant default probabilities was shown and finally, the relevant aspects of validation were discussed.

From the previous chapters and paragraphs it becomes apparent that statistical models lack the elegance and sophistication of Merton-based or reduced-form models and that their design relies heavily on the individual modeller. Neither are they perfectly discriminating tools that can exactly differentiate between good and bad borrowers as accuracy ratios hover around sixty to seventy percent for good models. However, in a borrower landscape dominated by small and middle-sized enterprises where any reliable market data is absent, they become a very attractive alternative to simple expert-based system where the credit-granting and pricing decision is made solely by the bank's credit expert. In addition, as outlined in Chapter 2, Basel II requires banks to have sophisticated internal ratings in order to take advantage of more risk-sensitive equity requirements. Statistical scoring models, especially logit/probit based ones, satisfy these conditions placed by the Basel Committee and as a result, banks wishing to implement the internal ratings based approach turn to such models for credit risk quantification of their commercial loan portfolios in the segments of small and middle-sized enterprises.

Hence, while academic literature dealing with credit risk has turned away from credit scoring and focused on more sophisticated models – and rightfully so, especially since the emergence of credit derivatives, the market for which surpasses several billions of dollars in notional amounts, which need to be priced correctly – the old way of estimating default risk via the use of statistical regressions will continue to play a major role in banking practice.

Appendix A – German Abstract

Statistische Kreditrisikomodelle (Zusammenfassung)

Die Implementierung des auf internen Ratings basierenden (IRB) Ansatzes für die Eigenkapitalunterlegung für das Kreditrisiko gemäß Basel II birgt zahlreiche Herausforderungen für Banken, darunter die interne Einschätzung des Kreditrisikoprofils eines Schuldners in Form von Ausfallswahrscheinlichkeiten. Dadurch, dass für die meisten Firmenkunden einer Bank keine externen Ratings und auch keine Marktdaten vorliegen, müssen Banken Modelle entwickeln, die mit Buchhaltungsdaten als Eingangswerten eine Evaluierung des Kreditrisikoprofils durchführen können. Sowohl in der Bankenpraxis als auch in der akademischen Literatur werden statistische Kreditrisikomodelle häufig verwendet um die Anforderungen von Basel II zu erfüllen und das Kreditrisiko von Firmenkunden zu quantifizieren. Das Grundprinzip statistischer Kreditrisikomodelle besteht darin, dass Bilanzdaten von lebenden und ausgefallenen Firmen über einen gewissen Zeithorizont benötigt werden um eine Analyse bezüglich des Einflusses verschiedener Faktoren – Bilanzkennzahlen – auf die Kreditwürdigkeit durchführen zu können. Anschließend werden die Faktoren mit dem stärksten Einfluss mittels statistischer Verfahren gewichtet und kombiniert um einen Score zu produzieren mittels dessen eine Aussage über die Ausfallswahrscheinlichkeit des Schuldners getroffen werden kann. In der vorliegenden Diplomarbeit werden die wichtigsten Aspekte solcher Modelle behandelt. Es werden die möglichen statistischen Verfahren, die den Modellen zugrunde liegen, vorgestellt und es wird gezeigt warum in der heutigen Bankenpraxis Logit- und Probitmodelle am häufigsten verwendet werden. Der Prozess der Variablenselektion wird dargestellt und Modelle sowohl aus der Theorie als auch aus der Praxis mit ihren verwendeten Inputvariablen präsentiert. Dabei ist zu erkennen, dass die Modelle sehr unterschiedlich sind und dass keine exakten Richtlinien existieren nach denen man ein perfektes Modell bauen kann und dass die endgültige Form eines Modells von dem Segment, für das man ein Modell entwickelt, vom verwendeten Datensatz und auch vom Entwickler selbst abhängt. Es wird jedoch deutlich, dass die meisten Modelle auf ein Land fokussiert sind, nicht nach Industrien differenzieren und die Inputvariablen hauptsächlich die Dimensionen Profitabilität, Fremdfinanzierungsgrad (Leverage)

und Fremdkapitalabdeckung (Debt Coverage) beinhalten. Weiters wird gezeigt wie man im Rahmen der Modellkalibrierung die Ergebnisse der Modelle an Ausfallswahrscheinlichkeiten knüpft. Dieser Prozess ist vor allem im Kontext von Basel II von hoher Relevanz. Da jedoch auch sichergestellt werden muss, dass auch nach der Entwicklung das Modell aussagekräftige Ergebnisse liefert, ist ein Kapitel dem von Basel II oft betonten Themenbereich Validierung gewidmet – es werden die in der Praxis verwendeten Techniken und Methoden zur Beurteilung der Modelltrennschärfe sowie der Kalibrierungsgenauigkeit vorgestellt und behandelt. Abschließend wird ein Überblick über alternative Ansätze zur Vorhersage von Unternehmensausfällen gegeben. Diese haben jedoch entscheidende Nachteile, zum Beispiel die fehlende Transparenz von Neuronalen Netzen oder die Notwendigkeit von Marktdaten für strukturelle Modelle. Deshalb bleiben auch in Zukunft statistische Kreditrisikomodelle eine äußerst beliebte Methode zur Quantifizierung des Kreditrisikos im großen Segment der Firmenkunden für die weder ein externes Rating, noch Marktdaten zur Verfügung stehen.

Appendix B – Curriculum Vitae

Persönliche Daten

Name:	Peter Racko
Geburtsdatum:	8. März 1983 in Bratislava/Slowakei
Staatsangehörigkeit:	Slowakei
Adresse:	Plankenbüchlergasse 12, 1210 Wien
Telefon:	+43 (0) 699 124 78 671
E-Mail:	peter.racko@gmail.com

Ausbildung

Seit Okt. 2001	*Internationale Betriebswirtschaftslehre,* Universität Wien Absolvierte KFKs: *Banking* (KFK Prüfung) und *Corporate Finance* Alle Kurse absolviert
1993 – 2001	Gymnasium Mercury in Bratislava Matura mit Auszeichnung bestanden

Berufserfahrung, studienbegleitende Tätigkeiten

Seit Apr. 2007	Referent in der Risikomanagementgruppe der Abteilung für Bankenrevision der Oesterreichischen Nationalbank Tätigkeit in der Begutachtung von IRB-Systemen österreichischer Banken
Okt. 2006 – Jan. 2007	Praktikum bei Siemens Transportation Systems in Peking, China Mitarbeit an Projekten innerhalb der Strategy Controlling Abteilung von Siemens Transportation Systems China
Mär. 2006 – Sept. 2006	Newole & Partner GmbH Analysen und Recherchen betreffend verschiedener Industrien in der Slowakei und Tschechien
Okt. 2005 – Juni 2006	Studienassistent, Faculty IT-Support BWZ Uni Wien Interne Datenanalyse/Auswertung, Betreuung der PC-Räume an der Universität
Mai 2003 – Sept. 2006	Webmaster, Mitglied bei i-Network (Studentenorganisation)

Weitere Qualifikationen

Sprachen	Slowakisch	Muttersprache
	Englisch	Fließend
	Tschechisch	Passiv
	Französisch	Grundkenntnisse
EDV	Microsoft Office	
	Adobe Photoshop/Illustrator	
	HTML, Java Grundkenntnisse	

References

Aldrich, John H. and F.D. Nelson, *Linear Probability, Logit, and Probit Models*, Sage Publications, 1992

Altman, Edward I., *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, The Journal of Finance, Volume 23, 1968, pp. 589-609

Altman, Edward I. and G. Sabato, *Modelling Credit Risk for SMEs: Evidence from the US Market*, 2006

Altman, Edward I., *Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta Models*, 2000

Altshuler, Alex, A. E. Kocagil, P. Escott and E. Glormann, *Moody's RiskCalc for Private Companies: The Netherlands*, Moody's KMV, 2002 http://www.moodyskmv.com

Backhaus, Klaus, B. Erichson, W. Plinke und Rolf Weiber, *Multivariate Analysemethoden*, 10. Auflage, Springer-Verlag, 2003

Balcaen, Sofie, H. Ooghe, *Alternative Methodologies in Studies on Business Failure: Do They Produce Better Results than the Classical Statistical Methods?*, 2004

Bank Austria Creditanstalt AG, *Unternehmensfinanzierung im Wandel: Rating als neuer Marktfaktor*, 2005

Basel Committee on Banking Supervision, *An Explanatory Note on the Basel II Risk Weight Functions*, 2005, http://www.bis.org

Basel Committee on Banking Supervision, *International Convergence of Capital Measurement and Capital Standards*, 2005, http://www.bis.org

Basel Committee on Banking Supervision, *Studies on the Validation of Internal Rating Systems*, 2005, http://www.bis.org

Beaver, William H., *Financial Ratios as Predictors of Failure,* Journal of Accounting Research, 1966, pp. 71-111

Bech, Jens, P. Escott, F. Glormann and A. E. Kocagil, *Moody's RiskCalc for Private Companies: France*, Moody's KMV, 2001 http://www.moodyskmv.com

Bovas, Abraham and J. Ledolter, *Introduction to Regression Modeling*, Thomson Brooks/Cole, 2005

Box, G. and P. Tidwell, *Transformation of Independent Variables*, Technometrics 4, 1962, pp. 531-550

Brannath, Werner and A. Futschik, *Statistik für Wirtschaftswissenschaftler*, WUV Universitätsverlag, 2001

Cox, David R. and D. Oakes, *Analysis of Survival Data*, Chapman Hall, London, 1984

Engelmann, Berndt, E. Hayden and D. Tasche, *Measuring the Discriminative Power of Rating Systems*, 2003

Erste Bank der Oesterreichischen Sparkassen AG, *Das neue „Basel II"-Rating*, 2003

Escott, Phil, E. Glormann and A. E. Kocagil, *RiskCalc for Private Companies: The German Model*, Moody's KMV, 2001 http://www.moodyskmv.com

Escott, Phil, A. E. Kocagil, P. Yapallo and M. Yague, *Moody's RiskCalc for Private Companies: Spain*, Moody's KMV, 2001 http://www.moodyskmv.com

Falkenstein, Eric, A. Boral, and L. Carty, *RiskCalc for Private Companies*, Moody's KMV, 2000 http://www.moodyskmv.com

Falkenstein, Eric, A. Boral, and A. E. Kocagil, *RiskCalc for Private Companies II: More Results and the Australian Model*, Moody's KMV, 2000 http://www.moodyskmv.com

Fernandes, João Eduardo, *Corporate Credit Risk Modelling: Quantitative Rating System and Probability of Default Estimation*, 2005

Giesecke, Kay, *Credit Risk Modeling and Valuation: An Introduction*, 2004

Grant, Mark, *Basics of Bayes' Theorem*, 2002

Halling, Michael and Evelyn Hayden, *Statistische Methoden zur Vorhersage von Unternehmensausfällen*, 2004

Hodrick, R. and E. Prescott, *Post-War U.S. Business Cycle: An Empirical Investigation,* Journal of Money, Credit and Banking 29, 1997

Hosmer, D. and S. Lemeshow, *Applied Survival Analysis*, John Wiley & Sons, Inc., New York, 1999

Henneke, Jan, S. Trück, *Asset Correlations and Capital Requirements for SME in the Revised Basel II Framework*, 2005

Jarrow, Robert A. and S. Turnbull, *Pricing Derivatives on Financial Securities Subject to Credit Risk*, Journal of Finance Vol. 50, 1995, pp. 53-85

Kocagil, Ahmet E. and J. Akhavein, *Moody's RiskCalc for Private Companies: Japan*, Moody's KMV, 2001 http://www.moodyskmv.com

Kocagil, Ahmet E., J. Akhavein and A. Reyngold, *Moody's RiskCalc for Private Companies: Mexico*, Moody's KMV, 2001 http://www.moodyskmv.com

Kocagil, Ahmet E., P. Escott, F. Glormann, W. Malzkorn and A. Scott, *Moody's RiskCalc for Private Companies: UK*, Moody's KMV, 2002 http://www.moodyskmv.com

Kocagil, Ahmet E., R. Imming, F. Glormann and P. Escott, *RiskCalc for Private Companies: Austria*, Moody's KMV, 2002 http://www.moodyskmv.com

Kocagil, Ahmet E. and A. Reyngold, *RiskCalc Singapore*, Moody's KMV, 2002 http://www.moodyskmv.com

Kocagil, Ahmet E., N. Seiberlich, Ö. Teymur, A. Grass, E. Parillon, P. Escott and F. Glormann, *Moody's RiskCalc for Private Companies: Nordic Region*, Moody's KMV, 2003 http://www.moodyskmv.com

Kocagil, Ahmet E., V. Vasudev, F. Glormann and P. Escott, *Moody's RiskCalc for Private Companies: Italy*, Moody's KMV, 2002 http://www.moodyskmv.com

Lehmann, Bina, *Is It Worth the While? The Relevance of Qualitative Information in Credit Rating*, 2003

Murphy, Adrian, A. E. Kocagil, P. Escott and F. Glormann, *Moody's RiskCalc for Private Companies: Portugal*, Moody's KMV, 2002 http://www.moodyskmv.com

Oesterreichische Nationalbank (OeNB) and Finanzmarktaufsicht (FMA)*, Neue quantitative Modelle der Bankenaufsicht*, 2004

Oesterreichische Nationalbank (OeNB) and Finanzmarktaufsicht (FMA)*, Ratingmodelle und -validierung*, 2004

Raiffeisen Zentralbank Österreich AG, *Checkliste-Rating,* 2005

Ronn, Ehud I. and A.K. Verma, *Pricing Risk-Adjusted Deposit Insurance: An Option-Based Model*, Journal of Finance 41, 1986, pp. 871-895

Royston, Patrick and D. G. Altman, *Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling*, Journal of Applied Statistics 43, 1994, pp. 429-467

Saunders, Anthony and L. Allen, *Credit Risk Measurement*, 2nd Edition, John Wiley & Sons, Inc., New York, 2002

Schmiemann, Manfred, *SMEs and Entrepreneurhip in the EU*, Eurostat, 2006

Shumway, Tyler, *Forecasting Bankruptcy More Accurately: A Simple Hazard Model*, 1998

Tasche, Dirk, *A Traffic Lights Approach to PD Validation,* 2003

Westenholz, Dina, W. Malzkorn, A. E. Kocagil, F. Glormann and P. Escott, *Moody's RiskCalc for Private Companies: Belgium*, Moody's KMV, 2002
http://www.moodyskmv.com