MOODY'S RISKCALC™ FOR PRIVATE COMPANIES:  
THE GERMAN MODEL

Rating Methodology

Overview

In recognition of the growing need for credit risk benchmarks for unrated companies, Moody's Risk Management Services is creating its RiskCalc™ models for estimating firm probabilities of default using financial statements and other objective, non-financial data. Following the release of RiskCalc™ models designed specifically to fit the US, Canadian, and Australian economies, Moody's is releasing RiskCalc™ models designed for the Spanish and German economies. These are the first in a suite of European models that are being developed. The RiskCalc™ models are powerful, objective tools that serve the interests of institutions, borrowers, and investors alike. This document describes the model for German middle-market companies, called RiskCalc™ Germany. It contains:

— a description of the database of financial information, on which RiskCalc™ Germany was developed,
— a description of the methodology used to develop the model,
— a comparison of the relationship of various financial ratios to default,
— a report on empirical tests of the model’s predictive power.

This document is meant to be a self-contained description of the development and validation of RiskCalc™ Germany; however, some details may be omitted. A more detailed handling of some of the methodology is contained in: RiskCalc™ for Private Companies: Moody’s Default Model (2000).1

---

1 See: Falkenstein, Carty and Boral (2000).
Introduction

Experience has shown that a key determinant of lending performance is the ability to correctly assess the credit risk within a portfolio. Default models, including objective quantitative models, are increasingly being used to assist in this effort. While we refer the interested reader elsewhere for a complete listing of the uses of default models, a selected list of applications includes the following:

- **Capital allocation:** in their efforts to ensure the soundness of the financial system and to encourage appropriate behaviour, regulators are increasingly looking for objective, hard to manipulate measures of risk to use in capital allocation.

- **Credit process optimisation:** whilst a single number may prove inferior to the judgement of a credit expert, the default model can help to pinpoint those cases where this expertise adds the most value.

- **Pricing:** without an accurate measure of the risks involved in lending to middle-market companies, shareholder value may be destroyed through sub-optimal pricing.

- **Securitisation:** banks are increasingly looking to offer their clients a full range of services, without holding the capital this would require. At the same time, investors are seeking new classes of risk prompting a need for a transparent, objective credit risk standard.

Not only do all of these needs require a powerful, efficient tool that allows unambiguous comparison of different loans and companies, but also accurate pricing and trading of credit risk demands that any such tool is calibrated to a Probability of Default (PD). RiskCalc™ is designed to provide an independent benchmark for most credit decision needs. We believe that in order for any tool to qualify as a benchmark it must satisfy the following conditions. It must be:

1. **Easily Understandable**
   
   Customers consistently indicate that it is critically important for them to understand how and why a model works. The ratios driving a particular assessment and their relationship to the model’s outcome should be clear and intuitive.

2. **Powerful**
   
   A model that is unable to differentiate between good and bad companies is clearly of little use in credit decisions. A consequence of a powerful tool is the willingness of experienced personnel to use it in pricing and making decisions.

3. **Calibrated To Probabilities Of Default (PDs)**
   
   Whilst an un-calibrated model can be used to refuse or approve credit applications, it is of little use in ensuring that any risk assumed is appropriately priced and adequately capitalised. Furthermore, it will be of little use in trading debt.

4. **Empirically Validated**
   
   Without documented performance on large out-of-sample datasets, prudence dictates models must be viewed sceptically. Such testing also gives the user confidence that the model has not been "overfitted".

   If a model does not satisfy at least these four criteria then, whilst it may be a useful tool, it cannot be considered as a benchmark for the market.

RiskCalc™ Germany has been developed in cooperation with Oliver Wyman & Company, the leading global strategy consulting firm dedicated to the financial services industry. RiskCalc™ Germany’s development takes full advantage of Oliver Wyman & Company’s extensive experience in developing similar models for many of the largest banks in Europe, the deep accounting and modelling understanding of the people at Oliver Wyman’s affiliate, the Baetge Oliver Wyman Rating Network, and MRMS’ experience in the provision of financial software, credit training and both quantitative and judgmental risk assessment models.

Data Description

The intention with the RiskCalc™ suite of products is to provide credit risk benchmarks for firms that do not currently have a rating. MRMS has developed default probability models for public firms in the United States, Canada, and Europe.² The goal for RiskCalc™ Germany is to provide a Probability of Default (PD) for private firms in Germany, with an annual turnover of more than €0.5m.

---

However, we do not attempt to cover all private firms in Germany with this tool because of the very different nature of some firms. Thus, in preparing data for development of RiskCalc™ Germany, we eliminated the following types of companies from our analysis:

- **Small Companies** - as discussed elsewhere, the credit quality of the smallest firms is often as dependent on the finances of a key individual as on the firm itself. For this reason we have excluded those firms with an annual turnover of less than € 0.5m.3

- **Financial Institutions** - in our experience, the balance sheets of financial institutions are significantly different to those of other firms (e.g., they have relatively high leverage/gearing ratios compared to other industries). Furthermore, the fact that the "soundness" of a financial institution is carefully monitored by regulators means that they are best considered separately.

- **Public Institutions** - judging the credit risk of firms owned by a community, county, state or a similar public institution is complicated by the fact that the owners/users have occasionally been unwilling to see them fail.

- **Real Estate Companies** - the success or failure of real estate firms often hinges on a particular development which may not be captured within any one year’s financial statement.4

- **Affiliates** - these have been excluded from development because their probability of default is often dependent on that of the parent firm.

We further cleaned the database to ensure that we did not select a model based on spurious power driven by poor data.5 Thus, we excluded financial statements from our database based on plausibility checks of particular positions in financial statements (e.g., assets less than zero). We also excluded financial statements covering a period of less than twelve months.

Table 1 illustrates the sample used in developing the RiskCalc™ Germany model, and compares it with the samples used in developing other RiskCalc™ models. It is clear that the sample available in Germany is smaller than those used in other models; however, the size and nature of the validation sample (as described below) and the performance of our model on this sample6 reassure us that the model we have developed is very robust.

<table>
<thead>
<tr>
<th>Country</th>
<th>Time Span</th>
<th>Unique Firms</th>
<th>Unique Firms</th>
<th>Financial Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany (Development)</td>
<td>1987-1992</td>
<td>4,866</td>
<td>485</td>
<td>11,427</td>
</tr>
<tr>
<td>Spain</td>
<td>1992-1999</td>
<td>140,790</td>
<td>2,265</td>
<td>569,181</td>
</tr>
<tr>
<td>US</td>
<td>1989-1999</td>
<td>33,964</td>
<td>1,393</td>
<td>139,060</td>
</tr>
<tr>
<td>Germany (Validation)</td>
<td>1992-1999</td>
<td>20,000</td>
<td>1,000</td>
<td>100,000</td>
</tr>
</tbody>
</table>

The greatest contrast is between Germany and Spain, a difference driven by the different levels of development of the business information markets in these two countries. In Spain, as in the UK and France, there are firms that provide databases of financial statements for most companies, often bundling such information with credit information in the form of commercial reports. However, our experience in Germany is that the publicly available data sources do not provide the same coverage or depth of reporting. Thus we have developed the model using proprietary data collected through Baetge Oliver Wyman Rating Network’s previous work with banks in Germany.

Figure 1 demonstrates that the financial statements that were used to develop RiskCalc™ Germany date from 1987 to 1992 with a peak between 1988 and 1991. The validation sample covers a much more recent period, providing a true out-of-time validation of the model’s performance. The strong performance of RiskCalc™ Germany on this out-of-time data set leads us to conclude that the age of the development sample has had little impact on its quality.

---

3 Defined as Net Sales for companies that report according to the "Umsatzkostenverfahren". For companies that report according to the "Gesamtkostenverfahren" we considered the sum of sales + stock movements + own work capitalised.

4 This is the case for many types of "project finance" firms, e.g., ship building firms, and we would recommend use of a separate model for them.

5 It is an unfortunate fact that as the sample size decreases, the importance of having clean data increases.

6 See the Empirical Tests section for a description of the results of these tests.
Although the development sample is not as current as our validation sample, it was our development sample of choice since the data are of higher quality, with the defaulting and non-defaulting companies clearly identified,\(^7\) and dates of default accurately recorded. The data in our validation sample, which have been acquired more recently, have not undergone such a deep level of cleansing and validation. As such, they are more representative of the actual state of credit information within banks and provide a more realistic test of model performance.

\(\text{Figure 2}\) shows that the distribution of firms by industry differs between our development and validation samples, with a higher proportion of manufacturing companies in our development sample. Yet again, the performance of the tool on the validation sample has reassured us that it is robust across industries.

---

\(^7\) A common problem faced when developing rating tools surrounds the correct identification of defaulted firms. This can cause problems when using statistical techniques in the creation of a rating tool, resulting in the creation of a model that is sub-optimal.
Another difference between the development sample and the validation sample can be observed when considering the distribution of financial statements by sales classes. Within the development sample more than 70% of all financial statements are from firms with sales of more than €5m, whilst less than 5% of the statements are for firms with less than €1m. However, more than 20% of the statements in the validation sample are from firms with turnover of less than €1m, providing a good test of the applicability of our tool to smaller firms. This provides a further element of out-of-universe testing.

It is clear from the graphs and discussion above, that the validation of the performance of RiskCalc™ Germany is out-of-sample, out-of-time and out-of-universe. It is RiskCalc™’s performance on this extremely tough validation sample that gives us the greatest confidence that it is robust across industries, time and firm size.

In the development of quantitative default models, the number of companies in the development sample, and particularly the number of defaulted companies in that sample, is very important. The higher those numbers, the more likely one is to develop a powerful model. For our German model we considered companies as having defaulted, if they were identified as having entered or undergone:
- bankruptcy,
- debt composition proceedings,
- debt moratorium, or
- cheque or bill protest.

Aggregate Probability Of Default Assumptions

The estimation of long-term aggregate probabilities of default is important because it serves as an anchor point for the model. An increase in the long-term probability of default will raise all predicted probabilities of default and vice versa. In this section we describe the data-sources consulted to triangulate a central tendency estimate. Whilst most of the credit events that we used when developing the RiskCalc™ Germany model were bankruptcy related, outputs for the RiskCalc™ suite of tools are focussed on probabilities of company default, where default is defined as 90 days past due to banks. Thus when calibrating the models, we aim to use a central tendency estimate which measures the probability of 90 days past due bank default, not the probability of bankruptcy.

Ideally, one would use actual bank default experience. However, in the absence of such data, it is necessary to take other approaches (in the long term the reporting requirements recommended under BIS II should ultimately lead to much more accurate estimates). In deriving our estimate of the aggregate probability of default, we have used a couple of approaches which produced results in line with our previous experience: the first uses reported bank provisioning data; the second uses reported bankruptcy data and then adjusts this towards a default figure.

The first approach for determining the population default rate uses banks’ net loan loss provisions, which, over time, will tend to equal actual loan losses and hence reflect the underlying default rate. Loss rates and default rates are related to each other by the loss given default rate (LGD) as shown below:

\[
\text{Volume of Losses} = \text{Volume of Loans} \times \text{Probability of default} \times \text{LGD}
\]

\[
\Rightarrow \text{Probability of Default} = \frac{\text{Volume of Losses}}{(\text{Volume of Loans} \times \text{LGD})}
\]

The foundation approach to capital allocation as described in BIS II uses a loss given default rate of 50%, and in the absence of significant evidence to the contrary in Germany, we have used this rate to derive an estimate of the default rate.

The OECD\(^9\) reports that net loan loss provisions of German commercial banks averaged 0.77% between 1989-1999. Using a 50% loss given default rate this corresponds to an average default rate of 1.54%.

The second approach used data for the number of bankruptcies/insolvencies in Germany and adjusted this to reflect the fact that we would generally expect more firms to default than go through bankruptcy/insolvency proceedings. Figures from the Bundesamt indicate that the average rate of bankruptcies\(^10\) between 1992 and 1999 was approximately 0.8%. Our initial hypothesis was that this figure is low because it includes sole proprietorships where bankruptcy/insolvency figures are lower.

---

8 If there were multiple instances of default for a firm, we used the earliest date as the date of default. The bulk of incidences defaults in our development sample were in fact bankruptcies.

9 Various editions of “OECD: Bank Profitability - Financial Statements Banks”

10 This was calculated as: number of insolvencies/bankruptcies divided by number of companies.
To check our hypotheses, we purchased data from CreditReform where the insolvency/bankruptcy data is broken down to a more detailed level (e.g., by industry and size). These more detailed figures indicate that the average bankruptcy rate for a sample with similar size characteristics to ours was around 1.2% to 1.4%.

As mentioned above, the bankruptcy/insolvency is a “hard”/late definition of default and we would expect a higher rate of missed payments to banks than of bankruptcies. Since we are targeting a bank missed payment default rate, we adjusted the 1.2%-1.4% upwards. We also made adjustments to account for the calibration sample we used and the period covered by the CreditReform data.

Following these adjustments we derived a figure of around 1.6%. Combined with the estimate derived from provisioning data and our experience in developing other RiskCalc™ models, we felt that 1.6% was an appropriate rate to use for an aggregate one year probability of default estimate.

The model is also calibrated to a cumulative five-year horizon, which captures the model’s ability to predict that a company defaults at some point during the five-year period following publication of its financial statement. However, very few sources for such data exist, especially within the public domain. So we have used the experience of the Moody’s rated universe, which suggest that, on average, the five-year cumulative default rate is approximately four times the one-year rate.

**Model Description**

As any experienced modeller would argue, the best models are those that combine theory and science with experience and intuition (even in a world with perfect data, relying solely on statistical procedures is unwise, and in a data poor environment it could prove disastrous). As a result, our model would be classified as a non-structural model in that it does not use an explicit specification based on theory, but is highly informed by the many years of default modelling experience of Moody’s, Oliver Wyman & Company and the Baetge Oliver Wyman Rating Network. There is a trade-off between in-sample fit and out-of-sample robustness and our bias is towards a simple functional form and a small number of inputs. Our modelling approach can be briefly summarised in the following three steps:

**Single Factor Analysis:** it is the aim of single factor analysis to study the individual relationship of potentially relevant factors to default, creating a short list of the most important factors. As part of this process we also mini-model the relationship between these factors and default rates.

**Model Specification and Estimation of Factor Weights:** once individual factors have been analysed, the next step is to specify a model using a subset of the most important factors. These factors are combined in a logistic model and their weights are optimised.

**Calibration:** finally, once the model had been specified and its weights estimated, we need to map the output of the model, a score, to a probability of default.

The following sections provide more detail on these three steps.

**Single Factor Analysis And Transformation**

A specific characteristic of PD models based on financial information is the large number of ratios that could be calculated and included in the model building process. Although some of the financial ratios that can be derived will be useful to predict default, others are likely to be spuriously related to the default variable. Thus the way information is used to build the model is crucial in determining the capability and robustness of the final model in predicting default.

It is important when testing the predictive power of a particular ratio to have a prior expectation of how it will be related to default; otherwise one runs the risk of selecting variables based on statistical quirks. These relationships are generally monotonic. That is slope is either always positive, or always negative. If positive, a higher ratio value indicates a higher probability of default (e.g., Cost/Sales). If negative, a lower ratio value indicates a higher probability of default (e.g., Equity/Assets). If a ratio does not fit with our prior belief, then we exclude it from further analysis.

The next step is to test the discriminatory power of each factor. There are several possible measures one could use, and we have chosen the accuracy ratio which measures how rapidly defaulting companies are identified by the ratio. In the case of RiskCalc™ Germany we excluded all factors that had an accuracy ratio of less than 5%.

---

11 The calculated 5-year cumulative default rate for private firms in Germany is 6.4%.
12 A description of the logistic model is provided later in the document.
13 The best-documented class of non-monotonic ratios are growth ratios, which often exhibit a U-shaped relation with default.
14 For further details on the concept of accuracy ratio see Section IV below and the Appendices.
15 A value of 0% would indicate a completely random relationship to default, whilst a value of 100% would indicate a perfect relationship to default.
The next step is to mini-model the relationship between the selected factors and default, fitting a function to the data that smooths the raw data and "caps" the extreme values. These "caps" not only eliminate the impact of outliers in the estimation of the parameters of the final model, but they also ensure that the final model does not produce a PD that is driven by a statistical quirk. It also reflects the fact that beyond a certain level, there is little additional information provided about default.

Finally, the transformed ratios are normalised to ensure that they have the same average and standard deviation. This makes the values of very different ratios comparable and permits a very intuitive definition of the final model by simply assigning weights to each factor.

**Model Specification and Estimation Of Factor Weights**

In the second step, the selected transformed factors undergo a process of multivariate analysis, which looks at the predictive power of combinations of these ratios. Starting with a list of 20 ratios there would be over 1 million possible models which could be created, so it is important to use statistical selection procedures such as forward and backward regression to further reduce the set of factors, and hence possible models.

Including highly correlated ratios when estimating the optimal weights for a model without careful attention to address this issue can result in unstable estimates of these weights, and poor performance of a model when applied outside of the development sample. Furthermore, the weights assigned to these factors can often be counter-intuitive, e.g. it might be possible to have a model in which higher profitability led to higher default rates. So in selecting the ratios to be in the final model, we also analysed the correlations of ratios, excluding those factors that were highly correlated.

There is no hard and fast rule in determining how many ratios a particular PD model should contain: too few and the model will not capture all the relevant information; too many and the model will be powerful in-sample, but unstable when applied elsewhere and will most likely have onerous data input requirements. When deciding on the final model to use we combined an analysis of the power of the different models, as measured by the accuracy ratio, with our experience. Some of the considerations that went into the selection of the final ratios and model are:

- data requirements for the user should be as low as possible,
- the number of factors within the final model should be as low as possible,
- the factors and their weights should be intuitive, and
- the model should have the highest possible explanatory power

**Calibration**

The final part of the modelling process consists of mapping the output of the model to probabilities of default. This exercise can conceptually be divided into two parts. The first one serves to ensure that the average default rate predicted by the model equals our best estimate of the population default rate over the economic cycle. The second part is the mapping of scores to probabilities of default, as detailed below.

The basic methodology for generating the 1-year calibration curve was the same as the approach adopted in RiskCalc™ Australia, where the power curve is created and used to generate a calibration curve. The calibration curve is then adjusted so that the implied population default rate matches our assumption of the long-term aggregate default rate. Finally, the calibration curve was smoothed to reduce the impact of outliers and to achieve a monotonic relationship by fitting an exponential function to the data.

A problem encountered with many data sets is that there is a sample selection bias that implies a higher default rate amongst larger companies, an implication, which doesn’t sit well with our experience. Some of this bias is corrected by the fact that large firms generally have "better" financial statements, in so far as their ratios generally indicate better credit quality. However, financial statements fail to capture the diversification and management sophistication benefits enjoyed by many of the larger firms, and so we have adjusted the final calibration for the larger companies. This adjustment is gradually applied to the larger firms, resulting in a 1 to 2 grade class improvement for the very largest firms.

---

16 For non-monotonic ratios, this mini-modelling includes a step to capture the non-monotonicity.
18 See the Appendices for further details.
19 Our assumption, as mentioned above, is of a 1-year default rate of 1.6%. The adjustment consists of an upward shift of the curve.
20 In the RiskCalc Australia model, the data was smoothed using the Hodrick-Prescott filter. However, any method that maintains the exponential nature of the extremums is sufficient.
When developing the model we were aware of the fact that trading companies generally generate higher levels of sales for a given level of assets. Our expectation was therefore that the ratios being used in our final model were likely to produce lower scores for trading firms, and hence result in a lower PD for such firms.21 When we examined the distribution of grades awarded to firms across industries, it became clear that this was indeed happening.

Since the data that we have used when deriving the aggregate probability of default figure does not support a lower PD for trade firms, it is clear that we needed to adjust the predicted PDs for these firms. Given the power of the model, we felt this is best handled within the calibration step, realigning the grade distribution for trading firms with that of the overall population. As part of the mapping process from a score to a PD we apply a bonus to trading companies.22

To summarise, the transformation and normalisation of input factors constitute a transparent way of capturing the information that each ratio carries about the likelihood of default. The binary probabilistic logit model is an efficient method of determining the optimal weights for combining the input ratios. Finally, the calibration mapping transforms score output into easily interpretable probabilities of default, which in turn are mapped to Moody’s historical bond default rates in order to provide an easily recognised representation of the level of riskiness.

Ratios And Their Relation To Default

The RiskCalc™ Germany model uses nine factors which fall within the following broad categories: leverage/gearing, profitability, debt coverage, growth, activity and productivity. This section provides a description of these ratios and how they have been calculated. For simplicity we have provided short names for the ratios which capture the essence of what they measure (precise definitions of these ratios, and of the German language equivalents, can be found in the Appendices).

Leverage/Gearing Ratios

Within RiskCalc™ Germany we use three different leverage or gearing ratios: Equity, Net Indebtedness and Liability Structure.

The Equity ratio measures the ratio of a company’s equity to its assets, and simpler versions are widely used in credit models. We have made the following adjustments to a simple Equity / Assets ratio to counter creative accounting practices, and to try to generate a better measure of company credit strength:

---

21 Consider two firms, one a trading company, the other a manufacturing company, with the same levels of profits, equity and assets: because the trading company would generate more sales from the same level of assets, ratios such as Profit on Sales would be lower for the trading company, resulting in a worse score, and hence PD.

22 At most this bonus might improve the grade of a trading company by 1 grade.
- Subtracted intangible assets from equity and assets - our experience indicates that when a firm defaults, the value generated from the intangible assets is considerably lower than the accounting value, providing little protection to a creditor.
- Subtracted cash and equivalents from assets - one course of action for a firm wishing to improve its reported liquidity is to raise a short-term loan at the end of the year and hold it as cash.
- Subtracted land and buildings from assets - this is done to negate the effect of sale-and-lease-back arrangements.

Our initial belief, that firms that subsequently defaulted would have lower values was borne out by the data, as can be seen below. Not only do these firms have lower values on average, but also the high value classes are almost exclusively populated by firms that have not subsequently defaulted.

Net indebtedness measures the level of short-term liabilities not covered by a firm’s most liquid assets as a proportion of its total assets. Thus in addition to measuring the short-term leverage for a firm, it also provides a measure of the liquidity of a firm. We believe that in general firms that subsequently default will have higher values for this ratio as it indicates that the firm may face liquidity problems in the short-term.

![Net Indebtedness Ratios](image1)

![Liability Structure Ratios](image2)

The Liability Structure ratio considers the structure of a company’s liabilities and measures the proportion of trade liabilities, notes payable and bank debts to its total liabilities. The key element that has been excluded from liabilities in the top line is the liabilities to group and affiliate firms. We expect defaulting firms to have higher values for this ratio, reflecting a greater reliance on external creditors for funding. This can be clearly seen within the following graphs.
**Profitability**

It will surprise no one that profitability is clearly related to default, and hence appears in many credit models. However, there are many different measures of a firm's profitability, and the question is which profitability measures should be used. Since they are normally fairly powerful, we have chosen two ratios, EBITD and Profit on Sales, which differ in a number of respects, reducing the correlation between the ratios and picking up different elements of balance sheet policy.

EBITD measures a company’s profit per unit of its total assets. This profit has been adjusted to add back the interest expenses and depreciation costs. The reason for adding interest expenses is to evaluate a company’s profitability independently of the structure of its capital. Adding depreciation neutralises attempts to “window-dress” the balance sheet by re-defining depreciation. Our assumption is, that on average defaulting companies have lower values.

Profit on sales differs from EBITD in that it measures the amount of profit per unit of sales after subtracting ordinary expenses. Unsurprisingly, those firms who are failing to generate sufficient sales to cover their ordinary expenses, tend to default more frequently (i.e. defaulting firms have a lower value for this ratio).
Debt Coverage

Debt coverage measures the cash flow relative to a firm’s liabilities. We have adjusted the liabilities by subtracting advances from customers in order to account for industry specificities (e.g. general construction and plant construction), where advances from customers traditionally play an important role in financing. Without this adjustment, such companies would be disadvantaged by this ratio. In general firms which go on to default have lower values for this ratio.

Growth

The relationship between the rate at which companies grow and the rate at which they default is not as simple as that between other ratios and default. The reason is that whilst it is generally better to grow than to shrink, companies that grow very quickly often find themselves unable to meet the management challenges presented by such growth (especially within smaller companies). Furthermore, this growth is unlikely to be financed out of profits, resulting in a possible build up of debt and the associated risks this poses. We have found that Sales Growth is a consistent predictor of default and that it does not suffer from some of the problems of profit growth ratios (e.g. low/negative profitability and manipulation of reported figures).
**Activity**

Our experience has shown that the level of trade debt held by a firm is a good indicator of their creditworthiness, so our model includes the Trade Creditor ratio, which measures how many days it takes a company to pay its trade liabilities from its sales. The higher the value of this factor, the longer it takes the company to pay its trade liabilities from its sales, suggesting that the firm is more likely to default on its debts. Thus our expectation is that defaulting firms will have generally higher levels for this ratio. This is clearly demonstrated in the following figure, which shows that whilst 70% of non-defaulting firms have a value of less than 40, only 30% of defaulting firms have a value of less than 40.

**Productivity**

A firm’s productivity will clearly have some bearing on its likelihood of default, and we have included in our model a ratio, Personnel expenses on sales, which measures the level of sales a firm is able to generate from its staff costs. The charts below show that on a stand-alone basis, this ratio is much less predictive than the other ratios that we have used. However, in combination with the other ratios it leads to a significant increase in the accuracy ratio for the entire model.
The Weights

The output of the model (scores) is not only determined by the inputs, i.e., the factor values, but also by the weights assigned to the factors. Thus, one will get a better understanding of the relation between a particular input and a particular output by looking at the weights. Table 2 shows the relative contributions of the factors in RiskCalc™ Germany for a typical firm.

<table>
<thead>
<tr>
<th>Category</th>
<th>Factors</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage/Gearing</td>
<td>Equity Ratio, Net Indebtedness</td>
<td>38%</td>
</tr>
<tr>
<td></td>
<td>Liability Structure</td>
<td></td>
</tr>
<tr>
<td>Profitability</td>
<td>EBITD, Profit on Sales</td>
<td>25%</td>
</tr>
<tr>
<td>Debt Coverage</td>
<td>Debt Coverage</td>
<td>9%</td>
</tr>
<tr>
<td>Growth</td>
<td>Sales Growth</td>
<td>7%</td>
</tr>
<tr>
<td>Activity</td>
<td>Trade Creditors</td>
<td>10%</td>
</tr>
<tr>
<td>Productivity</td>
<td>Personnel Expenses, Profit on Sales</td>
<td>11%</td>
</tr>
</tbody>
</table>

An Extended Version Of RiskCalc™ Germany

As previously stated, in developing RiskCalc™ Germany we tried to minimise the number of input positions that are needed to calculate the ratios, in order to reduce the data burden on end users. Our aim was to include in the ratios only those financial line items that must be reported by German middle-market companies. We have also tried to limit the number of items a user needs to input, although this is less of a concern when the tool is being used as part of a centralised batch process.

Our experience in Germany has shown that many banks and auditors have access to more detailed company accounts (for example, this additional level of detail is available in the data sets we have used in development and validation). We have therefore developed an extended version of the model, which makes more comprehensive adjustments for creative accounting. The extended tool is slightly more powerful at a portfolio level, however we believe that the real benefit will be derived in specific cases.

The structure of the model, in terms of the types of ratios used and the way, in which individual ratios are combined, does not differ between the normal and extended versions of the model. The change is in the definition of some of the ratios, namely: the Equity ratio, the Net Indebtedness ratio, the Liabilities Structure ratio, the Profit on Sales ratio and the Debt Coverage ratio.23

Table 3

Extended Versions Of RiskCalc Factors

<table>
<thead>
<tr>
<th>Category</th>
<th>Standard Definition</th>
<th>Extended Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage / Gearing</td>
<td>(Equity - Intangible assets) / (Total assets - Intangible assets - Cash &amp; Equivalents - Land &amp; Buildings)</td>
<td>(Equity + 50% of special items with equity character* - Intangible assets) / (Total assets - Intangible assets - Cash &amp; Equivalents - Short term financial investments - Land &amp; Buildings)</td>
</tr>
<tr>
<td></td>
<td>(Trade liabilities + Notes payable + Bank liabilities) / (Liabilities - Advances)</td>
<td>(Trade liabilities + Notes payable + Bank liabilities) / (Liabilities + 50% of special items with equity character - Advances).</td>
</tr>
<tr>
<td></td>
<td>(Current Liabilities - Cash &amp; Equivalents) / Total assets</td>
<td>(Current Liabilities - Cash &amp; Equivalents - Short term financial investments) / Total assets</td>
</tr>
<tr>
<td>Profitability</td>
<td>Ordinary profit / Sales</td>
<td>Ordinary profit - Other taxes / Sales</td>
</tr>
<tr>
<td>Debt Coverage</td>
<td>Cash Flow / (Liabilities - Advances)</td>
<td>(Cash Flow - Other taxes) / (Liabilities + 50% of Special items with equity character - Advances).</td>
</tr>
</tbody>
</table>

* The name of this item in German is “Sonderposten mit Rücklageanteil”

Empirical Tests

The primary testing tools we use for assessing statistical power, i.e., the ability to rank-order defaulters and non-defaulters, are power curves. Power curves graphically illustrate the ability to exclude defaulters for arbitrary cut-off points and can be aggregated into a single statistical number, the accuracy ratio, which allows for numerical comparisons among models.

23 We have also used different mini-modelling and transformations for the ratios, as well as a separate calibration for the extended model.
The power curve maps the fraction of all companies with the worst score (horizontal axis) onto the fraction of defaulting companies within that group (vertical axis). If the sample contained 10% defaulters, then a perfect model would exclude all those defaulters at 10% of the sample excluded: the 10% of companies with lowest ranks would consist of the defaults. Purely uninformative and perfectly informative models are illustrated in Figure 13.

In reality, defaulters are not perfectly discriminated, creating a curvilinear function: thus at 10% of the sample excluded, 30% of defaulters would be excluded, at 20% of the sample 50% of the defaulters would be excluded, etc. This creates a line that is bowed towards the upper left (Northwest) of the chart: the more bowed the power curve, the better the model.

The information contained in the power curve can be summarised in a single number, known as the accuracy ratio. Its values are between 0, for a completely random model, and 1, which would be the value for the perfect model. The higher the accuracy ratio of a PD model, the better the model in predicting default.24

After completing the development of RiskCalc™ Germany we tested its statistical power by calculating its power curve and accuracy ratio on a previously unused validation sample.25 At the same time, we compared its performance with that of the Z-score, a benchmark chosen for its popularity in major accounting textbooks.26 As Figure 14 shows RiskCalc™ Germany is considerably more powerful than the simple Z-score.

---

24 For details please refer to Sobehart, Keenan and Stein (2000).
25 The structure of this validation sample is described in the Data Description section.
26 The Z-Score is defined as follows: Z-Score = 6.56 * [(Working Capital) / Assets] + 3.26 * [Retained Earnings / Assets] + 6.72 * [EBIT / Assets] + 1.05 * [Net Worth / Liabilities].
In addition to showing the accuracy ratios for RiskCalc™ Germany and the Z-score, Table 4, also includes results from an analysis of the performance of the two models on a "cleaned" version of the validation sample. These results have been included to demonstrate the effect which data quality can have on the performance of a model. The better the quality of the data given to the tool, the more powerful it is. As a result, it is important when comparisons are made between two models that these comparisons are based on the same data sample.

The cleansing of the validation sample was done in order to demonstrate the overall impact of eliminating some typical data problems that can be encountered when using a model, and the resulting improvement in power can be seen in Figure 15:

![Figure 15](image)

**Table 4**

<table>
<thead>
<tr>
<th>Model</th>
<th>Data Sample</th>
<th>Accuracy Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskCalc Germany</td>
<td>Validation Sample</td>
<td>59.7%</td>
</tr>
<tr>
<td>Z-Score</td>
<td>Validation Sample</td>
<td>30.2%</td>
</tr>
<tr>
<td>RiskCalc Germany</td>
<td>Validation Sample (cleaned)</td>
<td>70.9%</td>
</tr>
<tr>
<td>Z-Score</td>
<td>Validation Sample (cleaned)</td>
<td>36.7%</td>
</tr>
</tbody>
</table>

- Misclassification Of Defaulters\(^{27}\) - it is often harder to identify defaulters than one might expect and a significant number of defaulters are classified as "non-defaulters". A powerful model gives such firms a low ranking. However, when the accuracy ratio / power curve is calculated, they would be considered as non-defaulters, resulting in a lower accuracy ratio.
- Missing Data\(^{28}\) - despite our efforts to ensure that we only use readily available data, some items will inevitably be missing. Where data is missing we normally replace the affected ratios with an average score, and calculate the rest of the ratios. The results produced when data is missing are still powerful (a consequence of the fact that we generally select individual ratios which are in themselves powerful). However, there is some drop in the power.
- Use On Non-Target Firms\(^{29}\) - using a PD model on a firm for which it has not been designed will inevitably result in a reduction of the power.

\(^{27}\) Whilst we believe that our validation sample has defaulters correctly identified, our experience is that many German banks consider those firms with negative equity to be insolvent, and hence treated as a default within our definition. For the purposes of generating the "clean" validation sample we have removed such firms.

\(^{28}\) We could have removed all firms where ratios cannot be calculated. However, we have removed only those where sales growth could not be calculated due to a lack of the previous year’s sales figure.

\(^{29}\) For the "clean" validation sample, we removed those firms with annual sales of less than 0.5m Euros, as the model is not intended for use on such firms.
Implementation Tips

There are a few points which one should bear in mind when using the RiskCalc™ Germany model. As with other RiskCalc models, we have not included every element that we believe could conceivably impact a firm’s probability of default. For example, we have not included qualitative factors such as management quality, or considerations of a firm’s position within an industry, the competitive environment in which it operates, and future industry outlook. We realise we cannot hope to understand the specific circumstances surrounding every firm. A similar problem faces many financial institutions, and it is not surprising that more sophisticated banks use tools like RiskCalc™ to determine which firms or loan applications require further attention.

Another potentially counterintuitive result of RiskCalc™ is that its implied PDs often appear ‘too low’. Many users are used to higher default rate projections for individual companies. The average probability of default for middle-market firms of 1.6% per year seems to make perfect sense until one actually sees the individual credits to which this is applied, and one considers that this is consistent with Ba2 default probability (most people consider private credit in the B2-B1 range on average, not the Ba2 range). While we have used a 1.6% figure, it should be recognised that we want the model to be unbiased. That is, it represents our best statistical estimate of the future probability of default. In contrast, a natural inclination of an underwriter is to be pessimistic, as the cost to being too optimistic is high. In practice you may wish to adjust the results by multiplying the default probability by a constant such as 1.5, in order to better accommodate your prudent bias, to accommodate ‘gaming’ or simply because your old scale had a mean implied 1-year probability of default of, e.g. 3.0% and you feel the new scale should be modified slowly.30

It goes without saying that it is important that a tool like RiskCalc™ is not used blindly. For example, thoughtlessly inputting the numbers for a firm that has just divested a large part of its business could produce misleading results. If one used the sales figures for that firm from the year before and the year after its merger, the levels of sales would change dramatically leading to negative sale growth and a lower PD. In such a case, one should aim to use the most comparable figures available.

Target for RiskCalc™

It is also important to bear in mind that, whilst we have attempted to build a robust tool which can be used on most companies, it would be inappropriate to use it on all companies. Clearly, where less, or erroneous, information is used, the tool will have difficulties in differentiating how risky a firm is. It can still be used, but the results should be screened for plausibility.

The types of firms we do not believe the tool to be appropriate are: financial institutions; public sector firms; firms whose shares are traded/listed;31 firms whose performance is dominated by a couple of specific projects (e.g. real estate firms); and the youngest firms where the little information that is available is rarely stable or a true reflection of the status of the firm. Inaccuracies in the PDs for these firms will creep in, not only because their financial statements do not capture reality, but also because the aggregate probability of default for these types of firm may well be significantly different from the population norm.

Conclusions

The RiskCalc™ methodology is true to the essence of applied econometrics; it is based on sound theory and years of practical experience. The model is non-structural, well understood, and sophisticatedly simple, relying on well-established risk factors. By transforming (or “mini-modelling”) the input ratios and then combining them into a multivariate model, we capture and integrate a non-linear problem, yet retain transparency. The final mapping process takes into account our ‘top-down’ view of default rates.

We see default modelling as a forward-looking problem and so we are careful to check for robustness, through cross-validation, out-of-sample tests and an emphasis on simplicity. For our German model, careful attention has been paid to how financial ratios could differ between German and other European countries, considering the particularities of the German economy both from a micro and macro perspectives, how these ratios relate to default and how best to integrate them into a powerful model. The final result is a model that we believe is well tuned to forecast tomorrow’s defaults, not just explain yesterday’s.

Using RiskCalc™ Germany should help improve profitability through the credit cycle, from credit decisioning to pricing to monitoring to securitising. RiskCalc™ is not intended as a sufficient measure of risk; it should be viewed as a very powerful aggregator of financial statement information into a meaningful and validated number that allows consistent comparison of portfolio risks.

30 An adjustment of 1.5 across the board, would give a portfolio PD of 2.4% (1.5 x 1.6%), compared to say 3.0% for the in-house estimate and 1.6% for the RiskCalc model.
31 For the publicly traded European companies we would recommend using RiskCalc Public - Europe, the RiskCalc model for public European firms instead. For details see: Falkenstein et al (2001).
<table>
<thead>
<tr>
<th>Category</th>
<th>Definition (English)</th>
<th>Definition (German)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage/Gearing</td>
<td><strong>Leverage/Gearing</strong> &lt;br&gt; Liabilities structure (Trade liabilities + Notes payable + Bank liabilities) / (Liabilities - Advances)</td>
<td>(Verbindlichkeiten aus Lieferungen und Leistungen + Akzepten + Bankverbindlichkeiten) / Fremdkapital - Erhaltene Anzahlungen</td>
</tr>
<tr>
<td></td>
<td><strong>Net indebtedness</strong> &lt;br&gt; Current Liabilities - Cash &amp; Equivalents / Total assets</td>
<td>(Kurzfristiges Fremdkapital - Schecks, Kassenbestand) / Bilanzsumme</td>
</tr>
<tr>
<td></td>
<td><strong>Equity ratio</strong> &lt;br&gt; Equity - Intangible assets / (Total assets - Intangible assets - Cash &amp; Equivalents - Land &amp; Buildings)</td>
<td>(Eigenkapital - Immaterielle Vermögensgegenstände) / Bilanzsumme</td>
</tr>
<tr>
<td>Profitability</td>
<td><strong>EBITD</strong> &lt;br&gt; Net profit + Interest expenses + Income taxes + Depreciation / Total assets</td>
<td>Jahresüberschuss + Zinsaufwendungen + Steuern vom Einkommen und Ertrag + Abschreibungen / Bilanzsumme</td>
</tr>
<tr>
<td></td>
<td><strong>Profit on Sales</strong> &lt;br&gt; Ordinary profit / Sales</td>
<td>Ordentliches Betriebsergebnis / Umsatz</td>
</tr>
<tr>
<td>Debt coverage</td>
<td><strong>Debt coverage</strong> &lt;br&gt; Cash Flow / (Liabilities - Advances)</td>
<td>Ertragswirtschaftlicher Cash Flow / Fremdkapital - Erhaltene Anzahlungen</td>
</tr>
<tr>
<td>Growth</td>
<td><strong>Sales</strong> &lt;br&gt; Growth Sales(t) / Sales(t-1)</td>
<td>Umsatz(t) / Umsatz(t-1)</td>
</tr>
<tr>
<td>Activity</td>
<td><strong>Trade creditors ratio</strong> &lt;br&gt; (Notes payable + Trade liabilities) *360 / Sales</td>
<td>(Akzepten + Verbindlichkeiten aus Lieferungen und Leistungen) *360 / Umsatz</td>
</tr>
<tr>
<td>Productivity</td>
<td><strong>Personnel expenses on sales</strong> &lt;br&gt; Personnel expenses / Sales</td>
<td>Personalaufwand / Umsatz</td>
</tr>
</tbody>
</table>
Appendix B: Testing Metrics

Power Curves

A power curve\(^{32}\) is constructed by plotting, for each threshold, the proportion of defaults excluded at various levels of sample exclusion. The vertical axis measures the percent of defaults excluded conditional upon excluding various percentage levels of the sample. Thus if using a score to exclude 50\% of the sample caused it to lose 80\% of the defaulting companies, the power curve would go through a line corresponding to \(x=0.5\) and \(y=0.8\). Accuracy is indexed more precisely by measuring the area under the curve, an area that increases as the curve bends.

\[
\text{power (b)} = \frac{\sum_{t=1}^{b} p(t)}{\sum_{t=1}^{B} p(t)} = \frac{\text{defaults excluded at b}}{\text{total defaults}} \quad (1)
\]

Here \(B\) is the total number of bins (often 10 for illustrative purposes), and \(b\) is a particular bin. The power at bin \(b\) represents the sum of all the defaults in the ‘worst’ fraction \(b/B\) of the scores, as ranked by the metric \(M\).\(^{33}\)

The net result is Figure 16 below, which shows the probability of default for a level of \(M\), and statistical power, which pertains to the nature of the data up to a level of \(M\). In this case we rank-order the firms from risky (left) to less risky (right), so that the \(P(M)\) and \(\text{Power}(M)\) correspond. The graph shows a particular case. This type of model would quickly have excluded most of the bad companies: a 20\% exclusion of the worst companies according to the \(M\) score would exclude 70\% of the future defaulters.

There is a one-to-one correspondence of power and probability of default by rank order, in that for any point \(t\) along a default metric:

\[
p(t) = \bar{p} \cdot \frac{\partial}{\partial t} \text{power}(t) \quad (2),
\]

where \(\bar{p}\) is the mean probability of default.

---

\(^{32}\) Also known as Gini curve, CAP plot, Lorenz curve, ordinal dominance graph, or ROC curve.

\(^{33}\) Since defaults excluded ‘at bin \(b\)’ is ambiguous - it could mean ‘up to bin \(b\)’ or ‘up to and including bin \(b\)’ - we calculate the area using the average of the two methods. Nevertheless this adjustment makes practically no difference.
While the graphical or tabular display of power is informative, and has the advantage of allowing one to examine power at a variety of thresholds, it is useful to aggregate the power curve information into a single number that allow unambiguous comparison. One such measure is the area under the power curve. A model more bowed out towards the left will have a greater area, and be more powerful on average. Using the area under the power curve implies that there can exist threshold levels such that a model with a smaller total area has a momentary advantage. Thus the area is not a measure of global or complete dominance, just an intuitive measure of dominance on average. The area can be calculated using equation (1) above, specifically

\[
\text{Area} = \frac{1}{B} \sum_{b=1}^{B} \text{power}(b) \quad (3),
\]

where \( B \) is the total number of bins. If the Area is greater for one model than another, it is more powerful.

Of course to really compare models, one not only needs an aggregate measure of power, but also a standard error on this metric. The standard error of this area is approximated well by the following formula:

\[
\sigma_{\text{Area}} = \sqrt{\frac{1}{BD} \sum_{b=1}^{B} p(b) * (1 - p(b))} \quad (4),
\]

where \( D \) is the number of defaults in the sample.
Appendix C: Calibration Curve Construction Details

The construction of the probability of default curve, which we use to calibrate the model, is done in the following way. We take each defaulting firm, and find its score 18 months prior to the default date. If a score does not exist on this date, we move back in time, up to a maximum of 41 months before default. When no score is available, we exclude the observation. When testing the power of a tool one would normally use the rank of a score within the calendar year from which the score was taken. However, as we want the tool to take into account business cycle movements reflected in the score, we rank the scores across all years for calibration purposes. Each defaulter is then mapped to a percentile, and this collection of percentiles is the basis from which the calibration power curve is created.

Specifically, given a collection of percentiles of defaulting firms $\{\phi_j\}_{j=1}^J$, where $J$ is the total number of defaulting firms, the power for each bin is simply:

$$\text{power}(b) = \frac{1}{J} \sum_{j=1}^{J} \left\{ 1 \mid \phi_j \leq \frac{b}{B} \right\}$$

where $\{ 1 \mid \phi_j \leq \frac{b}{B} \}$ is an indicator function equal to 1 if the defaulting firm, $j$, was in a percentile lower than $b/B$. For example, for a one-year probability of default curve, we would take a default in 10/98, and move back to 4/97 to find the percentile of the RiskCalc score using that month. As is most probable, the statement date is not exactly at 4/97, and so we must go back in time, to 3/97, then 2/97, etc., until we find the date at which we have a financial statement.

The data we had at our disposal did not allow us to calibrate the model directly to a cumulative five-year horizon. Instead our calibration has been based on observing the changes in the calibration curve as we moved from a one year prediction horizon, through to a cumulative three year horizon. Using this information, and combining it with our previous model building experiences, we constructed a five year cumulative calibration curve. The implied drop in the accuracy ratio in moving from a one year to a five year horizon was approximately 10%. Finally, the intercept of the calibration curve was adjusted in order to obtain our long-term five-year average default rate of 6.4%.

By using defaulting firms once in the creation of a set of percentiles of defaulted firm scores, we avoid double counting firms. Double counting can cause problems, especially when considering standard errors, which usually assume independence within the sample.

---

34 In Australia a 12 month period was used; in the case of Germany we have used 18 months because the data which we have used refer to bankruptcy, which is a later definition of default than the defaults used in Australia, generally happening about 6 months after a typical bank default event.
References


Falkenstein, E., Boral, A., and A.E. Kocagil, 2000, RiskCalc™ for Private Companies II: More Results and the Australian Model, Moody’s Investors Service Special Comment, December.


Moody's RiskCalc™ For Private Companies: The German Model

To order reprints of this report (100 copies minimum), please call 800.811.6980 toll free in the USA. Outside the US, please call 1.212.553.1658.

Report Number: 72043