Rating Methodology

Overview

In recognition of the growing need for benchmarks in the rating of middle market companies, Moody’s KMV is expanding the RiskCalc™ network of models for estimating probabilities of default for individual firms. RiskCalc™ models are estimated based on actual company-specific financial statement data.

The previous version of RiskCalc™ model for private Australian companies was released at the end of 2000 as one of the first RiskCalc™ private firm default models. At the time of this writing, the new version of the Australian RiskCalc™ model joins RiskCalc™ private firm models for the US, Canada, Mexico, Japan, Germany, Spain, France, UK, Netherlands, Belgium, Portugal, as well as a US bank model. Each model in this suite is designed to allow users to attach probabilities of default to private firms throughout the world in a consistent manner and framework. This allows users to attach default probabilities to a private firms and to rank them in terms of their default risk.

Since the launch of the earlier version of RiskCalc™ Australia, we have almost doubled the number of Australian defaults in our Credit Research Database (CRD), further refined techniques for improving data reliability, and augmented our dataset with an industry code field. This additional data information along with refined estimation techniques have allowed us to create a notably more powerful and consistent version of the RiskCalc™ model for Australian private companies.

Highlights

1. We describe Moody’s RiskCalc™ for Australian companies, a predictive statistical model of default, the factors in the model, the modeling approach, and the accuracy of the model.

2. Innovations over the original version of RiskCalc™ for Australian firms include:
   - an expanded data set of over 2500 defaults, an increase of about 75% in the default data set;
   - improvements in modeling and validation methodology; and
   - the inclusion of industry specific risk factors.

3. We find that in testing, RiskCalc™:
   - outperforms publicly available alternatives by a significant margin both in- and out-of-sample;
   - exhibits significant gains in overall power, versus the previous version of RiskCalc™;
   - uniformly has more power across:
     - industry sectors,
     - geographic regions,
     - firms sizes,
     - financial statement types, and
     - historical time periods.

4. Moody’s RiskCalc™ for Australian companies was developed using over 90,000 financial statements from almost 30,000 Australian middle market borrowers observed between 1990 and 2001. This dataset includes over 2,500 defaulted and private firms.

1. For the most up to date list of available models, the reader is referred to the Moody’s website www.moodyskmv.com
The following is a self-contained description of the development and validation of the new version of the Moody's RiskCalc™ for Australian private companies. However, some details are omitted as a more detailed handling of some of the methodology is contained in RiskCalc™ for Private Companies: Moody's Default Model and RiskCalc™ for Private Companies: Australia documents.

We have organized the remainder of this report as follows:

- Introduction
- Model Factors
- Model Framework
- Validation and Empirical Test
- The Dataset
- Implementation Tips
- Conclusion
- Appendices
- References

**Introduction**

Experience has shown that a key determinant of lending performance is the ability of institutions to correctly assess the credit risk within their portfolios. Objective quantitative default models are becoming increasingly vital in this effort.

A selected list of the applications of default models includes:

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

- **Monitoring and Credit process optimization**: while a single number may prove inferior to the judgement of a credit expert, the default model can help to pinpoint those cases where human judgement adds the most value.

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

- **Securitization**: banks are increasingly seeking to offer their clients a full range of lending services, without desiring to hold the full capital that this would require. At the same time, investors are seeking new classes assets, prompting a need for a transparent, objective measure for evaluating bank loans and other obligations. Tools such as RiskCalc™ provide such measures and can facilitate the evaluation of non-Agency rated credits within Collateralized Debt Obligations (CDOs). As a consequence Moody's Investors Service is currently a major user of RiskCalc™ in the evaluation of non-Agency rated credits within CDOs.

In addition to these uses, which require a powerful, efficient tool that allows unambiguous comparison of the credit quality of different loans and companies, the accurate pricing and trading of credit risk also demands that any such tool is calibrated to a probability of default. RiskCalc™ is designed to provide such an independent benchmark, tied directly to a probabilistic interpretation, for most credit decision needs.

We believe that in order for any tool to qualify as a benchmark it must satisfy the following conditions:

1. **It must be understandable**, 
   Customers consistently indicate that it is as important for them to understand why a model works as it is for the model to provide marginal improvements in accuracy. The ratios driving a particular assessment should be clear and intuitive.

2. **Powerful**, 
   A model, which 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 decision making.

---

2. CDOs include CBOs and CLOs.
3. **Calibrated to probabilities of default (PDs), and**

While an uncalibrated model can be used to decline or accept credits, it is of little use in ensuring that any risk assumed is accurately priced and capitalized. Furthermore, it will be of little use for trading debt. Thus, a benchmark must be tied directly to probability measures through empirical calibration.

4. **Empirically validated.**

Without documented performance on large datasets, prudence dictates that a third-party model must be viewed skeptically. Such testing also gives the user confidence that the model is stable and has not been “overfitted”.

If a model does not satisfy these criteria then, while it may be a useful tool, it cannot be considered a benchmark for the market. For example, market participants could not use a more powerful tool in secondary market transactions if the model outputs had not been calibrated. While we are confident that the model we have developed for Australian private companies is very powerful, we concede that more powerful models could exist. Nevertheless, the products that form the RiskCalc™ suite are capable of being true benchmarks: they are easy to use, intuitive, powerful, calibrated, and validated.

**An Introduction To The Modeling Approach**

The estimation of a second-generation model for Australia is a natural consequence of our extended data acquisition and improved global knowledge of quantitative credit risk assessment. This effort aims to incorporate the aforementioned new information into the RiskCalc™ model, and improve its overall performance as well as across different industry sectors, regions and firm sizes.

There are steps to the core RiskCalc™ modeling process: transformation, modeling, and mapping. In the transformation stage variables are converted statistically from sometimes noisy raw data into more useful and powerful representations that aid in default prediction. In the second step, these transformed variables are combined statistically to yield a risk score. In the third step, the score is mapped to an empirical default probability. Prior to these steps, significant analysis takes place to select a subset of variables that predict default well, and after the model is built significant work is done to validate the model and give confidence that it is robust and will perform well in the future. Further details of our methodology can be found in the **Modeling Framework** section.

**Model Factors**

RiskCalc™ for Australian private companies uses seven broad categories of risk: size of the firm, leverage, profitability, debt coverage, liquidity, activity and growth. This section provides a description of these ratios and explains how they are calculated, and how they are related to default behavior.

In this section, we present the results of the transformation stage of our analysis graphically. In the x-axes of Figures 1 through 6, the population is sorted by percentiles (according to the variable), and then the corresponding observed default rate (for each bucket) is measured. A steeper slope clearly indicates a higher discriminatory power of a given variable, whereas a continuous (upward or downward) slope throughout suggests that the power of the variable is not restricted to a subsegment of the population only. In interpreting the graphs it is useful to recall that, by definition the ranges of the underlying financial ratio that corresponds to each bucket vary from one variable to another.

As we discuss, these relationships are also an important component of the variable selection process.

**Profitability**

In developing the new version of RiskCalc™ Australia, we evaluated a number of alternative measures of profitability. Following our analysis, we found the current and previous year’s values of net income minus extraordinary items over sales ratio as powerful measures representing the profitability factor. In addition, EBITDA over interest expense was included as a factor to measure the debt coverage of firms. As we report in the **Empirical Tests** section, we checked for possible correlations between the selected model variables and verified that the correlations between the chosen variables were within reasonable limits.$^3$

Our hypothesis that firms with lower profitability would subsequently default more frequently was confirmed by the data, and can be seen in **Figure 1** for all of the different measures of profitability.
We observe that both Return on Assets (ROA) as well as Net Income Less Extraordinary Items over Sales (NIXS) perform relatively well. Nevertheless, overall NIXS performs better: even in segments where the power of ROA is minimal NIXS seems to demonstrate power, as it can be seen comparing the slopes of the two curves for different intervals in Figure 1. Another reason for choosing profit to sales ratios is that it is less correlated with other factors used in the model (Cash to Assets, Tangible Net Worth to Tangible Assets) than ROA.

In addition to the two factors above, we include the previous year's net income to sales (NIS) in the model. This is done to address three issues: First, a track record of consistent profitability is likely to be more indicative of stable profits and low default risk than high profits in one year. In that context, it may be argued that profit growth ratios (in our growth category) may capture part of the dynamics for a given firm. Nevertheless, the u-shaped relationship between growth variables and default rates means that both high and low rates of profit growth are associated with higher likelihood of default. In order to further distinguish whether a given percentage profitability was realized starting from a higher or lower starting point we also need the previous year's NIS value. Second, analysts are often suspicious of extraordinary items and therefore reluctant not to consider them in computing profits – extraordinary items are real expenses. By including NIS of the previous year, we penalize a company that reports extraordinary items consistently year after year. Third, the measure works well empirically in a multivariate context – it is consistently statistically significant and increases the power of the model across multiple specifications.

**Leverage/Gearing**

Leverage is an important measure of the credit risk of a firm since it measures the firm's ability to withstand unforeseen circumstances. Within the current version of the RiskCalc™ model for Australian private firms the leverage/gearing of a firm is captured by the ratio of tangible net worth to total tangible assets and the retained earnings to total assets ratio.

The first ratio is an important indicator of a company's financial stability because tangible net worth (expressed as a percentage of tangible assets) can be thought of as a cushion in a downturn, or during a period when the company is making losses. Similarly, retained earnings expressed as a fraction of total assets can be thought in a similar vein as a proxy: the larger the value of reported retained earnings for a given firm, the less will be its likelihood of defaulting.

The univariate default relationships of the aforementioned ratios and that of the liabilities to assets ratio are presented in Figure 2. We found that tangible net worth to tangible assets outperformed liabilities to assets in a multivariate context, which is likely due to the uncertainty of the value represented by intangible assets.

---

4. Note that unlike the case of an ordinary linear model, the profitability measures are introduced in a non-linear context to the model (through the transformations). Thus, this choice does not lead to collinearity problem per se.
Liquidity

There are many different liquidity ratios in common usage, but at heart they measure similar things, differing generally in how much value they place on different types of current assets. The current version of RiskCalc™ Australia model uses the current ratio (current assets over current liabilities), and cash and marketable securities expressed as a portion of total assets. Figure 3 demonstrates that companies with lower current ratios and smaller holdings of cash and marketable securities tend to have higher default probabilities.

Activity

We examine the activity of firms by looking at their inventory to sales ratio. We expect that given the same level of sales higher inventories would indicate a higher level of riskiness for the firm. Thus, one would predict that the default probability would decline, as the magnitude of this ratio becomes smaller. Figure 4 illustrates that our expectation is verified by empirical data and unlike the current liabilities/sales, or accounts receivables ratio, inventory ratio is powerful both at the upper and lower end of the spectrum.
It is prudent to examine statistically whether the dynamics of the firm behavior is related to future default behavior. In this context, the current version of the model incorporates two variables to account for the dynamics of the firm. The first one measures the growth of its sales, and the other variable measures the change in the return on assets. As in the case of any other growth variable, we expect a u-shaped behavior in the probability space, so that very low and very high growth values are associated with elevated levels of default risk. Figure 5 illustrates the selected growth ratios and their behavior vis-à-vis observed default probabilities. As it can be seen sales growth tends to have the largest slope in the top 25%, which is followed by net income growth (NIGR). Change in return on assets is found to fare better than NIGR in a multivariate context and thus is included instead of NIGR in the model.5

5. Change in ROA tends to be a more stable representation of how income changes between two fiscal years. For example, consider a firm whose ROA went from 1% to 2%. Most analysts would regard this as a modest increase in earnings. The Change in ROA is 1% (a modest increase in earnings), whereas Net Income Growth is 100% (a dramatic increase in earnings).
Size

Size, by its nature, is a variable that is correlated with many financial statement inputs and the quality of financial statements. In empirical literature there are numerous studies testing the impact of size on various performance measures of a given company. One typically proxies firm size by value of its total assets or its sales volume, and expects to observe a negative relationship between size and default probability. Figure 6 exhibits some alternative size measures and their relationship to default based on Australian data. We note that sales and current asset variables are rather constant over different percentiles, whereas total assets series exhibits the anticipated negative relationship.

Comparison Of Model Variables

It may be interesting to compare the model variables that are chosen in the current version of the Australian model with those of the previous version: Figure 7 summarizes the comparisons graphically by factor category.

Inspecting the graphs we can make a few observations. First, we note that size now exhibits a stronger non-linear relationship to default risk: impact of size on default risk is strong up to a certain threshold (approximately 25\textsuperscript{th} percentile), beyond which it has a notably smaller impact. Similarly, we find that the inventory variable now exhibits a strong effect for the lowest 40 percentiles or so and only then does its effect become relatively ‘flat’ (until it slopes upwards again in the upper 1/3 of the sample). Finally, we note that the new version employs two leverage ratios in contrast to the previous version, which included only a single measure of leverage.

It is important to note that, on average we observe that the univariate power of the model variables have not degraded notably. This shows that the model selection process we implemented when estimating the earlier version yielded robust-performing model variables. The individual factor weights however are somewhat different than the previous version as discussed in the next section. Another key difference between the two versions is the fact that the new version of the model also utilizes information about the industry/sector of the analyzed firm, thus addressing issues regarding industry specific variation in default risk.
Figure 7

Model Variables

Assets

Growth Measures

Percentiles

Assets New Version
Assets Old Version

Percentiles

Sales Growth New Version
Sales Growth Old Version
Change in ROA New Version

Activity Measures

Liquidity Measures

Percentiles

Inv to Sales New Version
Inv to Sales Old Version

Percentiles

Cash and Equivalents to Assets New Version
Current Ratio New Version
Cash and Eq to Assets Old Version
Current Ratio Old Version

Leverage Measures

Profitability Measures

Percentiles

Retained Earnings to Assets New Version
Tangible Net Worth to Tangible Assets New Version
Liabilities to Assets Old Version

Percentiles

Net Income to Sales New Version
Net Income to Sales Previous Year New Version
Net Income to Assets Old Version
EBITDA to Interest Expense New Version

Moody's Rating Methodology
Model Framework

RiskCalc™ for Australian private companies is a non-structural, empirical model that is estimated utilizing country-specific empirical default and financial statement data. Our modeling approach values parsimony in both functional form and the number of inputs. Our modeling approach can be briefly summarized in the following three steps: transformation, modeling, and mapping.

- **Univariate Analysis and Transformation**: the aim of single factor analysis is to study the individual relationship to default of a set of potentially relevant factors that could be regarded a priori as independent variables in the final model. In this step we also mini-model the factors and transform them.

- **Model Specification and Estimation**: once the individual factors have been analyzed, the next step is to specify a model, using a subset of the most powerful factors. These factors are combined in a logistic model and their weights are optimized.

- **Calibration**: finally, once the model has been specified and its weights estimated it is necessary to map the output of the model, a score, to a specific probability of default.

In this section, we discuss these steps in detail and then describe the characteristics of the final model that they produce.

Univariate Analysis And Transformation

A specific characteristic of credit scoring models based on financial statement information is the large number of variables that could potentially be used to predict default. It is very easy to define several hundred financial ratios, combining all of the information contained in the financial statements of a company in very different ways to assess its credit worthiness. However, this would likely lead to a model that would be over-fit: it would perform well on development data but not in the real world.

The way financial statement information is used to build the model is crucial in determining the capability and robustness of the final model in predicting default. In particular, some of the financial ratios that can be derived will be useful to predict default, but others are likely to be spuriously related to the default variable. Furthermore, some of the ratios can take extremely high or low values for some companies, without adding any information for default prediction purposes. These two facts highlight the importance of the variable selection and transformation processes that are performed during the single factor analysis phase.

Given the large number of possible ratios, it is important to reduce the list of ratios that enter the final model selection process. This screening of ratios is based on the following criteria:

- **They must be intuitive**: If the final model is to be intuitive and make business sense, it must include factors that are intuitive and make business sense.
- **They must be powerful**: We want to keep in our set of potential regressors those factors that have a high discriminatory power between defaulted and non-defaulted companies.
- **There must be enough observations**: Developing and validating a model requires a large number of observations. Furthermore, a large number of missing values typically indicates that the information is difficult to obtain, and hence it would not be prudent to include it in the final model.

We test the predictive power of each ratio using the accuracy ratio (AR) which measures the ability of a metric to differentiate between firms that later went on to default from those that did not. Where a ratio has little predictive power we exclude it from further analysis. Nevertheless, since it is possible for some relatively weak factors (in terms of univariate power) to add new and useful information to the model in a multivariate context, we typically choose a small subset of candidates in each category and examine their performances in a multivariate context (in terms of AR as well as collinearity) as well before we finalize our variable set.

Having excluded counter-intuitive or uninformative ratios in the previous steps, we transform the data by mini-modeling the selected factors to capture their relationship to default. As shown in Figure 8, this relationship is generally monotonic, meaning that the slope is either always positive, so that a higher ratio value indicates a higher probability of default, or always negative, so that a lower ratio value indicates a higher probability of default. It is also apparent from Figure 8 that in this example, the relationship is a non-linear one, which is typically the case.

---

6. A model that is complicated may not be very useful for the practitioners as they would like to understand the drivers and the intuition of the model.
7. The x-axis shows the percentile in which a particular ratio value lies, and the y-axis shows the default frequency that is observed for firms with ratios in that percentile.
8. Growth ratios typically exhibit a U-shaped relationship with default and thus are non-monotonic, per se.
Given this monotonicity we model the relationship to default so that we capture it in a smooth manner and “cap” extreme values as part of our transformation process. This “capping” not only eliminates the impact of outliers in the estimation of the parameters of the final model, but ensures that the final score calculated for a firm is not distorted by the impact of a small number of observations in the “tails” of our data set. Moreover, it also reflects the fact that beyond a certain level, most ratios provide little additional information about default.

Model Specification And Estimation
In the second step, the selected transformed factors undergo a process of multivariate analysis, to determine the predictive power of different 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 follow some basic guidelines and limit the number possible models.

There is no hard and fast rule in determining how many ratios a particular rating 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 should not exceed the typical risk factors we have found to matter in default modeling.
- The factors and their weights should be intuitive.
- The model should have high explanatory power.

This step involves using the transformed inputs within a multivariate model so that the weights assigned to the multivariate model are appropriately adjusted not only for their univariate power, but for their power in the presence of other, often correlated information. Thus the model accounts for correlations, just not through any direct interaction terms such as net income × sales growth.

9. Furthermore, from a statistical point-of-view, a large number of ratios increases the error/variance in the estimates of the weights for each factor. As the size of a development data set decreases the confidence one has about the significance risk of factors selected by a statistical procedure also decreases.

10. We are not categorically against such extensions, but we are very wary of the degrees of freedom they bring forth. Moreover, there are no obvious interactions within the data, and the number and type of non-obvious interactions is sufficiently numerous to introduce more error from overfitting than nuanced enhancement at this point.
We then use these transformations as the input to a binary model that predicts default. In our case we estimate a probit model, which uses the normal or Gaussian cumulative distribution function, specifically:
\[
y = \text{Prob}(\text{default} \mid x; \beta, \sigma) = \Phi(\beta' T(x)) = \int_{-\infty}^{\beta' T(x)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt
\]

The advantage of the probit model, as opposed to, say, ordinary least squares, is that it specifically accounts for the fact that the output, being binary, is restricted to be between 0 and 1.

The resulting model once estimated is a generalized linear model, in that it is a nonlinear function of a linear model:
\[
y = \Phi(f(x, \beta)), \quad \text{where the linear part is simply}
\]
\[
f(x, \beta) = \beta_0 + T_1(x_1) \beta_1 + T_2(x_2) \beta_2 + \ldots + T_{10}(x_{10}) \beta_{10}
\]
the T(X)'s are the transformations (as in Figure 8, above).

\textit{Calibration Mapping}

The final part of the modeling consists of mapping the output of the model to probabilities of default. This exercise can conceptually be divided into two parts. The first, serves to ensure that the average default rate predicted by the model will equal our best estimate of the average population default rate, over an 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 other RiskCalc™ models, where the overall sample is used to create a calibration curve (which maps a score to a PD). This calibration curve is then adjusted so that the implied population default rate matches our best estimate of the anchor point default rate: 1.7%.

\textbf{Figure 9} below shows how the output of the model is mapped to empirical default probabilities. Note that the ordinal ranking along the x-axis implies that the output of the model could be in arbitrary units. We estimated the relationship between model output and sample default probability using a smoothing algorithm. As it turns out, this smoothing process is done identically that of the input transformations described above.\textsuperscript{12}

\textsuperscript{11} Specifically, a Probit function.

\textsuperscript{12} We use local regression techniques (loess) in this process.
The ultimate adjustment is simply

\[ PD_{\text{Population}} = PD_{\text{Sample}} \times \frac{\text{population default rate}}{\text{sample default rate}} \]

To summarize, the transformation and normalization of input ratios constitute a transparent way of capturing the information that each ratio carries about the likelihood of default. The probit model is an efficient method of determining the optimal weights for combining the input ratios coupled with expert judgement review of the raw statistical output. Finally, the mapping transforms the score output into easily interpretable probabilities of default, which in turn are mapped to a rating grade scale\textsuperscript{13}.

The RiskCalc™ Australia Weights

One way to understand how the model works is to consider the approximate weightings\textsuperscript{14} on various factors. If we group the inputs into their broad risk categories we get the following weight distribution for the old and new versions of the RiskCalc™ model for Australian private companies:

<table>
<thead>
<tr>
<th>Factor</th>
<th>Australia (Version 1.0)</th>
<th>Australia (Version 1.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquidity</td>
<td>33%</td>
<td>26%</td>
</tr>
<tr>
<td>Leverage</td>
<td>35%</td>
<td>21%</td>
</tr>
<tr>
<td>Profitability</td>
<td>15%</td>
<td>19%</td>
</tr>
<tr>
<td>Activity (Inventories)</td>
<td>7%</td>
<td>18%</td>
</tr>
<tr>
<td>Growth</td>
<td>5%</td>
<td>9%</td>
</tr>
<tr>
<td>Size</td>
<td>5%</td>
<td>7%</td>
</tr>
</tbody>
</table>

**Top Three Risk Factors Remain The Same, Even Though The Exact Weightings Are Different**

We note that the top three risk factors have remained the same between the two versions. Nevertheless, the exact weights of the variables are somewhat different. We observe that the current version puts relatively less weight on the liquidity and leverage ratios, whereas it applies a larger weight on profitability, size and the other categories. This is probably due to the fact that: a) in this version we include industry effects in the model, and b) the transforms for size, activity and some other categories exhibit stronger power. Overall, we note that the weights are more evenly distributed in the new model, which is a desirable feature in an empirical prediction model for robustness purposes.

As mentioned earlier, a key difference between the two versions is the fact that the new version of the model utilizes information regarding the industry/sector of the firm under analysis. If we were to examine the weights of the risk factors net of industry effects, we obtain the figures as reported in **Table 1a** above. On the other hand, same set of factor weights inclusive of industry effects is displayed below in **Table 1b**.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquidity</td>
<td>21%</td>
</tr>
<tr>
<td>Industry</td>
<td>21%</td>
</tr>
<tr>
<td>Leverage</td>
<td>16%</td>
</tr>
<tr>
<td>Profitability</td>
<td>15%</td>
</tr>
<tr>
<td>Activity</td>
<td>14%</td>
</tr>
<tr>
<td>Growth</td>
<td>7%</td>
</tr>
<tr>
<td>Size</td>
<td>6%</td>
</tr>
</tbody>
</table>

We observe that inclusion of industry takes away weight from all categories; in particular, from leverage, profitability and liquidity categories. This is not very surprising, as it is known that firms in different industries tend to operate at differing leverage levels. In a similar way average profitability and liquidity cushion may also vary across industries.

\textsuperscript{13} See Appendix E for a description of the relationship between the “pd” rating grade scale used with RiskCalc™ models and the widely recognized Moody’s Investor Services ratings grade scale.

\textsuperscript{14} It is important to note that because RiskCalc™ is a non-linear model, the weights given here are only approximations for average firms.
To control for variations in the default rate across industry groups, in the new version of the model the constant term in the regression is adjusted for the industry. This calls for a positive adjustment in some industries, and negative in others. Figure 10 summarizes the impacts of different industries on the estimated PD. Consistent with our previous modeling experience in several other countries, the largest positive adjustment to the PD is observed in the case of construction industry. In the case of mining industry, on the other hand the model reduces the estimated PD level. In addition, we note that in the long-term model the industry effects tend to increase the PD levels by a relatively larger amount (or reduce them by a lesser fraction).

**Validation And Empirical Tests**

The primary goals of validation and testing are to:

- determine how well a model performs;
- ensure that a model has not been overfit and that its performance is reliable and well understood;
- confirm that the modeling approach, not just an individual model, is robust through time and credit cycles.

Model validation is an essential step to credit model development. We take care to perform tests in a rigorous and robust manner while also guarding against unintended errors. For example, it is important to compare all models on the same data. We have found that the same model may get different performance results on different datasets, even when there is no specific selection bias in choosing the data. To facilitate comparison, and avoid misleading results, we use the same dataset to evaluate RiskCalc™ and competing models.

**Model Power**

Historically, the primary model performance (power) measure used by academics has been the level of misclassification errors for a model, which assumes that there is some score/cut-off below which firms are rejected (and above which they are accepted) and then measures the percentage of misclassifications. This measure is calculated based on the percentage of defaulting firms that are accepted, and the percentage of non-defaulting firms that are rejected, and depends upon the cut-off selected. Essentially, power curves extend this analysis by plotting the cumulative percentage of defaults excluded at each possible cut-off point for a given model.

One way of interpreting the power curve is that it illustrates the percentage of defaulting firms that would be excluded as one excluded more and more of the worst “rated” firms in a data set. Thus one could interpret a power curve which went through (10%, 50%) as meaning that if one excluded the 10% of firms with the worst scores, one would exclude 50% of all firms which subsequently defaulted. In comparing the performance of two models on the same data set, the more powerful model will “exclude” a higher percentage of defaults for a given percentage of firms excluded (so the power curve will appear more bowed towards the top left corner of the chart).

---

15. Statistically speaking, the Type I and Type II error rates, where Type I error indicates accepting a firm which defaults, and Type II indicates rejecting a firm which does not default.

16. In fact, if you know the default rate in a sample, you can calculate the Type I and Type II error rates for a particular cut-off from the power curve.

17. It should be noted that most performance statistics are sensitive to the underlying data set and hence that a meaningful comparison can only be made between two models if the same data set is used (see Stein 2002 for a discussion and empirical examples).
Based on this interpretation, one can also conceive of a “perfect” model which would give all defaults worse scores than non-defaults, and a “random” or uninformative model, which would “exclude” defaults at the same rate as non-defaults. Figure 11 shows what the power curves for a typical model, a “perfect” model and a “random” model would look like.

The accuracy ratio (AR) summarizes the power curve for a model on a data set, and compares this with that of the “perfect” and “random” model. The accuracy ratio measures the area under the power curve for a model and compares this with the area under the “perfect” and “random” models, as shown in Figure 11 above. Thus the “perfect” model would have an accuracy ratio of 100%, and the random model would have an accuracy ratio of 0%. When comparing the performance of two models on a data set, the more powerful model on that data set will have a higher accuracy ratio.

In addition to checking the absolute performance of the rating tool, we compared the performance of our tool to that of Altman’s Z-score\(^{18}\), a benchmark chosen for its popularity in accounting and financial analysis texts. As one can see from Figures 12a and 12b the tool we developed significantly outperforms the modified Z-score. Moreover, as Figure 12a illustrates for the 1-year horizon the overall power of the model is notably better in relation to the previous version of the model, as evidenced by the dominance of the curve for the new model of the curve for the previous version. On the other hand, in the 5-year horizon the overall power of the model is marginally better than the old model. That said, even at the five year horizon, significant improvement is observed in the industry- and size bracket specific tests as discussed below.

Table 2 presents overall accuracy ratios for both versions of the RiskCalc™ Australia as estimated on the new and larger dataset. We report the accuracy ratios for both short and long horizons and compare them with the Z-Score.

<table>
<thead>
<tr>
<th>Model</th>
<th>1-Year AR</th>
<th>5-Year AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskCalc™ Australia (version 1.5)</td>
<td>49.3%</td>
<td>38.9%</td>
</tr>
<tr>
<td>RiskCalc™ Australia (version 1.0)</td>
<td>43.8%</td>
<td>36.1%</td>
</tr>
<tr>
<td>Z-score</td>
<td>31.9%</td>
<td>24.5%</td>
</tr>
</tbody>
</table>

\(^{18}\) Altman’s Z-score (Altman, 1968) is defined as: \( Z = 6.56\times\frac{\text{Working Capital}}{\text{Assets}} + 3.26\times\frac{\text{Retained Earnings}}{\text{Assets}} + 6.72\times\frac{\text{EBIT}}{\text{Assets}} + 1.05\times\frac{\text{Net Worth}}{\text{Liabilities}}. \)
Both models display significant power, and RiskCalc™ is a significant improvement on the Z-score benchmark.

Model performance is a function of two factors: the data set used to test a model and the model itself. Thus the performance of a model should only be compared with the performance of another model on the identical dataset, as only this allows for a true comparison of model power alone. Table 2, which exhibits model power results of three different models on the same dataset, makes two major points. In all cases both RiskCalc™ versions dominate the Z-Score, a benchmark chosen for its popularity in accounting and CFA texts. Furthermore, we note that the new version of RiskCalc™ also performs better than the older version on both horizons.

Having discussed the accuracy ratio comparison of the old and the new versions of RiskCalc™ model for Australian private companies, and having established that the new version clearly dominates the previous version, it may also be useful to examine the default probability graphs (final PDs) in percentile space. Figure 13 shows the final model PDs of both versions, where the y-axis is in logarithm of PD and x-axis is in percentiles. Accordingly, the steepness of this curve in slope represents the ability of the model to differentiate between the good and bad firms. The relatively more powerful models will start at a point below the alternative curve (at 0%), and end up at a point that is higher than the alternative (at 100%). We note that is exactly the case here, and the new version exhibits a steeper slope as described above.
In the next several sub-sections, we evaluate the power of RiskCalc™ for Australian companies for specific subsets of the population. In particular, we examine the performance of the model for different audit quality, industry, geographic region, size and year.

**Audit Quality**

As shown in Table 3, only about 28% of the underlying data that RiskCalc™ Australia model is estimated on is known to be audited or accountant prepared. It is important, therefore, that power be consistent across both audited versus unaudited statement data, as otherwise the usefulness of the model would be severely constrained.

Table 3: Audit Quality Of Data

<table>
<thead>
<tr>
<th>Type Of Statement</th>
<th>Percent Of Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal preparation</td>
<td>40.0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>30.8%</td>
</tr>
<tr>
<td>Accountant Prepared</td>
<td>14.8%</td>
</tr>
<tr>
<td>Audited</td>
<td>13.2%</td>
</tr>
<tr>
<td>Tax Return</td>
<td>0.8%</td>
</tr>
<tr>
<td>Other</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Figure 14 below shows the relative performance of RiskCalc™ across these data subsamples for Australian firms. While performance does degrade slightly for unaudited/incomplete statements, the model retains its discriminatory power on these statements. One can assert that audit quality may be thought to be positively correlated with firm size. If that is true, a model that controls for size would exhibit smaller or perhaps no significant AR deviation for the two types of financial statements. Recalling that the 5-year model controls for size, whereas the 1-year model does not utilize that variable explicitly, the results in Figure 14 are consistent with this assertion, as the 5-year AR deviation is smaller than the one observed in the 1-year case.
A performance test in which many practitioners are interested is the power of the model in different industry segments. Put differently, the question is whether the model is capable of providing powerful estimates across different industries. Practitioners are also concerned with the model’s performance across different size brackets. In order to address these issues, we performed a series of model power tests by industry and size groups using the same dataset and the old and the current versions of the model. Specifically, we measure the power of both versions of the model on the new data set. We also include the power of the Z-score for reference. The results are summarized for short and long term horizons in Tables 4a-4b and Tables 5a-5b, respectively.

### Table 4a: Model Power By Industry – 1-Year Model

<table>
<thead>
<tr>
<th>Industry</th>
<th>Percent Of Defaults</th>
<th>AR - New Version</th>
<th>AR - Old Version</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>7.0%</td>
<td>43.1%</td>
<td>41.0%</td>
<td>18.1%</td>
</tr>
<tr>
<td>Construction</td>
<td>12.0%</td>
<td>44.4%</td>
<td>40.4%</td>
<td>27.1%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>16.5%</td>
<td>57.3%</td>
<td>53.6%</td>
<td>40.7%</td>
</tr>
<tr>
<td>Mining</td>
<td>0.3%</td>
<td>54.3%</td>
<td>68.4%</td>
<td>92.5%</td>
</tr>
<tr>
<td>Other</td>
<td>3.1%</td>
<td>51.5%</td>
<td>54.0%</td>
<td>47.2%</td>
</tr>
<tr>
<td>Service</td>
<td>26.2%</td>
<td>47.7%</td>
<td>38.2%</td>
<td>29.8%</td>
</tr>
<tr>
<td>Trade</td>
<td>30.5%</td>
<td>51.2%</td>
<td>48.0%</td>
<td>35.6%</td>
</tr>
<tr>
<td>Transportation</td>
<td>4.4%</td>
<td>44.6%</td>
<td>37.8%</td>
<td>22.5%</td>
</tr>
</tbody>
</table>

### Table 4b: Model Power By Industry – 5-Year Model

<table>
<thead>
<tr>
<th>Industry</th>
<th>Percent Of Defaults</th>
<th>AR - New Version</th>
<th>AR - Old Version</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>6.9%</td>
<td>26.6%</td>
<td>26.5%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Construction</td>
<td>13.3%</td>
<td>30.3%</td>
<td>28.9%</td>
<td>20.7%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>15.9%</td>
<td>37.2%</td>
<td>36.3%</td>
<td>26.6%</td>
</tr>
<tr>
<td>Mining</td>
<td>0.4%</td>
<td>67.2%</td>
<td>81.9%</td>
<td>65.7%</td>
</tr>
<tr>
<td>Other</td>
<td>3.1%</td>
<td>37.5%</td>
<td>41.0%</td>
<td>29.7%</td>
</tr>
<tr>
<td>Service</td>
<td>25.8%</td>
<td>39.0%</td>
<td>34.1%</td>
<td>24.4%</td>
</tr>
<tr>
<td>Trade</td>
<td>29.7%</td>
<td>43.0%</td>
<td>41.2%</td>
<td>28.2%</td>
</tr>
<tr>
<td>Transportation</td>
<td>4.9%</td>
<td>47.3%</td>
<td>47.3%</td>
<td>23.6%</td>
</tr>
</tbody>
</table>
Tables 5a&5b show that the new version of the model outperforms the older version in all industry groups (in both the 1-year and 5-year horizons), except mining and ‘other’ categories. Nevertheless, one notes that both categories have a very small number of defaults, so that it is hard to make any statistically meaningful inferences in those industries. If we put industries with very small default observations aside, we observe that the model exhibits highest ARs in the following industries: manufacturing, trade, transportation and services in both the 1-year and 5-year models. The performance of the Z-score model is consistently and notably lower than either version of the model. The only cases where its performance appears to be higher than RiskCalc™ at first glance are really statistical artifacts of small number of defaulters, e.g. in the mining industry.

The performance of the model by asset size exhibits a similar pattern (Tables 5a and 5b). Namely, the new version of the model outperforms the old version in all size classes. Moreover, it should be noted that in some categories the differences in ARs are quite notable and there are no exceptions to this finding in either time horizon. Finally, we note that both RiskCalc™ versions dominate the Z-score by a wide margin.

Recall that the industries with the highest ARs in Tables 4a and 4b are the industries that account for the majority of our database. Thus, given the relatively wide coverage of our Australian CRD efforts,19 under the assumption that our data reflects a portrayal of a typical middle market portfolio, we have suggestive evidence that the new version of the RiskCalc™ model is a powerful tool in evaluating the credit risk of middle market portfolios across various industries and/or size groups.

Power Performance By Year And Region

One may be interested in evaluating the performance of the model across time or region. Thus, we also conducted model power tests for different years (Tables 6a and 6b), as well as different regions (Tables 7a and 7b). Examining the model power results by year shows that the new version of the model dominates the old version of the model in both short as well as long-term horizons. Thus, this test reveals that the model's performance is robust across time.

---

19. For details on CRD please see: http://riskcalc.moodysrms.com
Similarly, the results in Tables 7a and 7b illustrate that the model power is also robust across different regions, and once again the new version dominates the old version of RiskCalc™ and the Z-Score. Seemingly, the only exception to this statement is the region of Tasmania, but given the low number of defaults and the very slight difference in performance, it is difficult to make conclusive statistical comparisons between the ARs of the two versions in this case.

### K-Fold Power Tests

We present the in-sample performance of the model in Figure 12, above. The findings show that the model exhibits high degrees of power in distinguishing good credits from bad ones. However, these are power statistics based on the entire sample. An immediate question one may raise is whether these performance statistics would hold for different segments of the overall sample. Put differently, we would like to know whether these results are robust throughout the sample and are not influenced by a particular subsample of it.

A standard test of robustness is the so-called “k-fold test.” In order to implement this test, we divided the failing and non-failing banks into k-equally-sized segments. This yielded k equally-sized observation subsamples that exhibit the identical overall default rate and are temporally and cross-sectionally independent. Accordingly, we estimate the model on k-1 sub-samples, and score the k-th subsample. We repeat this procedure for all possible combinations, and put the k scored “out-of-sample” subsamples together and calculate an accuracy ratio on this combined data set. Note that this test is always out of sample for the data being tested.
Table 8a summarizes the $k$-fold test results (with $k=5$). The reported figures are the accuracy ratios by the corresponding sample and time horizons.

<table>
<thead>
<tr>
<th>Validation Sample</th>
<th>1 year AR</th>
<th>5 year AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subsample 1</td>
<td>46.8%</td>
<td>44.2%</td>
</tr>
<tr>
<td>Subsample 2</td>
<td>45.7%</td>
<td>41.1%</td>
</tr>
<tr>
<td>Subsample 3</td>
<td>48.4%</td>
<td>41.7%</td>
</tr>
<tr>
<td>Subsample 4</td>
<td>50.1%</td>
<td>45.2%</td>
</tr>
<tr>
<td>Subsample 5</td>
<td>49.7%</td>
<td>43.3%</td>
</tr>
<tr>
<td>K-Fold Overall AR</td>
<td>49.1%</td>
<td>39.9%</td>
</tr>
<tr>
<td>In sample AR (Table 2)</td>
<td>49.3%</td>
<td>38.9%</td>
</tr>
</tbody>
</table>

Examining Table 8a, one observes that the overall $k$-fold AR results are very close to the in-sample test results, which provides evidence that there do not appear to be subsegments of the dataset that influence the model parameters overly, and thus the model performance appears robust across our data.

**Walk-Forward Tests**

Another typical issue of concern in evaluating a prediction model is assessing whether it may be overfit to the development dataset: if that is the case the model may appear to be very powerful based on in-sample statistics, nevertheless its power out of sample may be drastically worse. The best way to address this issue is to see how the model would have performed in the past against future data by comparing its predictions against what actually happened. This can be accomplished by means of a so-called walk-forward test.

In the walk-forward test, we estimate the model on the data up to a certain point in the past and score the future year (relative to that point) with that model. Next, the cut-off for the estimation data is advanced a year and the process is repeated. The process is continued in this manner until there is no future data available. Then all the scored out-of-sample subsamples are combined and the accuracy ratio (AR) and power curves on the combined set is calculated. Note that this test is always out-of-sample and out-of-time for the data being tested since no prediction is ever made using a model estimated on the data being tested.

The 1- and 5-year walk-forward power curves are displayed in Figures 15a and 15b, respectively.
As Figures 15a and 15b illustrate, the model performs relatively well in an out-of-sample context for both short and long-term horizons yielding ARs of 43.2% and 43.1%, respectively (the corresponding in-sample ARs of the new model are 49.3% and 38.9%). Thus, the results suggest that the new version of the RiskCalc™ model provides rather robust results in an out-of-sample context and it does not present any symptoms of overfitting.

One may be tempted to compare the walk-forward AR figures with the reported (in-sample) figures of the old model, which are also around 43%. Nevertheless, that is not a fair test since the old model was estimated on a significant portion of that dataset: the old model was developed on a data set that ran through 1999. Therefore using a sample of 1999 financial statements to predict subsequent defaults provides a true out-of-sample test on the old model. Thus, our walk forward results from 1999 provide an out-of-sample power comparison between the two models. Both models display significant out of sample power, but the new version of RiskCalc™ is an improvement over the previous version in both the short and long-term time horizons.

Table 8b summarizes the findings on the out-of-sample performances of the two versions. Accordingly, the new version outperforms the old version, which also confirms that it is a relatively more robust and powerful prediction model in an out-of-sample context.

In sum, the findings affirm that our model performs better than any publicly available alternative on an overall basis. Furthermore, model performance analysis by industry, size category, year and region also reveals that the new version of the model is notably better than the old version of RiskCalc™ model for Australian private companies. Finally, k-fold and walk-forward results re-affirm that the model performance is robust with respect to sample and is very powerful as measured in an out-of-sample context.

Correlations
Although less directly related to model predictive power, an important check of the model, and one of the first tests we performed, is an evaluation of the correlation structure between the model variables. Clearly, one would like to avoid creating a model that includes highly correlated variables as this would imply the presence of collinearity in the model and thus the undesirable econometric consequences related to this problem.
Fortunately, we find that in general the fact that variables enter the model pursuant to a non-linear transformation process can serve to alleviate this. Thus, the introduction of factors, which may be correlated with each other in levels (non-transformed form), does not necessarily pose a problem for this model, as it is estimated on the transformed data.

We estimated the correlation matrices for the entire set of model variables. The results are presented below in Table 9a. As the estimated correlation coefficients show there is no pair of variables that is of alarming concern in the model. Typically correlations are well below 50%, with only few exceptions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Assets</th>
<th>NI - XI / Sales</th>
<th>Current Ratio</th>
<th>Cash To Total Assets</th>
<th>RE To Total Assets</th>
<th>Debt Coverage</th>
<th>Inv To Sales</th>
<th>Sales Growth</th>
<th>Change In ROA</th>
<th>NI To Sales Last Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Assets</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NI - XI / Sales</td>
<td>0.051</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Ratio</td>
<td>0.081</td>
<td>0.087</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cash To Total Assets</td>
<td>0.108</td>
<td>0.118</td>
<td>0.263</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RE To Total Assets</td>
<td>0.192</td>
<td>0.282</td>
<td>0.208</td>
<td>0.166</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debt Coverage</td>
<td>0.072</td>
<td>0.518</td>
<td>0.138</td>
<td>0.185</td>
<td>0.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inv To Sales</td>
<td>0.161</td>
<td>0.202</td>
<td>0.293</td>
<td>0.013</td>
<td>0.019</td>
<td>0.002</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales Growth</td>
<td>0.051</td>
<td>0.033</td>
<td>0.022</td>
<td>0.008*</td>
<td>0.07</td>
<td>0.078</td>
<td>0.045</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tangible Net Worth To</td>
<td>0.234</td>
<td>0.265</td>
<td>0.298</td>
<td>0.172</td>
<td>0.664</td>
<td>0.283</td>
<td>0.007*</td>
<td>0.028</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Tangible Assets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in ROA</td>
<td>0.148</td>
<td>0.05</td>
<td>0.003*</td>
<td>0.019</td>
<td>0.067</td>
<td>0.009*</td>
<td>0.012</td>
<td>0.158</td>
<td>0.062</td>
<td>1</td>
</tr>
<tr>
<td>NI to Sales Last Year</td>
<td>0.059</td>
<td>0.426</td>
<td>0.088</td>
<td>0.065</td>
<td>0.227</td>
<td>0.242</td>
<td>0.124</td>
<td>0.055</td>
<td>0.211</td>
<td>0.048</td>
</tr>
</tbody>
</table>

\* All figures except the ones marked by an asterisk are significant at 99 percent level.

Variance Inflation Factors

The diagonal elements of the inverse correlation matrix (i.e., -1 times the diagonal elements of the sweep matrix) for variables that are in the equation are also sometimes called variance inflation factors (VIF; e.g., see Neter, Wasserman, Kutner, 1985). If the predictor variables are uncorrelated, then the diagonal elements of the inverse correlation matrix are equal to 1.0; thus, for correlated predictors, these elements represent an "inflation factor" for the variance of the regression coefficients, due to the redundancy of the predictors. As Woolridge (2000) shows VIF is inversely related to the tolerance value, such that a VIF of 10 corresponds to a tolerance value of 0.10. Clearly, any threshold is somewhat arbitrary. Nevertheless, if any of the R^2 values are greater than 0.75 (so VIF is greater than 4.0), we suspect that multicollinearity might be a problem. If any of the R^2 values are greater than 0.90 (so VIF is greater than 10) then conclude that multicollinearity is a serious problem.

Table 9b illustrates, in all cases the estimated VIF values are below 2.5, i.e. notably below the threshold levels of 4 and 10 that are commonly used in VIF analysis when testing for presence of multicollinearity. Thus, we conclude that the findings strongly indicate that the model variables do not present any substantial multicollinearity.

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets</td>
<td>1.11</td>
</tr>
<tr>
<td>NI - XI / Sales</td>
<td>2.11</td>
</tr>
<tr>
<td>Current ratio</td>
<td>1.28</td>
</tr>
<tr>
<td>Cash to Total Assets</td>
<td>1.13</td>
</tr>
<tr>
<td>RE to Total Assets</td>
<td>1.84</td>
</tr>
<tr>
<td>Debt Coverage</td>
<td>2.05</td>
</tr>
<tr>
<td>Inventories to Sales</td>
<td>1.26</td>
</tr>
<tr>
<td>Sales Growth</td>
<td>1.06</td>
</tr>
<tr>
<td>Tangible net worth to tangible assets</td>
<td>1.81</td>
</tr>
<tr>
<td>Change in ROA</td>
<td>1.09</td>
</tr>
<tr>
<td>NI to Sales Last Year</td>
<td>1.22</td>
</tr>
</tbody>
</table>

20. The sampling distribution variance for OLS slope coefficients can be expressed as: \( \text{Var}(b_j) = (1-R_j^2)^{-1} \sigma^2 (n-1)^{-1} (\sigma^2) \). In this formula, \( R_j^2 \) is the explained variance we obtain when regressing \( x_j \) on the other \( x \) variables in the model, and \( \sigma^2 \) is the variance of \( x_j \). Recall that the variance of \( b_j \) is used in constructing the t-ratios that we use to evaluate significance. This variance is increased if either \( \alpha_x^2 \) is large, \( \alpha_y^2 \) is small, or \( R_j^2 \) is large. The first term of the expression above is called the variance inflation factor (VIF). If \( x_j \) is highly correlated with the other \( x \) variables, then \( R_j^2 \) will be large, making the denominator of the VIF small, and hence the VIF very large. This inflates the variance of \( b_j \), making it difficult to obtain a significant t-ratio. To some extent, we can offset this problem if \( \alpha_x^2 \) is very small (e.g., there is little noise in the dependent variable—or alternatively, that the x’s account for most of the variation in y). We can also offset some of the problem if \( \alpha_y^2 \) is large. Increasing the variance of \( x_j \) will also help generate more noise in the regression of \( x_j \) on the other x’s, and will thus tend to make \( R_j^2 \) smaller.

21. As Woolridge (2000) shows VIF is inversely related to the tolerance value, such that a VIF of 10 corresponds to a tolerance value of 0.10. Clearly, any threshold is somewhat arbitrary. Nevertheless, if any of the R^2 values are greater than 0.75 (so VIF is greater than 4.0), we suspect that multicollinearity might be a problem. If any of the R^2 values are greater than 0.90 (so VIF is greater than 10) then conclude that multicollinearity is a serious problem.

22. See Woolridge (2000) for details.
The Dataset

The one of the design objectives of each model in the RiskCalc™ network of models is to provide credit risk benchmarks for those firms for which well understood credit measures are not available. The goal of the current RiskCalc™ model is to provide a probability of default (PD) for private Australian companies in the middle market.

However, the use of a single model to cover all company types and industries is often inappropriate due to the very different nature of some firms. In order to create a more powerful model for the Australian middle market, we eliminated the following types of companies from our data set:

- **Small companies** – the future success of the smallest firms is often as dependent on the finances of the key individuals as that of the company. For this reason, we excluded companies that never had assets of more than 100,000 Australian Dollars.
- **Financial institutions** – the balance sheets of financial institutions (banks, insurance companies, and investment companies) exhibit significantly different characteristics than those of other private firms. (For example, they tend to exhibit relatively high gearing/leverage.) Furthermore, the fact that financial institutions are generally regulated, and often required to hold capital, suggests that they should be considered separately.
- **Real estate development companies** – the success or failure of a real estate development and investment company often hinges on a particular development, so that the annual accounts often provide only a partial description of the dynamics of the firm and its likelihood of default.23
- **Public Sector and Non-Profit institutions** – estimating default probabilities for government run companies can complicated by the fact that the states or municipalities which use/own them have historically often been unwilling to allow them to fail.

It is widely accepted in the financial analysis and accounting communities that the financial statements of smaller companies, such as those in the middle market, can be on average less accurate and of lower quality than those of larger companies. Therefore, we further cleaned the database to ensure that we did not include financial statements with highly suspect accounting. For example, we excluded financial statements from our database based on plausibility checks of particular positions in financial statements (e.g. assets less than zero) or where the financial statement covered a period of less than twelve months.

**Descriptive Statistics Of The Data**

Moody’s KMV’s proprietary Credit Research Database (CRD) is critical to the development of RiskCalc™ models in the markets we serve. Due to the opacity of private firm financial and default histories, the primary sources of CRD data are the active portfolios of domestic financial institutions through CRD Participation. In Australia, the seven largest domestic commercial lenders have participated in our Credit Research Database default study since 1999. The result is the largest known repository of Australian obligor financial statement and credit performance data.

Table 10 provides a summary of the data set used in development, validation and calibration of RiskCalc™ model for Australian private companies and compares it with those used in developing other RiskCalc™ models, such as the US, UK as well as the previous version of the Australian model. As the table shows the new version of the model is estimated based on a dataset with more financial statements and substantially larger number of defaulters than the previous version.

| Table 10 Information On Private Firm Sample Data |
|-----------------|----------------|----------------|-----------------|----------------|
| **Country**     | **Time Span**  | **Unique Firms** | **Unique Firm Defaults** | **Financial Statements** |
| Australia (V 1.5) | 1990-2001 | 29,636 | 2,519 | 93,701 |
| Australia (V 1.0) | 1990-1999 | 27,712 | 1,447 | 79,877 |
| United Kingdom  | 1989-2000 | 64,531 | 4,723 | 283,511 |
| United States   | 1989-1999 | 33,964 | 1,393 | 139,060 |

---

23. This is also the case for many types of "project finance" firms, and we would recommend use of separate models for these. At the time of writing, this characteristic is explicitly recognized within the proposals for the new Basel capital accord.
Figure 16 shows that our Australian financial statement data peak around 1998-1999, where the majority of statements are from the 1997-2000 period in the sample. Default behavior also exhibits a peak in 1998 in frequency and to a lesser extent in 2000. We suspect that the low default figures in early 1990s merely reflect a data collection problem rather than a near-zero default environment.24

Figure 17 shows the relative industry concentrations in our dataset. Note that the largest categories are trade (25%), services (25%), and manufacturing (14%). Unlike the early version of our Australian dataset, companies with unknown sector information constitute only about 12% the dataset.25 The remaining 23% are accounted for by construction, agriculture, transportation, mining and other type of firms. This industry coverage represents a significant improvement from the previous dataset. This improvement has permitted the inclusion of industry information in the new version of the model.

Figure 18 below exhibits the distribution of our financial statement data by size groups. As can be seen from the chart, the majority of companies (about 75%) in our dataset have assets in the range of 200,000 to 5,000,000 Australian Dollars. Thus, our dataset reflects the size distribution of a typical middle market portfolio, with the lower and upper tails accounting for about 11% and 14% of the data coverage in the sample.

24. We should mention in passing that this issue does not pose a problem for our modeling strategy as we correlate past financial statements with future default events (with a separate constant window size for the 1- and 5-year models).

25. Industry information was unknown for 75% of the sample that the earlier version of RiskCalc™ Australia was built on.


**Definition Of Default**

Since most companies do not default, defaulting companies are more rare and thus more valuable from an information perspective. Much of the dearth in default data is due to the vagaries of data storage within financial institutions. Defaulting companies are sometimes purged from the system after their troubles begin which results in a sample bias in that the default probability implicit in current bank databases is invariably low.

Our intention in developing RiskCalc™ models is to provide assistance to banks and other institutions or investors in determining the risk of incurring losses as a result of company defaults, missed payments or other credit events. The proposals for the new Basel Capital Accord (BIS II) have stimulated debates about what constitutes an appropriate definition of default. As there are no definitions that can be applied universally across all countries we have found that the criteria outlined below applies to most of the advanced economies in the world.

In the current study default is defined as any of the following events:

- 90 days past due
- bankruptcy
- placement on internal non-accrual list
- write-down

If we determine that any of these actions being taken we consider the obligor defaulted as of that date.

**Aggregate Default Probability Assumptions**

As we also reported in the December 2000 RiskCalc™ Australia model document, Table 11 illustrates two indicators of bank credit quality that help guide our top-down opinion of aggregate default probabilities in varying countries. In general the banking sectors in the various countries have roughly similar ratios of non-performing loans as a percent of total assets, which implies that asset quality is approximately equal across the countries. The different proportions of consumer and small business lending in these figures clearly make aggregate comparisons difficult, nonetheless they are still informative. Another top-down note within Table 11 is the median price-earnings ratio (P/E) of banks, especially the P/E ratio of banks relative to other corporates in each country. The ratio is highest in Australia, followed by Canada, and finally the US. This fact is consistent with the view that the riskiness of Australian banks is lower than that of Canada or US banks. Moreover, we note that the average P/E for all corporates in Australia is generally lower than for Canada and the US, so that, relatively, banks in Australia are a less risky investment than in Canada or the US.

---

We also performed an informal survey of data contributors, and invariably the expected charge-off rate, the same one used in provisioning and pricing models, was somewhere between 30 and 50 basis points for middle market lenders in each country. Loan recovery rates, also from informal survey, as well as what is documented in Gupton (2000), are approximately 70-80%. When combined with our US loan recovery point estimate of 75%, this again gives us comfort that the Australian experience is quite similar to the US experience. These survey points suggest a reasonable default probability boundary between 1.2% and 2.0%. Finally, the PDs estimated by the previous version of the model (which assumes a central tendency rate of 1.7%) have been found generally consistent with the long-term credit quality estimates of the local banks. Thus, it appears most appropriate to assume that Australia has a long-term central tendency default rate of 1.7%, the same rate we used for the previous version of the Australian model.

**Implementation Tips**

Our aim in developing the RiskCalc™ suite of products is not merely to provide a set of powerful tools, but also to ensure that they can be used without imposing onerous data requirements on users. As a result we have chosen to use information that is reliable and readily available. Based only on information in the annual accounts, RiskCalc™ Australia produces very powerful results. However, prudence dictates that if an analyst has access to additional important information he or she should consider it. For example, if the analyst were aware that there are strong ties between the firm being evaluated and a subsidiary, and that the subsidiary is experiencing difficulties, then this information should be considered when making pricing or lending decisions. As recognized in the new Basel capital accord, successful analysis depends not just on having high quality information and powerful tools, but also on how these are implemented into an overall credit process.

However, as acknowledged in proposals for the new capital accord, and demonstrated by our validation results above, information contained in a firm's financial statements can prove a very powerful predictor of default. Thus, in addition to its use as a validated objective measure of default probability, we also see significant scope for use of RiskCalc™ within an internal credit rating system along with the bank's own expertise to take into account some of the non-financial elements mentioned above.

Many users are used to higher default rate projections for individual companies. The average probability of default for middle market firms of 1.7% per year may be at variance to the individual credits to which the model is applied, particularly when one considers that this average default probability is consistent with Ba2 probability of default. (Many lenders consider private credit to be in the B2-B1 range on average, not the Ba2 range.) While we have used a 1.7 % figure, users should recognise that our objective is to make the model unbiased. That is, we have designed it to represent 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, and extending a loan to a firm which subsequently defaults, is generally higher than the lost revenue from rejecting a customer.

It is widely accepted that in using financial statement information to assess the credit-worthiness of a firm, it is desirable to use the most recent and representative information. However, while it may therefore be desirable to use information from interim statements, it is important to bear in mind that any P&L (income statement) figures must be carefully annualized and that such statements are usually unaudited.

---

27. Recall that charge-offs = default rate x (1 – recovery rate).

28. Moody’s Risk Adviser is another Moody’s KMV product which has been used by many banks to capture and combine non-financial elements within an internal rating system, and can be used to combine the outputs of RiskCalc™ with non-financial elements.

29. Failure to annualize an interim statement might well lead to very poor profitability and debt coverage ratios while poor annualization (e.g. simply multiplying income statement items by 4 for a quarterly statement) could be misleading in cyclical/seasonal industries.
Similarly, while RiskCalc™ is powerful at a variety of horizons, and while we believe that using a score based on the previous year's statement would generally be preferable to not using a quantitative score at all, the user should consider the extent to which an older financial statement reflects the current situation of a firm. For example, if the user knows that a firm has undergone significant re-structuring since publishing their last annual statement (e.g. a merger or divestiture) thoughtlessly inputting these numbers could produce misleading results. In such a case, one should aim to use the most comparable figures available.

**Target For RiskCalc™ For Australian Private Companies**

It is also important to bear in mind that, while 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 available, the tool will have difficulties in differentiating how risky a firm is, but it can still be used.

The types of firm where we would recommend that users treat the results with caution are: financial institutions; public sector firms; firms whose shares are actively traded/listed; firms whose performance is dominated by a couple of specific projects (e.g. real estate development firms); firms with assets of less than $100,000; 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 ratings for these firms will creep in, not only because their financial statements may not capture the whole picture, but also because the aggregate probability of default for these types of firm may well be significantly different from the population norm.

**Conclusions**

In this document we describe the new version of Moody’s RiskCalc™ model for Australian private companies, a predictive statistical model of default, the factors in the model, the modeling approach, and the accuracy of the model.

The RiskCalc™ methodology is true to the essence of applied econometrics: 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-modeling”, 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 modeling as a forward-looking problem and so we are careful to check for robustness, both through cross-validation and out-of-sample tests, and through an emphasis on simplicity. For our Australian model, careful attention has been paid to how financial ratios could differ between Australia and other countries considering the particularities of the Australian economy both from a micro and macro perspective. Careful attention has also been paid to how these ratios relate to default and to selecting the most parsimonious, yet robust, way 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.

The new version of the RiskCalc™ model for Australian companies was estimated utilizing a dataset with over 90,000 financial statements from almost 30,000 Australian middle market borrowers observed between 1990 and 2001 and over 2500 defaults. In addition to the financial risk factors that were included in the original version, the new version also includes industry-specific risk factors.

We found that the new version of the model outperforms the older version of the model and other publicly available alternatives by a significant margin in in- and out-of sample testing, exhibits significant gains in overall power, versus the previous version of RiskCalc. Moreover, it uniformly has more power across industry sectors, geographic regions, size brackets, financial statement types, and historical time periods.

Using the RiskCalc™ model for Australian private companies should help improve profitability through the credit cycle, be it through use in decisioning, pricing, monitoring or securitization. As a powerful, objective model, it serves the interests of institutions, borrowers and investors alike. While RiskCalc™ is not intended as the ultimate measure of risk, it should be viewed as a very powerful aggregator of financial statement information, which generates a meaningful and validated number that allows for the consistent comparison of portfolio risks.

30. For example, as a result of the careful regulation of financial institutions, the default rates for these firms are generally very low.
Appendix A: Testing Metrics

A power curve\textsuperscript{31} is constructed by plotting, for each score, $m$, the proportion of defaults with a score worse than $m$, against the proportion of all firms with a score worse than $m$. In order to plot the power curve for a model, one should do the following:

- Score all the firms with the model.
- For each score, $m$, calculate the percentage of all firms with scores worse than $m$ - this is the x-axis value\textsuperscript{32}.
- For each score, $m$, calculate the percentage of defaulted firms with scores worse than $m$ - this is the y-axis value.

Thus, if a particular model or metric $M$, gave 5\% of all firms a score worse than $m$, and 10\% of all defaults a score worse than $m$, then its power curve would go through the point (0.05,0.1). This could be interpreted as meaning that if one were to reject all credits with a score in the worst 5\% (based on $M$), then one would exclude 10\% of all firms who go on to default.

If we consider a particular metric $M$, for which we bucket the scores into $B$ different bins, then the height of the power curve in a particular bin, $b$, would be calculated as follows:

\[
\text{power} \left( b \right) = \frac{\sum_{i=1}^{b} D(i)}{\sum_{i=1}^{B} D(i)} = \frac{\text{defaults excluded at } b}{\text{total defaults}} \tag{A.1}
\]

where, $\text{power}(b)$ is the height of the power curve in bin $b$ and $D(b)$ is the number of defaults in bin $b$.

The result is Figure 19 below which plots the power curve for a metric $M$ (the line $\text{Power}(M)$, which relates to the left hand axis). In this case we rank-order the firms from risky (left) to less risky (right). This 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.

---

31. Also known as the CAP plot.
32. Here “worse than” is taken to indicate that the firm is higher risk, i.e. more likely to default.
33. We use percentage on the x-axis rather than the score output so that two models, with possibly different ranges of scores, can be compared to one another on the same data set.
Figure 19 also demonstrates the fact that a power curve, together with a default rate, implies a particular calibration curve (this is plotted as Calib(M) which relates to the right hand axis). The default rate for a particular percentile, is equal to the slope of the power curve at that point, multiplied by the average default rate for the sample. Thus, for any point $m$ along a default metric:

$$\text{Calib}(m) \quad p(t) = \bar{p} \times \frac{\partial \text{power}(m)}{\partial m},$$

where $\bar{p}$ is the mean probability of default, and $\frac{\partial \text{power}(m)}{\partial m}$ is the slope of the power curve at point $m$.

**Accuracy Ratio**

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 allows unambiguous comparison. The metric that we use, called the Accuracy Ratio, compares the area under the power curve for the model with the area under the random and perfect models. A more powerful model will be bowed out towards the left, and will have a larger area, resulting in a higher accuracy ratio.

The accuracy ratio is defined as the ratio of the area between the actual model and the random model to the area between the perfect model and the random model (see Figure 11 in the Empirical Tests section for a graphical demonstration). Thus the perfect model would have an accuracy ratio of 100% and a random model would have an accuracy ratio of 0%.

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 accuracy ratio is not a measure of global or complete dominance, just an intuitive measure of dominance on average.

It should be noted that it would be inappropriate to compare the accuracy ratios for two models on two different data sets, since any model tested on two different data sets will get different accuracy ratios on the data sets. The accuracy ratio does however allow one to compare the performance of two models on the same underlying data set.
Appendix B: Power Curve Construction Details

The testing approach is as follows. Typically we have annual financial statements for each firm until they default. If a firm defaults within 90 days of the financial statement that firm-year observation is dropped (see next paragraph for justification). We assign a default score to each financial statement. If the difference between the default date and the financial statement date (days until default) is within the default window (90 to 730 days for the 1 year model and 90 to 1825 days for the 5 year model) that firm-year observation is labeled a bad. Likewise, if the firm does not default within the window that firm-year observation is labeled as a good. We retain all good firm-year observations, but only one bad firm-year observation per firm. Specifically, we retain the earliest bad firm-year observation for each firm. Each remaining firm-year observation is then mapped into a percentile according to their score, and this collection of percentiles is the basis by which the power curve is created.

We exclude firms that default within 90 days of the financial statement date to avoid the misleading results that come from model performance over irrelevant time periods, such as 60 days after a statement date. Predicting defaults of very short horizons, such as less than 90 days, is basically useless, as very few statements are completed within this time. Many lenders take 6 months to be confident that most of their middle market exposures have delivered their latest annual statements. 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 also cause problems, especially with standard errors that usually assume independence within the sample.
Appendix C: The Relation Between RiskCalc™ PDs And Dot-PD Ratings And Moody’s Investor Services Long-Term Bond Ratings

RiskCalc™ PDs and Moody’s long-term bond ratings are not directly comparable. They are two different, though related, credit risk measures. Exhibit 1 compares many aspects of the two systems side-by-side, highlighting similarities and differences.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>RiskCalc™ PDs</th>
<th>Moody’s Long-Term Bond Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of Study</td>
<td>Obligor</td>
<td>Obligation and/or Obligor</td>
</tr>
<tr>
<td>Time Horizon</td>
<td>Specific, one or five years</td>
<td>Non-specific, long term</td>
</tr>
<tr>
<td>Risk Dimension</td>
<td>One dimensional: Probability of default</td>
<td>Multi-dimensional: Probability of default, severity of default &amp; transition risk</td>
</tr>
<tr>
<td>Information Requirements</td>
<td>Large, reliable, electronic datasets</td>
<td>Robust to poor quality or missing data</td>
</tr>
<tr>
<td>Volatility</td>
<td>High</td>
<td>Low - maintained through the cycle</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Support</td>
<td>Technical</td>
<td>Technical + Analyst contact &amp; insight</td>
</tr>
<tr>
<td>Scale</td>
<td>Continuous/Absolute</td>
<td>21 Risk Buckets/Relative</td>
</tr>
<tr>
<td>Structure</td>
<td>Simple, codified analysis of few variables</td>
<td>Flexible as situation may require</td>
</tr>
</tbody>
</table>

Despite the important differences between RiskCalc™ PDs and Moody’s long-term bond ratings, some users of one or both risk nomenclatures find it helpful to compare them. Moody’s bond default study provides a basis for such a comparison. This study rigorously correlates Moody’s long-term corporate bond ratings with ex-post default frequency, allowing us to calculate historical average bond default rates for each rating category. By mapping a firm’s PD into the historical average bond default rates, we create dot-PD ratings (e.g., Aaa.pd, Aa1.pd, Aa2.pd, Caa2.pd, … Caa3.pd, Ca.pd, C.pd), which facilitate comparison with long-term bond ratings. Moody’s bond default study is available over Moody’s KMV’s web site at [http://www.moodyskmv.com](http://www.moodyskmv.com). The details of the PD mapping to historical average long-term bond default rates are described in the May 2000 Special Comment, “Moody’s Default Model for Private Firms: RiskCalc™ for Private Firms”, also available from the web site.

Dot-pd ratings carry no additional information beyond PDs and are not long-term bond ratings for all of the reasons highlighted in Exhibit 1. They are, rather, a re-statement of the PDs and provide a short-hand nomenclature for probabilities of default. Our clients have found that, for some purposes, communicating risk levels in terms of alpha-numeric ratings rather than probabilities, is more intuitive. For example, for many, the difference between two companies with 0.0075 and 0.0131 probabilities of default is not as easily understood as the difference between an A3.pd company and a Baa1.pd company.

While dot-pd ratings are not the same as long-term bond ratings, there is a correlation between them. The correlation, by construction, is not exact. Ratings, as indicated in Exhibit 1, are functions of not only PD, but also of the severity of loss in the event of default (which incorporates key structural differences in instruments such as senior vs. subordinate, secured vs. unsecured, external supports) and an issuer’s risk of sudden, large changes in credit quality. Moody’s bond default study correlates ratings with only one of these risk dimensions, probability of default, while holding constant the severity of loss and ignoring transition risk. For this reason, by construction, the correlation between the two systems is imprecise.

An analogous situation is the relationship between a person’s weight to their height and girth. There is a strong enough correlation between weight and height that we may draw the conclusion that taller people, on average, weigh more than shorter people. However, we could more accurately predict weight if we knew not only height but also girth. Analogously, we could more accurately predict Moody’s bond ratings if, in addition to PD, we know the severity of loss, the transition risk, and the other differences outlined in Exhibit 1.

The intent of Moody’s RiskCalc™ models is not to substitute or predict Moody’s bond ratings. They are designed to calculate expected probabilities of default for defined time horizons. The output of these models, combined with correlation estimates, will facilitate quantification of risk at the obligor and portfolio level. In contrast to PDs, which are produced by a formula that relates information in selected financial ratios to probabilities of default, Moody’s analyst ratings are based on a more flexible and focused review of qualitative and quantitative factors, distilled by an analyst (and rating committee) with sectoral expertise and in-depth understanding of an issuer’s competitive position and strategic direction.

Despite the structural difficulties in directly comparing PDs with long-term bond ratings, many of our customers will find the systems complementary and valuable in different ways as part of a risk management solution.

---

34. The severity of loss can be captured through use of LossCalc, another Moody’s KMV product which provides a measure of the expected loss in the event of default.
References


