Moody's RiskCalc™ For Private Companies: Korea

Overview

In a continuing effort to provide benchmarks for private middle market companies, Moody's KMV has built a model for estimating firm default probabilities for private Korean companies which utilizes local company financial statements. This model joins a network of international models covering private firms of the United States of America, Canada, Mexico, Germany, Spain, France, the United Kingdom, Belgium, the Netherlands, Portugal, Italy, Austria, Japan, Australia, Singapore, and the Nordic Region, as well a sectoral model for US privately-held banks.¹ The RiskCalc network allows one to consistently attach probabilities of default to firms throughout the world. As a powerful objective model based on data provided by our local Credit Research Database (CRD) initiative, RiskCalc Korea serves the interests of institutions, borrowers and investors alike.

Highlights

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

2. We find that in testing, RiskCalc Korea:
   - outperforms publicly available alternatives by a significant margin in both in- and out-of-sample testing, and
   - uniformly has more power across
   - industry sectors,
   - firm sizes, and
   - historical time periods.

3. Moody's RiskCalc for Korean companies was developed using Moody's Credit Research Database, which contains over 115,000 financial statements from over 32,000 Korean middle market borrowers observed between 1986 and 2001. This dataset includes about 3,300 defaulted Korean private firms.

The following is a self-contained description of the development and validation of the Moody’s RiskCalc™ model for Korean 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 II: More Results and the Australian Model documents.

We have organized the remainder of this report as follows:

1. Introduction
2. Model Factors
3. Modeling Framework
4. Validation and Empirical Tests
5. The Dataset
6. Implementation Tips
7. Conclusions
8. Appendices
9. References

¹ For the most up to date list of available models, the reader is referred to Moody’s KMV website at http://www.moodyskmv.com.
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.

These applications require a powerful, efficient tool that allows unambiguous comparison of the credit quality of different loans and companies. Furthermore, 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.

3. **Calibrated to probabilities of default (PDs)**
   
   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 Korean private companies is very powerful, we concede that theoretically more powerful models could exist that are tailored for some specialized portfolios. Nevertheless, the products that form the RiskCalc network are capable of being true benchmarks: they are easy to use, intuitive, powerful, calibrated, and validated.

There are three steps that form the core of the RiskCalc modeling process: transformation, modeling, and mapping. In the transformation stage variables are converted statistically from typically 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.

---

2. CDOs include CBOs and CLOs.
3. Of course, the level of validation that can be performed depends on the amount of data that is available.
Model Factors

Moody’s RiskCalc™ for Korean private companies uses five broad categories of risk: cash flow/profitability, capital structure, liquidity, activity, and size. This section provides a description of these ratios and explains how they are calculated, and how they are related to default behavior. As we discuss below, these relationships are an important component of the variable selection process.

In this section, we present the results of the transformation stage of our analysis graphically. In the x-axes of Figures 1 through 7, the population is sorted by percentiles (according to the variable), and then the corresponding observed default rates (for each bucket) are measured and reported on the y-axes. In Figures 1-7, a steeper slope 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 correspond to each bucket vary from one variable to another.

Cash Flow: Profitability

Clearly, one of the most intuitive risk factors when modeling default behavior is the cash flow, e.g. profitability and debt coverage of the firm. One expects that higher profitability ratios should be positive signals for a firm and thus correspond to lower default probabilities.

In developing RiskCalc™ Korea, we evaluated a number of alternative measures of profitability and examined their distinguishing power of the good companies from the risky ones as described above. Following our analysis, we found that gross profit over current assets was a powerful ratio in predicting default behavior. We note that net income over total assets perform also reasonably well. Nevertheless, we chose the first ratio because it performs better: as it can be seen in Figure 1 it is an overall steeper curve.

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 both measures of profitability.

![Figure 1: More Profitable Firms Are Less Likely To Default](image-url)


**Cash Flow: Debt Coverage**

The ability of a given firm to fulfill its borrowing obligations based on its cash flow/income (debt coverage) is one of the key parameters of default prediction. Similar to the discussion above, we also evaluated a number of alternative debt coverage ratios. We find that Earnings (EBITDA) expressed as a fraction of interest expense gives a fairly accurate indication of how well a firm will be able to service its debt in the event of a downturn.

![Figure 2: Firms With More Debt Coverage Are Less Likely To Default](image)

**Capital Structure: Leverage**

The credit literature depicts leverage as an important measure of the credit risk of a firm since it measures the firm’s ability to withstand unforeseen circumstances. Following the transformation analysis, we decided to capture the leverage effect of a firm by its liabilities to assets (TL/TA) ratio in RiskCalc Korea model.

This ratio is an important indicator of a company’s financial stability because the more of the liabilities that cannot be covered by liquid assets (expressed as a percentage of total assets) the worse the company will fare in a downturn.

The univariate default relationship is presented in **Figure 3**. As seen in the figure, TL/TA has a steeper slope than total liabilities minus retained earnings over fixed assets, and thus is a better predictor of default behavior.

![Figure 3: Higher Leveraged Firms Default More Frequently](image)
**Capital Structure: Retained Earnings**
Retained earnings expressed as a fraction of liabilities (or assets) can be thought in a similar vein, as a capital structure measure signaling about the cushion the company will have in downturn.

**Figure 4: Firms With More Retained Earnings Are Less Likely To Default**

![Graph showing relationship between retained earnings and default likelihood.]

**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 RiskCalc™ Korea model uses cash and marketable securities expressed as a portion of current assets. **Figure 5** demonstrates that companies with lower current ratios and smaller holdings of cash and marketable securities tend to have higher default probabilities.

**Figure 5: Firms With Poor Liquidity Default More Frequently**

![Graph showing relationship between liquidity ratios and default likelihood.]

- Retained Earnings to Total Assets
- Retained Earnings to Current Assets
- Retained Earnings to Total Liabilities
- Cash and Marketable Securities to Current Assets
- Current Assets less Inventories to Current Liabilities
**Activity**

We examine the activity of firms by looking at their inventories to sales ratio. We expect that for a given level of inventories, higher sales would indicate a lower 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 6 illustrates that our expectation is verified by empirical data and that current liabilities to sales is less powerful than inventories to sales ratio.

![Figure 6: Firms With Poor Inventories/Sales Ratios Default More Frequently](image)

**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. It should be noted that correlation of size with other variables is not of concern in RiskCalc Korea (see: Multicollinearity Check section in Validation and Empirical Tests). One typically proxies firm size by value of its various assets, such as fixed or total assets. Figure 7 exhibits two alternative size measures - total and fixed assets in real values, i.e. adjusted for inflation, and their relationship to default behavior in Korea. Note that total assets variable exhibits stronger overall differentiating power in terms of a firm's default likelihood.

![Figure 7: Firm Size Is Inversely Related To Default Probability](image)

It may be interesting to view the transformation functions that the model utilizes all at once. Figure 8 summarizes these functions graphically by risk factors. One notes that the “good” ratios (such as profitability, coverage ratios, size, etc.) exhibit a negative slope so that a higher ratio value is associated with a reduced probability of default. In contrast, ‘bad’ ratios (such as leverage ratios, Inventories/Sales ratio etc) tend to exhibit a positive slope so that a higher ratio value is associated with a higher probability of default.
Figure 8: RiskCalc Korea Model Variables

- **Gross Profit to Current Assets**
  - Default Likelihood vs. Ratio Percentile
- **EBITDA to Interest Expense**
  - Default Likelihood vs. Ratio Percentile
- **Total Liabilities to Total Assets**
  - Default Likelihood vs. Ratio Percentile
- **Retained Earnings to Total Assets**
  - Default Likelihood vs. Ratio Percentile
- **Inventory to Net Sales**
  - Default Likelihood vs. Ratio Percentile
- **Real Assets**
  - Default Likelihood vs. Ratio Percentile
Modeling Framework

RiskCalc™ for Korean 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 individual relationships 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 hundreds of financial ratios, combining all of the information contained in the financial statements of a company in different ways to assess its credit worthiness. However, this would likely lead to a model that would be overfitted. In other words, such a model would perform well on the development dataset but it would perform poorly outside of it, thus inhibiting its potential applicability for credit analysis 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 univariate 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 demonstrates extremely small or no predictive power we exclude it from further analysis.

Having excluded counterintuitive or uninformative ratios in the previous steps, we transform the data by minimoolding the selected factors to capture their relationship to default. As shown in Figure 9, 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 9 that in this example, the relationship is a non-linear one, which is typically the case.

---

4. 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.

5. 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.

6. 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 also 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 that could be created, so it is important to follow some basic guidelines and limit the number possible models.

There is no set rule in determining how many ratios a particular rating model should contain: a model with too few variables will not capture all the relevant information, whereas another one with too many variables 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 ratios, 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.
- The model should not exhibit any detectable econometric problems, such as multicollinearity.

The modeling 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 also 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.

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:

$$y = \text{Prob} (\text{default} | x; \beta, \sigma) = F(\beta ' T(x)) = \int_{-\infty}^{\beta ' T(x)/\sigma} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{t^2}{2} \right) dt$$

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.

---

7. 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.

8. 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.
The resulting model, once estimated, is a generalized linear model, in that it is a nonlinear\(^9\) function of a linear model: \(y = \Phi(f(x,b))\), where the linear part is simply

\[
f(x, \beta) = \beta_0 + T_1(x_1) \beta_1 + T_2(x_2) \beta_2 + \ldots + T_{10}(x_{10}) \beta_{10}
\]

and the \(T(X)\)'s are the transformations (as in Figure 9, above).

**Calibration Mapping**

The final part of the modeling consists of mapping the output of the model to probabilities of default.

As Figure 10 shows, 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 estimate the relationship between model output and sample default probability using a smoothing algorithm.\(^{10}\)

Finally, we adjust the mean sample default probability to our projected population default probability by simply multiplying by a constant so that, over the entire sample, our average firm had 1 and 5 year default probabilities of 2.1% and 8.4%, respectively.\(^{11}\)

In other words, the adjustment can be presented as:

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

To summarize, the transformation 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 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 "pd" rating grade scale.\(^{12}\)

**The RiskCalc Korea Weights**

One way to understand how the model works is to consider the approximate weightings\(^{13}\) on various factors. If we consider broad risk categories we get the following weight distribution for the RiskCalc model for Korean private companies. Top three risk factors are debt coverage, leverage and retained earnings, which are closely followed by profitability, size and activity factors. Finally, liquidity accounts for about 8% of the overall weight in the model.

---

9. Specifically, a probit function.
10. We use local regression techniques (loess) in this process.
11. See the discussion on the population default rate in the Dataset section.
12. See Appendix C for a description of the relationship between the "pd" rating grade scale used with RiskCalc models and the widely recognized Moody's Investors Service ratings grade scale.
13. The reported figures are weights that correspond to the average firm in the sample.
To control for variations in the default rate across industry groups we adjust the constant term in the regression for the industry. This calls for a positive adjustment in some industries, and negative in others. Figure 11 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.

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 overfitted 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 employed by academic researchers 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.

Table 1: Relative Weights Of Risk Factors

<table>
<thead>
<tr>
<th>RiskFactor</th>
<th>RiskCalc Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debt coverage</td>
<td>24%</td>
</tr>
<tr>
<td>Leverage</td>
<td>16%</td>
</tr>
<tr>
<td>Retained Earnings</td>
<td>14%</td>
</tr>
<tr>
<td>Profitability</td>
<td>11%</td>
</tr>
<tr>
<td>Size</td>
<td>10%</td>
</tr>
<tr>
<td>Activity</td>
<td>7%</td>
</tr>
<tr>
<td>Liquidity</td>
<td>8%</td>
</tr>
</tbody>
</table>

Top Three Risk Factors Are Debt Coverage, Leverage And Retained Earnings

14. 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.

15. 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.
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).

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 12 shows what the power curves for a typical model, a “perfect” model and a “random” model would look like.

The accuracy ratio 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.

Table 2 presents overall accuracy ratios for RiskCalc Korea for the 1- and 5-year horizons. In order to provide some benchmark, we also report the accuracy ratios for the Z-Score.

As the table illustrates, RiskCalc Korea clearly outperforms the Z-Score by a wide margin, both for the 1- and 5-year horizons.

---

**Table 2: Accuracy Ratios As Measured On The Korea Dataset**

<table>
<thead>
<tr>
<th>Model</th>
<th>1-year AR</th>
<th>5-Year AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskCalc™ Korea</td>
<td>53.6%</td>
<td>47.4%</td>
</tr>
<tr>
<td>Z-score</td>
<td>28.3%</td>
<td>22.9%</td>
</tr>
</tbody>
</table>

---

16. 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.

17. Altman’s Z-score (Altman, 1968) is defined as: 0.717*[Working Capital/Assets] + 0.847*[Retained Earnings/Assets] + 3.107*[EBIT/Assets] + 0.420*[Net Worth/Liabilities] + 0.998*[Sales/Assets]
In the next several sub-sections, we evaluate the power of RiskCalc for Korean companies for specific subsets of the population. In particular, we examine the performance of the model for different industries, sizes and years.

**Model Power By Industry And Size Groups**

A performance test in which many analysts are interested is the power of the model in different industry segments. Put differently, the question is whether the model is capable in providing powerful estimates across different industries. Practitioners are also concerned with a 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. We also include the power of Z-score for reference. The results are summarized for short and long term horizons in Tables 3a and 3b and Tables 4a and 4b, for model power by industry and by size, respectively.

### Table 3a: Model Power By Industry - 1-Year Model*

<table>
<thead>
<tr>
<th>Industry</th>
<th>Percent Of Defaults</th>
<th>RC Korea</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td>62%</td>
<td>50.3%</td>
<td>30.7%</td>
</tr>
<tr>
<td>Construction</td>
<td>27%</td>
<td>48.2%</td>
<td>32.7%</td>
</tr>
<tr>
<td>Trade</td>
<td>7%</td>
<td>53.7%</td>
<td>33.1%</td>
</tr>
<tr>
<td>Services</td>
<td>2%</td>
<td>64.7%</td>
<td>37.4%</td>
</tr>
<tr>
<td>Mining, Transportation, and Utilities</td>
<td>2%</td>
<td>53.4%</td>
<td>36.2%</td>
</tr>
<tr>
<td>Other or Unknown</td>
<td>1%</td>
<td>76.4%</td>
<td>45.4%</td>
</tr>
</tbody>
</table>

*Industries with very few defaults (less than 1% of overall sample) are omitted in Tables 3a and 3b.

### Table 3b: Model Power By Industry - 5-Year Model

<table>
<thead>
<tr>
<th>Industry</th>
<th>Percent Of Defaults</th>
<th>RC Korea</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td>60%</td>
<td>44.1%</td>
<td>21.1%</td>
</tr>
<tr>
<td>Construction</td>
<td>27%</td>
<td>41.7%</td>
<td>26.8%</td>
</tr>
<tr>
<td>Trade</td>
<td>7%</td>
<td>50.3%</td>
<td>24.6%</td>
</tr>
<tr>
<td>Services</td>
<td>3%</td>
<td>61.2%</td>
<td>33.6%</td>
</tr>
<tr>
<td>Mining, Transportation, and Utilities</td>
<td>2%</td>
<td>44.1%</td>
<td>27.7%</td>
</tr>
<tr>
<td>Other or Unknown</td>
<td>1%</td>
<td>66.0%</td>
<td>17.1%</td>
</tr>
</tbody>
</table>

### Table 4a: Model Power By Size - 1-Year Model

<table>
<thead>
<tr>
<th>Size</th>
<th>Percent Of Defaults</th>
<th>AR – RC Korea</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 bil - 2 bil</td>
<td>11%</td>
<td>44.0%</td>
<td>20.3%</td>
</tr>
<tr>
<td>2 bil - 5 bil</td>
<td>17%</td>
<td>45.9%</td>
<td>15.4%</td>
</tr>
<tr>
<td>5 bil - 10 bil</td>
<td>19%</td>
<td>53.9%</td>
<td>23.3%</td>
</tr>
<tr>
<td>10 bil - 20 bil</td>
<td>27%</td>
<td>58.4%</td>
<td>23.9%</td>
</tr>
<tr>
<td>&gt; 20 bil</td>
<td>26%</td>
<td>57.5%</td>
<td>30.5%</td>
</tr>
</tbody>
</table>

### Table 4b: Model Power By Size - 5-Year Model

<table>
<thead>
<tr>
<th>Size</th>
<th>Percent Of Defaults</th>
<th>AR – RC Korea</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 bil - 2 bil</td>
<td>11%</td>
<td>37.0%</td>
<td>11.7%</td>
</tr>
<tr>
<td>2 bil - 5 bil</td>
<td>17%</td>
<td>44.4%</td>
<td>12.1%</td>
</tr>
<tr>
<td>5 bil - 10 bil</td>
<td>23%</td>
<td>50.0%</td>
<td>17.4%</td>
</tr>
<tr>
<td>10 bil - 20 bil</td>
<td>27%</td>
<td>54.6%</td>
<td>23.1%</td>
</tr>
<tr>
<td>&gt; 20 bil</td>
<td>22%</td>
<td>52.3%</td>
<td>24.8%</td>
</tr>
</tbody>
</table>
The performance of the model by asset size exhibits a similar pattern as above: The ARs tend to increase with asset size of the firms. In addition, we note that RiskCalc Korea outperforms Z-score in all size classes: it should be noted that in all categories the differences in ARs are quite notable (20-30% for both 1-year and 5-year horizons), and there are no exceptions to this finding in either time horizon.

**Power Performance By Year**

Since models are designed to be implemented at various points in time over a business cycle, one may be interested in evaluating the performance of the model at different points in time. In order to address this issue, we conducted model power tests by year (Tables 5a and 5b). This way, one can observe whether the model performance is excessively time dependent and exhibits big swings in power depending on time.

Examining the model power results over time (by year) illustrates once again that RiskCalc Korea statistically dominates Z-score in both short and long-term horizons. Moreover, as the volatility of AR is rather low for RiskCalc we conclude that the model performance is robust across time.

### Table 5a: Model Power By Year - 1-Year Model

<table>
<thead>
<tr>
<th>Year</th>
<th>Percent Of Defaults</th>
<th>RC Korea</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>2%</td>
<td>58.2%</td>
<td>29.1%</td>
</tr>
<tr>
<td>1992</td>
<td>5%</td>
<td>50.6%</td>
<td>20.9%</td>
</tr>
<tr>
<td>1993</td>
<td>6%</td>
<td>54.8%</td>
<td>18.4%</td>
</tr>
<tr>
<td>1994</td>
<td>5%</td>
<td>54.7%</td>
<td>19.2%</td>
</tr>
<tr>
<td>1995</td>
<td>9%</td>
<td>48.6%</td>
<td>24.8%</td>
</tr>
<tr>
<td>1996</td>
<td>25%</td>
<td>52.0%</td>
<td>35.1%</td>
</tr>
<tr>
<td>1997</td>
<td>14%</td>
<td>56.7%</td>
<td>46.9%</td>
</tr>
<tr>
<td>1998</td>
<td>10%</td>
<td>51.6%</td>
<td>34.4%</td>
</tr>
<tr>
<td>1999</td>
<td>19%</td>
<td>51.7%</td>
<td>18.7%</td>
</tr>
<tr>
<td>2000</td>
<td>5%</td>
<td>55.9%</td>
<td>22.7%</td>
</tr>
</tbody>
</table>

### Table 5b: Model Power By Year - 5-Year Model

<table>
<thead>
<tr>
<th>Year</th>
<th>Percent Of Defaults</th>
<th>RC Korea</th>
<th>Z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>1%</td>
<td>52.9%</td>
<td>15.4%</td>
</tr>
<tr>
<td>1989</td>
<td>2%</td>
<td>46.7%</td>
<td>27.9%</td>
</tr>
<tr>
<td>1990</td>
<td>4%</td>
<td>50.3%</td>
<td>24.5%</td>
</tr>
<tr>
<td>1991</td>
<td>5%</td>
<td>49.7%</td>
<td>25.2%</td>
</tr>
<tr>
<td>1992</td>
<td>7%</td>
<td>45.9%</td>
<td>16.7%</td>
</tr>
<tr>
<td>1993</td>
<td>10%</td>
<td>47.0%</td>
<td>21.7%</td>
</tr>
<tr>
<td>1994</td>
<td>12%</td>
<td>50.4%</td>
<td>26.2%</td>
</tr>
<tr>
<td>1995</td>
<td>14%</td>
<td>48.6%</td>
<td>23.4%</td>
</tr>
<tr>
<td>1996</td>
<td>18%</td>
<td>48.9%</td>
<td>29.0%</td>
</tr>
<tr>
<td>1997</td>
<td>12%</td>
<td>45.3%</td>
<td>24.2%</td>
</tr>
<tr>
<td>1998</td>
<td>9%</td>
<td>44.1%</td>
<td>14.5%</td>
</tr>
<tr>
<td>1999</td>
<td>7%</td>
<td>50.8%</td>
<td>19.0%</td>
</tr>
<tr>
<td>2000</td>
<td>2%</td>
<td>54.1%</td>
<td>22.7%</td>
</tr>
</tbody>
</table>

**K-Fold Tests**

As the in-sample performance evaluation of the model shows (in Table 2), the model exhibits a high degree of power in distinguishing good credits from bad ones. A natural follow-up question that one may raise is whether these performance statistics would hold for different segments of the sample as well. Put differently, in order to increase our confidence in the model we would like to know whether the model performance is robust throughout the sample and it is not driven by a particular subsample of it.

A standard test for evaluating the robustness of a model is the so-called “k-fold test.” In order to implement this test one divides the defaulting and non-defaulting companies into $k$ equally-sized segments. This yields $k$ 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.

Table 6 summarizes the $k$-fold test results (with $k=5$). The reported figures are the accuracy ratios by the corresponding sample and time horizons.

Examining Table 6, 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 any subsegments of the dataset that influence the model parameters excessively, and thus the model performance appears to be robust across the data set.

**Walk-Forward Tests**

Once the robustness of the model is established, a subsequent issue of concern 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 curve on the combined set are 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 walk-forward power curves are displayed in **Figures 13** for the 1-year horizon.

As **Figure 13** illustrates, the model performs relatively well in an out-of-sample context. This holds both for short and long-term horizons with ARs of 52.2% and 49.7%, respectively (the corresponding in-sample ARs are 53.6% and 47.4%). Thus, the results suggest that the model provides rather robust results in an out-of-sample context and it does not present any symptoms of overfitting.

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, and year also reveals that the model is robust across different subpopulations. Finally, $k$-fold and walk-forward results re-affirm that the model performance is stable with respect to sample and is very powerful as measured in an out-of-sample context.

---

18. For details on walk-forward testing procedure please see Stein (2002).
Multicollinearity Check
Although less directly related to model predictive power, an important check of the model, and one of the first tests we performed, is an econometric evaluation of the model where we check for the potential presence of multicollinearity. One would like to avoid constructing a model that includes highly collinear variables as this would imply the presence of multicollinearity in the model and thus the undesirable econometric consequences related to this problem.

Variance Inflation Factors
The diagonal elements of the inverse correlation 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.19

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash and Marketable Securities to Current Assets</td>
<td>1.16</td>
</tr>
<tr>
<td>EBITDA to Interest Expense</td>
<td>1.41</td>
</tr>
<tr>
<td>Inventory to Net Sales</td>
<td>1.27</td>
</tr>
<tr>
<td>Total Liabilities to Total Assets</td>
<td>1.74</td>
</tr>
<tr>
<td>Gross Profit to Current Assets</td>
<td>1.14</td>
</tr>
<tr>
<td>Retained Earnings to Total Assets</td>
<td>1.69</td>
</tr>
<tr>
<td>Real Total Assets</td>
<td>1.11</td>
</tr>
</tbody>
</table>

As Table 7 illustrates, in the case of all model variables the estimated VIF values are notably below the threshold levels of 4 and 10 that are commonly used in VIF analysis when testing for presence of multicollinearity.20 Thus, we conclude that the findings strongly indicate that the model variables do not present any substantial multicollinearity.

The Dataset
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 Korean 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 Korean 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 1 billion won.
- **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.

---

19. The sampling distribution variance for OLS slope coefficients can be expressed as:

\[
V(b_j) = \frac{1}{1 - R_j^2} \cdot \frac{\sigma^2}{(n-1)S_j^2}
\]

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 \( S_j^2 \) 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 \( \sigma^2 \) is large, \( S_j^2 \) 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 \( \sigma^2 \) 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 \( S_j^2 \) 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.

20. As Wooldridge (2000) shows VIF is inversely related to the tolerance value (1-\( R_j^2 \)), 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 \( R_j^2 \) values are greater than 0.75 (so that VIF is greater than 4.0), we suspect that multicollinearity might be a problem. If any of the \( R_j^2 \) values are greater than 0.90 (so that VIF is greater than 10) then we conclude that multicollinearity is a serious problem.
• **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.\(^{21}\)

• **Public Sector and Non-Profit institutions** – estimating default probabilities for government run companies can be 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 not equal liabilities plus net worth) 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 the case of Korea, KIS (Korean Information Service) has granted us the rights to use their database -- the largest known repository of Korean obligor financial statement and credit performance data.

Table 8 provides a summary of the data set used in development, validation and calibration of RiskCalc™ model for Korean private companies and compares it with those used in developing other RiskCalc™ models, such as the US, Canada as well as the new version of the Australian model.

<table>
<thead>
<tr>
<th>Country</th>
<th>Time Span</th>
<th>Unique Firms</th>
<th>Unique Firm Defaults</th>
<th>Financial Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea</td>
<td>1986-2001</td>
<td>32,228</td>
<td>3,297</td>
<td>115,723</td>
</tr>
<tr>
<td>Australia (v 1.5)</td>
<td>1990-2001</td>
<td>29,636</td>
<td>2,519</td>
<td>93,701</td>
</tr>
<tr>
<td>Canada</td>
<td>1989-1999</td>
<td>8,115</td>
<td>501</td>
<td>27,274</td>
</tr>
<tr>
<td>United States</td>
<td>1989-1999</td>
<td>33,964</td>
<td>1,393</td>
<td>139,060</td>
</tr>
</tbody>
</table>

Figure 14 shows that our Korean financial statement data peak around 1999-2000, where the majority of statements are from the 1997-2000 period in the sample. Default behavior also exhibits a peak in frequency in 1998-2000. We suspect that the low default figures in early 1990s merely reflect a data collection problem rather than a near-zero default environment.\(^{22}\)

---

21. 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.

22. 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).
Figure 15 shows the relative industry concentrations in our dataset. Note that the largest categories are manufacturing (49%), construction (22%), and trade (13%). Unknown sector information constitutes only about 1% the dataset. Services and mining-transportation-utilities firms account for the remaining 15%.

**Figure 15: Industry Distribution Of Financial Statements**

![Industry Distribution Of Financial Statements](image)

*Most Korean Firms Are In Manufacturing, Construction And Trade*

Figure 16 below exhibits the distribution of our financial statement data by size groups. As can be seen from the chart, close to half of the companies in our dataset have assets in the range of 2 billion to 10 billion Wons.

**Figure 16: Size Distribution Of Financial Statements**

![Size Distribution Of Financial Statements](image)

**Definition of Default**

Since most companies do not default, defaulting companies are relatively rare and thus more valuable from an information perspective. Much of the dearth in default data is due to the data storage issues 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 apply to most of the advanced economies in the world.
In the current study default is defined as any of the following events:

- bankruptcy
- placement on internal non-accrual list
- troubled-debt restructure
- write-down

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

**Aggregate Default Probability Assumptions**

The estimate of an aggregate probability of default is important because it serves as an anchor point for the model. Changing it upward will move all predicted probabilities of default upwards and *vice versa*. In deriving this estimate, it is important to consider the structure of the sample used in developing and calibrating a rating tool as well as its intended use. Thus, a model that was developed for use only on the very largest Korean firms would have a very different mean default rate than one developed for use only on the very smallest. Users should therefore bear in mind that the figure we use as the mean default rate has been selected because we believe that it is an appropriate figure for the model development portfolio we have used in estimating the RiskCalc Korea model.

In order to determine the mean default rate figure we calculated the annual default rates for cohorts over several years and calculated the average of them. When computing the default rates, we used the defaulters that had a financial statement in the previous year (if previous year is missing the year prior to that). For the denominator we included companies that had a financial statement or a default flag in the following year. The motivation for this is to avoid the pitfalls of over-counting (as in a case of a merger), or under-counting (as in a case of a company not filing a financial statement in the year before default), as well as having incompatible numerator and denominator in the default rate ratio. This procedure yielded us the table below. Taking the average of the observed default rates we arrived at 2.1%.

Consequently, we consulted with Moody’s analysts familiar with Korean market about the default rate: these discussions suggested that the appropriate window is approximately between 1.9% and 2.3% over long run. As our estimate happens to be in the middle of this range, this further reinforces our view that the 2.1% default rate seems appropriate for the Korean middle market.

<table>
<thead>
<tr>
<th>Year</th>
<th>Annual Default Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>2.8%</td>
</tr>
<tr>
<td>1994</td>
<td>2.4%</td>
</tr>
<tr>
<td>1995</td>
<td>2.8%</td>
</tr>
<tr>
<td>1996</td>
<td>1.0%</td>
</tr>
<tr>
<td>1997</td>
<td>1.8%</td>
</tr>
<tr>
<td>1998</td>
<td>3.0%</td>
</tr>
<tr>
<td>1999</td>
<td>1.0%</td>
</tr>
<tr>
<td>2000</td>
<td>2.0%</td>
</tr>
<tr>
<td>Average</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

In deriving the central tendency rate for the cumulative 5 year PD, we have again faced challenges given the relative lack of publicly available data. In developing the North American private model, Moody’s KMV spent considerable time in examining the relationship between the 1 and cumulative 5 year PDs and the result of this work provides the initial basis for deriving a 5 year cumulative anchor point. The benefit of this work is that it covers a substantial period of time, and can be used to supplement the information provided by our database. The results of these analyses suggest that 5 year cumulative default rate is, on average, approximately 4 times the level of the 1-year default rate. Thus in calibrating RiskCalc model for Korean private companies for the 5-year horizon, we have used an anchor point of 8.4%.

---

23. When we expand the definition of default to include “90 days past due,” the model performance drops marginally to 51.3% and 45.7% in one and five years, respectively.
24. For more details on this work, we refer the reader to the description in “RiskCalc™ For Private Companies: Moody’s Default Model”.
25. Bond default studies (e.g. Moody’s Special Comment, January 2000, “Historical Default Rates of Corporate Bond Issuers, 1920-1999”), and experience working with bank loan portfolios, show that the relationship between 1 and 5 year cumulative default rates varies by credit quality. Thus, whilst the “average” is a factor of 4, the average 5 year cumulative default rate for Aa rated bonds is more than 10 times higher than the average 1-year default rate. This variation is caused by credit migration, whereby the credit quality of highly-rated firms tends to deteriorate, whilst poorly-rated firms, if they survive, improve in credit quality.
Implementation Tips

Our aim in developing the RiskCalc™ network 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™ Korea produces very powerful results. However, prudence dictates that if an analyst has access to additional relevant information, which is not captured by the model but yet may have credit risk implications, 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.26

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 annualized27 and that such statements are usually unaudited.

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 diversification) clearly using these numbers alone could produce misleading results. In such a case, one should aim to use the most comparable figures available.

Target For RiskCalc™ For Korean 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 the riskiness of a firm.

The types of firms 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 $1 billion; 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 firms may well be significantly different from the population norm.28

Localization Issues

In developing and evaluating credit risk models for private (non-quoted) companies across different countries there are three important potential risk factors: (i) idiosyncratic factors, (ii) market-related (systematic) risk, and (ii) local (country-specific) differences.

In the private firm universe one does not have access to market price information and other related measures of riskiness that can be implemented to track the riskiness of a company. Moreover, unlike the public companies, the importance of idiosyncratic factors by far outweighs the systematic factors in the case of private companies.29 Thus, models that are developed based on public firm universe need to make particular assumptions to circumvent the missing variable problem, and they need to be specially tailored to account for this heightened level of idiosyncratic risk.

Another technical issue, which becomes prevalent in modeling credit risk across different geographies is presence of local differences that are specific to the country that is being modeled, i.e., due to tax code, regulations, local business law, and so forth. Clearly, variations in regulations and the business and tax law may affect and alter the behavior of local companies, and thus the sensitivities (coefficients) of different factors vis-à-vis default behavior.

26. Moody’s Risk Advisor 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.
27. 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.
28. For example, as a result of the careful regulation of financial institutions, the default rates for these firms are generally very low.
One may argue that firms in a given industry behave the same way irrespective where the firm may be located: this argument may hold from a technological point of view. Nevertheless, from the point of view of financing the operations and credit worthiness of a firm, differences due to localization is a crucial issue, and a model which scores companies in the same industry the same way irrespective of their locale, is likely to yield erroneous results. Moreover, if it were a model, which was originally estimated on a non-local dataset, the extent of this error would be further exacerbated.

To illustrate the importance of localization, we test the (AR) performance of three alternative models on the Korean data: RiskCalc Korea, RiskCalc North America and the Z-Score, where the latter two are estimated based on non-local (North American) default experience.

As Figure 17 illustrates graphically, RiskCalc Korea model which addresses idiosyncratic as well as country-specific risk, exhibits a significant improvement over the Z-score benchmark and RiskCalc North America, which is a model that was estimated on non-local (non-Korean) data. Since RiskCalc network of models, by definition address the idiosyncratic and local factors of risk, this finding is not surprising.30

30. As RiskCalc network of models are estimated for each country individually, the modeling framework by definition captures any potential differences local business environment, regulations and so forth.
Conclusions

In this document we describe the RiskCalc™ model for Korean 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 Korean model, careful attention has been paid to how financial ratios could differ between Korea and other countries considering the particularities of the Korean 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 RiskCalc model for Korean companies was estimated utilizing a dataset with over 115,000 financial statements from over 32,000 Korean middle market borrowers observed between 1986 and 2001 and almost 3,300 defaults.

We found that the RiskCalc model for Korean private companies outperforms other publicly available alternatives by a significant margin in- and out-of-sample testing. Moreover, it uniformly has more power across industry sectors, size brackets, and historical time periods.

Using the RiskCalc™ model for Korean 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.
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\textsuperscript{32} \( m \), 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{33}.
- 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:

\[
power(b) = \frac{\sum_{i=1}^{b} D(i)}{\sum_{i=1}^{B} D(i)},
\]

where, \( power(b) \) 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 18 below, which plots the power curve for a metric \( M \) (the line 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.

\textsuperscript{31} Also known as the CAP plot.
\textsuperscript{32} Here “worse than” is taken to indicate that the firm is higher risk, i.e. more likely to default.
\textsuperscript{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 18 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:

\[
p(m) = \bar{p} \times \left\{ \frac{\partial \text{Power} (m)}{\partial m} \right\},
\]

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 12 in the Validation and 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 default 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.

Exhibit 1: Similarities & Differences Between RiskCalc PDs And Moody’s Long-Term Bond Ratings

<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. 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 restatement of the PDs and provide a shorthand nomenclature for probabilities of default. Our clients have found that, for some purposes, communicating risk levels in terms of alphanumeric 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. subordinated, 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 knew 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


