Moody's Investors Service
Global Credit Research

June 1999

Special Comment

Contact	Phone
New York	
James V. Herrity	1.212.553.1653
Sean Keenan	
Jorge R. Sobehart	
Lea V. Carty	
Eric G. Falkenstein	

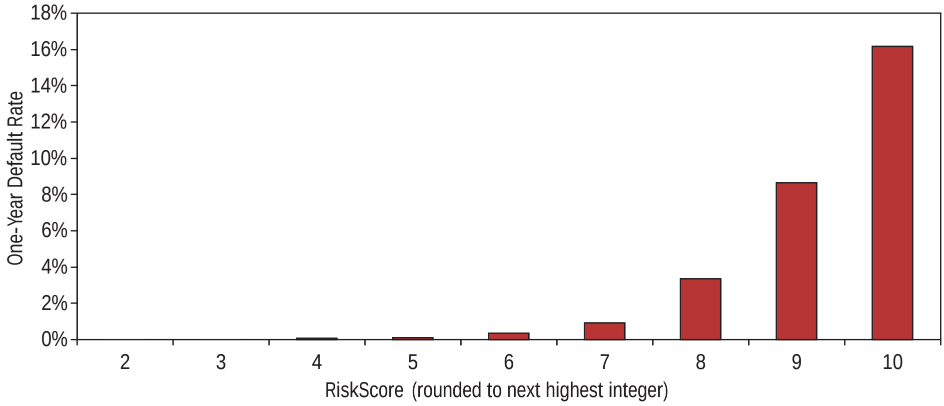
Measuring Private Firm Default Risk

Summary

This report describes and documents the performance of Moody's Risk Management Services' Riskscore™ credit risk assessment model. The model analyzes financial statement data – a principal source of credit risk information for privately held companies – to produce a firm-level credit risk measure. We analyze the model's accuracy and internal consistency with respect to its ability to predict default, as well as changes in credit quality generally. The results we report are based on our unique Credit Risk Database (CRD) of financial statements and defaults for over 21,000 commercial borrowers compiled with the help of several commercial lenders. Our main observations and conclusions are:

- The Moody's Riskscore™ expert system model produces a reliable quantitative measure of credit risk. Riskscores™ are distributed more evenly and are less subject to bunching at the lowest "pass" grades than are banks' internal risk ratings.
- The Riskscore™ model provides useful default estimates for privately held borrowers at time horizons from 1 to 5 years.
- We estimate default rates for commercial borrowers at the one-year time horizon as 0.00%, 0.00%, 0.06%, 0.11%, 0.36%, 0.92%, 3.35%, 8.65% and 16.17% for integer rounded Riskscore™ values of 2-10, respectively. (See chart below).
- Using different performance measures and two different data sets we compared the Riskscore™ model to the Linear Discriminant Analysis (LDA) model, widely used as a benchmark for risk ratings. By all measures, the Riskscore™ model performance was significantly better than LDA.
- An analysis of one-year Riskscore™ transitions reveals dynamics similar to those associated with Moody's corporate bond ratings. Current and future Riskscores™ are shown to be correlated, demonstrating that current values predict deterioration in credit quality other than default.

Estimated Default Rates by RiskScore



continued on page 5

Author

James V. Herrity

Senior Production Associate

Ruth Manso-Rivera

Table of Contents

Introduction

For a variety of reasons, banks have become increasingly interested in quantifying the credit risks associated with lending activities. Risk quantification is a pre-requisite for increasing the efficiency, consistency and accuracy of risk management initiatives. Credit approval, credit management, risk-based pricing, loan securitization, and loan portfolio management all may benefit from the use of quantitative risk assessment models. While other risk management tools and techniques will continue to be developed and loan securitization will continue to expand, the extent to which any lender will benefit from this progress will depend on the lender's ability to apply quantitative risk assessment techniques.

Although banks have internally developed risk rating systems, the associated risk definitions are not consistent from institution to institution, thereby, hindering their use in multi-party securitization activities. Additionally, most internal bank rating systems do not provide quantitative middle market default estimates based on data outside of the particular institution. While financial institutions and regulators alike desire empirically derived middle market default estimates, the essentially private nature of this debt market has restricted broad access to the borrower credit information and the associated default experience necessary to develop such estimates. Recently however, Moody's, in co-operation with several financial institutions, has compiled sufficient middle market credit and default data to support credit research efforts in that market. This report, in part, describes that effort.

Measuring borrower credit risk across multiple institutions' portfolios requires an independent credit opinion that is objectively and consistently applied to each obligor. In Moody's studies of corporate bond defaults, our credit ratings have served the bond market by providing a measure of the probability of default and of the likely severity of loss to the creditor in the event of a default. In this study of commercial loan obligors, we use the Riskscore™ risk assessment model developed by Moody's Risk Management Services, Inc. as the indicator of credit risk. The model was designed to support lending activities. It is an expert system that embodies a best-practices approach to commercial credit evaluation.

This report analyses the performance of the Riskscore™ model as a quantitative indicator of credit quality and default risk, and compares its performance against a widely used risk rating benchmark. Performance is measured by a set of tests which provide unambiguous measures of predictive accuracy allowing for direct model comparison. These tests are described in the report, with detailed explanations contained in the appendix. We also estimate the default risk associated with Riskscore™ values using a data set constructed specifically for that purpose.

Moody's Risk Management Services' Riskscore™ Model

Moody's Risk Management Services' Riskscore™ model is an expert-rule based system designed to assess a company's financial condition on the basis of an analysis of its financial statements. The model produces obligor-level risk ratings between 1 and 10 (by 0.01 increments) that are intended to provide credit professionals with an accurate, consistent, and reliable measure of credit risk. The Riskscore™ is a weighted summation of four key component scores: Profitability Score, Liquidity/Cash Flow Score, Capital Structure Score, and Sales Growth Score which are themselves expert-rule based systems engineered to evaluate the borrowers' performance in each key area. The weights of the component scores have been statistically optimized with respect to default detection.

Many credit risk analysis techniques including those commercially available that have been characterized as purely quantitative are in fact, partly subjective. In practice, the relationships between financial variables included in a model and the relative importance of these variables are often determined by expert judgment — judgment which can be the key to a model's performance. Expert judgment can be assimilated into quantitative models in a variety of ways and determining the optimal way of processing this information is a challenge. To support this aspect of model building, our staff in conjunction with and an advisory group of banking and credit professionals investigated various approaches that have the potential to enhance conventional quantitative risk analysis. They translated the expertise of credit officers and analysts into a set of adaptive expert-knowledge rules. The following detailed statistical analysis of the model's performance demonstrates that the rules we developed contain significant predictive power over the wide range of obligor financial conditions that is present in our samples.

Data Overview

We analyze two datasets in this report. We use our Credit Risk Database (CRD) to assess the Riskscore™ model's ability to differentiate commercial borrower credit quality with an emphasis on default prediction. The second dataset is used to facilitate the estimation of default rate by RiskScore. In order to provide meaningful out-of-sample default rate estimates, the CRD's default rate should be the same as the population's default rate. Since that is not the case[1], we rely on a second dataset of publicly held firms for which Moody's default and bankruptcy database contains a complete reckoning of defaults & bankruptcies. We developed this database with the intention that its firms be similar in size to the privately held borrowers in our CRD database. In this way, we created a dataset that is complete with respect to the overall default rate. This dataset is further described in the Small Capital Market (SCM) Data section.

Although we use the SCM data in some performance tests, the Riskscore™ model was originally developed for commercial banking applications. For that reason, the most relevant analyses are those based on actual financial statement and default information from commercial and industrial loan portfolios. Moody's CRD was compiled expressly to provide the basis for these tests, and is the primary data set used in this study. We will continue to develop and expand the CRD, which will serve as a platform for future research.

Credit Risk Database (CRD)

DEFAULT DEFINITION

In the CRD, defaults were identified by the contributing institutions. The different bank policies classify as defaults instances in which:

1. the credit became 90 days or more past due in either principal or interest payments for reasons other than processing delays caused by the bank;
2. the credit was placed on non-accrual status;
3. the credit was fully or partially charged off; or
4. the borrower filed for bankruptcy.

DATA COMPOSITION

The commercial borrower financial data used in this study to generate Riskscores™ are drawn from a proprietary database compiled from the commercial loan portfolios of 14 lending institutions. Several of these institutions are holding companies that provided data from multiple lending institutions. All financial data contributors were US institutions, save one Canadian institution, and all borrowers are located in either the US or Canada. Our entire database comprises financial statements from over 80,000 commercial borrowers.

Our raw database includes records with different degrees of information concerning both the borrowers' financial conditions and their default status. For this study we used only those data records associated with confirmed borrowers.[2] Additionally, the data were subject to a cleaning process that eliminated, for example, companies with unbalanced financial statements. As a result of these filters, data from only 8 of the 14 contributing institutions were used. The final "clean" data set used in this study consists of 43,735 valid Riskscores™ drawn from over 85,000 financial statements for 21,575 confirmed borrowers. Of these borrowers, 353 (1.6% of the sample) defaulted. These observations span the period from 1988 to 1998, which includes the period of high default rates in the early 1990s and the following lull in default activity. The data are concentrated in the early nineties, with half the observations coming

---

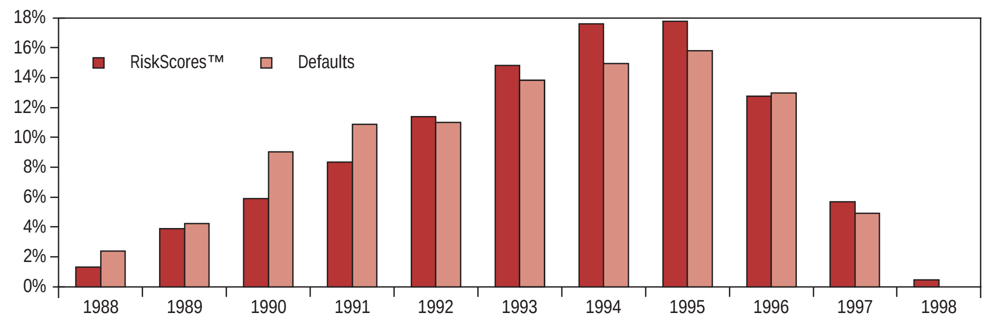
[1] CRD data were provided according to the capabilities of the contributing lenders' internal systems. Institutional factors sometimes create impediments to the construction of a representative and balanced data set. For example, as credits deteriorate and/or default they may be transferred to a special division. This often means that these data are maintained in a different format and in a different physical location than data for performing credits. Additionally, data collection and storage procedures may vary over time and with the geographic location of branch offices. Thus, the CRD is not a representative sample of the population, nor is it the product of a single sampling scheme. While it provides a sound basis for evaluating the model's performance, it does not provide a reliable basis for drawing conclusions about the entire population of middle market borrowers.

[2] Some of the financial statement data represented non-borrowers, for example: prospects, declined requests and withdrawn requests. Not all institutions confirmed actual borrower records.

from the 1993-95 period. Defaults are also similarly distributed, although in relative terms the high default years of 1990 and 1991 produced more defaults per Riskscore™ observation. These distributions are shown in **Exhibit 1** below.

Exhibit 1

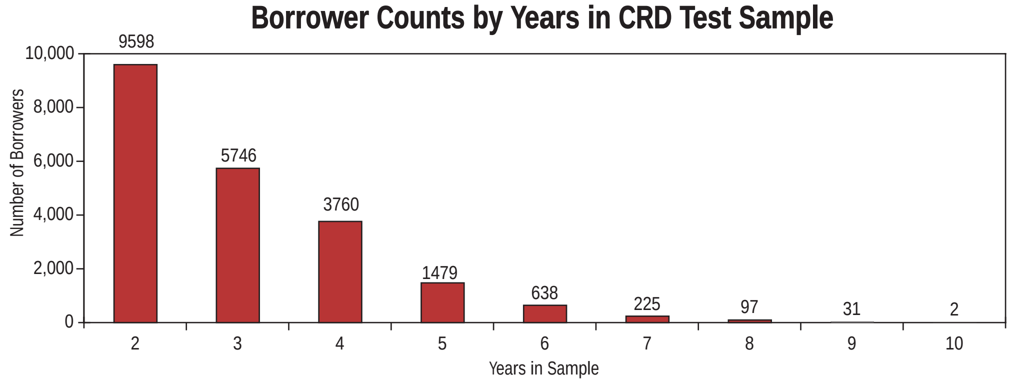
Time Distribution of RiskScore™ Observations and Defaults



For 9,598 borrowers we had only one scoreable observation (comprising at least three consecutive years of financial information), while for 11,977 borrowers we had multiple observations at different points in time. Multiple observations allow us to evaluate the extent to which Riskscores™ tend to change over time for a given borrower. The distribution of the number of observations per borrower is presented in **Exhibit 2** below.

Exhibit 2

Borrower Counts by Years in CRD Test Sample



GEOGRAPHIC DISTRIBUTION OF DATA

Borrowers in this study's CRD sample are located in 47 states, the District of Columbia, and Canada, though for a significant portion, 26%, we could not identify a unique borrower state or province due to missing information. The bulk of the borrowers were domiciled in the mid-Atlantic region of the U.S., with Pennsylvania as the largest single contributor of borrower data. Only about 500 borrowers were domiciled in the western and southwestern regions of the U.S. The distribution of borrowers by location is presented in **Exhibit 3** below.

INDUSTRIAL COMPOSITION OF DATA

For 34,288 of the 43,757 observations in our sample, we could associate each borrower with a four digit SIC code. For analytical simplicity, we rolled SIC codes up into 13 broad industry groups.

This distribution, presented in **Exhibit 4,** looks similar to the industrial composition of Moody's-rated public bond issuers (see Keenan, Carty, and Shtogrin, 1998) in that about 35% of commercial borrowers (manufacturing, mining & contractors) are industrial concerns, and roughly 15% are in retail and consumer related services. The biggest difference is that financial institutions, who comprise roughly 17% of Moody's-rated bond issuers, have been excluded from the CRD. This is offset by a proportionally higher presence in the CRD of wholesale, retail, contracting and service companies.

Exhibit 3

Geographic Distribution of Borrowers

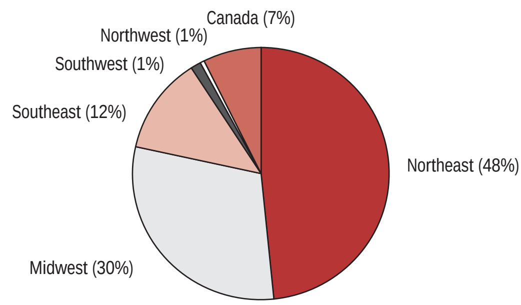
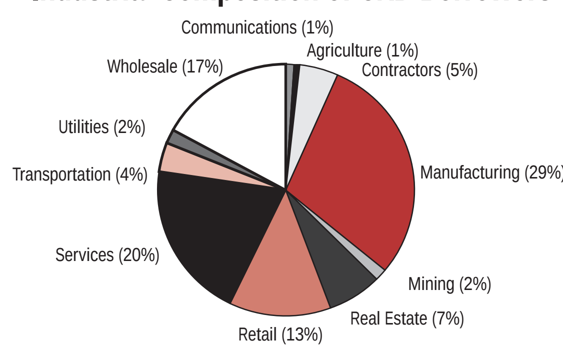


Exhibit 4

Industrial Composition of CRD Borrowers



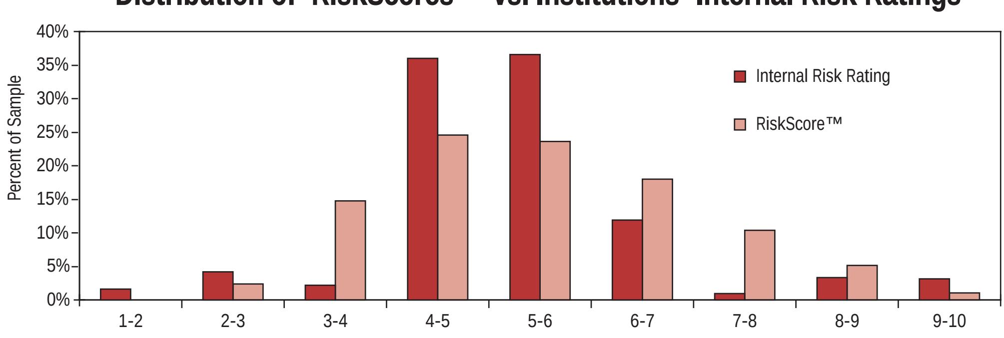
COMPARISON TO FINANCIAL INSTITUTIONS' INTERNAL RISK RATINGS

Exhibit 5 serves to compare the dispersion of Riskscores™ and financial institution risk ratings *relative to their own scales only*.[3] The fact that each is based on a scale of one to ten does not mean that a Riskscore™ of 5 is equivalent in risk terms to a institution's risk rating of 5 or vice versa.

The distribution of Riskscores™ exhibits greater dispersion than that of the institutions' own internal risk ratings[4], which are concentrated around the values of 4 and 5. This corroborates the most common criticism of lending institutions' risk rating processes: that incentive structures lead to a lack of credit quality differentiation with a concentration of risk ratings at the lowest levels acceptable for booking new credits. In fact, risk grades of 4 and 5 are typically the lowest "pass" grades of a lending institution's rating scale. Loans rated below 5 either require additional scrutiny or may be rejected. On the other hand, loans rated above 3 may require pricing or fee concessions. Both forces work to concentrate risk ratings at the low end of the pass categories.[5]

Exhibit 5

Distribution of RiskScores™ vs. Institutions' Internal Risk Ratings



3 This comparison is based on those obligors for which the financial institution provided internal risk ratings. For this comparison, each financial institution's internal ratings were ordinally ranked and then mapped to a common 1 to 10 scale based on each institution's risk rating definitions.

4 The comparison examines the distribution across two entire scales. It is not intended to imply that Riskscore™ "10" is equal to a bank's "10" which is usually reserved for "LOSS". Additionally, note that the granularity of the Riskscore™ model is understated by the comparison as the above distribution uses integer rounded Riskscores™.

5 See Treacy & Carey (1998).

SMALL CAPITAL MARKET (SCM) DATA

In order to estimate the default likelihood associated with Riskscore™ values, and to estimate out-of-sample the default risk for commercial borrowers, we constructed a second data set which provides wider, more balanced and more complete coverage of borrowers and default events. We refer to this as our Small Capital Market data set (SCM). This data set was constructed using publicly available sources of financial statement information and from Moody's proprietary default database. We sought to construct a data set that, while providing balance and completeness, particularly with respect to default, preserved the essential credit characteristics of commercial loan borrowers.

The SCM data set is comprised of the smallest firms represented in COMPUSTAT for which we were able to calculate valid Riskscores™. This subset was created using a two-variable size filter, intended to identify firms with similar characteristics to those in our CRD database.[6] Like our private firm sample, the COMPUSTAT universe consists of US and Canadian borrowers. The SCM data set consists of 53,832 observations on 7,156 separate obligors.

Defaults were identified using Moody's proprietary default database, and default events are defined as in Moody's corporate bond default study.[7] Moody's default database includes both rated and unrated defaulters, insuring a high degree of completeness with respect to default/bankruptcy information.

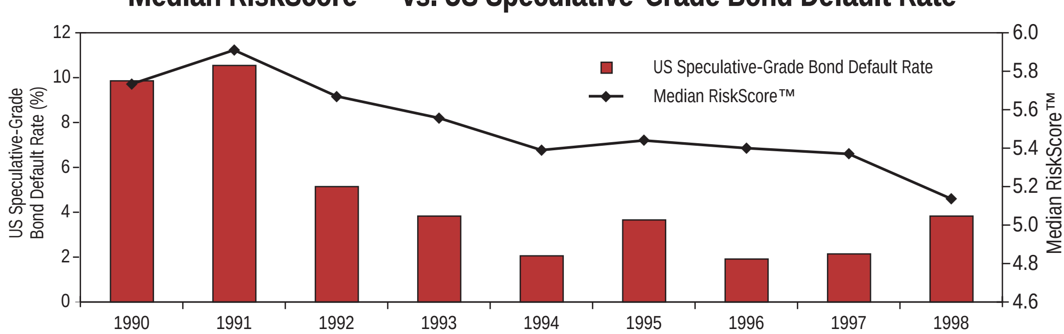
Riskscores™ as Indicators of Credit Quality

CREDIT QUALITY TRENDS

After rising in 1991, the median Riskscore™ in our CRD data sample has shown a gradual downward trend, indicating a trend toward improving credit quality through 1998. This improvement in the median has been a result of improvement at all levels of credit quality, not merely a reduction in the proportion of problem credits. Over the period 1992-1997, the proportion of Riskscores™ above 7 has fallen from 8.12% to 5.54%, while the proportion of Riskscores™ below 3 has risen from 1.63% to 2.96%. Improving average scores are consistent with a more general, aggregate improvement in credit quality, as shown in **Exhibit 6**, which plots the median Riskscore™ along with the U.S. Speculative-grade bond default rate. The only meaningful deviation between the two series is 1998, a year for which our risk score sample size is extremely small.

Exhibit 6

Median RiskScore™ vs. US Speculative-Grade Bond Default Rate



OBSERVED DEFAULT RATES BY RISKSCORE™

The primary test of the model's ability to differentiate the credit quality of borrowers and potential borrowers is whether it accurately predicts default events over the near term. We evaluated this aspect of the model by grouping together Riskscores™ by rounding up to the nearest integer and calculating one-year default rates for each bucket. These default rates are presented in **Exhibit 7** below.

---

6 Specifically, we rejected firms for which .0707*log(tangible assets) + .0562*log(net sales)>1.23.

7 See Keenan, Shtogrin, & Sobehart, (1999), pp. 9-10.

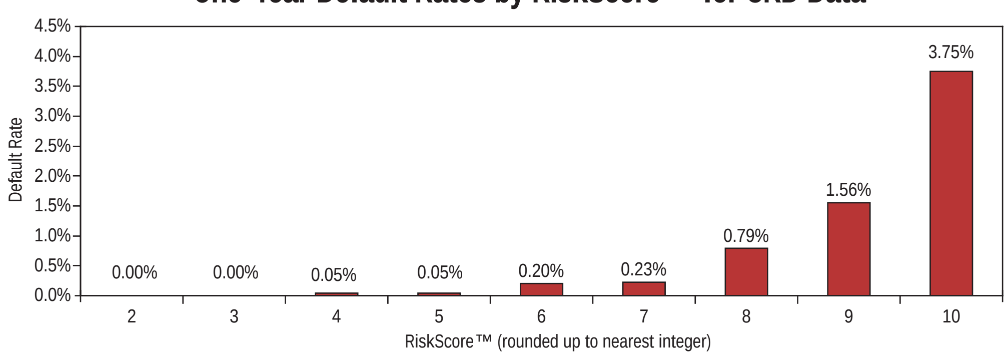
The smooth increase in default rate observed as the Riskscore™ increases indicates that the model is performing quite well with respect to a one-year default criterion. We obtain an approximate doubling of the default rate as we move across successively riskier buckets. This approximate doubling of default rates for higher risk buckets is a characteristic of Moody's bond ratings as well.

Importantly, these one-year default rates, which reach a maximum of only 3.75% for scores in the 9-10 range, cannot be interpreted as "default risk" measures for the score ranges indicated. They are not default risk measures because they are only measuring the observed default rate within our sample. The sample has an aggregate default rate of just 1.6% over the entire 11 year study period — a rate significantly lower than that for speculative-grade corporate bonds and clearly too low for private companies as a whole.

Exhibit 7 demonstrates the model's ability to differentiate the *relative* credit risk associated with different borrowers. The slope of the default rate reveals the model's ability to differentiate borrowers' credit risks. We estimate the level of out-of-sample default risk associated with each Riskscore™ value using the SCM data set in a later section of this report.

Exhibit 7

One-Year Default Rates by RiskScore™ for CRD Data



MULTI-YEAR DEFAULT RATES

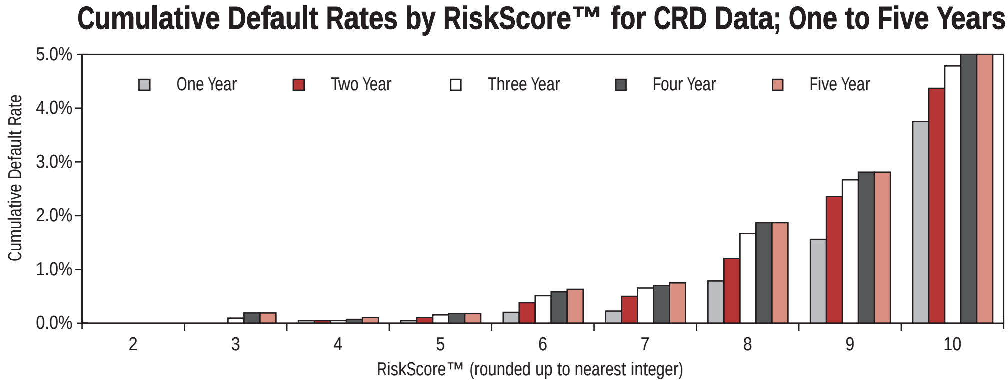
From a risk management perspective, the ability of the model to carry consistent predictive information about the likelihood of default over successively longer time periods is crucial. Near term default risk may be the primary concern for lenders, as it is for many types of bond investors; however, in many respects lenders have greater flexibility than bond investors. Such flexibility can be used to manage credit risk. Specifically, a bank's ability to structure individual credits, including setting the term of the facility and fee and spread levels, can allow the bank to obtain more effective compensation for the credit risk being borne. A key component in such a process is the ability to gauge not only the current credit risk for a given borrower, but the risk with respect to successively longer time horizons.

We evaluate the predictive power of the model over longer time horizons by calculating multi-year, or cumulative default rates for risk buckets formed by rounding Riskscore™ values up to the nearest integer. Internal model consistency requires that default rates, for a given risk bucket, at each time horizon, are no higher than those for the next riskiest bucket. In other words, cumulative default rate curves for a given credit quality should lie everywhere above curves for higher quality borrowers. Moreover, for each risk bucket, the curve should be smoothly increasing as the time horizon is increased.

Exhibit 8 plots cumulative default rates at horizons of one through five years by rounded RiskScore. **Exhibit 8** shows that the curves do lie above each other for higher scores. This demonstrates that the Riskscore™ values are, on average, correctly differentiating the forward default risk of borrowers over multi-year horizons.

In the middle of the range — in buckets 6 and 7, representing Riskscore™ values between 5 and 7 — the differentiation is less dramatic than for higher risk buckets. However in this range, the default rate is below 1% even at the five-year horizon, so there are very few observations on which to measure the discriminating power of the model over extended periods.

As expected, the default rate time profiles are smoothly increasing over longer horizons. The exceptions are for the riskiest borrowers, buckets 8, 9, and 10, as the time horizon goes beyond four years. The flattening out of these default rates reflects the exhaustion of the data as the time horizon is increased — as shown in **Exhibit 2**, only about 3% of borrowers can be observed for more than four consecutive years.



ESTIMATING DEFAULT PROBABILITY BY RISKSCORE™ VALUE

Within our CRD data set, the smooth upward-sloping relationship between Riskscores™ and default rates indicates the model's ability to differentiate borrower credit quality. However, because the CRD's aggregate default rate is low, it is difficult to gauge the relationship between Riskscores™ and out-of-sample default likelihood using the default frequencies presented above. Recall that just 1.6% of our borrowers were known to have defaulted any time during the 11 year study period. This represents a lower overall default rate than that of Ba3-rated bond issuers, and is likely to significantly understate the true default rate for privately held firms as a whole.

In order to estimate the default rate associated with Riskscore™ values, we calculated default rates using our SCM data set. Because of the comprehensiveness of Moody's corporate default & bankruptcy database, we can be confident that the default rates calculated for the SCM data set generally reflect the default risk associated with Riskscore™ values for borrowers fitting this profile. **Exhibit 9** shows one-year default rates calculated over the SCM data set (as does the chart on the cover of this report), along with the CRD default rates presented in **Exhibit 7**. While the model's ability to differentiate credit risk across both samples is comparable (i.e., the slopes of the two default rate curves are comparable), the default frequencies observed for the broader market data are an order of magnitude higher.

Exhibit 9

Default Rate by RiskScore™

(CRD Privately Held vs. Small Publicly Held Borrowers)

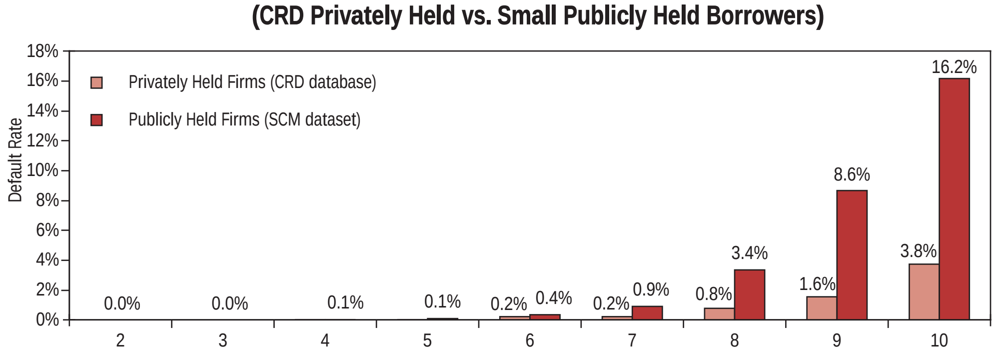
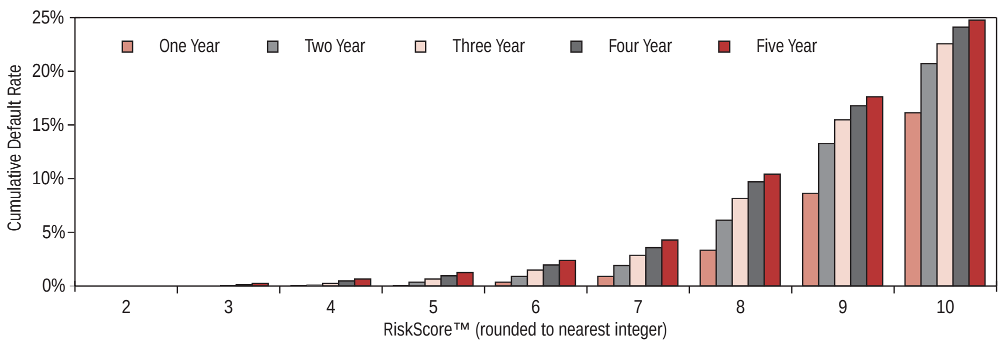


Exhibit 10

Cumulative Default Rates for Small Capital Market Borrowers, One to Five Years



Because the SCM data set comprises a broader sample of borrowers, and because we are confident that it contains a balanced, representative proportion of defaulters vs. non-defaulters, we believe the SCM default rates represent reasonable estimates of the default risk associated with different Riskscore™ values for privately held borrowers. As Moody's expands its **Exhibit 11** market loan and default database, we will be able to determine more precisely the default risk implied by a given Riskscore™ for a privately held borrower.

To produce estimates of middle market default risk over longer time horizons, we calculated multi-year default rates from one to five years for the SCM data set. These default rates are presented in **Exhibit 10** above, and in tabular form in **Exhibit 18** in the appendix. They are well behaved, rising smoothly and proportionally higher for higher Riskscores™ and as the time horizon is increased. For the riskiest bucket, the default rate rises from 16.2% at the one-year horizon to 24.8% at the five-year time horizon.

Evaluating the Performance of the Riskscore™ Model

This section presents additional statistical measures designed to gauge the relative predictive power of models when applied to the same data. Our purpose here is not to present an exhaustive comparison of the Riskscore™ model with other statistical models, but to provide objective measures of performance relative to a benchmark model which may be helpful when performance comparisons are required.

This analysis is essentially a comparison of model errors. Errors need to be defined with respect to a specific criterion and in this section we focus on the ability to predict a default event within a one-year time horizon. Default prediction models can err in one of two ways:

a) The model can indicate low risk when, in fact, the risk is high (Type I error).

- For Moody's ratings, this case corresponds to highly rated defaults such as Johns Manville's 1982 default while holding the A3 rating.
- Default rate studies are relatively good at identifying Type I errors.

b) The model can indicate high risk when, in fact, the risk is low (Type II error).

- Corresponds to a firm with a low rating that should, in fact, be rated higher. E.g.: were Moody's to rate the US government Caa1, we would be committing a Type II error.
- Default rate studies are less powerful at identifying Type II errors.

It is possible for some models to commit less of one type of error than another. For example, Moody's could rate every company Caa3 and would thereby minimize the potential for Type I errors. However, we would be committing egregious Type II errors. On the other hand, Moody's could rate every company Aaa, minimizing Type II errors but producing egregious Type I errors. As this example hints, success at minimizing one type of error necessarily comes at the expense of increasing the other type of error. Good models address both types of error. The next section provides a measure of accuracy which measures both Type I and Type II classification errors simultaneously.

For purposes of comparison, we contrast the performance of the Riskscore™ model with a Linear Discriminant Analysis (LDA) model. The LDA model has, in one form or another, been used for default prediction purposes for several decades, and as a benchmark for model comparisons. The most widely used publicly available model is Altman's Z-Score model (Altman, 1968, 1977, 1995).

The Riskscore™ model's private firm performance is compared to Altman's 4 variable Z-Score model[8] using the CRD data set. While our primary goal in constructing the SCM data set was to have a more reliable estimate of aggregate commercial borrower default risk, it also provides us with an additional data set to which we can apply the Riskscore™ model to obtain another set of performance test results. In this case, comparison is facilitated by fact that COMPUSTAT contains values for Altman's Z-Score model for all firms in the sample.[9] We provide performance test results for both data sets in the subsequent analyses.

CUMULATIVE ACCURACY MEASURES

Our primary performance analysis focuses on cumulative predictive accuracy – the extent to which the model's risk ratings identify future defaulters in a given sample through the order of risk ratings assigned. The two key related measures are the graphical cumulative accuracy profile (CAP), and the accuracy ratio (AR) which reduces the information in the CAP to a single number between zero and one. An AR of one indicates perfect accuracy and zero indicates that the risk ratings are no better than random numbers (These measures are described in greater detail in the appendix). CAP and AR measures are convenient because they evaluate both Type I and Type II errors in a straightforward way.

Exhibit 11 shows one-year CAP plots for both Riskscores™ and the LDA model scores, based on the CRD data. The Riskscore™ model is clearly outperforming the LDA model, lying everywhere above the LDA CAP. The uneven progression of both curves reflects the relatively low incidence of default in the sample, which is partly a result of the middle market data gathering process. Nevertheless, the fast rising Riskscore™ CAP indicates that with the first 20% of the data, roughly 70% of defaulters are identified, whereas only about 57% are identified by the Z-Score model.

Exhibit 11

One-Year Cumulative Accuracy Profiles (CRD Data)

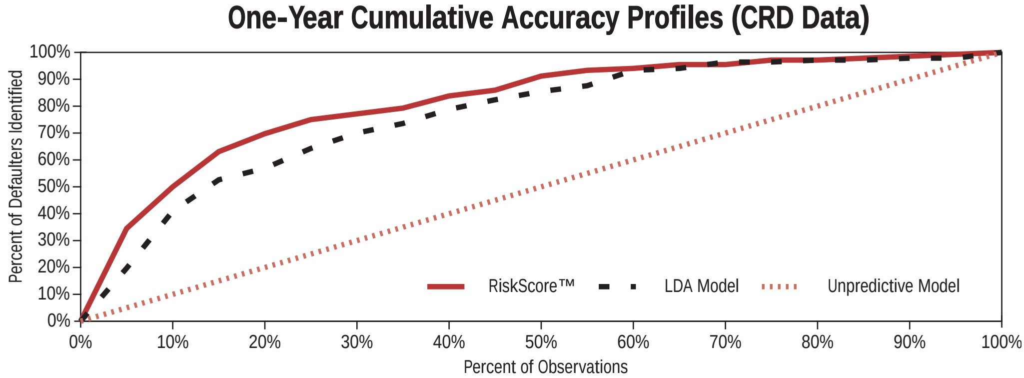


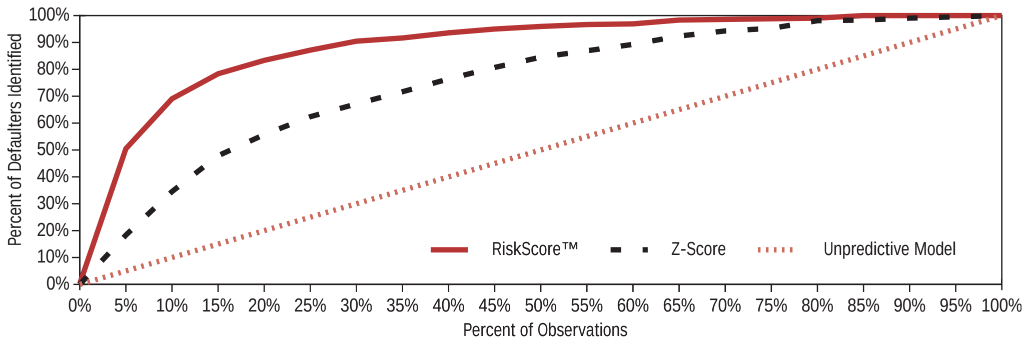
Exhibit 12 shows the one-year CAP plots for both the Riskscore™ and the Z-Score (LDA) models for the SCM data set. In this case, the LDA model is the five-variable Z-Score model. For these data, the Riskscore™ model is outperforming the LDA model by a wide margin. As can be seen from the CAP plot, over 83% of defaulters are identified with the first 20% of the data, whereas only about 56% are identified by the Z-Score model. This is particularly interesting because the Z-Score incorporates not only financial statement-derived information, but also information from equity markets through its reliance on the ratio of the market value of equity to the book value of debt.

8 These values are based on Altman's 1995, four-variable model. As explained in Altman 1995, this differs from the original model in that the book value of equity is used instead of the market value of equity, and that the capital-turnover ratio variable is excluded. See Appendix for details.

9 These values are based on Altman's 1968 five-variable model. See Appendix for detailed explanation.

Exhibit 12

One-Year Cumulative Accuracy Profiles (SCM Data)



The AR measures presented in **Exhibit 13** confirm that the Riskscore™ model outperforms the LDA model, and provides confidence measures for each measure of predictive power in the form of standard deviations for the AR statistics.

Exhibit 13

Inter Model Comparisons: One-Year Accuracy Ratios

Model	Data set	One-Year AR (Std. Deviation)
RiskScore™	CRD	0.64 (0.02)
LDA (4 Variable Z-Score)	CRD	0.53 (0.01)
RiskScore™	SCM	0.79 (0.01)
LDA (5 Variable Z-Score)	SCM	0.54 (0.01)

INFORMATION ENTROPY RATIO (IER)

We supplement the CAPs and ARs with a separate performance measure, the Information Entropy Ratio (IER). The association of the word information with the concept of entropy should be taken in a loose sense as it may carry a different connotation to the casual reader than the one intended. Intuitively, information entropy measures the overall "amount of uncertainty" represented by a probability distribution (see the appendix for a more detailed explanation). We can use it to assess how well a model predicts defaults by comparing the amount of "uncertainty" there is about default in the case where we have no model (a state of total ignorance) to the amount of "uncertainty" left over after we have introduced a model (presumably, a state of less ignorance). This is done via the IER. Because a risk assessment model has the singular goal of reducing uncertainty, the IER gives us an unambiguous measure of how well the model is performing.

To calculate the IER, we first calculate the information entropy of the sample. That is, without attempting to control for any knowledge that we might have about credit quality, we measure the "uncertainty" associated with the event of default. We then recalculate the information entropy after having taken into account the predictive power of the model. The IER is the ratio of the information entropy with the model to the information entropy without the model. Thus, the *lower* the IER, the *more uncertainty that is being eliminated* by the model and the more predictive is the model.

Exhibit 14

Inter Model Comparisons Using the Information Entropy Ratio

Model	Dataset	IER (Standard Error)
RiskScore	CRD	0.88 (0.02)
LDA (4 Variable Z-Score)	CRD	0.94 (0.02)
RiskScore	SCM	0.77 (0.02)
LDA (5 Variable Z-Score)	SCM	0.90 (0.02)

Exhibit 14 shows the information entropy ratio achieved by each tested model for both data sets, along with standard errors for each IER statistic. The table shows that the Riskscore™ model is outperforming the LDA model by a wide margin for both data sets. It also indicates that, for both models, the reduction in uncertainty is higher for the SCM data set than it is for the middle market data. At least part of the reason for this pattern is the higher quality and greater completeness with respect to defaults of the SCM dataset.

In all cases, however, the IER is fairly high (a perfectly predictive model would have an IER of zero). This is because there is relatively little uncertainty in either data set to begin with. For the CRD, a default occurred within one year for only 0.31% of the Riskscore™ observations, whereas for the SCM data set, only 0.85% of the Riskscore™ observations preceded defaults by one year or less. Thus, even if we know very little about who defaulted, we can say that each borrower has more than a 99% chance of being a non-defaulter over a one-year horizon.

Nevertheless, the IER is a clear measure of relative performance for a given data set, and the Riskscore™ model outperforms the LDA model for both sets.

Riskscore™ Volatility and Transition

Because Riskscore™ values are nearly continuous numerical measures, we can measure the volatility of the model output straightforwardly by calculating the score value variances. We first calculated variances over periods from one to five years, counting the first observation for each borrower as year one. By including every observation available for each transition, we obtained a sample which declined in size from 22,026 borrowers to 1,479 over the five-year period (see **Exhibit 2**). For 1,479 borrowers, we had observations over five consecutive years, and for this five-year cohort we calculated both the variance over periods from one to five years, and inter-year variances across successive years. Finally, counting as a one-year transition every Riskscore™ observation preceded by another for the same obligor, we calculated an overall one-year variance. These statistics are presented in **Exhibit 15**.

Exhibit 15

Riskscore™ Variances by Period and Sampling Technique

All Observations

Years	Variance	Observations
1-2	1.48	22,026
2-3	1.94	10,123
3-4	2.04	3,927
4-5	2.07	1,479

Five-Year Cohort

Years	Variance	Observations
1-2	1.51	1,479
1-3	1.88	1,479
1-4	1.99	1,479
1-5	2.07	1,479
1-2	1.51	1,479
2-3	1.47	1,479
3-4	1.44	1,479
4-5	1.36	1,479

Overall One Year

1-2	1.47	37,555

The striking feature of **Exhibit 15** is that the one-year variances are so close for the different sampling techniques, and, given the number of observations included, 22,026 for all first-year transitions and 37,555 for all one-year transitions, these statistics are robust. Thus, Riskscore™ volatility is quite steady at about 1.47 over a wide variety of borrowers over the period of time covered by our sample. The inter-year variances for the five-year cohort, decline steadily from 1.51 to 1.36 as the cohort ages so, while the model is fairly sensitive to changes in the financial information, a current Riskscore™ value contains significant information about the borrower's credit quality five years into the future.

To gauge the degree to which Riskscore™ values are predictive of changes in credit quality generally, not just default, we measure the extent to which current Riskscore™ values are correlated with future Riskscore™ values. This correlation can be identified and quantified by rounding both the initial score and score one year later into integer buckets and calculating transition frequencies. These frequencies form a transition matrix, similar to the rating transition matrices Moody's publishes to characterize the time-correlation of rating drift for corporate bond issuers.[10]

The one-year transition matrix for rounded score values presented in **Exhibit 16** shows the degree to which rating volatility is a function of current credit quality, and the likelihood of large changes in Riskscore™ over a one year time horizon. As with Moody's corporate bond ratings, there is a general tendency for more risky credits to drift, in either direction, more than less risky credits. Because the model was designed to signal deterioration of credit quality over a relatively short time horizon, Riskscores™ show a greater overall tendency to drift than do Moody's ratings on corporate bond issuers. The idiosyncratic 1-2 and 2-3 categories in **Exhibit 16** reflect the low number of observations in these categories.

The extent to which Riskscore™ values predict overall future credit quality can be seen by comparing one-year default rates by rounded score value with the one-year frequencies for being rated higher than 9.

[10] See Carty (1997).

These latter values are simply the last column of **Exhibit 16**, and the comparison is presented graphically in **Exhibit 17** below.

Exhibit 16

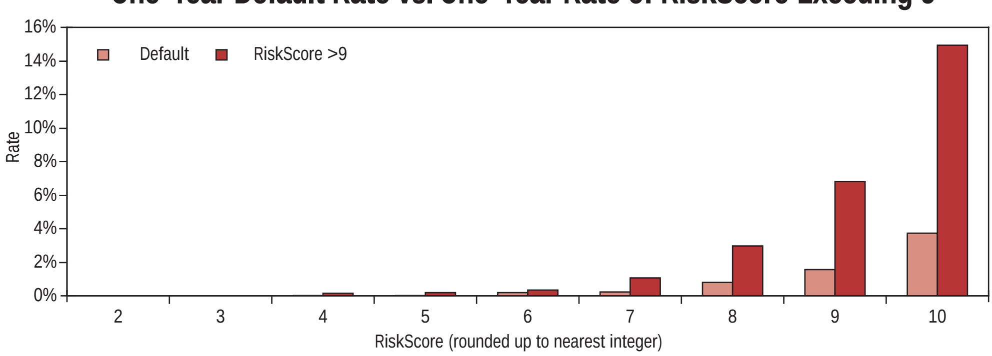
One-Year Transition Matrix for RiskScores™*

Initial Score	Score One Year Later: 2	3	4	5	6	7	8	9	10
2	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
3	0.81%	21.41%	46.06%	21.41%	5.25%	3.03%	1.21%	0.40%	0.00%
4	0.03%	8.64%	42.84%	32.14%	9.93%	4.30%	1.49%	0.39%	0.15%
5	0.02%	2.02%	21.02%	40.49%	23.11%	9.39%	2.98%	0.73%	0.18%
6	0.02%	0.44%	7.03%	25.99%	36.20%	19.88%	7.52%	2.48%	0.34%
7	0.00%	0.21%	2.95%	14.85%	27.77%	31.52%	15.28%	6.22%	1.05%
8	0.00%	0.13%	0.99%	7.23%	19.76%	27.80%	27.39%	13.43%	2.96%
9	0.00%	0.10%	0.39%	1.85%	11.49%	26.39%	26.48%	25.71%	6.82%
10	0.00%	0.45%	0.45%	0.00%	5.43%	15.38%	28.96%	30.32%	14.93%

**RiskScores™ rounded up to nearest integer.*

Exhibit 17

One-Year Default Rate vs. One-Year Rate of RiskScore Exeeding 9



Conclusion

Moody's Risk Management Services' Riskscore™ model is a rule-based expert system designed to assess a borrower's creditworthiness using financial statement information. The model was developed to embody the best-practices of commercial credit analysis. Its rule base has been enhanced using advanced statistical techniques to maximize the predictive power over the range of borrower types and financial conditions represented in our databases.

Using the CRD's private firm financial statement data, as well as small capital market data constructed to have more balanced and comprehensive market coverage, the Riskscore™ model has been demonstrated to consistently differentiate credit quality and predict default.

In this report, we have examined a variety of measures that indicate that the Riskscore™ model outperformed a widely used benchmark for default prediction – the LDA model. We estimated the default risk associated with Riskscore™ values using a small capital market data set constructed using COMPUSTAT and Moody's proprietary default database. This data set was designed to preserve middle market borrower characteristics while providing a sample more reflective of the aggregate commercial loan market default rate.

We have also demonstrated that middle market Riskscore™ values exhibit one-year transition dynamics with characteristics similar to those for Moody's corporate bond ratings. We have shown that Riskscore™ values are correlated with future Riskscore™ values, so that current values can predict deteriorations in credit quality that do not result in default.

Bibliography & Suggested Readings

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

Altman, Edward I., Haldemam, R., and Narayanan. P., (1977), "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations", *Journal of Banking and Finance*, June 1977.

Altman, Edward I., (1988), "Measuring Corporate Bond Mortality and Performance," *Journal of Finance*, February.

Altman, Edward I., Hartzell, J., and Peck, M., (1995), "Emerging Markets Corporate Bonds: A Scoring System", Salomon Brothers, Inc., New York, NY.

Asarnow, Elliot and Edwards, David, (1994), "Measuring Loss on Defaulted Bank Loans", Citibank, Portfolio Strategies Group.

Carty, Lea V., (1997), "Moody's Rating Migration and Credit Quality Correlation, 1920-1996", Moody's Special Comment, July.

Carty, Lea V., Hamilton, David T., Keenan, Sean C., Moss, Adam, Mulvaney, Michael, Marshella, Tom, and Subhas, M.G., (1998), "Bankrupt Bank Loan Recoveries", Moody's Special Comment, June.

English, William B. and Nelson, William R., (1998), "Bank Risk Rating of Business Loans", Federal Reserve Board Working Paper.

Federal Reserve System Task Force on Internal Credit Risk Models, (1998), "Credit Risk Models at Major U.S. Banking Institutions: Current State of the Art and Implications for Assessments of Capital Adequacy, Federal Reserve Board of Governors, May.

Hendricks, Darryll and Hirtle, Beverly, (1997), "Bank Capital Requirements for Market Risk: The Internal Models Approach", Federal Reserve Bank of New York, Economic Policy Review December.

Hurt, Lew, and Felsovalyi, (1998), "Measuring Loss on Latin American Defaulted Bank Loans: a 27-Year Study of 27 Countries", The Journal of Lending & Credit Risk Management, October 1998, pp. 41-46.

Kallberg, J.G, Edwards, David M., and Banasiak, Michael J., (1997), "Firm Failure: Exploring the Role of Non-Accounting Data", Dun & Bradstreet Business Credit Services Group.

Keenan, Sean C., Shtogrin, Igor, and Sobehart, Jorge R., (1999), "Historical Default Rates of Corporate Bond Issuers, 1920-1998", Moody's Special Comment, January.

Keenan, Sean C., Carty, Lea V., and Shtogrin, Igor, (1998), "Historical Default Rates of Corporate Bond Issuers, 1920-1997", Moody's Special Comment, February.

Lennox, Clive, (1998), "Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches", University of Bristol, Department of Finance Working Paper.

Roche, James, Brennan, William, McGirt, Derek, and Verde, Mariarosa, (1998), "Bank Loan Ratings", in Fabozzi, Frank Ed. Bank Loans: Secondary Market and Portfolio Management, Chapter 4, pp. 57-70.

Treacy, William F., and Carey, Mark S., (1998), "Credit Risk Rating at Large U.S. Banks," Board of Governors of the Federal Reserve System. Federal Reserve Bulletin. November, p. 897-921.

Exhibit 18

Cumulative Default Rates, One to Five Years (SCM dataset)

Riskscore™ Bucket	One Year	Two Year	Three Year	Four Year	Five Year
2	0.00%	0.00%	0.00%	0.00%	0.00%
3	0.00%	0.00%	0.07%	0.15%	0.26%
4	0.06%	0.15%	0.27%	0.47%	0.67%
5	0.11%	0.36%	0.66%	0.95%	1.26%
6	0.36%	0.93%	1.53%	2.00%	2.42%
7	0.92%	1.91%	2.85%	3.60%	4.29%
8	3.35%	6.13%	8.20%	9.72%	10.42%
9	8.65%	13.30%	15.52%	16.80%	17.63%
10	16.17%	20.74%	22.60%	24.15%	24.77%

CUMULATIVE ACCURACY PROFILES

Cumulative Accuracy Profiles (CAPs) are line plots based on each risk measure's ability to predict defaulters and non-defaulters (e.g., Type I and Type II errors). In this study, we focus on the one-year prediction horizon.

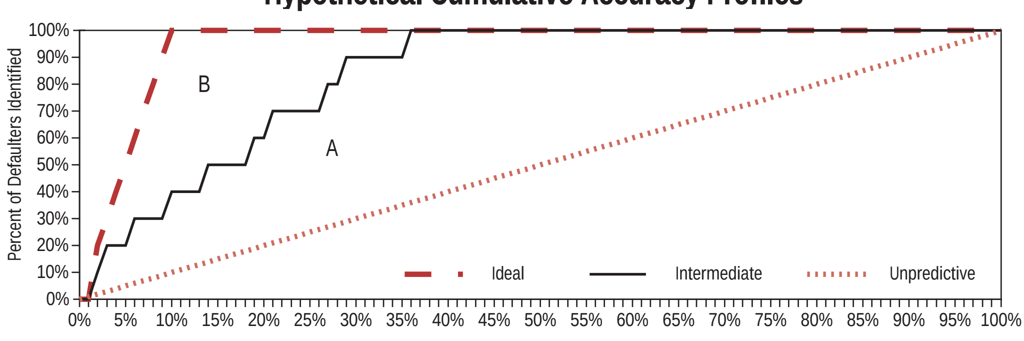
To plot Cumulative Accuracy Profiles, companies are ordered by risk rating from riskiest to safest. As one moves up the sorted companies from the riskiest to a company at or below which lies *x* percent of the total number of company risk scores, one calculates the percentage, *y*, of the total number of defaulters in the sample that have an equal or lower risk score. The CAP is the plot of the *y's* versus the *x's*.

A good model concentrates defaulters at lower risk scores and non-defaulters at higher risk ratings. In such a model the percentage of all defaulters identified increases quickly as one moves up the sorted sample. If the model were poor (i.e., assigned risk scores randomly) we would expect to capture a fraction of the defaulters proportional to the number of observations, generating a straight line. A useful property of CAPs is that they reveal information about the predictive accuracy of a risk rating over its entire range of values for a particular time horizon.

Hypothetical CAPs for ideal, intermediate and unpredictive (random) risk models are presented in **Exhibit 19**.

Exhibit 19

Hypothetical Cumulative Accuracy Profiles[11]



[11] While observed aggregate defaults rates are closer to 1%, for illustrative purposes only, we've assumed the aggregate default rate = 10%.

ACCURACY RATIOS (ARs)

It is convenient to have a summary measure that ranks the predictive accuracy of each risk measure. We obtain such a measure by comparing the CAP of any risk measure with both the ideal and random CAPs. The closer the CAP is to its ideal, the more area there is between it and the random CAP. The largest amount of area possible to enclose is identified by the ideal CAP.

The ratio of the area between a risk measure's CAP and the random CAP to the area between the ideal CAP and the random CAP is the Accuracy Ratio (AR). The AR is a global ratio which measures the proportion of defaulters/non-defaulters in a sample that can be identified per increment of the risk score that is being evaluated. The AR is a fraction between 0 and 1. Risk measures with ARs close to 0 display little advantage over a random assignment of risk scores while those with ARs near 1 display almost perfect foresight. For the CAPs shown in **Exhibit 19**, the AR is A/(A+B). Most of the models we tested have ARs in the range from 50% to 80%.

INFORMATION ENTROPY

We begin this introduction to information entropy with a mathematical definition of information. There are many possible such definitions. However, the one that follows has useful properties in analysis and a sensible interpretation in that it does not lead to contradictions. The association of the word information with the concept of entropy should be taken in a loose sense because it often carries a different connotation for the casual reader. Intuitively and in the sense in which we define it below, the entropy measures the overall "amount of uncertainty" represented by a probability distribution.

Let us assume the occurrence of an event with only two possible outcomes: (A) issuer defaults with probability p, and (B) issuer does not default with probability 1-p. The probabilities of the outcomes provide partial information of the event. The amount of additional information required to completely determine whether or not the first outcome occurred is defined as

$$\text{Information} = -\log_2(p)$$

where $\log_2(p)$ is the logarithm[12] of p in base 2.

If only the first outcome were possible, then p would equal 1 and the information required is quantified as $-\log_2(p) = 0$. In this case, there is no uncertainty about the outcome and, therefore, there is no relevant information that was not previously known. If the two events are equally likely (uninformative case), then $p = ½$ and the amount of information required reaches a maximum value of $-\log_2(p) = 1$ (binary unit or "bit"). This is a state of total ignorance which requires the maximum amount of information to resolve. Exactly 1 bit of information (the equivalent to a yes-no type of answer) tells us which of the two equally likely possibilities have occurred.

Having defined information in this precise way, we can precisely define information entropy. Consider two mutually exclusive outcomes of an event: (A) issuer defaults, and (B) issuer does not default; one of which must be true. Given a set of risk scores $S = \{R_1, .., R_n\}$ produced by a model, the entropy with respect to propositions A and B for a specific risk score R is

$$h(R) = -(P(A/R)\log_2 P(A/R) + P(B/R)\log_2 P(B/R))$$

where P(A/R) is the probability that the issuer defaults given that the risk score is R. The average over all possible risk scores is the information entropy

$$H = h(R_1)P(R_1) + h(R_2)P(R_2) + \ldots. + h(R_n)P(R_n).$$

The IER is simply the ratio of the entropy based on the model scores to the entropy given only the data itself. If the model were perfectly predictive, the IER would be 0 (no uncertainty). If the model held no predictive power, the IER would be 1.

[12] The use of 2 as the logarithmic base has certain advantages for this example but any base can be used. Usually, the natural logarithms are used for convenience. Note, however, that the amount of information depends upon what logarithmic base is used which determines the unit of measure of information.

Thus, the IER is an absolute measure of the reduction in uncertainty associated with a particular model *for a given data set*. As such, it provides an unambiguous measure of model performance as long as all the models considered are applied to the same data set. Technically, the properties that make this true are:

a) if the risk score set S contains more information about A and B than another set S', then H(S) < H(S').

b) acquisition of new information can never increase the value of H.

THE LDA MODEL

The model consists of a linear function of financial ratios that produced the best classification of firms into either distressed or non-distressed categories, based on a representative learning set of data. For each variable, a separation line is found which maximizes the separation between distressed and non-distressed firms. The line is found simultaneously producing a multidimensional linear function. This function is then used to classify out-of-sample companies as belonging to either the distressed or non-distressed group. Importantly, then, the magnitude of the LDA score can be interpreted in a probability sense only indirectly, as an indicator of the probability of belonging to the distressed group, not as the probability of default itself.

Altman's Z-Score model is a particular implementation of the LDA model and is widely used as a benchmark. The estimated linear discriminant function embedded in the original 1968 Z-Score model is:

> Z = 1.2*(working capital/total assets) + 1.4*(retained earnings/total assets) + 3.3*(earnings before interest & taxes/total assets) + 0.6*(market value of equity / book value of total debt) + 0.99*(sales/total assets).

Our comparison for the small capital market firms (SCM) Z-Score uses values published in COMPUSTAT, which are based on this formulation. The LDA model scores reported for the CRD data were calculated using the 1995 version of the model. As explained in Altman, 1995, this model, Z''-Score, differs from the original in two ways. First, the book value of equity is used in place of market value of equity. This allows the model to be applied to non-public firms. Secondly, the capital-turnover variable (sales/total assets), has been dropped from the equation. This is done to minimize the potential industry effect which is more likely to take place when such an industry sensitive variable is used. We tested the Z'' model against the original Z-Score model for the SCM data set and found that the Z'' did carry more predictive power than the original Z-Score model. The revised model is:

> Z'' = 6.56*(working capital/total assets) + 3.26*(retained earnings/total assets) + 6.72*(earnings before interest & taxes/total assets) + 1.05*(book value of equity / book value of total debt).

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

Report Number: 45768