Special Comment

April 2003

Contact	Phone
New York	
Richard Cantor	1.212.553.1653
Christopher Mann	

Measuring The Performance Of Corporate Bond Ratings

Summary

- A rating system's performance should be measured by its success or failure in meeting its objectives. This *Special Comment* discusses various metrics which may be used to judge rating systems and the application of these metrics to Moody's historical performance.

- Moody's corporate bond ratings are intended to be "accurate" and "stable" measures of *relative* credit risk, as determined by each issuer's relative fundamental creditworthiness and without reference to explicit time horizons. Moody's performance should therefore be measured by both rating accuracy (the correlation between ratings and defaults) and rating stability (the frequency and magnitude of ratings changes).

- To measure the accuracy of its corporate bond ratings:

  - Moody's tracks the cumulative accuracy profiles and accuracy ratios of its ratings, since relative rating accuracy is a primary objective of its rating system.
  - Moody's also tracks investment-grade default rates and the average rating of defaulting issuers prior to their defaults. These metrics measure Moody's success at meeting a secondary *cardinal* or *absolute* rating system objective, namely that ratings be useful to investors who employ simple rating "cutoffs" in their investment eligibility guidelines.
  - Moody's does not target specific default rates for the individual rating categories. For rating systems that do have this type of cardinal objective, performance can be measured by the *stability* of realized default rates by rating category around target default rates.

- To measure ratings stability Moody's tracks:

  - the frequency of rating changes,
  - the frequency of large rating changes,
  - and the frequency of rating reversals.

- Moody's ratings should change only when *relative* fundamental creditworthiness changes. Since relative fundamental credit risk generally changes quite slowly, ratings should be stable, especially when compared with other market based risk measures. Infrequent reversals, and stability more generally, facilitate the use of ratings in connection with investment eligibility guidelines and performance benchmarks.

- The accuracy and stability of Moody's ratings has fluctuated over time. Moody's rating accuracy ratios are currently within historical norms, which range from the lows of the early 1980's to a high in the mid-1990's. Current rating stability is low by historical norms.

- Changes in ratings performance may reflect either changes in the quality of the ratings process or changes in the environment that make defaults more or less difficult to predict, or make fundamental creditworthiness more volatile, thus making rating stability harder to maintain. To control for changes in the economic environment, we also compare the accuracy and stability of Moody's ratings to the accuracy and stability of other credit risk measures, such as bond market-implied ratings inferred from observed credit spreads.

  - Over the past four years, compared to Moody's ratings, bond market-implied ratings have been more correlated with defaults at one-year horizons, but their accuracy ratios are comparable as the horizon lengthens to three years.

  - Moody's ratings have been much more stable than market-based credit opinions. In any given twelve-month period, there is about one rating action for every four issuers and only one out of every 100 issuers experiences a rating reversal (a rating change followed by a rating change in the opposite direction). In contrast, based on market-implied ratings (inferred from monthly bond spreads), nine out of ten issuers experience rating changes each year and seven out of ten experience rating reversals.

- This study is limited to the assessment of the accuracy and stability of Moody's ratings; however, a complete analysis of Moody's rating system also requires an analysis of its outlooks and Watchlist designations, which will be a subject of Moody's future research.

Table of Contents

Page

Introduction

The meaning and quality of Moody's corporate bond ratings have recently come under scrutiny by market participants and regulators. Market participants variously state that Moody's ratings react too slowly to market information, react too quickly to market information, are issuer driven and therefore too high, or are overly conservative and therefore too low. These contradictory views reflect the tension amongst the users of ratings about the objectives of our rating system, and ultimately the appropriate metrics that should be employed to measure the success or the failure of meeting these objectives.[1]

A fuller understanding of how we and others measure quality may be useful for market participants and regulators who need to make judgments about the usefulness of different credit risk measurement systems. Important systems include but are not limited to credit ratings from rating agencies and credit scores based on market prices such as the Merton variable and the Moody's KMV's EDF™. Other systems may be developed to respond to changes in the market. Each of these systems, old and new, will be more or less useful for different purposes and will have different intended uses. It is important that they have transparent objectives and that their performance be measured against their stated objectives so that market participants can make informed decisions about which to use.

This *Special Comment* discusses the objectives of Moody's rating system and the metrics we use to measure its performance over time. Moody's intends its ratings to be "accurate" and relatively "stable" measures of credit risk, correlated with subsequent default and loss experience. The accuracy and stability of a rating system can be measured in a number of ways, but the appropriateness of each metric depends on the objectives of the rating system.

Moody's primary objective is for its ratings to provide an accurate *relative* (i.e., ordinal) ranking of credit risk at each point in time, without reference to an explicit time horizon. As such, the most appropriate measure of Moody's accuracy is the "power" of its ratings, the information content of their rank orderings of companies *at specific points in time* (i.e., specific rating cohorts) with respect to expected credit losses as realized over a long horizon. Credits that have low ratings today should on average prove to be more risky than credits that have high ratings today.

In addition to a relative ranking of risk *at a point in time*, some investors desire a consistent relative ranking of credits *across time*, so that the riskiness of a credit today can be compared to similarly rated debt instruments in the past. To measure the accuracy of Moody's ratings across time, the most appropriate metric is the "power" of a pool of ratings assigned to multiple credits, and possibly even the same credits, observed at different points over time.

Moody's believes that as a consequence of its *relative* rating approach, the meaning of its ratings should be highly consistent over time. Since the relative creditworthiness of bond issuers does not, on average, change rapidly, there should not generally be any need to change average rating levels sharply over time. As a practical matter, therefore, Moody's does not manage its ratings to achieve *cardinal* accuracy or to maintain constant default rates by rating category. Doing so would require Moody's to change its ratings en masse in response to changes in cyclical conditions. Rather, ratings are changed "one-at-a-time" as needed in order to improve the current rank ordering of credit risk.

1. *The specific objectives of the rating system and the methods used to achieve those objectives are laid out in the May 2002 Special Comment entitled, "Understanding Moody's Corporate Bond Ratings and Rating Process."*

Many investors are also concerned about *absolute* or *cardinal* attributes of the rating system; i.e., the specific expected default or loss rates of a rating system over different investment horizons. An appropriate metric for a cardinal rating system is the variability of realized loss rates around idealized or historical average default rates by rating category, as measured by mean absolute error or root mean squared error statisitics. Moody's does not, however, target specific expected loss rates for its individual rating categories in the corporate sector.[2] Nevertheless, for the same reason expressed above, the combination of Moody's through-the-cycle rating approach and the long-term stability of the economy implies that future *long-horizon* expected loss rates are likely to be similar to historical loss rates for specific rating categories.[3]

Although Moody's does not target specific loss rates by rating category, Moody's does maintain certain cardinal risk objectives for its corporate bond rating system. In particular, Moody's aspires to assign low (at least speculative-grade) ratings to defaulting credits well in advance of their default. Hence, to measure its absolute ratings accuracy (the cardinal information content of its ratings), Moody's emphasizes two simple metrics — investment-grade default rates over multiple horizons and the average rating of defaulting issuers up to three years prior to their defaults. These metrics are designed to assess the usefulness of Moody's ratings to investors who employ simple rating "cutoffs" in the design of their investment eligibility guidelines.

Moody's also intends its ratings to be stable measures of relative credit risk. To a large extent, Moody's desire for stable ratings reflects the view that more stable ratings are "better" ratings. Moody's ratings should change when relative credit risk — as measured by fundamental credit analysis — changes. Fundamental credit analysis incorporates an evaluation of franchise value, financial statement analysis, management quality and plausible scenario analysis. Since relative fundamental credit risk generally changes quite slowly, if ratings are highly volatile, they are unlikely to closely track relative fundamental creditworthiness.

The stability of Moody's ratings also facilitates their use in the market for various applications. Financial market participants regularly use ratings in security selection and portfolio composition and have embedded ratings throughout the economy's financial architecture — in portfolio governance, in performance attribution, in the regulation of financial markets and institutions, and in financial contracts and covenants. Rating changes can have substantial economic consequences for a wide variety of debt issuers and investors. As a result, Moody's changes ratings only when an issuer's relative fundamental creditworthiness has changed and the change is unlikely to be reversed in a short period of time.

To measure rating stability, Moody's tracks the following volatility metrics, each measured over twelve-month intervals:

- the frequency of rating changes,
- the frequency of large (i.e., 3 or more rating notch) rating changes, and
- the frequency of rating reversals (rating actions in the opposite direction of a previous rating action),

2. *In contrast, Moody's employs a cardinal system in rating structured finance securities, where each rating category is associated with a schedule of expected loss rates that are derived from historical information and vary with the length of the investment horizon.*

3. *Some investors, however, are very concerned with the expected default rates associated with corporate ratings over short horizons, particularly in the high-yield sector. To meet these investors' needs, Moody's provides a model-based, monthly forecast of the one-year-ahead speculative-grade default rate. This model, and models like it, can be used by investors to translate Moody's relative rating system into a cardinal rating system.*

These metrics for accuracy and stability can answer the question, "how accurate and how stable are ratings today, compared to previous years?" However, these measures do not control for changes in the economic environment. They indicate whether or not performance has changed, not why it has changed. Changes in performance may be due to changes in the quality of the ratings process, or may simply reflect a change in the environment that makes defaults more or less difficult to predict or makes rating stability easier or harder to maintain. To control for changes in the economic environment, we also compare the accuracy and stability of Moody's ratings to the accuracy and stability of other credit risk measures, such as bond market-implied ratings inferred from observed credit spreads.

There may, of course, be a tradeoff between these two aspects of rating quality, accuracy and stability. It might be possible to increase the short-term correlation between ratings and defaults by making ratings more volatile and more responsive to new information, without regard for its potentially transitory nature. As part of the implementation of its rating system, Moody's limits the frequency of rating changes to perceived changes in fundamental characteristics, dampens their severity, avoids rating change reversals, dialogues with issuers, and ultimately makes decisions by committee. These actions are consistent with the desire to produce stable, long-term ratings with a minimum of false default predictions. They also presumably, though not necessarily, reduce the short-term prediction content of ratings. In recognition that the current rating may not always immediately reflect potential changes in an issuer's credit situation, Moody's provides the market additional signals in the form of changes in rating outlooks and Watchlist designations that, together with the rating more fully describe each issuer's likelihood of credit losses.[4]

The organization of the rest of this *Special Comment* is as follows. The first section outlines the theoretical aspects of measuring performance. The strengths and weakness of various metrics are discussed, first those for accuracy and then those for stability. The second section of the paper applies these metrics to Moody's performance between 1983 and 2002 and compares Moody's performance to the performance of hypothetical ratings inferred from bond market credit spreads observed between 1999 and 2002. The third section discusses the implications and our plans for further research.

Measuring Rating Accuracy And Stability

Ratings should be accurate and stable indicators of relative fundamental credit risk and thus should be highly predictive of the relative incidence within the rated universe of payment defaults, bankruptcies, and distressed exchanges.[5] Moreover, since relative fundamental credit risk generally changes only gradually, rating changes should also be generally infrequent, small in magnitude, and unlikely to be reversed within short periods of time. In the following sections, we explain the issuer-based[6] metrics available to measure relative accuracy at a) a point in time and b) across time. We then discuss metrics that can be used to measure the absolute/cardinal accuracy of a rating system. Lastly, we discuss some simple measures of rating stability.

4. *To fully measure the information content and timeliness of Moody's ratings, one would also need to take these additional signals into account. However, such an analysis is beyond the scope of this Special Comment.*

5. *Rating performance should be judged by both default and loss severity given default. However, our accuracy metrics focus only on default experience because data on loss severity are limited and the incorporation of severity into performance measures is analytically complex. See the Moody's Special Comment "LossCalc™: Moody's Model for Predicting Loss Given Default" by Greg Gupton and Roger Stein (2002).*

6. *All the metrics discussed in this report are "issuer-based" measures, meaning that each issuer's rating history and default experience is given equal weight in the analysis. Alternatively, performance metrics can be weighted by the number or dollar volume of each issuer's outstanding obligations. Such weighting schemes would be appropriate if the objective is to evaluate a rating system by its performance against the largest and most numerous rated debt obligations in the market. However, such schemes place extraordinary weight on the performance of the ratings of a few very large issuers, and their statistical power tends to be quite limited. In its February 2003 default study, Moody's highlights areas of similarity and differences between its issuer-weighted and volume-weighted default statistics.*

Relative Rating Accuracy

The key metric used by Moody's to measure the relative accuracy of its ratings is the "cumulative accuracy profile" (or power curve) and the accuracy ratio, which is one way to compress the information in the cumulative accuracy profile into a single number. These metrics can be used to measure (1) the performance of a single rating cohort at a point in time, (2) the average performance of a sequence of rating cohorts over time, or (3) the performance of a pool of ratings drawn from different cohorts over time.

Defining Cumulative Accuracy Profiles

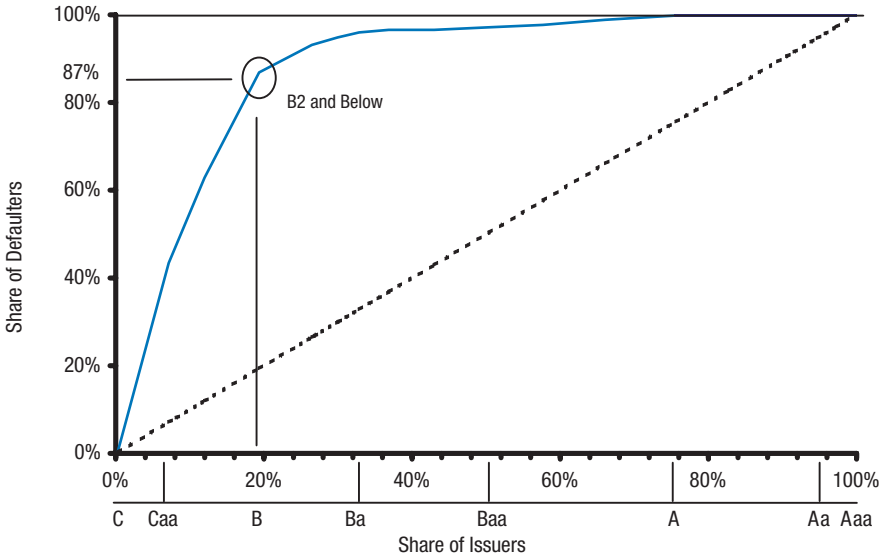
CAP curves are useful for making visual assessments of the information content embedded in the relative ranking of credit risk provided by a set of ratings. Each point along a CAP curve indicates the power of a specific rating category as a tool to discriminate between defaulters and non-defaulters over a specified investment horizon, such as one year, five years, or twenty years. A *single-cohort* CAP curve evaluates the performance of a group of ratings that were outstanding at the same specific point in time, whereas a *multiple cohort* CAP curve evaluates a pool of ratings that were outstanding at different points in time.

> *The cumulative accuracy profile is constructed by plotting, for each rating category, the proportion of defaults accounted for by firms with the same or a lower rating against the proportion of all firms with the same or a lower rating.*

The CAP curve is also known as a "power curve," because it shows how effective a rating system is at detecting defaults from the population. Exhibit 1 presents the one-year-ahead horizon CAP curve for the entire pool of Moody's unsecured corporate-bond issuer-ratings that were outstanding as of January 1, 2001. The percentage of defaulters during 2001 captured by each rating category and below is plotted against the vertical axis. The share of issuers with ratings at or below each rating category is plotted against the horizontal axis. For example, issuers rated B2 or below represent 19% of the overall population and 87% of the defaulting population.

The further the curve bows toward the northwest corner, the greater the fraction of all defaults that can be accounted for by the lowest rating categories. The closer the curve is to the 45% line, which is the power curve associated with randomly assigned ratings, the weaker is the information content of the rating system.

Exhibit 1: Moody's 2001 One-Year Cumulative Accuracy Profile



** The letter ratings on the horizontal axis are for illustrative purposes only and do not affect the CAP.*

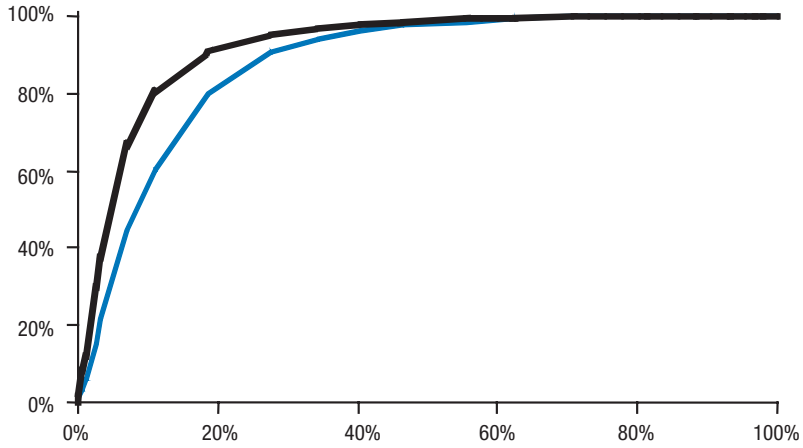
Comparing Cumulative Accuracy Profiles

CAP plots can be used to compare the accuracy of different rating systems or to compare the accuracy of a single rating system measured at different points in time for different rating cohorts. These comparisons sometimes produce unambiguous rankings, but often, the rankings are ambiguous because the different rating systems may have relatively more power in different parts of the rating scale.

Exhibit 2 presents the case in which one set of ratings is a uniformly more powerful predictor of default than another set of ratings, represented by the black and blue lines, respectively. In this comparison, regardless of the share of the ratings sample examined, the black CAP curve indicates a more powerful rating system, capturing a higher percentage of the subsequent defaulters with its lower ratings, and by implication, a higher percentage of the non-defaulters with its higher ratings. If the two CAP curves represent two different systems being applied to the same set of obligors over the same investment horizon, then the rating system described by lightly shaded CAP curve is clearly more accurate.

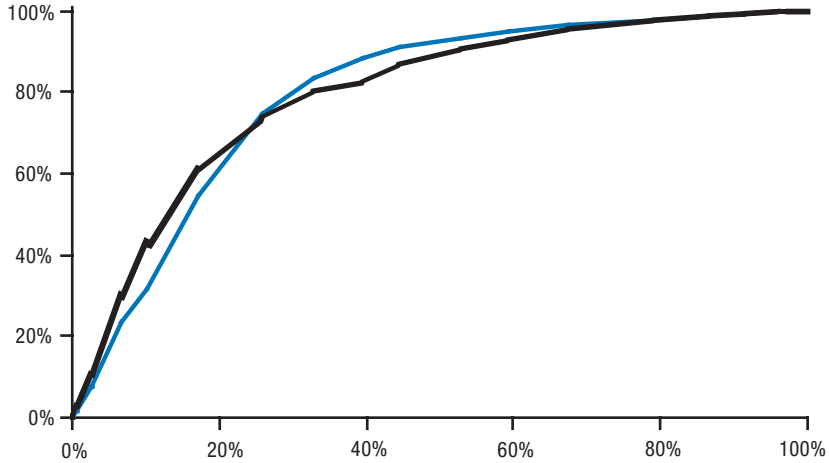
If these two CAP curves represent the performance of two different rating cohorts evaluated by a single rating system, it is also true that ratings associated with the black CAP curve proved to be more accurate. However, since the two curves are based on different samples, it is impossible to know whether the superior performance was the result of superior analysis or differences in the credit risk environment, which made it fundamentally easier to distinguish defaulters from non-defaulters.

Exhibit 2: Unambiguous CAP Curve Performance Comparison



Often, CAP curves cannot be compared unambiguously. Exhibit 3 presents two CAP curves that cross at roughly 25% of the overall population, which corresponds to roughly 75% of the defaulting population. The ratings characterized by the black CAP curve provide a better rank ordering of credit risk among the higher risk portion of the rating scale; whereas the blue CAP curve is more accurate when ranking credit within the lower risk portion of the rating scale. In this situation, the choice of which rating system to use would depend on which issuers were considered the most important.

Exhibit 3: Ambiguous CAP Curve Performance Comparison



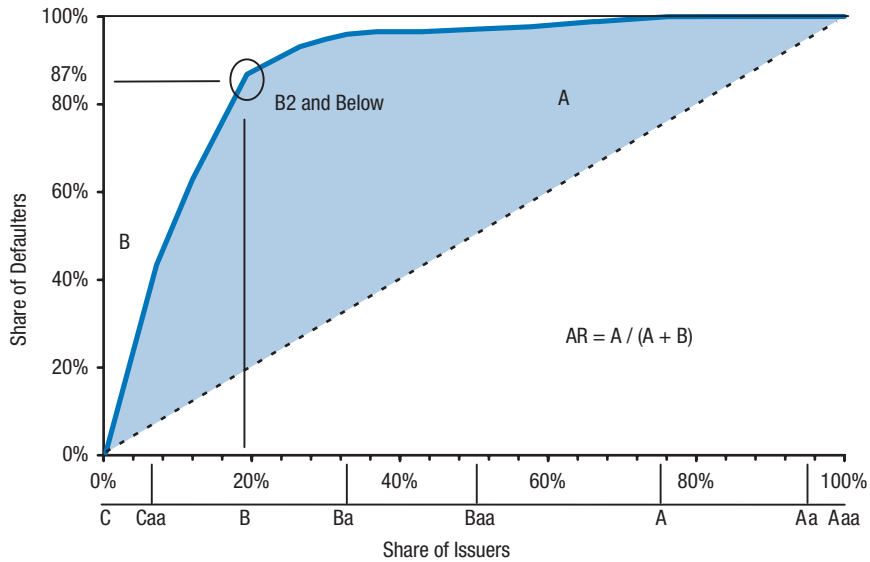
Accuracy Ratios

The accuracy ratio (AR) compresses the information depicted in the CAP curve into a single summary statistic.

> *The accuracy ratio is the ratio of the area between the CAP curve and the 45-degree line (area A in Exhibit 4 below) to the maximum possible area above the 45-degree line, which is ½ (areas A+B in the Exhibit 4 below)* [7]

7. This definition is very closely related to the accuracy ratio concept first proposed in the Moody's Special Comment "Benchmarking Quantitative Default Risk Models: A Validation Methodology" by Jorge R. Sobehart, Sean C. Keenan, and Roger M. Stein (2000). Those authors define the AR as the ratio of the area under the CAP curve relative to the area under the "perfect foresight" curve, rather than relative to the entire triangle above the 45-degree line. A feature of Sobehart, et. al.'s definition is that whenever the aggregate default rate changes, the AR changes, regardless of whether or not accuracy (as measured by the CAP curve) has changed. The benefit of not including the perfect foresight comparison is that the accuracy ratio will be invariant to changes in the aggregate default rate and will only change due to changes in the rating distribution and the distribution of default rates. In practical terms, though, there is little empirical difference between the accuracy ratio of Sobehart et al and that used in this Special Comment.

Exhibit 4: Deriving The Accuracy Ratio From The CAP Plot



** The letter ratings on the horizontal axis are for illustrative purposes only and do not affect the CAP.*

The AR can also be derived algebraically, without reference to the CAP plot.[8] Like CAP plots, ARs can be calculated for both individual and pooled cohorts. Some important implications of accuracy ratios are as follows:

1. *Although the accuracy ratio is a good summary measure, not every increase in the accuracy ratio implies an unambiguous improvement in accuracy.* When comparing two CAP plots that do not intersect, the AR that summarizes the CAP curve further to the northwest quadrant will indeed always be higher than the AR for the other CAP curve. If, however, two CAP curves intersect, the difference in their ARs measures their relative power under implicit assumptions about the relative importance of ratings in different parts of the rating scale.

2. *The accuracy ratio lies between minus one and positive one, similar to a correlation statistic.* If all defaulters were initially assigned the lowest rating category, the accuracy ratio would approach one. If all defaulters were distributed randomly throughout the population without regard to ratings, the accuracy ratio would be zero. And, if all defaulters were initially assigned the highest rating category, the accuracy ratio would approach minus one.[9] The accuracy ratio is therefore similar to a correlation statistic, which also ranges between one and minus one.

3. *The accuracy ratio measures only <u>relative</u> accuracy, not <u>absolute</u> accuracy, and is invariant to proportional changes in marginal default rates.* The marginal default rate is the percent of issuers in any given rating category that subsequently default. If the marginal default rates for all rating categories change proportionally, neither the CAP plot nor the accuracy ratio changes at all.

8. *Define D_{Ri} as the sum of all defaulters with initial ratings Ri or lower, where R_i = C, Ca, ..., Aaa and N_{Ri} is the number of issuers rated R^i. The area of under the CAP is given by $\sum_{Ri = C, Ca, ..., Aaa} (D_R + D_{Ri-1})(N_{Ri} - N_{Ri-1})/(2DN)$, where D and N are the total numbers of defaults and issuers, respectively. D_o and N_o are equal to zero.*

9. *Under the definition of the AR proposed by Sobehart, <u>et</u>. <u>al</u>. (see previous footnote), perfectly accurate or perfectly inaccurate ratings would lead to AR's precisely equal to one and negative one, respectively, rather than simply "approaching" those values as under the current definition.*

4. *The accuracy ratio weighs equally Type I and Type II error.* Type I error is defined as rating a defaulter too high, and Type II error is rating a non-defaulter too low. The accuracy ratio does not differentially "weigh" Type I or Type II errors: the AR derived from a CAP plot describing the share of defaulters predicted at every rating level is equivalent to the AR of a CAP plot describing the share of non-defaulters predicted at every rating level.

5. *The accuracy ratio requires a large number of in-sample defaults to be meaningful.* The accuracy ratio's power is low when there are few defaults in the sample. As a result, at least over short horizons, the metric most reflects the information content of speculative-grade ratings: the degree to which defaulters are highly concentrated in the Caa rating category and less so in the Ba rating category strongly influences the one-year-ahead accuracy ratio. The quality of investment-grade ratings is measured better by longer horizon accuracy ratios, because more investment-grade issuers default over longer horizons.

6. *The accuracy ratio is only one of several methods for reducing the CAP plot to a representative number.* The accuracy ratio, being a linear function of the area under the CAP plot, tacitly places implicit relative weights, and therefore relative importance, on defaults and non-defaults at different points in the rating spectrum. These weights may be different than those desired by a user, who, for example, might only be interested in performance in distinct portions of the rating spectrum.

Accuracy At A Point In Time Versus Across Time

Relative accuracy can, therefore, be measured using the cumulative accuracy profile and its associated accuracy ratio. However, CAP plots and ARs can be constructed using different aggregation methods and the interpretation of the results depends significantly on the method used. If the objective of the rating system is to create an ordinal ranking of credits at a point in time without constraining the default rates of the rating categories, an appropriate measure would be to average the AR's of cohorts from each point in time. If, however, the objective of the rating system is to constrain each rating category to have a specific default rate then it may be reasonable to pool the ratings from each cohort and to calculate the AR of the pooled sample. In the second case, there is also value to examining the average of the average of the AR's for each cohort because absolute rating systems can also be used as relative or ordinal systems. The converse is not true; examining a pooled cohort AR may not be meaningful for an ordinal rating system. In practice, though, the relative meaning of Moody's ratings is consistent enough that there is little difference between the two aggregation methods. More volatile rating systems, such as ratings derived from market-based information, can have substantial differences.

When calculating the average of AR's over the different cohorts, one must also choose whether to use an equally-weighted average, an issuer-weighted average, or a default-weighted average. The first corresponds most closely to what is presented in a graph of AR's from different time periods (such as presented in Figure 6.) The last two correspond more closely to the most powerful econometric weighting; cohorts with more issuers and more defaulters have smaller standard errors around the estimations. Further, the issuer- and default-weighted statistics are more comparable to the pooled cohort result because the pooled cohort automatically puts higher weight on larger cohorts.

Absolute/Cardinal Rating Accuracy

In addition to the extent that a rating system has power to discriminate between defaulters and non-defaulters on a relative basis, some investors are concerned about *absolute* or *cardinal* attributes, the specific expected default or loss rates, of a rating system over different investment horizons. If the objective of a rating system is to target idealized or historical average default rates by rating category, then the cardinal accuracy of a rating system can be measured by the variability of realized loss rates around those targets. Simpler metrics, however, can be used if the cardinal objective of rating system is more modest: that low ratings (i.e., speculative-grade ratings) be assigned to defaulting credits well in advance of their defaults. Success against this objective can be measured by the investment-grade default rate (which should be low) and the average rating of defaulting credits (which should also be low) prior to their defaults. These metrics assess the usefulness of ratings to investors that employ simple rating cutoffs in the design of their investment eligibility guidelines.[10] This section presents several of the metrics that Moody's has traditionally measured and reported.

Cumulative Default Rates By Rating Category

A simple intuitive test of the utility of a rating system is to measure whether or not realized cumulative default rates (CDRs) are higher for lower rating categories than they are for higher rating categories. However, there is little basis for selecting ideal target CDRs by rating categories. Moreover, CDRs themselves do not provide any information about the *power* of a rating system, because they do not incorporate any information concerning the distribution of ratings. For example, supposed two rating systems might have the same CDRs for their "high," "medium," and "average" rating levels. One agency may have a much more powerful rating system because its ratings are spread equally across the rating scale, whereas the other agency may have a less powerful rating system because its ratings are much more concentrated in the middle (less informative) portion of the rating category.

Variability Of Cumulative Default Rates Around Idealized Rates

As discussed above, *average* default rates themselves, while very meaningful measures of ratings performance, have some limitations. However, the *variability* of realized default rates around historical average or idealized default rates can be used to measure the utility of ratings for investors that use ratings to predict their actual loss experience on a portfolio basis. For example, international bank regulators have recently proposed that rating agency performance be judged based on the (upward) deviations of realized three-year cumulative default rates from certain idealized benchmark rates.[11] The major problem with this approach is that the permissible "deviations" must by necessity be very large because changes in default risk are highly correlated across obligors and hence default rates by rating category are highly variable over time.[12]

10. Moody's Special Comment "Default & Recovery Rates of Corporate Bond Issuers: A Statistical Review of Moody's Rating Performance, 1920 – 2002" by David Hamilton and Praveen Varma presents the realization of Moody's on several of the cardinal performance measures presented herein.

11. See Basel Committee on Banking Supervision's, Quantitative Impact Study 3 — Technical Guidance, October 2002.

12. These issues are analyzed extensively in Moody's Special Comment, "Testing for Rating Consistency in Annual Default Rates," February 2001, by Richard Cantor and Eric Falkenstein.

Investment-Grade Default Rates

Although investors may be disappointed whenever realized default rates exceed the historical average default rate for a given rating category, not all default rate deviations are equal. In particular, many investors are much more concerned about the average default rate and the variability of default rates for investment-grade than for speculative-grade securities. (The investment-grade default rate is the share of all companies previously rated investment-grade that subsequently default over some specified investment horizon, such as one year or five years.) Performance metrics based on the level and variability of investment-grade default rates are particularly important for investors that employ simple investment-grade-rating "cutoffs" in the design of their investment eligibility guidelines.

Average Rating Prior To Default

The investment-grade default rate measures a rating agency's success at avoiding gross Type 1 errors that would occur if they assigned very high ratings to companies that subsequently default. However, another tool is needed to measure a rating agency's success at assigning its lowest ratings to companies that subsequently default. The average rating prior to default is tool that can be used for this purpose. Other things being equal (including of course the agency's overall rating distribution), rating performance is improved whenever the average rating prior to default (looking back over any horizon) is lowered. In practice, Moody's regularly examines the average rating of defaulters at the time of default and over the 36 months prior to default. When these 36 points are averaged, one obtains a single statistic, which loosely corresponds to the average rating one-and-a-half years prior to default.

Rating Stability Measures

Many financial market participants — investors, regulators and issuers — desire stable ratings. However, while reflecting an aversion to volatility per se, their desire for rating stability also reflects the view that *more stable* ratings are *more accurate* ratings with respect to the relative fundamental credit risk of a borrower. Moody's ratings are derived through fundamental credit analysis, which incorporates an evaluation of franchise value, financial statement analysis, management quality, and scenario analysis. Since relative fundamental credit risk normally changes quite slowly, highly volatile ratings are therefore unlikely to be tracking relative fundamental creditworthiness closely.

An emphasis on relative fundamental creditworthiness is also consistent with a through-the-cycle rating approach. Though transitory shocks might temporarily increase the relative credit risk of a firm, these shocks are often short-lived and, once passed, have little affect on issuers as going concerns. A through-the-cycle approach places low weight on short-term credit shocks and thereby reduces rating volatility in general and the likelihood of rating reversals (upgrades followed by downgrades or vice versa) in particular.

Stability has also facilitated the use of ratings in the market for a variety of applications. Financial market participants regularly use ratings in security selection and portfolio composition and have imbedded ratings throughout the economy's financial architecture — in portfolio governance, in performance attribution, in the regulation of financial markets and institutions, and in financial contracts and covenants. As a result, rating changes can have substantial economic consequences for a wide variety of debt issuers and investors. Changes in ratings should therefore be made only when an issuer's relative fundamental creditworthiness has changed and the change is unlikely to be reversed within a short period of time.

By introducing a second objective, rating stability, into rating system management, some accuracy with respect to short term default prediction may be sacrificed.[13] However, it is also possible that an effort to maintain stable ratings may increase rating accuracy, particularly over longer horizons. Whether or not there is a tension between accuracy and stability is ultimately an empirical question.

While there are many conceivable ways to measure rating stability, Moody's tracks the following volatility metrics:

- the frequency of rating changes
- the frequency of rating changes of 3 or more rating notches, and
- the frequency of rating reversals (rating actions in the opposite direction of a previous rating action);

each measured over twelve-month intervals.[14]

Rating Actions

The frequency of rating actions is defined as the share of issuers that experience rating changes during the past twelve months. This metric tracks how often Moody's ratings are changed. Moody's ratings are continually monitored, but are only changed when a rating committee believes such a change is warranted. The frequency of rating actions will fluctuate from time to time, in response to changes in economic volatility.

Large Rating Changes

Like relative fundamental creditworthiness, ratings are expected to change slowly and gradually over time. While certain unexpected events may require multi-notch-rating adjustments, changes in credit quality will typically be reflected in a series of single-notch rating changes spaced out over extended periods of time. Accurate and stable ratings should anticipate changes in credit quality and adapt to new information in a controlled and judicious manner. A rise in the frequency of large rating changes (as measured by rating changes of three or more notches) may suggest that ratings have been too slow to incorporate changes in credit risk.

Rating Reversals

Rating changes can have dramatic consequences for both investors and debt issuers. As a result, Moody's does not change ratings with the arrival of each new piece of information, but instead looks for signs that an issuer's long-term creditworthiness has changed. Moreover, Moody's needs to be reasonably confident that there will be no need to reverse a rating change in the near future. To monitor rating reversals, we track the share of issuers with upgrades following downgrades, or vice versa, over a 12-month period. A rise in the rating reversal rate may indicate that ratings are becoming overly responsive to temporary shocks.

---

13. Two recent academic papers by Günther Löffler have shown that the application of a through-the-cycle approach and efforts to avoid rating reversals lead to ratings which are stable but serially correlated in rating changes. See "Avoiding the Rating Bounce: Why Rating Agencies Are Slow to React to New Information," Goethe-Universtät Frankfurt, Working Paper, July 2002, and "An Anatomy of Rating Through the Cycle," Journal of Banking and Finance, forthcoming.

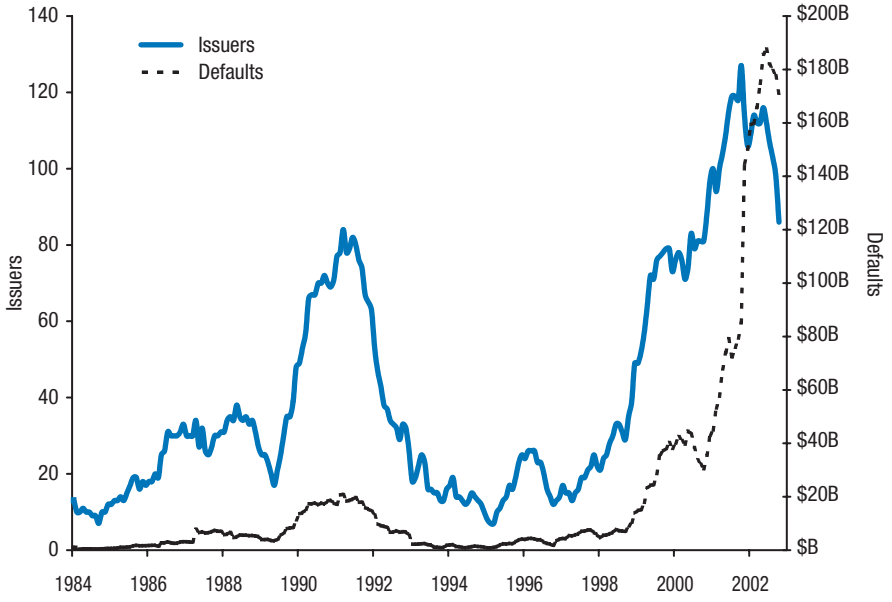
14. Market participants often consider the occurrence of fallen angels (rating changes from investment grade to speculative grade) to be more important than other rating changes. Moody's therefore tracks the frequency of fallen angels as well. This metric is not considered a performance measure as doing so might induce a stronger aversion to downgrading investment grade firms than that for downgrading firms in general.

Moody's Historical Performance

In this section, we measure Moody's rating performance based on rating histories and default experience of Moody's rated bond issuers between January 1983 through December 2002. We begin with a review of Moody's rating accuracy, measured at various points in time and averaged over time, for both one-year-ahead and five-year-ahead investment periods. We then report statistics on the variability of default rates around their historical means, investment-grade default rates, average rating prior to default, and the various rating stability metrics that were outlined in the previous section. Accuracy, as measured by the five-year accuracy ratio and average rating before default, appears to have peaked in the mid-1990s and is now holding steady at slightly lower levels. Volatility, by most measures, has risen since 2000.

When evaluating rating performance over time, it is important to keep in mind that the number of issuers is generally growing over time and the number of defaults within each period can be quite volatile. Exhibit 5 depicts the change in rated issuers and defaults over time. The difference between the two series drives differences in the sample averages for various metrics and is sensitive to when we calculate an average using a pooled cohort approach or an average of monthly cohorts, weighted either equally, by issuers or by defaults.

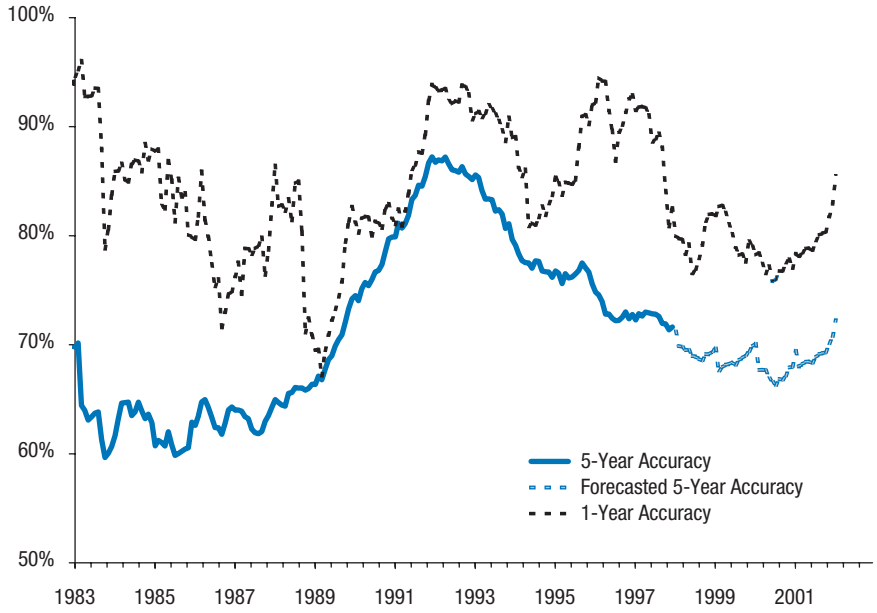
Exhibit 5: Default Statistics, 1982 – 2002



Tracking The Relative Rating Accuracy Of Moody's Ratings

Exhibit 6 depicts the time series of the one-year- and five-year-horizon accuracy ratios of cohorts formed at the end of each month. Rating accuracy measured at the one-year horizon is quite volatile because default rates by rating category are volatile. However, the chart indicates that one-year horizon rating accuracy was highest in the early and mid-1990's and weakest in the mid- and late 1980's.

Exhibit 6: Moody's One-Year And Five-Year Accuracy: Measured Monthly, 1983 – 2001



Since Moody's ratings are intended to provide stable long-term signals of relative credit risk, we are also interested in accuracy ratios measured over longer horizons than just one year. The difficulty with long-term performance measurement, however, is that, by definition, one can only assess the long-term performance of ratings that were assigned in the past. This information may be of limited use in understanding recent performance.

Accuracy ratios are naturally lower over long horizons than over short horizons because defaults in the distant future are more difficult to anticipate than defaults that are imminent. For example, in the most recent observation of a complete five-year horizon (the December 31, 1997 ratings cohort), the five-year accuracy ratio is 72.0% while the one-year accuracy ratio is 87.6% for the same cohort.

Exhibit 6 also depicts changes in Moody's five-year accuracy ratio over time. Compared to one-year statistics, we can see a more marked rise in performance during the early 1990s and a sharper subsequent decline. Although the last complete five-year accuracy ratio can only be measured against ratings that were outstanding five years ago, it is possible to project future performance based on the information we have about defaults since that time. Exhibit 6 includes our projection and is derived from the strong empirical relationship we observe between short-term default experience and five-year accuracy. The projection anticipates a further modest increase in accuracy.

Exhibit 7 presents accuracy ratios averaged over time for one-year-ahead and five-year-ahead investment horizons. The pooled statistics are calculated by aggregating ratings for all monthly cohorts and forming one "massive" cumulative accuracy profile and deriving the associated accuracy ratio. Weighted-average accuracy ratios are formed by calculating accuracy ratios associated with each monthly cohort and averaging across cohorts, weighted by the number of issuers in each cohort. (Equal-weighted cohort averages and default-weighted cohort averages produced very similar results.)

Exhibit 7: Accuracy Ratios Averaged Over Time (1983 – 2002)

	Pooled Cohorts	Issuer-Weighted Cohort Averages
One-Year Horizon		
All Credits	82.6	83.5
Investment-Grade Credits	63.8	66.6
Speculative-Grade Credits	52.8	54.0
Five-Year Horizon		
All Credits	71.0	72.9
Investment-Grade Credits	43.9	48.2
Speculative-Grade Credits	27.9	31.7

Exhibit 7 reveals that, as one might expect, accuracy is lower at longer horizons because uncertainty is higher over longer prediction horizons. Further, the issuer-weighted cohort averages are consistently higher than the results for the pooled cohorts because default rates by rating category vary over time. The pooled cohort mixes ratings from periods of relatively low and relatively high default rates. The issuer weighted cohort averages, on the other hand, provide an average measure of how well Moody's ranked its issuers at each point in time. Though the pooled cohort results are lower than those for the average cohort, they are not significantly lower. As stated previously, accuracy ratios are invariant to changes in the aggregate default rate but not to changes in the distribution of ratings and the distribution of defaults among the rating categories. The closeness of the pooled cohort results to the cohort averages imply that the relative meaning of Moody's ratings has been consistent over our period of study.

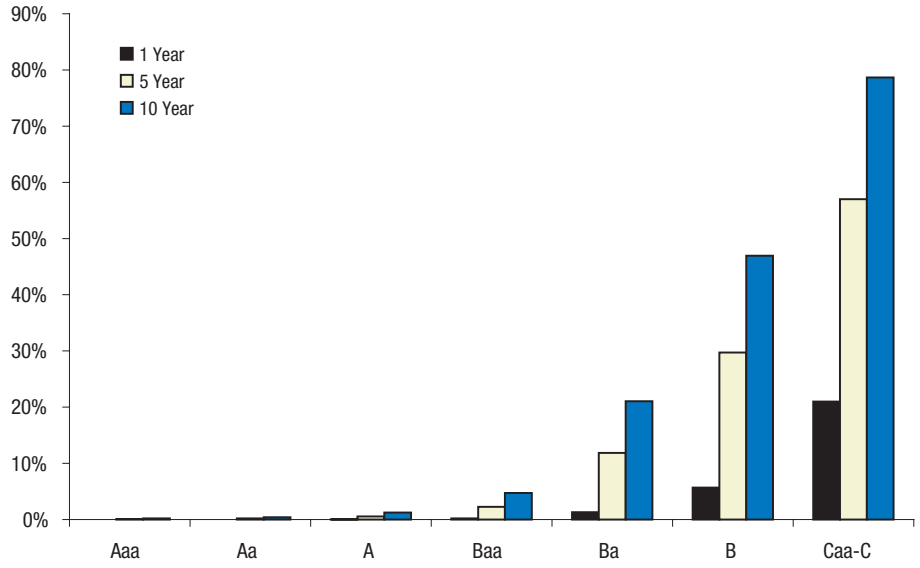
Exhibit 7 also compares the predictive power within the investment-grade and speculative-grade portion of the rating scale. Here we see that the investment-grade ratings are generally "more accurate" than speculative-grade ratings, which means roughly that the difference between Baa and Aaa default rates is greater than the difference between Caa and Ba default rates. Unsurprisingly, Moody's overall accuracy is much greater than its accuracy within the investment-grade and speculative-grade sector because much of the information content of the rating system comes from the ability to determine whether credits are investment grade or speculative grade.

Tracking Moody's Absolute/Cardinal Rating Accuracy

Cumulative Default Rates By Rating Category

Exhibit 8 displays cumulative defaults by rating category over different investment horizons. This chart is commonly used to demonstrate the power of credit ratings to discriminate relative credit risk. As expected, the default rates are monotonically increasing as the rating level decreases and as horizon increases.

Exhibit 8: Multi-Year Cumulative Default Rates By Rating Category, 1983 - 2001



Variability Of Cumulative Default Rates By Rating Category

Since Moody's does not target idealized default rates for its rating categories, we do not expect realized default rates to stay within any pre-specified band of variability. Nevertheless, Moody's annual default study does provide extensive detail on the variability of realized default rates by rating category over different horizons for different annual cohorts. To indicate the range of this variability, Exhibit 9 presents the mean and volatility of three-year default rates by rating category, where the three-year horizon has been highlighted because it has been the recent focus of attention of the Basel Committee (see Footnote 8). The "boxes" in the chart delineate the one standard deviation bounds and the "whiskers" show the minimum and maximum realizations. Each category displays a wide variation. For example, B-rated cohorts experienced three-year default rates that varied from 3.6% to 34.5%.

Exhibit 9: Three-Year Cumulative Default Rates By Rating Category

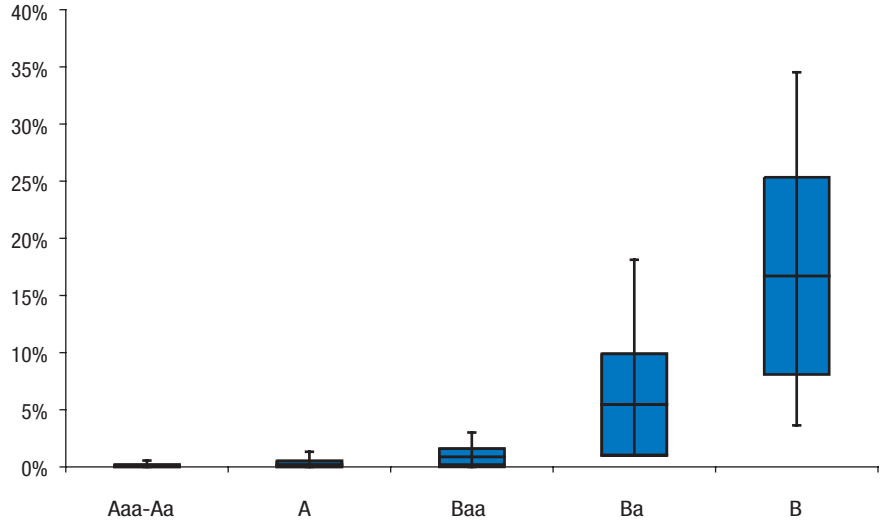
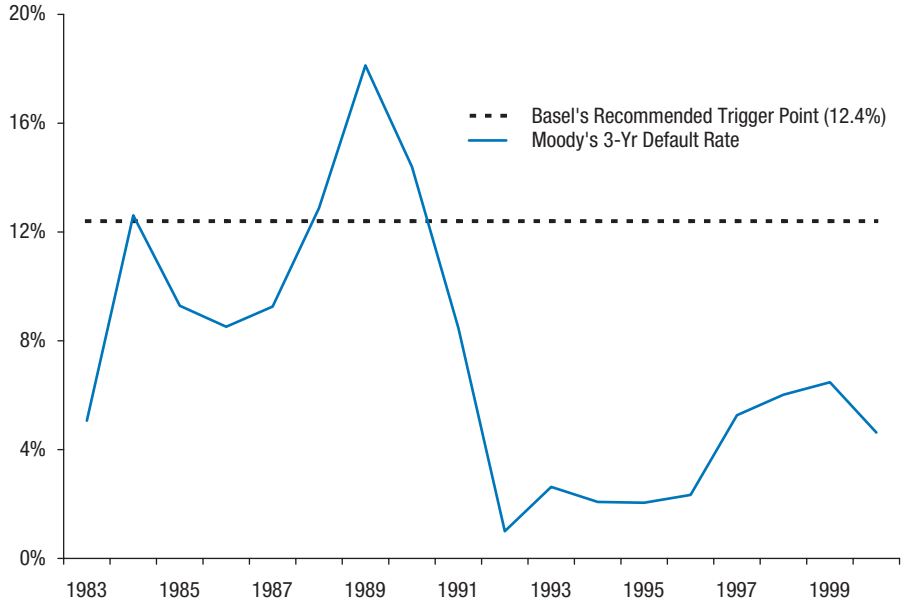


Exhibit 10 depicts the variation of 3-year cumulative default rates for Ba-rated cohorts over time and indicates the Basel Committee's recommended "monitoring" level for that rating category. Notice that between 1989 and 1991, Moody's Ba-rated default rate exceeded the Committee's suggested target rate. As discussed above, however, Moody's does not include the stability of specific idealized default rates in its performance objectives.

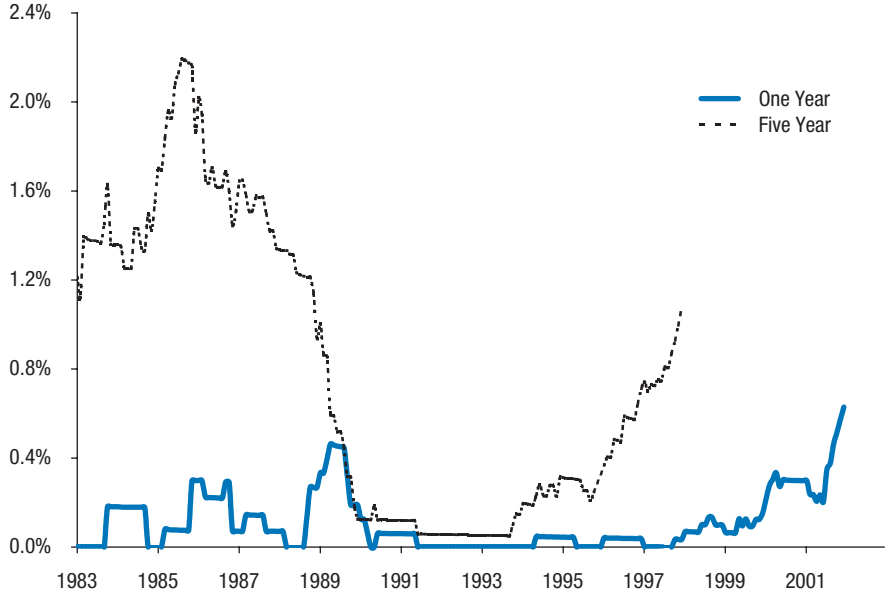
Exhibit 10: Three-Year Cumulative Default Rates For Ba-Rated Cohorts



Investment Grade Default Rates

Exhibit 11 displays the percentage of investment-grade issuers that subsequently defaulted over one-year and five-year horizons. The most recent cohort that has a measurable five-year default rate is the December 31st, 1997 cohort. One-year investment-grade default rates peaked during 2002, whereas, five-year investment-grade default rates have risen, but have not yet reached their historical maximums. The October 31st, 1985 cohort consisted of 1,333 investment-grade issuers. Twenty-nine of these defaulted during the subsequent five years. For comparison, the December 31st, 1997 cohort had 2,877 investment-grade issuers and only thirty-one subsequently defaulted. Investment-grade default rates have risen in recent years, but the five-year investment-grade default rate remains well below the highs observed in the mid-1980s.

Exhibit 11: Investment-Grade Default Rates



Average Rating Prior To Default

The solid line in Exhibit 12 presents the typical pattern of ratings over the three years before default for all issuers that defaulted between 1983 and 2001. As depicted, the typical defaulter was rated Ba3 three years prior to default, and its rating slowly declined to Caa2 at the time of default. If we average the thirty-seven monthly observations depicted in the chart, we see that the typical defaulter had a rating roughly halfway between B2 and B3 during the three years prior to its default.

Exhibit 12: Average Rating Prior To Default: 1983 – 2001

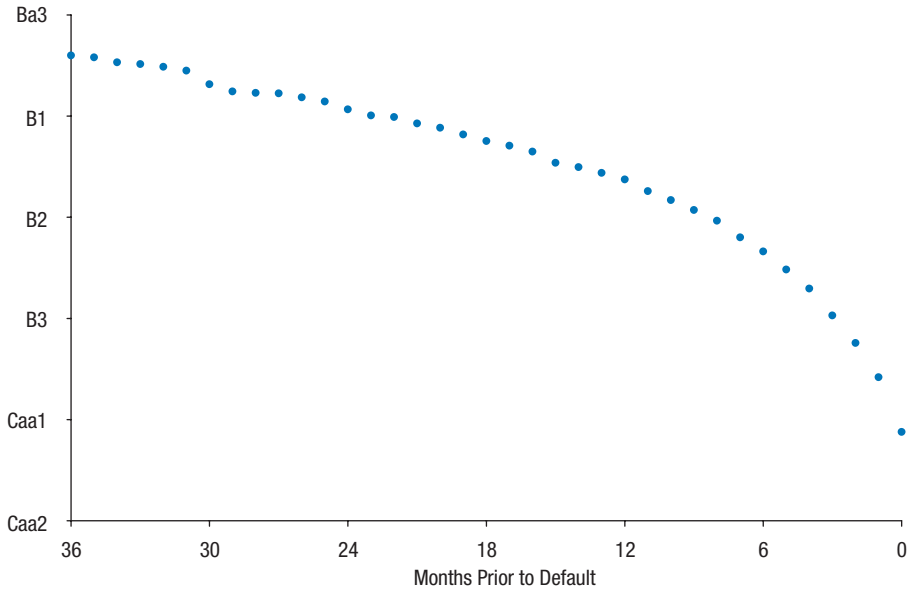
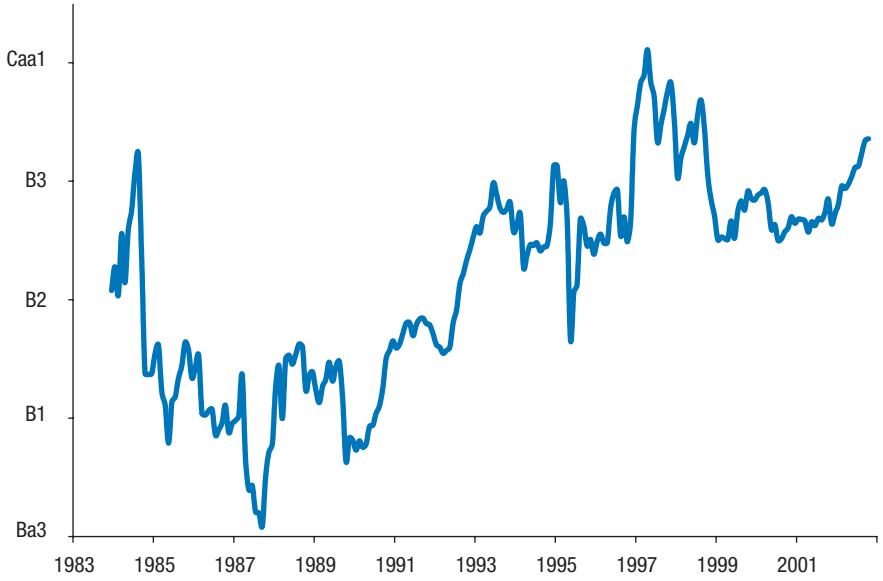


Exhibit 13 shows the progression of the average rating prior to default statistic, updated monthly, for Moody's universe of corporate and sovereign defaulters. Each monthly data point represents an average, across all defaulting issuers over the previous twelve months, of their average rating during the 36-month period preceding their defaults. Because the statistic looks back three years prior to the default, the peak in performance that we observe in 1997 is actually a composite of ratings performance from 1993 through 1997. Since 1997, Moody's has been less successful in assigning low ratings in advance of default; however, recent issuer-weighted performance remains superior to that of the 1980s and is consistent with that of most of the 1990s.

Exhibit 13: Average Rating Prior To Default



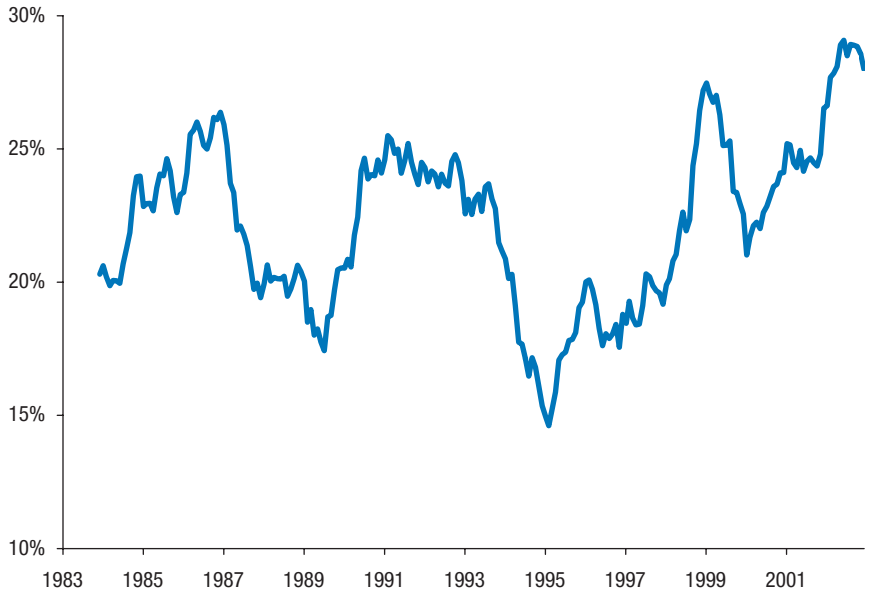
Tracking Moody's Rating Stability

Moody's ratings have been relatively stable over time. In any typical year, fewer than a quarter of issuers experience a rating change of any type, less than five-percent experience large rating changes, and only about one-percent experience a reversal of a previous rating change. The last two years have been atypically volatile but the absolute percentages are still relatively low (28%, 8%, and 0.8% respectively).

Rating Actions

Exhibit 14 provides a historical perspective on the number of issuers that experienced rating changes annually. The statistic is updated monthly and always looks forward one year from the calculation date. As shown in the chart, rating actions are more frequent around recessions, (1990-1991 and 2001-2002), and rating volatility, as measured by this metric, is currently at the highest level observed since 1982.

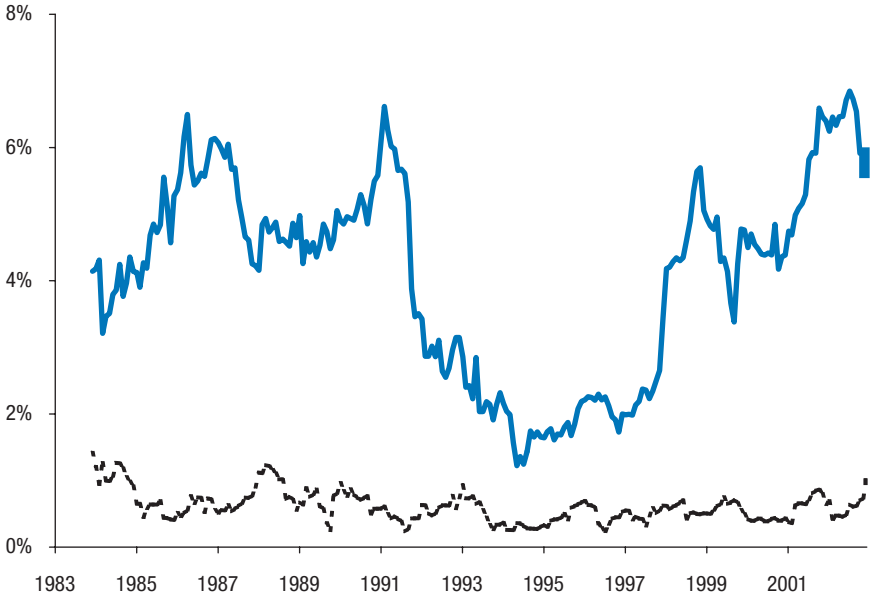
Exhibit 14: Share Of Issuers That Experience Rating Actions Within Previous 12 Months



Large Rating Changes

The solid line in Exhibit 15 depicts the share of issuers that experienced a rating change of three or more notches, based on a moving twelve-month window, measured monthly. Large rating changes are recorded whether they were due to one large change or multiple small changes during the twelve-month period. As depicted, the number of issuers with large rating changes was low during the majority of the nineties but has returned to levels seen in the 1980s. The series hit an all time high in 2001. Approximately 44% of the large movers originate from investment grade ratings and the remaining 56% originated from speculative grade ratings.

Exhibit 15: Share Of Issuers With Large Rating Changes (Solid Line) And Reversals (Dotted Line) Within Previous 12 Months



Rating Reversals

The dotted line in Exhibit 15 depicts changes in the rating reversal rate over time. The reversal rate has been plotted with the large rating change rate to emphasize how infrequent rating reversals are. As depicted, the share of issuers with rating reversals has historically ranged between 0.4 % and 1.2%. The rating reversal rate was generally higher in the 1980s than it was in the 1990s or the last few years. Nevertheless, the rating reversal rate has been slowly increasing over the past ten years and experienced a local peak during the past few months. *The number of reversals is small relative to the total number of rating actions and the number of large rating movements.*

Comparing Moody's Ratings And Bond Market-Implied Ratings

The metrics reviewed above address the question, has rating performance changed over time? Some changes, however, may reflect changes in the environment or in the ease or difficulty of rating credit risk, rather than changes in methods or practices.

To measure the cause of performance changes, Moody's regularly compares its performance to that of other credit risk measures. Other risk measures include the ratings of other agencies — such as Standard & Poors or Fitch — or ratings inferred from market based measures such as bond market credit spreads and Moody's KMV Expected Default Frequencies (EDF™). Of these, the most central point of reference is the bond market (through ratings inferred from credit spreads) because, at the broadest conceptual level, Moody's primary goal is to improve the transparency and efficiency of credit markets. The following discussion compares the accuracy and stability of Moody's ratings and ratings inferred from bond market credit spreads. (Since January 31, 1999, Moody's has derived bond market-implied ratings for over 6,500 corporate and sovereign bonds, covering over 2,300 unique issuers representing over 2,000 corporate families.[15])

As detailed in Exhibit 16, bond market market-implied ratings are more powerful than Moody's ratings at a one-year horizon, but the difference becomes less substantial at a three-year horizon. These conclusions hold, regardless of whether one measures the accuracy ratio based on a pooling of the cohorts or through an issuer-weighted average of the accuracy ratios of all the individual monthly cohorts.[16]

Exhibit 16: Accuracy Ratios For Moody's Ratings And Bond Market-Implied Ratings

	Pooled Cohorts	Issuer-Weighted Averages
Moody's Ratings		
1-Year Horizon	70.3	70.7
3-Year Horizon	61.0	61.3
Bond Market-Implied Ratings		
1-Year Horizon	77.6	77.1
3-Year Horizon	62.6	62.9

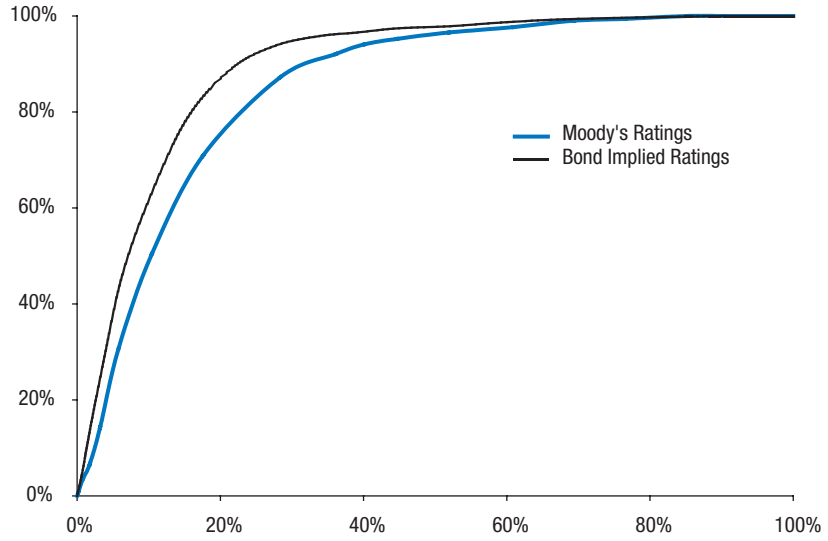
(Based On Matched-Sample, Monthly Cohorts, Jan 1999 – Jan 2002)

15. *Using end-of-month bid price and spread data, we relate option-adjusted spreads to option-adjusted durations on a single day for all straight and callable coupon bonds in our sample. Non-dollar denominated credit spreads are converted to dollar equivalent spreads based on corporate swaps in the dollar and non-dollar markets. We then derive a pricing "matrix" which maps the median credit spread by rating category to different option-adjusted spreads. A look-up table then allows us to infer the bond market's implied credit rating for any individual bond. We obtain issuer-level bond market-implied ratings by averaging the issuer's bond level ratings after adjusting for the standard notching differences between bonds of different seniority and security. Our approach abstracts from differences in credit spreads that result from differences in liquidity. Attempts to control for differences in liquidity, using standard regression analysis, actually produced a weaker relationship among ratings, spreads and durations than the simple median approach described above.*

16. *In other research (not shown here), we find that Moody's accuracy ratios (based on pooled cohorts) improve relative to those of market-based risk measures as the horizon lengthens. This finding should be unsurprising since over the past 20 years, market deviations from trend have proven to be short-lived and hence, compared with more stable credit ratings, market-based risk measures have provided less accurate measures of long-term credit risk.*

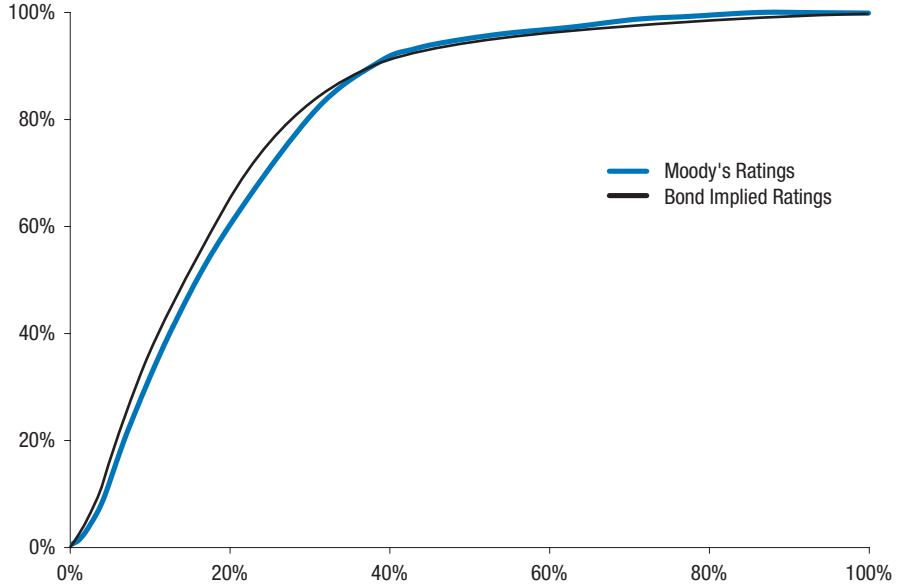
Exhibit 17 shows that the difference in one-year-horizon ARs reflects an unambiguous difference in CAP plots. However, Exhibit 18 indicates that, at a 3-year-horizon, the AR of bond market-implied ratings slightly exceeds the AR of Moody's ratings, although Moody's modestly outperforms the bond market in the higher (investment-grade) portion of the rating range.[17]

Exhibit 17: One-Year-Horizon CAP Plots For Moody's Ratings And Bond Market-Implied Ratings



** Based on pooled montly cohorts, Jan 1999 – Jan 2002*

Exhibit 18: Three-Year-Horizon Accuracy Ratios For Moody's Ratings And Bond Market-Implied Ratings



** Based on pooled montly cohorts, Jan 1999 – Jan 2002*

17. *Again, we have shown in other research, that as the horizon lengthens, Moody's comparative advantage viz. market-based measures improves, particularly within the higher portion of the rating scale.*

As one might also expect, bond market-implied ratings are much more volatile than Moody's ratings. Exhibit 19 presents the respective 12-month average volatility statistics, based on data from 1999 through 2002. Moody's ratings have been much more stable than market-based credit opinions. In any given twelve-month period, there is about one rating action for every four issuers and only one out of every 100 issuers experiences a rating reversal (a rating change followed by a rating change in the opposite direction). In contrast, the same statistics for market-implied ratings inferred from bond spreads are nine out of ten issuers with rating actions and three out of four experience rating reversals.

Exhibit 19: 12-Month Average Volatility Statistics (As a Percentage Of Issuers)

	Moody's	Bond Implied Ratings
Rating Changes	25%	91%
Large Rating Changes	7%	43%
Rating Reversals	1%	76%
Memo Item: Average Number of Rating Changes Over 12 Months For Each Issuer That Experiences a Rating Change	1.2	4.5

(Based On Monthly Cohorts, Jan 1999 – Jan 2002)

Concluding Comments: The Role Of Rating Outlooks And Watchlist Designations

While Moody's believes that long-term accuracy and short-term stability are positively correlated, there may be a tradeoff between these two aspects of rating quality, accuracy and stability. It might be possible to increase the short-term correlation between ratings and defaults by making more volatile and more responsive to new information, without regard for its potentially transitory nature. As part of the implementation of its rating system, Moody's limits the frequency of rating changes to perceived changed in fundamental characteristics, dampens their severity, avoids rating change reversals, dialogues with issuers, and ultimately makes decisions by committee. These actions are consistent with the endeavor to produce stable, long-term ratings with a minimum of false default predictions. They also presumably, though not necessarily, reduce the short-term prediction content of ratings.

In recognition that the current rating may not always fully reflect potential changes in an issuer's credit situation, Moody's provides the market with additional signals — changes in rating outlooks and Watchlist designations — that together with the rating, more fully describe each issuer's likelihood of credit losses. In future research, we plan to document how outlooks and watchlists can be used to extract incremental information content from the ratings system.

To order reprints of this report (100 copies minimum), please call 1.212.553.1658.

Report Number: 77916

Authors

Richard Cantor
Christopher Mann

Production Specialists

John Tzanos
Yung Chu