Special Comment

Contact	Phone
New York	
Richard Cantor	1.212.553.1653
Chris Mann	

Analyzing the Tradeoff Between Ratings Accuracy and Stability

Contents:

Introduction

Our conversations with investors, issuers and regulators have led us to conclude that many market participants have a strong preference for credit ratings that are not only accurate but also stable. They want ratings to reflect enduring changes in credit risk because rating changes have real consequences – due primarily to ratings based portfolio governance rules and rating triggers – that are costly to reverse. Market participants, moreover, do not want ratings that simply track market-based measures of credit risk. Rather, ratings should reflect independent analytical judgments that provide counterpoint to often volatile market-based assessments.

It may be possible, however, to increase the short-term predictive content of our rating system by increasing the responsiveness of Moody's ratings to new information about credit fundamentals. In other words, it may be possible to increase ratings accuracy while reducing, perhaps substantially, ratings stability. Since different market participants have different needs and those needs evolve over time, Moody's periodically conducts market surveys to understand how it can maximize the utility of the rating system. We hope that the discussion presented in this Special Comment will stimulate additional market feedback on the current balance between accuracy and stability reflected in Moody's credit rating system.

In the sections that follow, we discuss the potential accuracy and stability attributes of different rating management systems. We then illustrate the potential tradeoff by comparing the combinations of the accuracy and stability associated with the many different "rating systems" that can be derived by applying various filters to Moody's-KMV's expected default frequencies ("EDF^{TM}'s). The performance of Moody's traditional ratings – both adjusted and unadjusted for rating reviews and outlooks – is then compared to the performance of the various filtered versions of EDF-implied ratings.

Why Do Users of the Rating System Value Both Accuracy and Stability?

Users of Moody's rating system value both accuracy and stability. This point is obvious when one considers extreme situations. No investor would be satisfied with a perfectly stable rating system in which rating changes never occur, but ratings are wholly inaccurate. Similarly, no investor would be indifferent between two rating systems that had precisely the same levels of accuracy, but one had stable ratings and the other had ratings that gyrated wildly from day to day or even within a day.

While the desire for stable ratings may be intuitive, it is useful to be more precise about why stability is likely to be useful. Users of rating systems value stability because they sometimes take actions based in part on rating changes and these actions imply costs that may be unrecoverable if those actions need to be reversed in response to future rating changes. Stability is important because ratings affect behavior and the actions taken in response to rating changes have consequences.

The most widely cited example of actions based on ratings that are costly to reverse involves ratings-based bond portfolio composition guidelines. Bond fund managers do not, of course, base asset purchase and sale decisions on ratings alone; rather, they conduct their own credit research, evaluate non-credit elements of each investment, and evaluate each security's overall risk/return characteristics. However, ratings are used by investors as governance tools to monitor and to constrain the investment choices available to portfolio managers they employ. When, for example, a security is downgraded into the speculative-grade rating range, some portfolio managers are required to sell their holdings.[1]

In the presence of ratings-based governance rules, volatile ratings are costly. If ratings were to adjust wildly from day to day, asset managers might be required to purchase, sell, and then repurchase the same securities with great frequency, imposing large transaction costs upon the portfolio. Alternatively, investors could manage the rating volatility problem by modifying their rating-based governance rules or dropping them entirely.[2]

Ratings are also used as tools for mitigating principal-agent problems in other areas within credit markets, and, in these cases as well, the value of ratings as governance tools depends on both accuracy and stability. For example, credit committees and credit officers often require varying levels of review based on a borrower's credit rating before approving underwriter or loan officer recommendations. In addition, lenders often offer borrowers more favorable terms if they are willing to commit to ratings-based covenants that trigger debt repricing, refinancing, or collateralization. Moreover, some financial regulators vary capital requirements with the riskiness of an institution's assets, as measured in part by the credit ratings assigned to its investments. In all these cases, changes in ratings have real consequences for the users of ratings, so volatility reduces the efficiency of ratings as tools of governance. In many of these cases, certain rating changes lead an agent to take an action that is costly to undo if the rating change is subsequently reversed, particularly for less liquid securities.

Because of the ratings-based governance rules and financial triggers in place, issuers and financial regulators also have a strong preference for stable rating systems. When many market participants employ similar ratings-based triggers or governance rules, rating agencies to some extent end up playing a gatekeeper role in the capital markets. Rating upgrades can at times make it much easier for individual issuers to access large pools of debt capital and rating downgrades can have the opposite effect. The effects of such changes in market access can be difficult to reverse: easier access may lead to increased fixed investment, and more difficult access can lead to downsizing or, in extreme cases, default. In such a context, large and, perhaps transient, rating changes can be very harmful.

Why Is There a Tradeoff?

Certainly there are situations when there is no tradeoff, when agencies can improve accuracy or stability without worsening performance along the other dimension. For example, suppose a rating agency is not availing itself of the best available analytical methods. In this case, adopting a superior methodology may increase accuracy without impacting stability. Such an outward move from inside the accuracy/stability "frontier" is depicted in the chart below as a move from point A to point B. Similarly, a rating agency may be regularly changing ratings and then quickly reversing those actions in response to information that turns out to be wholly unrelated to credit risk. Such rating actions may have very little effect on accuracy but may substantially reduce stability. Avoiding such rating actions in the future can shift the rating system from point A to point C.

---

1. Ratings-based governance rules are common in the institutional money management sector because many investors do not want to evaluate individual credits and hire professional money managers to do that work for them. Recognizing, however, that investment managers may have an incentive to take excessive risk (due to the limited amount of their personal capital at risk), investors often require their managers to purchase only rated debt; indeed, they often require their managers to purchase and to hold only investment-grade-rated debt.

2. These issues are explored empirically in a recent paper by G. Löffler (2004), "Ratings Versus Market-Based Measures of Default Risk in Portfolio Governance," in Cantor, R. (ed.), Recent Research on Credit Ratings, special issue of the Journal of Banking and Finance, 28 (11).

Exhibit 1

The Accuracy/Stability Frontier
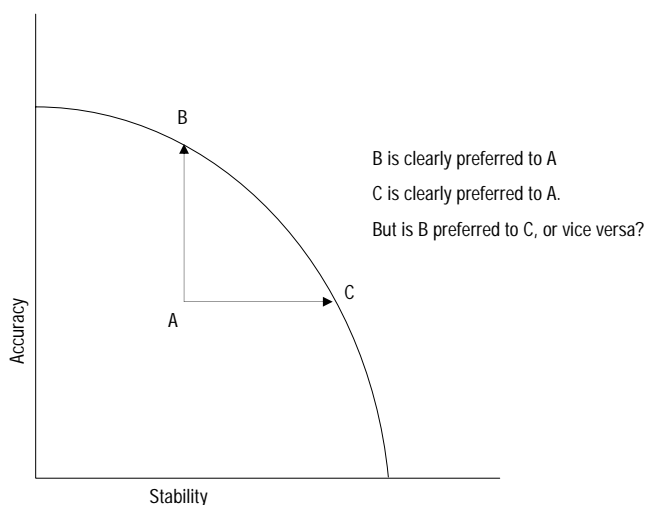
However, the desirability of changes in rating practices that improve accuracy but worsen rating stability or vice versa is harder to evaluate. For example, Moody's currently changes ratings on roughly 20% to 25% of all corporate issuers each year. If Moody's were to limit rating changes to just 10% of all issuers each year, changing ratings on only those issuers whose credit risk profiles had changed the most, then stability would obviously increase, but accuracy would presumably decline, resulting in a move along the accuracy/stability frontier such as from point B to point C. Similarly, if Moody's were to react more quickly to information released about potential changes in risk profiles, as revealed for example in daily stock price fluctuations, rating accuracy might increase, but stability would decline, as reflected in a move from point C to point B.

While it may be possible to use new rating analytics to push out the frontier (i.e., increase both accuracy and stability), the existence of a frontier – a tradeoff between accuracy and stability – is unavoidable. Rating system management practices imbed an implicit tradeoff between accuracy and stability because information about credit risk can be revealed and processed only gradually over time. When first revealed, each piece of new information suggests a potential change in the relative ranking of credit risk embodied in the rating system. If the objective were to simply maximize accuracy, it might be desirable to change many ratings every day. However, the long-term implications of each bit of new information can be fully understood only over time, as more clarifying information is obtained. The more time one takes to process the additional information, the more likely it is that the rating system will be stable, yet the accuracy of individual ratings may degrade between rating changes.

How Does Moody's Define Rating Accuracy and Stability?

Ratings accuracy refersdto the correlation between ratings and the risk of default or, more generally, credit losses. Ratings stability refers to the frequency and magnitude of rating changes, as well as the likelihood that rating changes will prove enduring.

As discussed in a previous Special Comment, Moody's defines accuracy in a number of different ways, but states that our primarily accuracy objective is that ratings should provide a powerful rank ordering of credit risk, reflecting both default probability and expected loss severity in the event of default over long horizons[3]. Moody's rating system is primarily intended to measure relative credit risk[4]. Our ratings do not necessarily imply fixed expected probabilities of loss; however, in recognition of the widespread use of ratings as coarse tools for separating low from high credit risk instruments, ratings are assigned with the objective of ensuring that investment-grade defaults are rare.

Moody's employs numerous metrics to assess rating accuracy and stability. For illustrative simplicity, in this Special Comment, we focus on just one measure of ratings accuracy, the one-year-horizon accuracy ratio[5], and one measure of rating stability, the share of ratings that remain unchanged over the course of a year.

---

3. See "Measuring the Performance of Moody's Corporate Bond Ratings," Moody's Special Comment, April 2003.
4. A relative credit rating system is by design more stable than an ordinal system, since macro-fluctuations in credit risk will not necessarily be reflected through rating changes.

Mapping an Accuracy/Stability Frontier Using Filtered EDF-Implied Ratings

In this section, we provide an example of how one can derive an accuracy/stability frontier based on Moody's-KMV's expected default frequencies ("EDFs")[6]. The analysis provides a systematic way to illustrate the potential tradeoff, under the condition that the only data used to develop a "rating system" are current and historical EDFs. This example is meant only to be illustrative, since results will vary depending upon the risk measure being used to construct the frontier, the time period of the analysis, the investment horizon selected as the focus of the analysis, and the metrics used to measure accuracy and stability.

To analyze the stability characteristics of EDFs in a manner comparable to credit ratings, EDFs need to be translated into the language of ratings. Fortunately, in addition to EDFs, Moody's-KMV publishes EDF-implied ratings, which are "credit ratings" mapped from EDFs, expressed using Moody's long-term rating symbols. Because EDFs (and therefore EDF-implied ratings) are systematically linked to stock prices, they can be quite volatile. The analysis is based on month-end EDF-implied rating data for all companies that also carry Moody's ratings from January 1999 through December 2005. There are 2,820 unique issuers in total, representing 158,201 monthly observations as well as 299 unique defaults[7]. For every Moody's rating change, EDF-implied ratings had 18 month-to-month rating changes. The mean absolute size per rating change, given a change occurred, was about a notch and a half for both systems. The EDF-implied ratings for this sample achieved a one-year-ahead accuracy ratio of 84.6%[8] and average rating stability rate of 13.7%[9].

Multiple rating systems, however, with different accuracy and stability characteristics, can be derived from these same EDF-implied ratings by applying "smoothing algorithms" or "filters." In particular, we considered three classes of filters, each of which limits the circumstances under which ratings can change, depending on the gap between the current filtered (EDF-implied) rating and the current actual (EDF-implied) rating.

In the first case, we consider only whether the gap is "large;" i.e., the filtered rating is changed only if the gap exceeds a certain threshold. If that threshold is one rating notch, then the filtered rating histories are exactly the same as the actual EDF-implied rating histories since the filtered rating changes any time the EDF-implied rating changes. If the threshold is two rating notches, then the filtered ratings change less often than the EDF-implied ratings; the filtered rating jumps to the current EDF-implied rating whenever the gap between them at month-end is two or more rating notches, and so on. At the limit, when the threshold is twenty-one, the filtered rating never changes and remains equal to the first EDF-implied rating observed in the sample. With only twenty-one possible rating categories (Aaa through C), the ratings gap can never exceed twenty notches.

In the second case, we consider only whether the gap has been "persistent," i.e., the filtered rating is changed only if the rating gap (between the filtered rating and the EDF-implied rating) is positive or negative for a given threshold number of consecutive months. If that threshold is one month, then the filtered rating histories are exactly the same as the actual EDF-implied rating histories since the filtered rating changes any time the EDF-implied rating changes. If the threshold is two, then the filtered ratings change less often than EDF-implied ratings; the filtered rating jumps to the current EDF-implied rating whenever a gap persists for two consecutive months. And so on. With only eighty-four months in our sample (January 1999 – December 2005), the limit is reached when the threshold is eighty-four; in which case, the filtered rating never changes and remains equal to the initial EDF-implied rating. Results for all months are presented, simulating how the system would have actually been used by analysts who adopted the filter in January 1999.
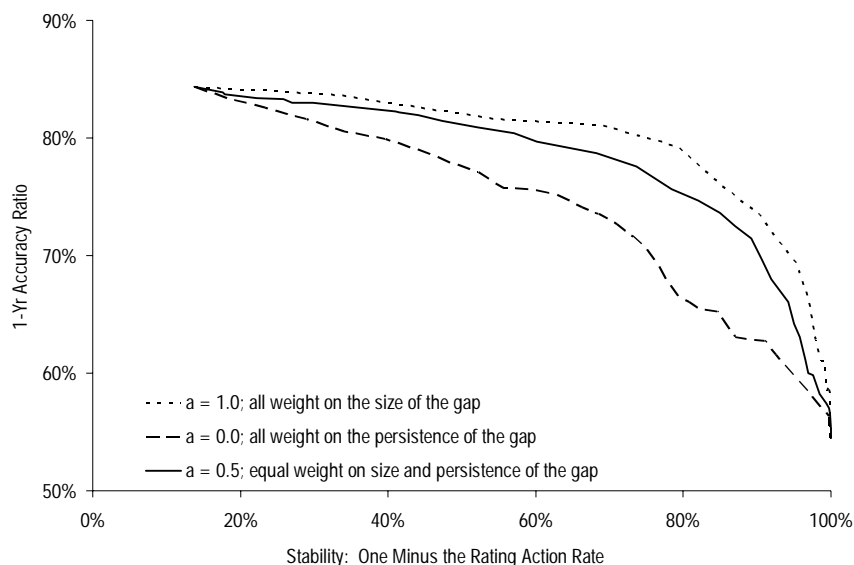
In the third case, we consider both the size of the gap and its persistence, with equal weight on each; i.e., the filtered rating is changed only if the combination of the ratings gap and its persistence exceeds a certain threshold number. If that threshold is one, then the filtered rating histories are again exactly the same as the actual EDF-implied rating histories. At higher threshold values, filtered ratings can change either because the ratings gap is large, or because a positive or negative gap has been persistent, or because of some combination of the two.[10]

---

5. While we focus here on a one-year horizon, Moody's rating system objectives emphasize longer horizon accuracy metrics, and Moody's ratings, in fact, perform better relative to other metrics at longer horizons. The enhanced power of Moody's ratings relative to other metrics at longer horizons is documented in "The Complementary Nature of Ratings and Market-Based Measures of Default Risk," by G. Löffler, unpublished paper, 2005 and "How Rating Agencies Achieve Rating Stability," by E. Altman and H. Rijken, published in Cantor, R. (ed.), Recent Research on Credit Ratings, special issue of the Journal of Banking and Finance, 2004. Other measures of rating accuracy include investment-grade default rates and average rating levels prior to default.

6. EDFs for public corporations are measures of one-year-ahead default probabilities based on each company's stock price and liability structure derived using a proprietary option-based model of the firm. Further information about the derivation of EDFs can be found on the Moody's KMV website. (See, for example, Peter Crosbie and Jeff Bohn, "Modeling Default Risk," December 18, 2004, http://www.moodyskmv.com/research.

7. Monthly cohorts are formed through December 2004, and their default experiences are tracked through December 2005.

8. The accuracy ratio of EDFs is slightly higher than that of EDF-implied ratings (84.6% versus 84.3 %) due to the greater granularity of the EDFs. These statistics are averages of the ARs of the monthly cohorts. In our sample period, the AR of the raw EDFs is slightly lower (also 84.3%) when calculated on a pooled basis across cohorts.

9. The rating action stability rate is the average fraction of issuers whose ratings are unchanged over any twelve-month period. In addition to rating action stability rates, we also analyzed large rating action rates (share of issuers experience rating changes spanning more than two refined rating categories) and rating reversal rates (share of issuers with both upward and downward rating actions within a twelve month period) as measures of rating volatility. We do not to report those results because they basically tell the same story.

For each of these three "rating systems," Exhibit 2 depicts precisely how much accuracy is lost and how much stability is gained as the threshold for rating changes is increased. Interestingly, the best performing filtered rating systems are those that decide whether or not to change ratings based exclusively on the size of the gap between the actual EDF-implied and the filtered rating, without placing any weight on the persistence of the gap. That is, any level of accuracy that can be achieved with either an "all persistence" or an "equal weight" filtering system can be achieved with greater rating stability (using perhaps a different threshold value) by focusing only exclusively on the gap and ignoring persistence. This is a critical finding: when using EDF-implied ratings to infer credit risk, there is no need to track how long a ratings gap has persisted. If one wants to reduce rating volatility, one should limit rating changes to situations when rating gaps are large, but one need not worry about how long the rating gap has persisted.

Exhibit 2

Accuracy and Stability Characteristics of
Various EDF-Implied Rating Systems

- - - · a = 1.0; all weight on the size of the gap
- - - a = 0.0; all weight on the persistence of the gap
——— a = 0.5; equal weight on size and persistence of the gap

Comparing the Performance of Moody's Ratings and Filtered EDF-implied Ratings

In Exhibit 3, we depict the performance metrics associated with Moody's ratings (represented by a single point) and the entire EDF-implied performance metrics curve associated with "all the weight on the size of the gap," since that weighting function dominates the performance of the others we considered. In this data sample, Moody's 1-year accuracy ratio is slightly more than 80% and its 1-year rating stability rate slight above 78%. Compared to the metrics of the unfiltered EDF-implied ratings (represented by the upper left hand point of tradeoff curve associated with the threshold value one), Moody's ratings appear less accurate but more stable. However, if one compares the filtered EDF-implied rating that has the most similar level of stability to Moody's ratings – the case in which the threshold b is set equal to 5 – then it turns out that Moody's ratings are slightly more accurate than EDFs. Note that the threshold b equal to 5 is quite high; i.e., to achieve a level of rating stability similar to that of Moody's ratings, filtered EDF-implied ratings should change only when the gaps between filtered and unfiltered EDF-implied ratings are five or more rating notches.

---

10. We modeled the accuracy and stability of an even larger set of EDF-implied rating systems; however, since all the intermediate cases are bracketed in intuitive ways by the three cases enumerated above, we have not reported the results. Formally, we define a sequence of monthly ($T = 0,1,2$, etc) filtered ratings R_T based on current and past EDF-implied ratings $\tilde{R}_T$ for any given company as follows:

$R_T = R_{T-1}$ when $|\tilde{R}_T - R_T|^a \cdot N^{1-a} < b$ and $\tilde{R}_T$ otherwise.

N = the number of consecutive months that $\tilde{R}_T - R_T$ has remained positive or has remained negative,
a = the relative weight on the magnitude of the gap relative to its persistence, and
b = the threshold beyond which rating changes are allowed to occur.

For each value of "a" (the relative weight assigned to the size and the persistence of the gap between the filtered and the actual EDF-implied rating), we calculate the accuracy ratio and rating stability rate associated with different values of "b" (the threshold for rating changes). Rating stability, of course, increases as b increases because fewer situations prompt rating changes.

Comparing the Optimal Filtered EDF-Implied
Rating Rules to Moody's Ratings
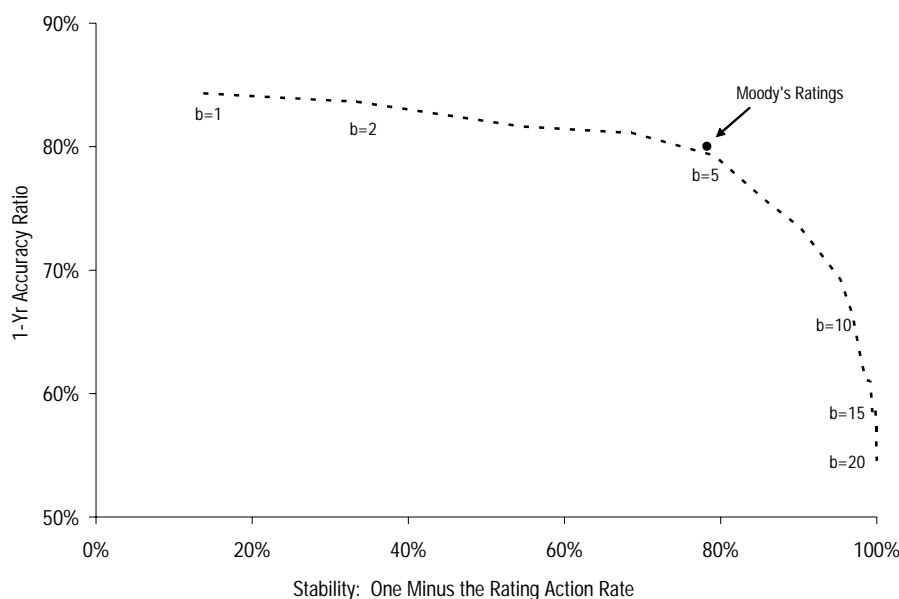
These results, however, may not be definitive since it is possible that more sophisticated decision rules could achieve a tradeoff more favorable to EDFs. Nevertheless, a recent academic paper confirms, using a very different framework, the potential accuracy gains from market-based metrics often fail to offset their associated volatility costs within the context of portfolio governance rules[11].

Using Rating Outlooks and Watchlists to Achieve a Different Mix of Accuracy and Stability

Moody's credit opinion is actually more complex than simply the rating. In particular, Moody's assigns positive or negative Outlooks and Watchlists to issuers that face asymmetric risks of upgrade or downgrade. These asymmetries are insufficient to prompt immediate rating changes because, in each case, there is a substantial probability a rating change might need to be reversed within a relatively short period of time, such as one year[12].

In prior research, we have shown that simple adjustments to Moody's ratings based on Outlooks and Watchlists substantially increase rating accuracy in predicting three-year default risk[13]. In a data set spanning 1996-2003, we observed that issuers on review for upgrade or downgrade experienced default rates similar to those of issuers with stable outlooks, rated two notches higher or lower. Similarly, issuers assigned positive or negative Outlooks experienced default rates equal to issuers rated one notch higher or lower, respectively. The accuracy and stability characteristics of a hypothetical rating system embodying these adjustments can be inferred by recreating issuer rating histories as if the rating changes were made at the time of outlook or rating reviews.

11. See "Ratings Versus Market-Based Measures of Default Risk in Portfolio Governance," by G. Löffler, published in Cantor, R. (ed.), Recent Research on Credit Ratings, special issue of the Journal of Banking and Finance, 28. Based on data from 1983 to 2002, Löffler identifies cases in which traditional ratings-based governance rules each offer higher returns to investors relative to the market-based rules - and vice versa. More generally, he finds that many statistical measures that are currently used to assess rating quality may be insufficient to judge the economic value of rating information in a specific context.

12. For example, an issuer that is "in play" is usually placed on review until its merger or acquisition is completed and its rating consequences are well understood. Similarly, issuers that experience above or below normal earnings growth are often assigned positive or negative outlooks during the period in which a rating committee awaits additional evidence that demonstrates whether the change in the firm's credit posture is temporary or enduring. During these periods of uncertainty, Moody's typically maintains the issuer's rating but signals the nature of the risk through Outlook and Watchlist assignments. At the same time, asset prices have generally shifted sharply, though it remains equally possible that the prices will soon shift back or shift even further away.

13. See "Understanding Moody's Corporate Bond Ratings and Ratings Process," Moody's Special Comment, May 2002, "Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History," Moody's Special Comment, February 2004, and "Rating Transitions and Defaults Conditional on Rating Outlooks Revisited: 1995-2005," Moody's Special Comment, December 2005.

Exhibit 4 characterizes the performance of Moody's ratings by themselves and Moody's ratings adjusted for Outlook or Watchlist status. As discussed elsewhere, the accuracy of Moody's ratings is increased if one treats ratings:

- on review for upgrade (downgrade) as if they were rated two notches higher or lower, and
- ratings with a positive (negative) Outlook as if they were rated one notch higher (lower)

than the ratings of issuers with stable outlooks. Under such conditions, accuracy increases and stability decreases because changes in Outlooks and Watchlists change the adjusted rating series. The performance of such a rating "system" is in fact fairly similar to that of the filtered EDF-implied rating rule when b equals 4; i.e., filtered EDF-implied ratings are changed only when the EDF-implied rating has moved four or more notches.

Exhibit 4

Comparing the Optimal Filtered EDF-Implied Rating Rules to Moody's Ratings Adjusted for Outlooks and Watchlists
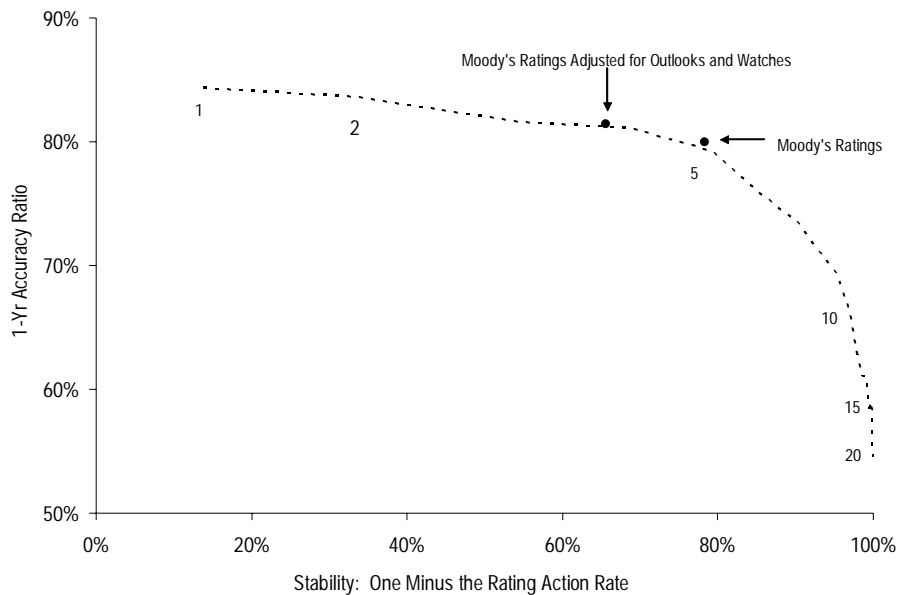
Does Moody's Achieve the Right Mix of Accuracy and Stability

The current rating system embodies a tradeoff between accuracy and stability that apparently meets the needs of many who use ratings as governance tools. At a one-year horizon, Moody's ratings are slightly less accurate than certain market-based credit ratings, and they are likely to be relatively more accurate at longer horizons. Moreover, work by others suggests it may be difficult to substantially increase ratings accuracy through the use of market-based information without sharply increasing rating volatility[14]. Furthermore, users of Moody's ratings can already obtain substantially superior accuracy performance with a relatively modest decrease in stability through the use of rating adjustments in response to Outlook and Watchlist information.

In January 2002, Moody's published a Special Comment ("The Bond Rating Process in a Changing Environment") that stated we were considering measures intended to improve rating timeliness, including shortening rating reviews, quicker reaction to material events, increased incidence of rating changes without formal reviews, and streamlining, or eliminating, rating Outlooks. The Special Comment, however, emphasized that "we will not make material changes to our rating process, nor will we move forward with any proposal without extensive market dialog."

---

14. See "The Complementary Nature of Ratings and Market-Based Measures of Default Risk," by G. Löffler, unpublished paper, 2005 and "How Rating Agencies Achieve Rating Stability," by E. Altman and H. Rijken, published in Cantor, R. (ed.), Recent Research on Credit Ratings, special issue of the Journal of Banking and Finance, 2004.

In May 2002, another Special Comment ("Understanding Moody's Corporate Bond Ratings and Rating Process") summarized the feedback we had received on the previous Special Comment and other initiatives as a result of over 35 meetings with issuer organizations, institutional investors, regulators and other market participants. We concluded that market participants desire ratings stability. They want ratings to be a view of an issuer's fundamental credit risk, which they perceive to be a measure of intrinsic financial capacity that is relative more stable when compared with other, more market-sensitive measures. Moreover, market participants are concerned that the use of quantitative inputs to the rating process will lead to greater volatility based upon transient market sentiment.

In response to this feedback, we decided not to materially change our rating practices. Instead, we have chosen to increase the transparency around those practices and their consequences. In particular, we developed a set of detailed metrics that can be used to measure rating system performance. We now publish quarterly updates on our performance based on those metrics. We have also published an extensive study of the meaning and implications of our rating Outlooks and Watchlists and another study documenting the cyclical characteristics of our ratings.

Nothing in this most recent study of the tradeoff between accuracy and stability suggests that investors would be well served by any changes to the management of our rating system. Moody's, nevertheless, recognizes that different investors have different needs, and those needs change over time. Accordingly, we welcome market feedback on the balance between accuracy and stability reflected in the rating system as presented in this report.