Profits first, or clients first?

Some lessons from Moody's stock price.

Gunter Löffler*

University of Ulm

February 2007

Abstract

Credit rating agencies enjoy high profit margins, but the quality of their products is often criticized. One explanation would be that agencies exploit their market power to sell low-cost products at premium prices. Agencies endorse the view that apparent shortcomings of their ratings are side effects of a rating policy that meets the expectations of rating users. To distinguish between the two hypotheses I examine the stock price of Moody's, the only listed firm among the three leading agencies. On days on which Moody's takes a rating action that the market objects (according to Moody's), Moody's stock price drops by more than 0.5%. The findings confirm Moody's claims about market preferences. Serving those preferences benefits shareholders and is therefore credible.

JEL classification: G2

Key words: credit rating agencies, rating quality, oligopoly, market discipline

* Gunter Löffler, Professor of Finance, Department of Finance, University of Ulm, Helmholtzstrasse 18, 89069 Ulm, Germany. Phone: ++49-731-5023597

e-mail: gunter.loeffler @ uni-ulm.de

1 Introduction

Moody's Corporation is listed on the NYSE with a current (December 2006) market value of $19 billion. Its main business has been to issue ratings of corporate and public debt.[1] The credit rating industry is largely oligopolistic. Moody's and its (unlisted) competitor S&P have an estimated market share of 80% as measured by revenues.[2] Over the years 1998 to 2005, the average profit margin (operating income/sales) of Moody's was 43%. On the other hand, dissatisfaction with the ratings produced by Moody's is widespread.[3]

This looks like a perfect example for a microeconomics textbook. An oligopolist reaps healthy profits by charging too much for a service that is far from perfect. One frequent criticism is that ratings are too slow in responding to new information, exemplified in the following statement from a group of financial economists (Danielson et al., 2001, p. 12):

> "While credit ratings provide some assessment of a company's riskiness, ratings generally lag market developments. This lag can be explained by the agencies' reliance on accounting data, their inability to monitor all issuers continuously and their willingness to change ratings only when their decision is unlikely to be reversed shortly afterwards (i.e. to avoid rating volatility)."

Moody's, in turn, implicitly concedes that it benefits from current market structure but also claims that its success rests on meeting market expectations for quality.[4]

It seems useful to distill these observations into two competing hypotheses:

PROFITS FIRSTS Rating agencies follow a high-margin policy of selling low-cost products at premium prices. Costs are low because agencies employ too few analysts to allow continuous monitoring of issuers; high prices can be sustained because of the oligopolistic market structure. To avert criticism, rating agencies put forward the seeming advantages of low rating volatility and a fundamentals-based analysis.

CLIENTS FIRST Investors and issuers prefer stable ratings. Important reasons for such preferences are rating triggers in bond contracts and rating-based portfolio governance rules, which lead to high transaction cost if ratings change frequently. To keep its client base,

---

[1] Over the years 1998 to 2005, traditional rating business contributed 45.9% percent of Moody's revenue. Other business segments are structured finance ratings, research and Moody's KMV. These figures (and the ones on profit margins) are obtained from Moody's annual reports.

[2] Cf. US Senate Report 109-326, 2006.

[3] See, for example, the survey evidence in Baker and Mansi (2002).

[4] Cf., for example, Moody's annual report for the year 2005 (p. 55 on the impact of regulation on profits, p.13 on rating quality).

Moody's therefore manages its ratings to reduce volatility, e.g. by avoiding rating reversals and by ignoring market-based information. Side-effects like a delayed response of ratings to new information are accepted by clients.[5]

From a market perspective, we could reformulate as follows: Does a major intermediary inhibit capital market efficiency, or does it ensure a sophisticated second-best solution? Is there a need for further regulatory intervention, or not?

Needless to say, the truth will lie somewhere in between, but where about we currently do not know. It could well be largely one, or the other. US legislators probably would not have requested a report on credit rating agencies if they did not regard the PROFIT FIRSTS hypothesis to be a possibility.[6] On the other hand, simulation studies (Löffler, 2005) tell us that CLIENTS FIRST is a consistent story: a policy of avoiding rating reversals *could* explain many of the empirical observations usually interpreted as rating deficiencies. One could turn to survey evidence to find out whether rating agencies satisfy client expectations or not. However, the evidence is disparate, or collected by Moody's and therefore associated with a doubt.

In this paper, I follow a novel approach to examine the viability of the CLIENTS FIRST hypothesis. I examine how Moody's stock price reacts to rating actions that could be indicative of inefficiencies, or a failure to achieve the goal of rating stability. I find that the market punishes rating reversals. On a day on which Moody's reverses a rating change made in the previous three months, Moody's stock price on average drops by more than 0.5%. The finding not only confirms Moody's view that the market actually wants to have stable ratings. Importantly, it supports CLIENTS FIRST because one does need to conjure up the image of a selfless agency whose overriding motive is to satisfy clients. Serving the market's interest is rational because it directly pays off to shareholders.

While the evidence increases our understanding of why rating agencies act in the way they do, it is not obvious that they should act in such way. Perhaps the market overemphasizes reversals and thus drives ratings agencies to a sup-optimal rating policy. Even if market participants benefit significantly from a low reversal rate, it does not follow that they should punish individual reversals to such an extent. One justification would be that reversals are indicative of low rating quality: When assigning ratings, agencies employ a long-term horizon

---

[5] For information on investment restrictions and rating triggers, see Cantor and Packer (1997) and Stumpp and Coppola (2002), respectively.

[6] See Securities and Exchange Commission (2003).

aiming to abstract from short-term, transitory fluctuations in credit quality;[7] high reversal rates then suggest that agencies followed transitory ups and downs without recognizing that they are transitory.

To judge whether Moody's reversal rate is indicative of low predictive ability, I determine the reversal rate of a hypothetical, crystal-ball rating system and find it to be significantly *higher* than the one of Moody's ratings. It therefore appears that the market erroneously judges individual reversals to be mistakes, similar to the way in which experimental subjects judge probabilistic forecast to be wrong if the state predicted to have a probability larger than 50% did not materialize (cf. Konold, 1989).

Related literature includes papers on reputational capital and the impact of forecast evaluation on the behavior of forecasters. The finding that the market value of a large financial firm can be affected by day-to-day business decisions is familiar from Nanda and Yun (1997), who show that lead-underwriters suffer losses when initial public offerings are overpriced. Laster, Bennett and Geoum (1999) suggest that macro forecasters bias their forecast in order to fare best according to the simplistic evaluation rule employed by some users.

Ambivalence in judging the performance of rating agencies is mirrored in the rating literature. The results of Blume, Lo and MacKinlay (1998) suggest that rating agencies changed their rating standards without informing the market, questioning the reliability of rating information. Jorion, Shi and Zhang (2005), however, extend the analysis and conclude that rating standards remained stable. Altman and Rijken (2006) show that statistical models beat ratings in terms of default prediction—but only for horizons of up to three years. Several papers (e.g. Holthausen and Leftwich, 1985) find that issuers' stock or bond prices respond to downgrades but not to upgrades, suggesting that some ratings are more informative than others. According to Jorion and Zhang (2005), this irregularity is due to a misspecification in the research design. Campbell and Taksler (2003) show that equity volatility explains as much cross-sectional variation in corporate bond spreads as do ratings. While this questions the relevance of rating information, Kisgen (2005) shows that ratings are relatively important for firms' capital structure decisions. In theoretical models, Ramakrishnan and Thakor (1984) and Lizzeri (1999) show that the signals provided by a financial intermediary can be uninformative; Boot, Milbourn and Schmeits (2004), however, show that agencies can play an

[7] According to Moody's, a rating is meant to provide "a signal that looks through cycles and immaterial events and focuses on long-term creditworthiness" (Mahoney, 2002, p.3).

important role by serving as a coordinating mechanism.

The remainder of the paper is structured as follows. In section 2, I describe the data and the definition of variables. Section 3 presents the findings on Moody's stock price reaction to rating actions. Section 4 puts the observed reversal rate in perspective, and section 5 concludes.

2 Data and definition of variables

Moody's was first traded on the NYSE on 06/19/1998. I collect daily returns and market values from Datastream. To describe Moody's returns, I consider the value-weighted CRSP market portfolio, industry and size portfolios as well as the Fama/French factors SMB (small minus big) and HML (high book-to market minus low book-to-market). Returns are in excess of the risk-free rate of return.[8] Information on ratings consists of daily information on Moody's issuer-level ratings of corporate and sovereign bond issuers; the available data ends in May 2005. To derive expected reversal rates under a perfect-foresight rating system, I will use Moody's KMV EDFs. EDFs estimate a firm's one-year probability of default through an extension of Merton's (1974) structural approach to corporate default.[9]

Moody's highlights the following criteria for judging the performance of its rating system (cf. Cantor and Mann, 2003): accuracy as measured by the ability to predict defaults; the average stability of ratings; the frequency of large rating changes; the frequency of rating reversals. An individual rating event can directly be judged to be large or to represent a reversal, and so we can test whether it has an impact on stock prices. By contrast, there is no meaningful way of classifying an individual rating action as inaccurate or unstable. Nevertheless, I propose to examine cases of defaults of issuers that were rated investment-grade (rating Baa3 or better) shortly before default. This is inspired by anecdotal cases in which investment-grade defaults (e.g. Enron) sparked criticism of rating agencies. How aggregate stability could be measured through individual events is not obvious.

Based on these observations, I define the following dummy variables:

$REVERSAL_t$ one if, on day t, Moody's upgrades an issuer that was downgraded (or downgrades one that was upgraded) in the previous three months, zero else.

[8] I am grateful to Ken French for making this data available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[9] I am grateful to Moody's Investors Service and Moody's KMV for providing the data.

$DRIFT_t$	one if, on day *t*, Moody's upgrades an issuer that was upgraded in the previous three months (or a downgrade preceded by a downgrade), zero else.
$LARGE_t$	one if, on day *t*, Moody's changes a rating by three notches or more (e.g. from Aa1 to A1), zero else.
$DEFAULT_t$	one if, on day *t*, an issuer defaults that was rated investment-grade twelve months before default, zero else.

DRIFT just complements REVERSAL. The definition of variables is based on Moody's publications. Fons (2002) mentions the three-month period for reversals; Cantor and Mann (2003) use three notches to identify large rating changes. The twelve-month horizon used for DEFAULT is very frequent in credit risk management. As described in the next section, alternative definitions do not materially affect the results. If any of the events captured in the dummy variables happened on a day that is not a trading day, I attribute the event to the next trading day after the event.

Over the 1728 days with data on both Moody's returns and ratings, there are 78 days with REVERSAL=1. The event day counts for the other dummy variables are: 470 (DRIFT), 696 (LARGE), 31 (DEFAULT). Figure 1 shows the distribution of the events over time. Most defaults occur in the years 2001 and 2002; the other three events are distributed fairly evenly.

The selection of an abnormal return model for Moody's stock return is based on the factor model regressions in Table 1. As factors explaining Moody's returns I consider the market portfolio; industry portfolios for business services (BUS_SER) as well as for banks (BANKS); portfolio returns of size deciles to which Moody's belongs according to monthly size breakpoints; SMB and HML returns. Since Moody's exhibits several extreme returns over the sample period (the minimum is -16.9%, the maximum is 10.8%), I also run the regressions with winsorized Moody's returns. Specifically, I pull extreme values to the 1% and 99% quantiles, respectively. Table 1 shows that the two industry factors add significant explanatory power to the regression with just the market portfolio; the other three factors do not. Results do not depend on winsorization. Based on these results, I will use the model containing the market return and the two industry portfolios.

3 Market reaction to rating actions and defaults

To study stock price reactions to rating actions, I employ multivariate regressions in which events are captured through dummy variables.[10] I start by estimating

$$\begin{aligned} MOODYS_t = \alpha + \beta_1 MARKET_t + \beta_2 BUS\_SER_t + \beta_3 BANKS_t \\ + \beta_4 REVERSAL_t + \beta_5 DRIFT_t + \beta_6 LARGE_t + \beta_7 DEFAULT_t + u_t. \end{aligned} \tag{1}$$

The coefficients are estimated with ordinary least squares (OLS), standard errors with the White-correction for heteroskedasticity. To assess robustness, I also run the regression with winsorized Moody's returns.

Results are reported in Table 2. Among the four variables capturing rating actions, the only significant one is REVERSAL. Its estimated coefficient is –0.60 (–0.56 with winsorization), statistically significant at a level better than 0.1%. Since the returns are entered in percent, the coefficient says that Moody's stock price drops by more than half percent on days with a rating reversal. With 78 occurrences, reversals therefore cause economically significant damage to Moody's stock price.

To check whether the effects of rating actions are different for actions that involve a crossing of the investment-grade barrier, I augment equation (1) by three further dummy variables. Let IG denote an investment-grade rating, SG a speculative-grade rating. REVERSAL-IG is unity only for rating reversals of the type [IG→SG→IG], or [SG→IG→SG]. DRIFT-IG is unity only for sequences of the type [IG→IG→SG] or [SG→SG→IG]. LARGE→IG is unity only for large rating changes of the type [IG→SG] or [SG→IG]. There is no differentiation of the DEFAULT variable because it is restricted to investment-grade defaults by definition. The results, which are also reported in Table 2, indicate that the investment grade boundary does not create significant special effects; this holds both for the individual variables and for a joint test of the significance of the three dummy variables (p-value = 0.55; 0.48 for winsorized Moody's returns).

To check whether results are robust to the definition of variables, I apply a 6-month horizon to define reversals and trends (rather than a 3-month horizon) as well as investment-grade defaults (instead of 12 months); large rating changes are taken to be those with a change of 6 or more notches (instead of 3). Results do not change materially. The coefficient on reversals falls to around –0.25 but is still significant on the 5% level; the coefficients of the other variables remain insignificant.

---

[10] Binder (1985) discusses the use of multivariate regressions for event studies. A recent application is Bittlingmayer and Hazlett (2000).

Some modifications are investigated for reversals only: (i) I add a dummy variable which is one if there is more than one reversal per day—it does not add explanatory power; (ii) I check whether reversals of the type [downgrade→upgrade] have other effects than reversals of the type [downgrade→upgrade]—they do not; (iii) I use a 12-month horizon to define reversals—the coefficient goes down to around –0.14 (significant on a level of around 15%). One would expect that the effect of reversals goes down as the time horizon increases. The further apart two rating changes are, the less likely is that the second one will be judged in relation to the first.

So far, the analysis was not designed to capture any delayed price reactions or anticipatory effects. One might think of conducting an event study around the rating action dates. However, the standard event study procedure of eliminating overlapping event windows would result in a drastic reduction of observations. Even if one restricts events to reversals and investment-grade defaults, the number of non-confounding events is only 21 (reversals) and 20 (defaults) when examining 21-day windows centered on the event. Overlapping estimation windows for abnormal return model parameters would have to be taken into account as well. Within a multivariate regression, overlapping windows also reduce the precision of the estimation, but there is no need to drop observations. By construction, regression coefficients are estimates of partial effects.

Since large rating changes and trends were not significant in the previous regression, I restrict the ensuing analysis to reversals and defaults. Defaults were not significant either, but one can surmise that anticipation effects are particularly strong for defaults because they are typically preceded by a noticeable deterioration of ratings.[11] To separate anticipation from delayed reaction, I define two new dummy variables for each event type. $REVERSAL_{-10}$ ($REVERSAL_{+10}$) equals one during the ten-day period before (after) any of the 78 days with reversals. $DEFAULT_{-10}$ and $DEFAULT_{+10}$ are defined accordingly.

Table 3 shows the results when these four variables are added to a regression that contains the three factors and the two dummies REVERSAL and DEFAULT. Individually, there is no significant abnormal stock price movement during the ten days before or after any of the two event types. The same result obtains from joint tests of each pair, ($REVERSAL_{-10}$, $REVERSAL_{+10}$) and ($DEFAULT_{-10}$, $DEFAULT_{+10}$), respectively. In a variation of the

[11] Out of the 31 defaulters that have an investment-grade rating twelve months prior to default, only six have an investment-grade rating one month before default.

regression (also reported in Table 3), I restrict the analysis to dummies related to reversals. The results do not change materially. Nor do they change if the periods before and after events are taken to comprise five days instead of ten (not reported).

The results therefore confirm the previous ones. Rating reversals hurt Moody's stock price while other conspicuous rating actions do not. The effects appear to be confined to one day. As there is no significant opposite stock price movement after the event, the effects do not appear to be transitory.

One may object the interpretation on the grounds that the average abnormal return over the 21-day window may not be significant either. Indeed it is not, as a joint test of the hypothesis[12]

$$H0: \quad (REVERSAL + 10 \times REVERSAL_{-10} + 10 \times REVERSAL_{+10}) = 0$$

shows. Its p-value is around 0.59 for the four specifications (winsorization or not, DEFAULT dummies included or not). Since the idiosyncratic risk of Moody's is fairly large (29% p.a.), however, this is what we would expect to see even if REVERSALS have permanent effects. If the expected abnormal return on a reversal day is –0.60%, and zero on all others, the expected average abnormal return over a 21-day event window around the reversal would be –0.60% / 21 = –0.03%. Let the standard error of the daily abnormal returns be $129\%/260^{0.5}$=1.8%, and assume that the 78 reversal events are independent. The average abnormal return over the 21-day windows then can be expected to have a standard error of 1.8% / $(21\times78)^{0.5}$ =0.04%, which is larger than the expected abnormal return of 0.3%.

How do the empirical findings square with official Moody's statements? In the process leading to Basel II and in the aftermath of the Enron and Worldcom collapses, Moody's has published several reports to clarify its rating policy. They serve as the basis for the following discussion. The avoidance of rating reversals is described as an important element of Moody's rating policy. It is one instrument of lowering the volatility of ratings; the other being the relative nature of the rating system and the through-the-cycle approach (e.g. Cantor, 2001). However, reversal avoidance is not only described as a useful instrument. Moody's mentions specifically that the market might interpret a single reversal as an indication of low rating quality:

---

[12] The coefficients of $REVERSAL_{-10}$ and $REVERSAL_{+10}$ are multiplied by ten because they are estimates of the abnormal *daily* return during the respective ten-day interval.

> "Most market participants would argue (rightly or wrongly) that a rating reversal—an upgrade followed by a downgrade, or a downgrade followed by an upgrade—over, for example, a three month period—would be evidence of a rating `mistake'" (Fons, 2002, p.6).

This contrasts with the views on other conspicuous rating actions. Rating drift is described as "a natural consequence of our rating system-management practices" (Fons, 2002, p.12). Apparently, this side-effect of the rating policy is accepted by the market:

> "These practices impart a deliberate, and often serial, behavior to rating changes, and they sometimes limit the information content of individual rating changes. Our discussions with users of ratings, however, indicate that, despite criticism about rating timeliness, investors and other users prefer the system as it currently operates" (Fons, 2002, p. 7).

Large rating changes are monitored by Moody's to an extent that is similar to reversals. However, Moody's emphasizes that they may be good reasons for large rating changes:

> "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" (Cantor Mann, 2003, p. 16).

Overall, Moody's seems to regard large rating changes as less problematic than reversals. If large rating changes were indeed a problem, Moody's could split them into a sequence of small rating changes, exploiting the fact that such sequences are not regarded as problematic by the market.

Similarly, default occurrences among investment grade issuers are not described as a problem but actively defended:

> "Default on a bond rated Aaa upon issuance does not prove that the original rating was wrong, any more than punctual payment of a bond initially rated Caa proves that rating judgment wrong. Such evidence is anecdotal at best" (Cantor, 2001, p. 176).

The investment grade barrier, finally, does not receive special attention in the publications. The justification (given in a footnote) is that this could lead to situations in which investment grade issuers are not downgraded even though they should:

> "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" (Cantor and Mann, 2003, p. 16).

Moody's statements are therefore consistent with the findings from Tables 2 and 3. Stock price reactions suggest that the market wants to see few reversals; and Moody's claims that this is an important part of its rating policy. Rating drift does not damage the stock price because it is a side effect that is accepted. Large rating changes and default occurrences, which do not matter for the stock price, are tracked by Moody's but viewed as less damaging than reversals. The investment grade barrier, finally, does not affect stock price reactions, but it does not receive special attention by Moody's either. Put together, the empirical evidence corroborates Moody's view that their rating policy is tailored to market wishes.

There is an open end to this interpretation, however. Even if the market values low reversal rates highly, it does not necessarily follow that individual reversals should be punished. Like defaults and large rating changes, reversals are bound to occur. In the next section, we therefore pick up the question raised by Fons (2002, p.6) whether individual reversals are "evidence of a rating 'mistake'" or not.

4 Which reversal rate should one expect?

One reason why individual reversals drive stock prices down could be that the market considers Moody's reversal rate to be too high. This raises the question which reversal rate one would expect from a perfect rating agency. Its magnitude will depend on several factors, including the time series properties of default risk, the forecast horizon implicit in ratings, the forecasting ability of the rating agency, and the costs that reversals entail for issuers and investors. Since costs of reversals are difficult to quantify, I concentrate on the first three factors.

For simplicity, assume that ratings map the average credit quality expected to prevail over the next *T* years into discrete grades; [13] one can think of credit quality as a credit score or a default probability transformed with an inverse cumulative distribution function. If credit quality

[13] Cf. "Ratings are forward looking in that the ordering is designed to hold across multiple horizons" (Moody's Investors Service, 2006, p.1).

follows a random walk, the forecast is today's credit quality regardless of the horizon T; with a random walk, credit quality can diverge or remain fairly stable, but it can also move up and down and up again, producing reversals in the rating system. If credit quality has a mean-reverting component, ups and downs are more likely to follow each other; due to predictability, however, the effect on the reversal rate is ambiguous. The larger the forecast horizon T, for example, the less responsive will the rating be to short-term shocks to credit quality because they will be expected to fade out over time. Related to this, the reversal rate will depend on the agency's ability to separate transitory shocks from permanent ones.

In addition, note that the discrete nature of the rating system will increase the probability of reversals. A rating change is triggered only if the average expected credit quality exceeds a critical level. With standard, bell-shaped distribution functions, it is more likely that the critical level is exceeded by a small rather than by a large amount. This makes it more likely that credit quality moves back across the threshold just past rather than exceed the next threshold.

Since there are so many effects at work, I refrain from deriving optimal reversal rates based on a parametric modeling of the rating process. Instead, I empirically derive properties of a hypothetical rating system that is perfect in the sense that it is built on perfect foresight. To determine this system I take actual one-year estimates of default probability, use statistical methods to derive long-term trends, and map these estimates into a rating system similar to the one used by Moody's. The short-term default probability estimates used for the analysis are Moody's KMV EDFs. The statistical method applied is the Hodrick-Prescott filter (Hodrick and Prescott, 1997). With this method, the series of logarithmic EDFS for issuer i is split into a trend-component *HPTREND* and a cyclical component *HPCYCLE*:

$$\ln(EDF_{it}) = HPTREND_{it} + HPCYCLE_{it}$$

The split is determined separately for each issuer i by minimizing:

$$\sum_{t=\tau(i)}^{T(i)} (HPCYCLE_{it})^2 + \lambda \sum_{t=\tau(i)+2}^{T(i)} \left[(HPTREND_{it} - HPTREND_{i,t-1}) - (HPTREND_{i,t-1} - HPTREND_{i,t-2}) \right]^2 \tag{2}$$

where $\tau(i)$ denotes the starting date of series i, $T(i)$ its end date. λ is a smoothing constant. The larger λ is, the smoother is the estimated trend. Results are reported for two alternative choices, λ=10,000 and λ=500,000. They encompass λ–values typically suggested in the literature.[14] Note that the Hodrick-Prescott Filter uses the entire information from t= $\tau(i),\ldots,T(i)$, to determine the trend at a particular date t. It is therefore built on perfect foresight, providing an appropriate benchmark for the forward-looking rating concept of Moody's.

Since the analysis does not require Moody's stock price, I use the entire available data which comprises the years 1982 to 2005. It includes matched monthly data on EDFs and Moody's ratings for more than 4021 corporate bond issuers. Several data requirements reduce it to a sample of 2,767 series: As default is a special situation in which borrowers leave the cycle I split the time series on the occasion of default. Observations before the default month and after emergence from default are treated as a separate series; observations in between are discarded. Missing observations also lead to a split-up in separate series. I disregard series shorter than 48 months in order to avoid situations in which the Hodrick-Prescott filter yields implausible results because of a lack of fluctuations.

With the data produced by this selection process, I estimate equation (2) for each series. To illustrate the characteristics of the filtered EDF trends, Figure 2 shows the time series of EDFs and the two Hodrick-Prescott trends HPTREND for one company from the sample. Before the estimated trends are mapped into discrete rating grades, I discard both the first and the last 18 months of each series. Since there is nothing to look into the future at the end of the series, the Hodrick-Prescott filter loses its forward-looking ability towards the end; analogously, backward-smoothing is less pronounced at the start. Requiring 18 months of future or past observations is meant to ensure a smooth, information-rich trend that serves as a good benchmark for rating agencies. The values are mapped into 19 rating grades using the idealized loss rates reported in Yoshizawa (2003); mapping is also done for the unfiltered EDFs.

Table 4 reports statistics on reversal rates. To compute the reversal rate, the number of all reversals (within three months) is divided by the overall number of rating changes. The reversal/drift statistics divides the number of reversals by the number of drifts (defined as in section 2). First note that the rating system based on raw EDFs shows high reversal frequencies. Adding forecasting ability through filtering reduces the reversals dramatically. The reversal rate, for example, goes from 46% to 3.8% and 2.7% for λ set to 10,000 and 500,000, respectively. But even the heavenly smoothed trend with λ=500,000 shows much

[14] λ=14,400 is a common choice for monthly data. Pedersen (2001) recommends λ >100,000..

more reversals than actual Moody's ratings, which have a reversal rate of less than 1%.

We thus observe that a hypothetical rating system with a significant amount of perfect foresight exhibits much higher rating reversals than Moody's. Of course, we cannot rule out that the rating system studied here misses important features of Moody's rating system. If we nevertheless conclude from the results that Moody's reversal rate is relatively low, there seem to be two reasons why the market could demand such a low reversal rate. One is deadweight costs of rating reversals that are borne by investors and issuers. Ratings can be reversed quickly, but—together with investment restrictions or rating triggers—a rating change can induce irreversible costs. Many of those institutional features are tied to the investment-grade boundary,[15] which is why one would expect the costs to be particularly high for reversals crossing the investment-grade boundary. Since the market reaction is not significantly higher if the investment-grade boundary is involved (see Table 2), deadweight costs are unlikely to be the correct explanation.[16]

Another possible reason hinted at in Fons (2002) is misjudgment on the side of the market. Despite the fact that reversals occur with a rate that is consistent with perfect forecasting, the market might consider individual occurrences as indicative of rating mistakes. Support for this view comes from experiments. Evaluations of probability forecasts have been shown to be biased in the sense that subjects tend to evaluate forecasts as either right or wrong based on single outcomes. For example, if no rain fell on a day for which a 70% chance of rain was predicted, a large number of subjects judged the forecast to be wrong based on just this observation (Konold, 1989).

5 Concluding remarks

We have started with two terse hypotheses: does the rating agency Moody's cite market expectations for stable ratings only to cover a high-profit policy of monitoring ratings only infrequently? Or is its policy of reducing rating volatility a genuine response to market demands, which entails some compromises—like rating drift—that are accepted by the market?

[15] Cf. Stumpp and Coppola (2002) and Cantor and Packer (1997).

[16] One might object that the estimated coefficient for investment grade reversals (REVERSAL-IG) in Table 2 is negative, and that the power of rejecting the null of a zero coefficient should be low because there is only a small number of such reversals (the number is four). However, if reversals are defined using a horizon of six or twelve months, the coefficient on the dummy REVERSAL-IG turns positive as the number of occurrences increases to nine or 21, respectively.

The fact that Moody's stock price reacts negatively to reversals provides support to the second hypothesis because avoiding reversals is a cornerstone of the volatility reducing policy. Not only does it confirm Moody's claim that the market actually wants such a policy. Since Moody's market value is affected by reversals, it is in the interest of Moody's shareholders to pursue it.

The evidence is comforting in the sense that an important financial intermediary is driven by reputational concerns despite the fact that it enjoys a comfortable position in an oligopoly. Possibly disquieting, however, is that the market appears to evaluate ratings in a way that is at odds with a rational evaluation of forecasts. Specifically, the market seems to interpret single rating actions as evidence of rating mistakes even though they could well be consistent with a perfect rating system. In the end, market discipline could lead to a rating policy that is sub-optimal.

References

Altman, Edward, and Herbert Rijken. (2006). "A point-in-time perspective on through-the-cycle ratings." *Financial Analysts Journal* 62, 54-70.

Baker, H. Kent, and Sattar A. Mansi. (2002). "Assessing credit rating agencies by bond issuers and institutional investors." *Journal of Business Finance & Accounting* 29, 1367-1368.

Binder, John J. (1985). "On the use of the multivariate regression model in event studies." *Journal of Accounting Research* 23, 370-383.

Bittlingmayer, George, and Thomas W. Hazlett. (2000). "DOS Kapital: Has antitrust action against Microsoft created value in the computer industry." *Journal of Financial Economics* 55, 329-359.

Blume, Marshall E., Felix Kim, and Craig A. MacKinlay. (1998). "The declining credit quality of US corporate debt: Myth or reality?" *Journal of Finance* 53, 1389–2013.

Boot, Arnoud W. A., Todd T. Milbourn, and Anjolein Schmeits. (2006). "Credit ratings as coordination mechanisms." *Review of Financial Studies* 19, 81-118.

Campbell, John Y. and Glen B. Taksler. (2003). "Equity Volatility and Corporate Bond Yields." *Journal of Finance* 58, 2321-49.

Cantor, Richard. (2001). "Moody's investors service response to the consultative paper issued by the Basel Committee on Banking Supervision and its implications for the rating agency industry." *Journal of Banking and Finance* 25, 171-186.

Cantor, Richard, and Christopher Mann. (2003). Measuring the performance of corporate bond ratings. Special comment, *Moody's Investors Service*.

Cantor, Richard, and Frank Packer. (1997). "Differences of opinion and selection bias in the credit rating industry." *Journal of Banking and Finance* 21, 1395–1417.

Daníelsson, Jon, et al. (2001). "An academic response to Basel II." Working paper, London School of Economics.

Fons, Jeremy. (2002). "Understanding Moody's corporate bond ratings and rating process." Special Comment, *Moody's Investors Service*.

Hodrick, Robert J., and Edward C. Prescott. (1997). "Postwar U.S. business cycles: an empirical investigation." *Journal of Money, Credit and Banking* 29, 1-16.

Holthausen, Robert W., and Richard W. Leftwich. (1986). "The effects of bond rating changes on common stock prices." *Journal of Financial Economic*s 17, 57-89.

Jorion, Philippe, Charles Shi, and Sanjian Zhang. (2005). "Tightening Credit Standards: Fact of Fiction?" Working paper, University of California at Irvine.

Jorion, Philippe, and Gaiyan Zhang. (2005). "Non-linear effects of bond ratings changes." Working paper, University of California at Irvine.

Kisgen, Darren. (2005). "Do firms target credit ratings or leverage levels?" Working Paper, Boston College.

Konold, Clifford. (1989). "Informal Conceptions of Probabilit." Cognition and Instruction 6, 59-98.

Laster, David, Paul Bennett, and In Sun Geoum. (1999). "Rational bias in macroeconomic forecasts." *Quarterly Journal of Economics* 114, 293-318.

Lizzeri, Alessandro. (1999). "Information revelation and certification intermediaries." *Rand Journal of Economics* 30, 214-231.

Löffler, Gunter. (2005). "Avoiding the rating bounce: Why rating agencies are slow to react to new information." *Journal of Economic Behavior and Organization* 56, 365-381.

Mahoney, Christopher (2002). "The bond rating process: A progress report." Special comment, Moody's Investors Service.

Merton, Robert C. (1974). "On the pricing of corporate debt: The risk structure of interest rates." *Journal of Finance* 29, 449-470.

Moody's Investors Service. (2006). "Moody's rating system in brief." Special comment.

Nanda, Vikram, and Youngkeol Yun. (1997). "Reputation and financial intermediation: An empirical investigation of the impact of IPO mispricing on underwriter market value." *Journal of Financial Intermediation* 6, 39–63.

Pedersen, Torben M. (2001). "The Hodrick-Prescott filter, the Slutzky effect, and the distortionary effect of filters." *Journal of Economic Dynamics and Control* 25, 1081-1101.

Ramakrishnan, Ram T.S., and Anjan V. Thakor. (1984). "Information reliability and a theory of financial intermediation." *Review of Economic Studies* 51, 415-432.

Securities and Exchange Commission. (2003). "Report on the role and function of credit rating agencies in the operation of the securities markets."

Stumpp, Pamela M., and Monica M. Coppola. (2002). Moody's analysis of US corporate rating triggers heightens need for increased disclosure. Special comment, *Moody's Investors Service*.

Yoshizawa, Yuri. (2003). "Moody's approach to rating synthetic CDOs." Special comment, *Moody's Investors Service*.

Table 1: Choosing the abnormal return model for Moody's

All returns are excess returns over the risk-free rate. MARKET is the value-weighted CRSP market index. BUS_SERV is the return on the business service industry portfolio minus MARKET; BANKS is defined similar with the banking industry portfolio. SIZE is built with portfolio returns of size deciles to which Moody's belongs according to monthly size breakpoints; SMB and HML are the Fama/French portfolios *small minus large* and *high book-to-market minus low book-to-market.* T-statistics (in parentheses) are estimated with the White-correction for heteroskedasticity.

	Dependent variable					
	Moody's return (in %)			Winsorized Moody's return (in %)		
CONSTANT	0.079	0.079	0.080	0.078	0.077	0.077
	(1.82)	(1.82)	(1.83)	(1.95)	(1.94)	(1.93)
MARKET	0.579	0.581	0.611	0.559	0.561	0.600
	(13.87)	(13.86)	(9.70)	(15.40)	(15.40)	(11.13)
BUS_SERV		0.291	0.318		0.276	0.299
		(3.50)	(3.64)		(3.69)	(3.81)
BANKS		0.164	0.106		0.169	0.104
		(2.96)	(1.71)		(3.27)	(1.82)
SIZE			0.003			0.051
			(0.02)			(0.39)
SMB			-0.137			-0.144
			(-1.53)			(-1.75)
HML			0.085			0.104
			(0.78)			(1.10)
p(SIZE=SMB=HML=0)			0.188			0.073
Adj. R^2	0.138	0.152	0.154	0.151	0.168	0.170
Durbin Watson	2.104	2.117	2.111	2.118	2.130	2.122
Observations	1728	1728	1728	1728	1728	1728

Table 2: Moody's stock returns on days with conspicuous rating actions

All returns are excess returns over the risk-free rate. MARKET is the value-weighted CRSP market index. BUS_SERV is the return on the Business service industry portfolio minus MARKET; BANKS is defined similar with the banking industry portfolio. REVERSAL is one on days with a rating reversal, DRIFT is one on days where a rating change follows a previous one in the same direction; LARGE is one on days with a rating change of three notches or more. DEFAULT is one on days with a default by an issuer rated investment grade twelve months before default. *VARIABLE* -IG equals *VARIABLE* if the rating sequence contains a crossing of the investment grade boundary (Baa1), zero else. T-statistics (in parentheses) are estimated with the White-correction for heteroskedasticity.

	Dependent variable			
	Moody's return (in %)		Winsorized Moody's return (in %)	
CONSTANT	0.078	0.077	0.080	0.079
	(1.27)	(1.25)	(1.46)	(1.44)
MARKET	0.581	0.580	0.561	0.560
	(13.89)	(13.88)	(15.43)	(15.38)
BUS_SERV	0.290	0.283	0.274	0.270
	(3.50)	(3.41)	(3.69)	(3.61)
BANKS	0.156	0.155	0.161	0.161
	(2.82)	(2.81)	(3.13)	(3.12)
REVERSAL	-0.604	-0.553	-0.562	-0.549
	(-3.77)	(-3.68)	(-3.84)	(-3.70)
DRIFT	-0.040	-0.014	-0.032	-0.004
	(-0.43)	(-0.14)	(-0.36)	(-0.04)
LARGE	0.094	0.130	0.075	0.111
	(1.06)	(1.30)	(0.92)	(1.21)
DEFAULT	0.053	0.042	0.061	0.054
	(0.21)	(0.17)	(0.25)	(0.22)
REVERSAL-IG		-1.025		-0.293
		(-0.76)		(-0.38)
TREND-IG		-0.169		-0.169
		(-0.85)		(-0.86)
LARGE-IG		-0.112		-0.123
		(-0.71)		(-0.85)
Adj. R^2	0.155	0.155	0.170	0.170
Durbin Watson	2.103	2.097	2.117	2.113
Observations	1728	1728	1728	1728

Table 3: A closer look: Moody's stock returns around reversals and defaults

All returns are excess returns over the risk-free rate. MARKET is the value-weighted CRSP market index. BUS_SERV is the return on the Business service industry portfolio minus MARKET; BANKS is defined similar with the banking industry portfolio. REVERSAL is one on days with a rating reversal. DEFAULT is one on days with a default by an issuer rated investment grade twelve months before default. $VARIABLE_{-10}$ equals one on the ten days before VARIABLE equals one, zero else; $VARIABLE_{+10}$ equals one on the ten days after VARIABLE equals one. T-statistics (in parentheses) are estimated with the White-correction for heteroskedasticity.

	Dependent variable			
	Moody's return (in %)		Winsorized Moody's return (in %)	
CONSTANT	0.054	0.056	0.055	0.058
	(0.74)	(0.80)	(0.87)	(0.93)
MARKET	0.581	0.581	0.561	0.561
	(13.88)	(13.86)	(15.43)	(15.41)
BUS_SERV	0.286	0.286	0.271	0.271
	(3.43)	(3.43)	(3.62)	(3.63)
BANKS	0.155	0.156	0.161	0.161
	(2.82)	(2.83)	(3.12)	(3.14)
REVERSAL	-0.608	-0.606	-0.566	-0.563
	(-3.79)	(-3.79)	(-3.87)	(-3.86)
DEFAULT	0.069		0.073	
	(0.27)		(0.29)	
$REVERSAL_{-10}$	0.068	0.068	0.061	0.060
	(0.76)	(0.76)	(0.74)	(0.74)
$REVERSAL_{+10}$	0.068	0.068	0.061	0.061
	(0.76)	(0.77)	(0.75)	(0.76)
$DEFAULT_{-10}$	-0.015		-0.023	
	(-0.11)		(-0.19)	
$DEFAULT_{+10}$	0.023		0.029	
	(0.19)		(0.26)	
Adj. R^2	0.154	0.156	0.170	0.171
Durbin Watson	2.103	2.097	2.117	2.113
Observations	1728	1728	1728	1728

Table 4: Empirical reversal frequency for Moody's ratings and hypothetical rating systems based on EDFs

EDFs and Hodrick-Prescott filtered trends in EDFs (HP, with smoothing parameter λ) are mapped into a 19-grade rating system using the probabilities of default given in Yoshizawa (2003). The reversal rate is the share of all reversals (rating change is reversed within three months) among all rating changes. The reversal/drift measure is the number of reversals divided by the number of rating drifts (two rating changes in the same direction within three months).

		HP-filtered EDF Ratings		
	EDF Ratings	λ=10,000	λ=500,000	Moody's
reversal rate	46.63%	3.85%	2.66%	0.94%
Reversals/drifts	3.13	1.61	1.00	0.16

Figure 1: Distribution of rating events over the observation period

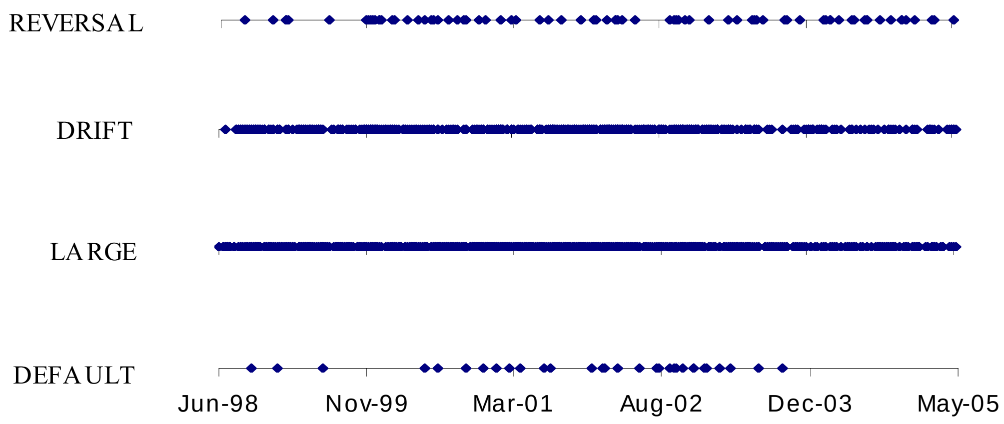

Figure 2: EDF and EDF trends over time for one company

HP denotes the Hodrick-Prescott filter with smoothing parameter λ. The ordinate is scaled logarithmically.