Forthcoming *The Journal of Financial Services Research.*

The final publication is available at springerlink.com

Can Market Discipline Work in the Case of Rating Agencies?

Some Lessons from Moody's Stock Price.

Gunter Löffler*

December 2011

Abstract

This paper examines whether the stock price of the rating agency Moody's reacts negatively to rating actions that could indicate low rating quality. The reaction to rating reversals, which Moody's describes as particularly damaging to investors, is economically significant. It suggests that market discipline has the potential to influence agency behavior. On the other hand, defaults of highly rated issuers do not consistently impact Moody's stock price. The focus on reversals and the neglect of default events are consistent with either collusion or with misconceptions of how rating quality should be evaluated. Both interpretations question whether market discipline can be sufficient to ensure a socially optimal rating policy within the current environment.

JEL classification: G2

Key words: rating agencies, rating quality, oligopoly, market discipline, reputational capital

* Gunter Löffler, Institute of Finance, University of Ulm, Helmholtzstrasse 18, 89069 Ulm, Germany. Phone: ++49-731-5023597, Fax: ++49-731-5023950.
E-mail: gunter.loeffler@uni-ulm.de

I am very grateful to an anonymous reviewer and the editor for many helpful comments, as well as to Moody's Investors Service for providing the data.

1 Introduction

It is widely believed that rating agencies contributed to the subprime crisis. Conflicts of interest, computational flaws and the neglect of risk factors were among the reasons why agencies issued overly optimistic ratings for structured finance products (see, e.g., Crouhy, Jarrow and Turnbull, 2008, and Coval, Jurek, and Stafford, 2009). Doubts about the reliability of ratings have sparked a discussion of how to increase rating quality through more and better regulation. An argument calling for a careful use of government intervention is that market discipline could effectively complement regulatory measures. Anecdotal evidence suggests that rating agencies can indeed lose significant reputational capital if the market receives negative information concerning the quality of their ratings. On May 21st, 2008, Moody's Investors Service, one of the leading rating agencies, stated that it was investigating errors in the rating process that could have led to structured finance securities wrongly receiving AAA ratings.[1] On the same day Moody's Corp., listed on the NYSE, lost 15%, or USD 1.7 billion in market value.

Apparently, reputational concerns were not strong enough to ensure a satisfactory quality of structured finance ratings in the years before the subprime crisis. This should not be taken as conclusive evidence against the effectiveness of market discipline, however. Although structured finance securities have been rated for more than 20 years, the subprime crisis was the first crisis to hit this market sector. From 1981-2006, the average annual default rate of structured finance transactions with investment-grade ratings was a mere 0.02%, with a maximum of 0.15% in 1998. By contrast, in 2008 the default rate jumped to 1.03%.[2] If market discipline fails before the first real test of a market, the conclusion that it cannot work later on is likely to be premature.

---

[1] News release "Moody's Confirms External Review of European CPDO Rating Process", available on http://ir.moodys.com/RELEASEDETAIL.cfm?releaseid=311726

[2] cf. Erturk (2009). In the mature corporate bond rating sector, by contrast, the 2008 investment-grade default rate of 0.41% compares to a 1981-2006 average of 0.10%, and does not exceed the previous maximum default rate in 2002 (cf. Vazza, 2009).

To assess the potential effectiveness of market discipline in the rating industry, it seems more instructive to examine the traditional corporate rating sector. This sector has a long history and has witnessed intense controversy over rating quality following the Enron bankruptcy. If market discipline can work, one should be able to detect its forces there. Note that the corporate bond segment is also part of the current debate as proposals on changes in rating regulation are not confined to structured finance ratings (e.g. European Commission, 2010).

To measure the effects of market discipline, I examine how the stock price of Moody's, one of the three leading rating agencies, reacts to rating events that might indicate low rating quality. To identify such events, I make use of surveys of investor preferences conducted by Moody's. The three research questions that I address are as follows:

1) Are the justifications that Moody's gives for its own rating policy consistent with the empirical evidence?
2) Is there evidence that market discipline can impose significant cost on a major rating agency?
3) If the answer to 2) is in the positive, does the evidence suggest that market discipline helps to ensure a satisfactory level of rating quality?

Answering the questions is important for assessing the potential role of market discipline and the market power of rating agencies. The answers are from obvious. Preferences of rating users are hard to measure, which makes it difficult to assess the accuracy of the account that agencies give of them. The oligopolistic structure[3] of the rating industry could lead to a situation in which rating agencies can afford to offer low quality services without being punished by the market. Finally, alignment of interests could lead to outcomes that are sub-optimal from a societal perspective even if the stock market can impose costs on agencies.

[3] Moody's and S&P have an estimated market share of 80% measured by revenue, cf. US Senate Report 109-326, 2006.

Starting with question 1), the answer is in the positive. According to Moody's, the rating event that investors deem most critical is a rating reversal, i.e. the downgrade of an issuer that was recently upgraded (or, alternatively, an upgrade following a recent downgrade). Reversals feature prominently because many issuers and investors prefer stable ratings. Rating triggers in bond contracts as well as rating-based portfolio governance rules lead to high transaction cost if ratings change frequently.[4] Even without consideration of such consequences, reversals can indicate rating errors. When assigning ratings, agencies employ a long-term horizon and aim to abstract from short-term, transitory fluctuations in credit quality.[5] High reversal rates suggest that agencies followed transitory ups and downs without recognizing their transitory nature. Empirical analysis demonstrates that if Moody's reverses a rating change made within the preceding three months, its stock price drops by an average of 0.41%. Over the 12-year period analyzed in this paper, such losses cumulate to a 31% reduction in market value.

The size of these effects suggests that market discipline has the potential to influence the behavior of rating agencies. The explanation is unlikely to be that low rating quality of Moody's strengthens its main competitor, Standard & Poor's. By market convention, most bond issues are rated by both rating agencies, which effectively turns a duopoly into a monopoly. Rather, low rating quality can lead to losses through the strengthening of other rating agencies or through an increase in the perceived likelihood of a regulatory crackdown. An excerpt from Moody's annual report 2003 (p. 7) illustrates the presence of this threat:

> "Other legislation and regulation relating to credit rating and research services has been considered from time to time by local, national and multinational bodies and is likely to be considered in the future. If enacted, any such legislation and regulation could significantly change the competitive landscape in which Moody's operates. The management of Moody's cannot predict whether these or any other proposals will be enacted, or the ultimate impact on the competitive position, financial position or results

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

[5] 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).

of operations of Moody's."

Coming to the third research question, the evidence raises doubts about the effectiveness of market discipline in securing a level of rating quality that is optimal from a societal perspective. Empirically, defaults of highly rated issuers are not consistently punished by the market. Together with the focus on reversals, this is consistent with collusion between issuers, fund managers and rating agencies (cf. Calomiris, 2009, and White, 2010). Inflated ratings allow fund managers to earn higher yields while still complying with regulatory limits; keeping reversal rates low at the cost of rating informativeness can be in the interest of issuers and fund managers if the cost associated with rating changes are large enough. An alternative interpretation consistent with the evidence is that the market overestimates the information that reversals contain about rating quality; such errors in assessing rating quality also question the effectiveness of market discipline.

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. Allen and Dudney (2008) conclude that negative publicity related to an anti-trust investigation reduced the influence of Moody's ratings on municipal bond yields. As exemplified in Penas and Tümer-Alkan (2010), documenting stock price reactions to firm news or actions is not sufficient for establishing that discipline is effective. In the case of Moody's, the assessment of the influence that market discipline has on the firm is facilitated by the fact that an important set of firm actions – rating actions – are easily observable.

The theoretical literature on the interplay of reputation, competition and quality offers diverse results. In the models of Klein and Leffler (1981) and Strausz (2005), low competition tends to increase the positive effects of reputation on product quality, while the reputational mechanism in Hörner's (2002) model is enhanced by more competition. Lizzeri (1999) concludes that signals provided by a monopolistic financial intermediary are uninformative. Doherty, Kartasheva and Phillips (2009), however, show that the Lizzeri result is a special case and that ratings issued by a monopolistic agency can be informative. Bolton, Freixas and Shapiro (2012) also obtain that competition is less conducive to rating quality than monopoly, a result that is empirically supported by Becker and Milbourn (2011). Mathis, Andrews and Rochet (2009) theoretically derive that reputational concerns can be a sufficient disciplining device if the agency does not derive a too large fraction of income from complex products (e.g. structured finance securities).

There is a great deal of empirical literature on ratings, some of which might suggest that rating quality is inadequate. Viewed as a whole, however, the literature does not lead to a clear-cut verdict on rating quality. The results of Blume, Lo and MacKinlay (1998), for example, suggest that rating agencies changed their rating standards without notifying the market, thereby questioning the reliability of rating information. Jorion, Shi and Zhang (2009), however, extend the analysis and conclude that rating standards remained stable. Altman and Rijken (2006) show that statistical models outperform ratings in terms of default prediction - but only for horizons of up to three years. This finding is consistent with the long-term horizon used by rating agencies. 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 (2007), this irregularity is due to a misspecification in research design. Campbell and Taksler (2003) show that equity volatility explains as much cross-sectional variation in corporate bond spreads as ratings do. While this questions the relevance of rating information, Kisgen (2009) shows that ratings are important for firms' capital structure decisions.

The remainder of the paper is structured as follows: In Section 2 I describe the data and the definition of variables. Section 3 presents findings on how Moody's stock price reacts to rating actions. Section 4 discusses the results, and Section 5 concludes.

2 Data and definition of variables

I focus on Moody's because of data availability and because the other two major rating agencies, S&P and Fitch, have been part of larger conglomerates, reducing the ability to discover effects specifically related to the rating business.[6] Between 1998 and 2010, Moody's dominant sector has been the traditional rating business comprising corporate finance, financial institutions, sovereigns, and public finance. An average annual share of 47.6% of Moody's total revenue was attributable to the traditional rating business.[7]

Moody's was first traded on 6/19/1998. I collect daily returns and market values from Datastream. To model Moody's stock returns, I consider the value-weighted CRSP market portfolio, industry and size portfolios, the Fama/French factors SMB (small minus big) and HML (high book-to market minus low book-to-market), and MOM, the average return of stocks with high prior returns minus the average return of stocks with low prior returns. Returns, taken from Ken French's website[8], are in excess of the risk-free rate of return (one-month Treasure bill rate, again from Ken French's database). Information on ratings consists of daily information on Moody's long-term, senior ratings of corporate and sovereign bond issuers; the data end in December 2010.[9]

---

[6] S&P, which itself pursues non-rating related activities, is part of McGraw-Hill; and Fitch is part of the France-based company Fimalac, which now focuses on risk management but was an industrial conglomerate before 2005. In 2004, halfway through the sample period, Fitch contributed 36% to Fimalacs' revenue while the financial services segment, which comprises the S&P rating business, contributed 39% to McGraw-Hill's revenue.

[7] Source: Moody's annual reports. Other business sectors are structured finance ratings, research, and Moody's KMV. From 1998 to 2010, the average revenue share of structured finance was 32.2%. The structured finance share exceeded the traditional rating business in the years 2005 to 2007.

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

[9] I examine estimated senior ratings as described in Gupta and Parwani (2009). Some changes in issuer ratings are due to changes in the methodology used to derive issuer ratings from individual bond ratings. I use a flag contained in the database as well as the detailed rating information provided on www.moodys.com to identify

Moody's highlights the following criteria for judging the performance of their rating system (cf. Cantor and Mann, 2003): accuracy as measured by the ability to predict default; average stability of ratings; frequency of large rating changes; and frequency of rating reversals. Since an individual rating event can be judged as large or as a reversal, its impact on Moody's stock price can be tested. By contrast, there is no meaningful way to classify 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 approach is inspired by anecdotal accounts of investment-grade defaults (e.g. Enron) that 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 an issuer that was upgraded) during the preceding 3 months, zero otherwise.

$DRIFT_t$ one if, on day *t*, Moody's upgrades an issuer that was upgraded during the preceding 3 months (or downgrades an issuer that was downgraded during the preceding 3 months), zero otherwise.

$LARGE_t$ one if, on day *t*, Moody's changes a rating by three notches or more (e.g. from Aa1 to A1), zero otherwise.

$DEFAULT_t$ one if, on day *t*, an issuer defaults after having been rated investment-grade within the 6 months preceding such default, zero otherwise.

The foregoing definition of variables is based on Moody's own publications. Fons (2002) mentions the 3-month period applicable to reversals; Cantor and Mann (2003) use three

actual rating actions as opposed to changes due to methodology, and use only actual rating actions in my analysis.

notches to identify large rating changes. DRIFT is examined because it complements REVERSAL. The motivation for the 6-month horizon used for DEFAULT is less direct; it is chosen to be the middle of the one-year time span that is usually considered to be short-term. As documented in the next section, alternative definitions do not materially affect the results. Where an event occurs on a non-trading day, I attribute it to the next following trading day.

Over the 3,154 days for which data was available for both Moody's returns and ratings, there are 90 days with REVERSAL=1. The event day counts for the remaining dummy variables are: 827 (DRIFT), 1090 (LARGE), and 42 (DEFAULT). Figure 1 shows how event days are distributed over time. Most defaults occurred in the years 2001-2002 and 2008-2009. The other three events are distributed more evenly, albeit with some clusters.

Table 1 gives a breakdown of rating actions across industries and calendar years. The statistics are based on individual rating actions rather than on the dummy variables from above, which are defined on a per-day level. As a means of comparison, the table also reports the breakdown of all rating actions. Some years and industries stand out by having a relatively large number of conspicuous rating actions. The rating actions with the largest variation across years or industries are reversals and defaults. For defaults, the figures reflect the historical default cycles, which peaked in 2001/2002 and 2008/2009, when telecom firms and financial institutions, respectively, were the most hard-hit firms. The large number of reversals in 2006 and 2007 is largely due to changes in rating methodologies implemented by Moody's. These caused a large number of small and clustered rating changes. On 4/10/2001, for instance, the ratings of 44 financial institutions (plus a number of their subsidiaries) were changed.[10] The changes led to 56 reversals, which also explains why financials have a

[10] Cf. „Moody's lowers bank ratings following refinement of methodology", Global Credit Research , 4/10/2007. The 2006 change in methodology is described in „Moody's publishes final Methodology and related research for Loss-Given-Default Assessments and Probability-Of-Default Ratings", Global Credit Research, 8/23/2006. Both reports are available on www.moodys.com.

relatively high share among reversals. If the reversals of financial firms that occurred on 4/10/2001 enter the calculation of the industry breakdown as just one reversal rather than 56, the share of financials goes down from 48.1% to 25.8%, which is in line with the 28.0% share of financials among all rating actions. Note that the clustering that is visible in Table 1 is not fully reflected in the dummy variables that will be used in the regressions. These variables take the value one if at least one rating action took place; they are insensitive to multiple rating changes due to changes in methodology or to a rating change of a parent company.

The selection of an abnormal return model for Moody's stock return is based on the regressions shown in Table 2. Factors used to explain Moody's returns include the CRSP value-weighted market portfolio MARKET; the industry portfolios BUS_SERV (return on the Fama-French "Business Services" industry portfolio minus MARKET) and BANKS (defined similarly with the Fama-French "Banking" industry portfolio); SIZE (portfolio returns of size deciles to which Moody's belongs according to monthly size breakpoints, minus MARKET); and SMB, HML and MOM returns. Since Moody's exhibits several extreme returns over the sample period (the minimum is -16.9% and the maximum is 15.9%), I also run regressions with winsorized Moody's returns. Specifically, I pull extreme values to the 1% and 99% quantiles, respectively.

Table 2 shows that the two industry factors as well as HML add significant explanatory power to the regression with only the market portfolio; the remaining three factors do not. Results do not depend on winsorization. Based on these results, I will use the model containing the market return, the two industry portfolios and HML returns.

3 Market reaction to rating actions and defaults

In the event study literature, multiple events pertaining to individual firms are commonly studied through a regression approach.[11] I follow this research and employ multivariate regressions in which events are captured through dummy variables. I start by estimating the following regression in which the daily stock return of Moody's (MOODYS, in excess over the risk-free rate) is explained by stock market indices as well as dummy variables for rating actions occurring on the same day:

$$\begin{aligned} MOODYS_t = \alpha + \beta_1 MARKET_t + \beta_2 BUS\_SER_t + \beta_3 BANKS_t + \beta_4 HML_t \\ + \beta_5 REVERSAL_t + \beta_6 DRIFT_t + \beta_7 LARGE_t + \beta_8 DEFAULT_t + u_t . \end{aligned} \tag{1}$$

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 presented in Table 3. The Durbin-Watson statistics are close to 2, indicating that there is no inference problem due to autocorrelation.[12]

Among the four variables that capture rating actions, REVERSAL and DEFAULT are significant. Results do not depend on winsorization. Since the returns enter the regressions as percentages, the coefficient indicates that Moody's stock price drops by an average of around 0.40% on days in which a rating reversal occurs. With 90 occurrences, reversals cause economically significant damage to Moody's stock price. According to the estimates, and assuming that the price change was indeed driven by reversals, Moody's market value at the end of the sample period is about 30.8% ($=(1-0.00409)^{90}-1$) lower than it would have been if the stock price had not shown negative reactions to rating reversals. For the DEFAULT event, the estimated stock price reaction is larger in magnitude (-0.997%); due to the lower number of DEFAULT events, the overall effect on Moody's stock price is similar to the effect

[11] Binder (1985) discusses the use of multivariate regressions for event studies. For recent applications see Cornett, Mehran and Tehranian (1998) or Bittlingmayer and Hazlett (2000).

[12] To further check robustness, partly motivated by the fact that reversal and default events appear to be clustered (cf. Figure 1), I estimate standard errors with the Newey-West estimator and a lag length of 21 (= 1 month). Results do not change conspicuously. For example, the t-statistics of the reversal dummy change from -2.34 to -2.28 (no winsorization) and from -2.65 to -2.64 (winsorization).

of REVERSALS. Through $(1-0.00997)^{42}-1$, it is computed to be −34.4%.

To check whether the effects of rating actions are different for actions that involve crossing the investment-grade barrier, I augment equation (1) with additional dummy variables. Let IG denote an investment-grade rating and 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 already restricted to investment-grade defaults. Since there are only two days on which REVERSAL-IG is one, there is no reliable way of estimating the associated stock price reactions. I therefore expand its definition to cover reversals in which the third rating of the sequence has a letter rating (on the seven letter grade scale as opposed to the 21 notch scale used elsewhere in the paper) different from the letter rating of the second rating in the sequence. The motivation is that many investors have rating-based benchmarks but that these benchmarks are usually based on letter grades, not on rating notches. Thus, a reversal in which the letter rating is not reversed should have relatively small consequences for institutional investors. I denote the variable which is one for reversals in which the letter rating is reversed by REVERSAL-LETTER; by construction, it also takes the value one if the investment-grade boundary is crossed. There are 33 days on which REVERSAL-LETTER is one.

Regression results after adding REVERSAL-LETTER, DRIFT-IG and LARGE-IG are also presented in Table 3. They indicate that neither the investment-grade barrier nor the letter-grade barrier in case of reversals result in significant special effects for these types of rating actions; this holds for both the individual variables and a joint test of the significance of the three dummy variables (p-value = 0.62; 0.63 for winsorized Moody’s returns).

Variation in stock price effects across issuers

It might be expected that events which involve large issuers lead to larger stock price reactions because they attract more attention from market participants and because the agency's fee income depends on the volume of rated bonds. For the same reasons, it might be expected that the stock price reaction would be stronger if several conspicuous events occurred on a given day. For the two events that showed a significant association with stock returns, I therefore define new dummy variables, EMINENT REVERSAL and EMINENT DEFAULT.

EMINENT REVERSAL is one if the issuer whose rating is reversed is a sovereign or a Fortune 500 company[13] *or* if more than one reversal occurs on the day in question. The dummy NON-EMINENT REVERSAL equals one if a reversal occurs that does not meet the foregoing criteria for eminence. EMINENT DEFAULT and NON-EMINENT DEFAULT are defined accordingly.[14] Of the 90 days on which reversals occurred, 26 are classified as eminent, while 16 of the 42 default events are classified as eminent.

Table 4 shows the results of regressions in which the REVERSAL dummy variable is replaced by the dummy variables EMINENT REVERSAL and NON-EMINENT REVERSAL, while the DEFAULT dummy is replaced by EMINENT DEFAULT and NON-EMINENT DEFAULT. Dummy variables for other rating actions are not included because they were not significant in prior analysis.

Non-eminent reversals lead to small, statistically insignificant stock price reactions of

---

[13] Fortune 500 lists are available on http://money.cnn.com/magazines/fortune/fortune500_archive. One could also use the volume of outstanding bonds to classify issuers, but this information is not contained in the database available to me.

[14] Defaults of subsidiaries are not taken into account when determining the number of defaults on a given day.

-0.129%. The stock price reaction to eminent reversals is stronger: -1.110% with a t-statistic of -3.03. The data therefore confirm the hypothesis that more important reversals lead to larger losses of reputational capital. Note that the estimated cumulated loss is very close to the one obtained above. Based on the regression from column 1, it is 31.1% ($= (1-0.00129)^{64} \times (1-0.01110)^{26} - 1$).[15]

Stock returns on days with eminent and non-eminent defaults, on the other hand, do not seem to conform to expectations. There is no significant effect on days with eminent defaults, whereas non-eminent defaults are significantly associated with lower returns. However, a closer look at the events shows that most of the negative association stems from defaults that occur during the subprime crisis. I define a dummy variable CRISIS, which equals one in the period from July 2007 on, and zero otherwise. In a new regression specification, I include both the interaction CRISIS×DEFAULT and the time dummy variable CRISIS. The motivation for the latter is that the crisis might have had a general impact on Moody's stock price that is not captured by the other variables.

Results are also shown in Table 4. Now, both EMINENT DEFAULT and NON-EMINENT DEFAULT are insignificant, while the estimated coefficient of CRISIS DEFAULT is both economically large (-2.414; -2.152 for winsorized returns) and statistically significant. As evidenced by the insignificance of the time dummy CRISIS, there is no detectable general effect of the crisis on Moody's stock price. A further test shows that EMINENT DEFAULT and NON-EMINENT DEFAULT are also jointly insignificant (p-value=0.94; 0.89 for winsorized returns), which implies that defaults of highly rated issuers did not lead to stock price changes before the subprime crisis.

[15] One could argue that one should ignore the estimated impact of non-eminent reversals because it is not statistically significant. Ignoring insignificant coefficients is ad hoc, though. On the other hand, it does not have a great effect. Ignoring the -0.00129 leads to a estimated cumulative loss of -25.2% ($=(1-0.01110)^{26}-1$), which is still economically significant.

Since most defaulters during the subprime crisis belong to the financial services sector, I alternatively examine a variable that takes the value one on days on which an issuer defaults which was rated investment grade by MOODYS six months prior to default and which MOODYS classifies as a financial services firm.[16] Estimated coefficients are very similar (-2.187; -2.242 for winsorized returns), making it difficult to decide whether the increase in stock price reaction is something that is special to the crisis, to financial institutions, or a combination of both.

Before further addressing time variation in stock price reactions, I use a classical event study framework to examine stock price returns around eminent events. The reason for not conducting such an analysis for the total set of events examined above in Table 3 was that the large number of events would lead to many overlapping windows, which would make it very difficult to interpret the results.[17] Given that the number of events that were found to be significant in Table 4 is considerable smaller, it appears sensible to complement the regression approach with an event study analysis. To estimate abnormal returns, I use the same multi-factor model as before (with MARKET, BUS_SERV, BANKS and HML as factors). The estimation period is taken to be the 90-day window ending 20 days before the event. For the events defined by the dummy variables EMINENT REVERSAL and CRISIS DEFAULT, Figure 2 shows average abnormal returns on the 20 days surrounding the event date. The figure also shows 95% and 99% confidence intervals determined according to the method of Brown and Warner (1980, App. A.3, with the market model being replaced here by the multi-factor model.)

---

[16] Classifications are available on www.moodys.com.

[17] Within a multivariate regression, overlapping event windows also reduce the precision of the estimation. However, by construction, regression coefficients are estimates of partial effects, which means that they can easily be interpreted in a standard way.

For EMINENT REVERSALS, the average abnormal return on the event date is one out of three returns that are located in the event window and that are significant at a 5% level. The only significant return that is recorded after day 0 occurs on day 1 and has the same, negative sign. Cumulative average abnormal returns are -2.07% [days -20 to 20, p-value=0.239] and -1.77% [days +1 to 20, p-value=0.576]. Thus, there is no evidence of a return reversal. On the other hand, none of the two stated cumulative returns is significant. Since the idiosyncratic risk of Moody's is fairly large (the average across the events is 1.89%) however, this is what we would expect to see even if eminent reversals have permanent effects. If the expected abnormal return on a reversal day is -1.110% and zero on all others, the expected cumulative return over a 41-day event window around the reversal would also be –1.110%. But the standard error of the 41-day cumulative abnormal return estimated with the Brown-Warner method is 1.74%, which is larger than the expected cumulative abnormal return under the hypothesis of no reversal and anticipation.

For CRISIS DEFAULTS, the picture is more irregular. The cumulative average abnormal returns are -0.52% [days -20 to 20, p-value=0.847] and 6.1% [days +1 to 20, p-value=0.026]. There is thus evidence of a return reversal. A careful conclusion could be that the event study does not question the robustness of the regression results for EMINENT REVERSALS, while it points out the turbulence surrounding many of the CRISIS DEFAULT events.

Variation in stock price effects across time

To examine how stock price reactions change over time, I move from the subset of eminent events back to the dummy variables REVERSAL and DEFAULT as well as DRIFT and LARGE. The reason is that we are now interested in learning whether there is time variation independent of the eminence of a specific event. For example, one could surmise that Moody's stock price suffers more in times of greater public criticism of rating agencies. To examine this issue, I run rolling-window regressions using specification (1). To achieve a sufficient number of observations, I choose a window size of 500 days, i.e. two years. For a given date τ shown on the x-axis of the figures, the coefficients β_5 to β_8 are thus determined through:

$$\begin{aligned} MOODYS_t &= \alpha + \beta_1 MARKET_t + \beta_2 BUS\_SER_t + \beta_3 BANKS_t + \beta_4 HML_t \\ &\quad + \beta_5 REVERSAL_t + \beta_6 DRIFT_t + \beta_7 LARGE_t + \beta_8 DEFAULT_t + u_t \\ t &= \tau - 250, \tau + 250 \end{aligned} \tag{2}$$

Separately for each rating action, Figure 3 shows the estimated coefficients along with their 95% confidence intervals. The dates displayed are the centers of the respective regression windows, i.e. the coefficient displayed for June 2002 is based on a regression using data from June 2001 to June 2003.

The estimated price impact of reversals is highest in the years after the Enron default (November 2001) and during the subprime crisis, times in which rating agencies faced heightened criticism. For example, the SEC held a hearing on credit rating agencies in November 2002 and the House of Representatives held a hearing in October 2008.[18] Due to the reduced number of observations in the rolling windows, confidence intervals are so wide that the change in the stock price impact of reversals is not significant at usual levels. However, the time pattern of estimates corresponds to the expectation, again supporting the interpretation that the estimated coefficient of the reversal dummy reflects the costs of losing reputation.

Table 1 showed that a large number of reversals occurred after 2005. To check whether results are constant over time, I ran regressions along the lines of Table 3 (first column), first with pre-2006 data, then with post-2006. Coefficients and t-statistics of REVERSAL are

[18] See: http://www.sec.gov/news/extra/credrate/credrate-sched.htm and https://house.resource.org/110/org.c-span.281924-1.pdf

virtually identical (pre-2006: -0.46 with a t-statistic of -2.07; post-2005: -0.47 with a t-statistic of -2.03).

In Table 3, rating changes in the same direction (DRIFT) as well as large rating changes (LARGE) did not show a significant association with stock price returns. The evolution of their coefficients over time does not show patterns that can be related to increased public awareness. Apart from a short period of time in the case of LARGE, estimated coefficients are insignificant and hover around zero. The time series of the estimated coefficients for DEFAULT illustrates the insights from Table 4. Most of the defaults that are associated with negative stock price changes are from the 2008-2009 period. Accordingly, there is no visible effect of defaults prior to the subprime crisis.

Additional analysis

The three-month horizon for the definition of REVERSAL and DRIFT was motivated by publications from Moody's; the six-month horizon for DEFAULT was chosen somewhat arbitrarily. To examine robustness, I augment regression (1) by three variables: (i) a variable that takes the value one if there is a reversal for a six-month horizon but not for a three-month horizon; (ii) a variable that takes the value one if there is rating drift for a six-month horizon but not for a three-month horizon; (iii) a variable that takes the value one if an issuer defaulted that was rated investment grade 12 months before default but not 6 months before default.

None of these three variables is significant, nor are they jointly significant (p-value= 0.587; 0.639 for winsorized returns), while the coefficients on REVERSAL and DEFAULT remain significant. This shows that the choice of the three-month horizon for REVERSAL and the six-month horizon for DEFAULT captures the association with stock price returns well. One would expect that the effect of reversals decreases as the time horizon increases. The further apart two rating changes are, the less likely it is that the second one will be judged in relation to the first. Similarly, a default will be judged less of a surprise the more distant the investment-grade rating of the issuer is.

I also examine whether reversals of the type *downgrade followed by upgrade* have effects different from reversals of the type *upgrade followed by downgrade*. Differences are insignificant, as is evident when a dummy variable that takes the value one for revisions of the type *upgrade followed by downgrade* is added to regression (1). Its p-value is 0.623 (0.415 for winsorized returns).

Further robustness checks relate to large rating changes. I differentiate according to whether the rating change is a downgrade or an upgrade, and whether the investment-grade boundary is crossed or not. This analysis is repeated for rating changes that are larger than three notches. The results, which are shown in the Appendix, do not change the conclusion that large rating changes show no significant association with Moody's stock returns.

Finally, it is interesting to check whether reversals of Moody's ratings have an effect on the stock prices of the two other main agencies' parent companies, McGraw-Hill (S&P) and Fimalac (Fitch). I run regressions along the lines of Tables 3 and 4 with Moody's stock return replaced by McGraw-Hill's and Fimalac's return, respectively. In the case of Fimalac, the set of factors is augmented by the MSCI France. Dummies for Moody's rating reversals are insignificant (p-value>0.1) in each regression. In the case of S&P, this is not surprising. Since most issues are rated by two agencies, S&P, being one of the two leading firms, does not necessarily gain if the other leading firm is weakened. Also, the power of detecting any spillovers is likely to be small because the rating businesses of both S&P and Fitch are parts of conglomerates. This can explain why there is no spillover from Moody's to Fitch even though Fitch could benefit from a weakening of Moody's. To check the validity of this argument, I run a regression of the daily stock return of Fimalac on the factors used previously as well as on Moody's stock return. Moody's return is insignificant (p-values are larger than 0.2) in each of the four possible combinations (Fimalac winsorized or not, Moody's winsorized or not). The fact that there is no detectable correlation between the idiosyncratic stock returns of Moody's and Fimalac indicates that the power of detecting cross-effects of Moody's rating actions is indeed very small for the case of Fitch.

The coefficients on DEFAULT are negative and significant when McGrawHill returns are studied; they are negative and insignificant when Fimalac returns are studied. This observation can be explained with the fact that the ratings of the major rating agencies tend to be similar and that most issuers are rated by Moody's, Fitch and S&P. Hence, the default of an issuer with a Moody's investment-grade rating before default will in most cases coincide with the default of an issuer with a S&P or Fitch investment-grade rating before default.

4 Discussion

To start the discussion, I will examine whether the empirical findings are consistent with statements by Moody's. In the process leading to Basel II and in the aftermath of the Enron and Worldcom collapses, Moody's published several reports designed to clarify their rating policy. They serve as the basis for the following discussion. The avoidance of rating reversals features prominently in Moody's documents. It is described as a key instrument used to meet investor preferences for low rating volatility, along with 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 an important instrument. Moody's mentions specifically that the market might interpret a single reversal as an indication of low rating quality:

> "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 also monitored by Moody's:

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

The statement does describe large rating changes as potentially problematic but this is not emphasized as in the case of reversals. This also holds for other documents, where reversals feature prominently (see the discussion above) but where I failed to find similar statements on large rating changes. In addition, the conclusion that Moody's regards large rating changes as less problematic than reversals is consistent with other investor preferences. If large rating changes were indeed very damaging to Moody's reputation, Moody's could split them into a sequence of small rating changes, exploiting the fact that serial rating changes are apparently not regarded as problematic by the market (cf. the above quote from Fons, 2002, p.7).

Default occurrences among investment grade issuers are not described as a problem, but are 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 does not receive special attention in these 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 be:

> "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 largely consistent with the findings presented in Tables 3 and 4. Stock price reactions suggest that the market punishes reversals, and Moody's claims that this is an important part of its rating policy. Rating drift does not damage the stock price, and Moody's describes it as a side effect that is accepted. Large rating changes, which do not significantly influence the stock price, are tracked by Moody's but are viewed as less damaging than reversals. The investment grade barrier neither affects stock price reactions to rating changes nor receives special attention by Moody's. Note, too, that Moody's does not differentiate between reversals of the type *downgrade followed by upgrade* or *upgrade followed by downgrade*, consistent with the finding that the stock price reaction does not differ between the two types. One might surmise that *downgrade followed by upgrade* is the truly damaging type because it leads to transaction costs for investors restricted to invest in bonds with a minimum rating. However, this is not the only reason why reversals are problematic. Many investors have rating-based benchmarks and would therefore sell upon an upgrade that leads to a removal from the benchmark index. Also, reversals can indicate a rating error, i.e. the agency followed transitory ups and downs without recognizing their transitory nature, irrespectively of their direction.

The consistence of Moody's statements with empirical findings relating to defaults is less clear. Until the subprime crisis, defaults of issuers rated investment-grade in the months before default did not entail negative effects for Moody's stock price (cf. Figure 3), consistent with Moody's view on the relevance of such events. The finding that negative reactions are observed during the subprime crisis could be due to different, non-exclusive reasons. Either the market has changed its interpretation of such events, i.e. views them as more damaging to Moody's reputation as it did before. Alternatively, the events could have signaled negative information for Moody's that is not necessarily related to reputation. On 17 days out of a total of 19 days on which highly rated firms default during the crisis, a financial services firm was involved. These firms were not only important clients of Moody's (through the issuance of bonds and structured finance instruments), their demise also signaled a bad market environment for financial services firms and structured finance activities in general. Even if Moody's reputation did not suffer from these defaults, its business position would be judged to have deteriorated because the defaults indicate a general reduction of issuing activities. Finally, note from Figure 2 that the robustness of the stock market reactions to defaults appears relatively small. Given these rival explanations, concluding that the stock price reactions to defaults are inconsistent with Moody's view of their relevance would stretch the evidence.

The first research question that was raised in the introduction can therefore be answered largely in the positive. Observed empirical patterns are consistent with Moody's summary of client preferences and their own rating policy. An alternative view - Moody's cites market expectations for stable ratings to cover a low-cost policy of monitoring ratings only infrequently – is not supported by the data.

The other two research questions concern the potential of market discipline and its actual effectiveness. In the following, I will discuss the implications of the empirical findings from section 3. To complete the discussion, I will also address arguments to which the present paper does not add new evidence.

A finding that supports the view that market discipline *could* be effective is the size of the stock market reactions associated with some rating actions. Both reversals and defaults have been shown to lead to a cumulative loss of around 30% in Moody's stock price. The finding suggests that market discipline can impose significant costs despite the oligopolistic structure of the rating industry. It is noteworthy because it might be surmised that an oligopolistic environment provides few monetary incentives to provide high-quality products. To understand why this does not need to be so, note that an increase in the probability of market structure changes can significantly alter cash flows expectations. The current market convention is to seek at least two ratings from different rating agencies. Since the positions of the leading two agencies are contested by other agencies[19], a switch of market preferences away from Moody's or Standard and Poor's is conceivable. Such a switch would greatly

[19] Currently, there are ten agencies that enjoy the "Nationally Recognized Statistical Rating Organization" status awarded by the SEC (cf. http://www.sec.gov/answers/nrsro.htm) - an important prerequisite for widespread investor use of an agency's ratings.

affect the franchise value of the agency that loses its status as one of the two agencies that are chosen by default. Furthermore, as became evident in the aftermath of the Enron scandal and the subprime crisis, widespread dissatisfaction with rating quality tends to increase the disposition of regulators to bring about changes in market structure. Finally, there are alternatives to agency ratings that might be substituted, e.g. ratings derived in an automated manner from accounting and stock market information.

A possible argument against the empirical effectiveness of market discipline is that rating accuracy apparently did not increase over the sample period, and that critical rating actions did not become less frequent. In order to be efficient, however, discipline need neither completely eliminate the cause for discipline, nor lead to an improvement over time. Just think of classical inspection games in which equilibria in mixed strategies are common. Of course, one could add that rating quality appears to have gone down over the sample period. With respect to the corporate ratings analysed in this paper, a deterioration is far from obvious. When measured through a commonly used accuracy metric, one-year accuracy in 2008 was not significantly smaller than in 1989; five-year accuracy as seen from 2005 was not worse than the one from 1984 (see Emery and Ou, 2010, Exhibit 15). The discriminatory power of corporate ratings during the subprime crisis therefore mirrors previous fluctuations of empirical rating accuracy with the business cycle. As to average rating quality, research conducted before the subprime crisis suggests that over long horizons, which rating agencies claim to focus on, rating accuracy is higher than the one of widely used alternatives such as logistic regression models (cf. Altman and Rijken, 2006).

Refocusing on the lessons that can be learned from this paper, doubts about the effectiveness of market discipline could be nourished by the patterns of stock price reactions documented in the previous section. In particular, one could expect that large rating changes and defaults of highly rated issuers cause equal or more damage to rating users than rating reversals, and therefore should go along with stock price reactions that are stronger than the ones documented here. Viewed from the opposite angle, one could fundamentally question the notion that rating stability should be of such importance to issuers and investors. It is hard to evaluate the information that individual rating actions convey about statistical rating quality. Investment-grade defaults, large rating changes as well as rating drifts and reversals are bound to occur even when ratings are efficient in the sense that they quickly and completely adjust to new information. It is also difficult to assess the cost that sub-optimal ratings entail for rating users. On the other hand, such assessments are important for the interpretation of the empirical results presented in the previous section.

To approach the problem, I focus on reversals, one reason being that Moody's itself hints at the possibility that they receive undue attention. In one of the publications cited above, Moody's mentions that investors believe "rightly or wrongly" (Fons, 2002, p.6) that reversals are evidence of a rating mistake. To assess the correctness of investor beliefs, one could compare the actual reversal rate with the rate that one would expect from a perfect rating agency. The frequency of reversals 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; one can think of credit quality as a credit score or a default probability transformed with an inverse cumulative distribution function. If credit quality 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*, 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 (cf. Löffler, 2005).

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:

$$\sum_{t=\tau(i)}^{T(i)} (HPCYCLE_{it})^2 + \lambda \sum_{t=\tau(i)+2}^{T(i)} \left[(HPTREND_{it} - HPTREND_{i,t-1}) - (HPTREND_{i,t-1} - HPTREND_{i,t-2})\right]^2 \tag{3}$$

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.[20] 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 EDF 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 (3) for each series. To illustrate the characteristics of the filtered EDF trends, Figure 4 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

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

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 5 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, changes from 46.6% 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 more reversals than actual Moody's ratings, which have a reversal rate of less than 1%. The large difference between actual and simulated reversal rates supports the previous interpretation that the stock price reaction to rating reversals is large enough to affect Moody's rating behavior.

There seem to be two reasons why the market could demand such a low reversal rate. One is 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 or to benchmark indices based on letter ratings, which is why one would expect the costs to be particularly high for reversals crossing the investment-grade boundary or a letter-grade boundary. Since the market reaction is not significantly higher for such reversals (see Table 3), the cost argument does not seem very compelling. 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).

Obviously, the misjudgment explanation raises doubts about the efficiency of market discipline, but so would the cost-based reversal avoidance story. If the preferences of investors and issuers have a strong influence on the behavior of rating agencies – and the evidence suggests that they do –, gaps between private costs and benefits and societal ones could impair efficiency. As pointed out by Calomiris (2009) and White (2010), not only issuers but also institutional investors such as pension fund managers can be interested in inflated (i.e., overly optimistic) ratings. In the face of investment restrictions, inflated ratings increase their flexibility. Inflated ratings also increase the fund managers' ability to achieve returns that sponsors (who trust the ratings) wrongly consider attractive on a risk-adjusted basis. Note that these observations are consistent with the empirical evidence of the paper. At least until the subprime crisis, Moody's was not penalized for investment-grade defaults; within the crisis, it is not obvious whether the stock price reactions reflect reputational losses or changes in the business environment. If sophisticated institutional investors benefit from inflated ratings, they have no incentive to punish rating agencies if ratings inflation is manifested. If the revenue of rating agencies is driven by the reputation they have among issuers and institutional investors rather than their reputation among retail investors or regulators, stock price reactions will reflect the private valuations of the former group rather than the societal costs of low rating quality.

5 Concluding remarks

For many years the leading credit rating agencies have enjoyed a comfortable position in an oligopolistic market. It might be surmised that such an environment provided few monetary incentives to provide high-quality products. However, the possibility of market structure changes can significantly affect an agency's firm value when the position of the incumbents is contested by competitors or threatened by government intervention.

Ultimately, it is an empirical question whether concerns about reputational capital affect an agency's behavior. In this study I have shown that the stock price of the rating agency Moody's reacts negatively to some corporate rating events, and that this reaction is economically significant. Over the 12-year period covered by this study, the cumulative reaction to rating reversals corresponds to a 30% loss in market capitalization. During the subprime crisis, defaults of highly rated issuers lead to similar losses, but due to the special situation during the crisis it is difficult to decide whether the losses should be attributed to a decline of reputational capital or contracting issuing activity.

The pattern of the observed stock price effects is consistent with Moody's own description of investor preferences, which focuses on the importance of rating stability for rating users. The evidence therefore dissipates possible concerns that the rating agency misrepresents investor preferences in order to provide a justification for monitoring ratings only infrequently. The evidence also shows that the market can impose significant costs. The 30% loss in market capitalization documented in the paper should be large enough to influence the agency's behavior.

Market discipline therefore seems to have the potential to work despite the oligopolistic structure of the rating industry. Whether it leads to outcomes that are satisfactory from a societal perspective is a different question. Several observations raise doubts. Based on simulations, the observed rating reversal frequency is very low compared to what one would expect if ratings made efficient use of information. As hinted at by Moody's itself, the market could falsely interpret single reversals as evidence of rating mistakes even though they could well be consistent with a perfect rating system. On the other hand, stock prices did not react to defaults of highly rated securities until 2008. The focus on reversals and the neglect of defaults of highly rated issuers are consistent with collusion between investors, issuers and rating agencies. As pointed out by Calomiris (2009), it is not only issuers but also fund managers who can benefit from inflated ratings, which increase their investment opportunity set while leaving unsophisticated investors uninformed about the risks.

To conclude, the observed market reactions appear to be consistent with either collusion or with misconceptions of how rating quality should be evaluated. Though market discipline is not absent in the rating market, the evidence therefore suggests that it is not sufficient to ensure a rating policy that is optimal from a societal point of view.

Hence, regulation appears necessary, and—especially when considering the rating agencies' role in the build-up of the subprime crisis—so do changes in regulatory structure. Switching from an issuer-pay to an investor-pay model is unlikely to be panacea because large institutional investors, too, can collude with rating agencies. Increased monitoring of rating quality by regulators is another option. When assessing its likely effectiveness, one should take into account that regulators may not be able to perfectly identify rating quality. As pointed out by White (2010), finally, it is not obvious that the appropriate reaction is simply more regulation. Another, possibly complementary strategy could be to reduce the central role that regulators assign to rating agencies in the regulation of financial institutions. This could help to reduce unwanted effects such as buy side demand for inflated ratings. With a different market structure, it is also conceivable that market discipline becomes more effective than it is in the present environment.

APPENDIX

Robustness checks related to large rating changes

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 Fama-French "Business Services" industry portfolio minus MARKET; BANKS is defined similarly with the Fama-French "Banking" portfolio; HML is the value factor. 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 equal or larger to the number of notches stated in the column header. NEG is one on days with a negative rating change, INV is one on days with a rating change that spans the investment grade boundary

	Dependent variable			
	Moody's return (in %)		Winsorized Moody's return (in %)	
	LARGE: ≥3 notches	LARGE: ≥6 notches	LARGE: ≥3 notches	LARGE: ≥6 notches
CONSTANT	0.057	0.058	0.061	0.060
	(1.25)	(1.44)	(1.45)	(1.59)
MARKET	0.972	0.974	0.856	0.858
	(22.24)	(22.25)	(26.21)	(26.26)
BUS_SERV	0.601	0.605	0.571	0.574
	(7.55)	(7.61)	(7.75)	(7.81)
BANKS	0.182	0.183	0.197	0.198
	(4.16)	(4.20)	(5.31)	(5.35)
HML	0.402	0.404	0.297	0.299
	(5.26)	(5.27)	(4.34)	(4.36)
REVERSAL	-0.412	-0.410	-0.440	-0.438
	(-2.37)	(-2.36)	(-2.67)	(-2.67)
DRIFT	-0.024	-0.031	-0.026	-0.035
	(-0.28)	(-0.36)	(-0.35)	(-0.46)
DEFAULT	-0.996	-1.021	-0.828	-0.857
	(-2.26)	(-2.30)	(-2.30)	(-2.38)
LARGE×(NEG=1)	0.062	0.203	0.047	0.205
	(0.67)	(0.49)	(0.55)	(0.50)
LARGE×(NEG=0)	-0.106	0.069	-0.102	0.042
	(-0.76)	(0.15)	(-0.75)	(0.09)
LARGE×(NEG=1)	-0.145	0.028	-0.131	0.041
× (INV=1)	(-0.86)	(0.05)	(-0.83)	(0.08)
LARGE×(NEG=0)	0.313	0.148	0.253	0.075
× (INV=0)	(1.33)	(0.28)	(1.09)	(0.14)
p-value (all LARGE variables)	0.620	0.821	0.736	0.901
R^2	0.356	0.356	0.345	0.344
Durbin Watson	2.028	2.027	2.035	2.037
Observations	3154	3154	3154	3154

References

Allen, Arthur C., and Donna M. Dudney. 2008. The impact of rating agency reputation on local government bond yields. Journal of Financial Services Research 33, 57-76.

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

Becker, Bo, and Todd Milbourn. 2011. How did increased competition affect credit ratings? Journal of Financial Economics 101, 2011-2012.

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.

Bolton, Patrick, Xavier Freixas, and Xavier Joel D. Shapiro. 2012. The credit ratings game. Journal of Finance, forthcoming.

Brown, Stephen J, and Jerold B. Warner. 1980. Measuring security price performance. Journal of Financial Economics 8, 205–258.

Calomiris, Charles W. 2009. A recipe for ratings reform. The Economists' Voice 6, 5.

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.

Cornett, Marcia Millon, Hamid Mehran, and Hassan Tehranian. 1998. The impact of risk-based premiums on FDIC-insured institutions. Journal of Financial Services Research 13, 153-169.

Coval, Joshua, Jakub Jurek, and Erik Stafford. 2009. The economics of structured finance. Journal of Economic Perspectives 23, 3–25.

Crouhy, Michel G, Robert A. Jarrow, and Stuart M. Turnbull. 2008. The subprime credit crisis of 2007. Journal of Derivatives 16, 81-110.

Doherty, Neil A., Anastasia V. Kartasheva, and Richard D. Phillips. 2009. Competition among rating agencies and information disclosure. Working Paper.

Emery, Kenneth, and Sharon Ou. 2010. Corporate default and recovery rates, 1920-2009. Moody's Investors Service.

Erturk, Erkan 2008. Global structured finance default and transition study—1978-2008. Standard and Poor's.

European Commission. 2010. Proposal for a regulation of the European parliament and of the council on amending Regulation (EC) No 1060/2009 on credit rating agencies.

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

Gupta, Vivek, and Kusum Parwani. 2009. Moody's senior ratings algorithm & estimated senior ratings. 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 Economics 17, 57-89.

Hörner, Johannes. 2002. Reputation and competition. American Economic Review 92, 644-663.

Jorion, Philippe, Charles Shi, and Sanjian Zhang. 2009. Tightening credit standards: Fact or fiction? Review of Accounting Studies 14, 1573-7136.

Jorion, Philippe, and Gaiyan Zhang. 2007. Information effects of bond rating changes: The role of the rating prior to the announcement. Journal of Fixed Income, Spring, 45-59.

Kisgen, Darren. 2009. Do firms target credit ratings or leverage levels? Journal of Financial and Quantitative Analysis 44, 1323 -1344.

Klein, Benjamin and Keith B. Leffler. 1981. The role of market forces in assuring contractual performance. The Journal of Political Economy 89, 615-641.

Konold, Clifford. 1989. Informal conceptions of probability. Cognition and Instruction 6, 59-98.

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.

Mathis, Jérôme, James McAndrews, and Jean-Charles Rochet. 2009. Rating the raters: Are reputation concerns powerful enough to discipline rating agencies? Journal of Monetary Economics 56, 657–674.

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.

Penas, María Fabiana, and Günseli Tümer-Alkan. 2010. Bank disclosure and market assessment of financial fragility: Evidence from Turkish banks' equity prices. Journal of Financial Services Research 37, 159-178.

Strausz, Roland. 2005. Honest certification and the threat of capture. International Journal of Industrial Organization 23, 45-62.

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.

Vazza, Diane. 2009. 2008 Annual global corporate default study and rating transitions. Standard and Poor's.

White, Lawrence J. 2010. Markets: The credit rating agencies. Journal of Economic Perspectives 24, 211-26.

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

Table 1: Conspicuous rating actions by year and industry

Based on industry information from www.moodys.com, issuers are assigned to one of the Fama/French 12 industries, or to the group Sovereigns/Sub-Sovereigns. The table shows how many of the conspicuous rating actions analyzed in this paper occurred in a given year or in a given industry. As a means of comparison the breakdowns of all rating actions in the database are also given. 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 six months before default. The defaulted Sub-Sovereign is Guangdong Enterprises

Year / Issuer group	REVERSAL	DRIFT	LARGE	DEFAULT	All rating changes
1998	0.0%	5.3%	3.7%	2.0%	4.8%
1999	0.5%	4.3%	7.8%	3.9%	7.2%
2000	2.7%	4.7%	10.1%	7.8%	7.5%
2001	3.3%	13.3%	13.9%	7.8%	9.8%
2002	3.8%	14.7%	14.3%	25.5%	10.0%
2003	1.6%	5.1%	7.0%	3.9%	7.2%
2004	4.4%	2.4%	4.3%	0.0%	6.9%
2005	4.4%	4.6%	4.6%	2.0%	7.0%
2006	15.3%	5.0%	4.7%	0.0%	6.9%
2007	35.5%	4.2%	6.3%	0.0%	6.0%
2008	4.4%	13.1%	7.4%	23.5%	8.5%
2009	23.0%	18.8%	11.6%	19.6%	12.2%
2010	1.1%	4.5%	4.3%	3.9%	6.1%
Business Equipment	3.8%	2.5%	3.0%	0.0%	3.2%
Chemicals	3.8%	2.6%	3.3%	0.0%	2.9%
Cons. Durables	3.8%	6.6%	3.6%	0.0%	4.7%
Cons. NonDurables	7.1%	5.2%	7.0%	0.0%	7.2%
Energy	3.3%	11.0%	11.2%	19.6%	7.7%
Health	0.5%	1.1%	3.5%	0.0%	2.2%
Manufacturing	2.2%	6.6%	6.0%	5.9%	7.0%
Finance	48.1%	29.2%	26.7%	49.0%	28.0%
Other	19.7%	18.0%	17.9%	5.9%	18.6%
Shops	3.3%	4.8%	3.5%	3.9%	4.9%
Telecom	2.2%	7.1%	9.4%	7.8%	6.5%
Utilities	1.6%	2.2%	3.3%	5.9%	4.4%
(Sub-)Sovereigns	0.5%	3.0%	1.4%	2.0%	2.7%

Table 2: 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 Fama-French "Business Services" industry portfolio minus MARKET; BANKS is defined similarly with the Fama-French "Banking" 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*. MOM is the momentum factor. T-statistics (in parentheses) are estimated with the White-correction for heteroskedasticity

	Dependent variable					
	Moody's return (in %)			Winsorized Moody's return (in %)		
CONSTANT	0.034	0.034	0.033	0.033	0.034	0.032
	(0.95)	(0.96)	(0.95)	(0.99)	(1.06)	(1.00)
MARKET	0.980	0.974	0.957	0.874	0.857	0.844
	(24.13)	(22.18)	(21.61)	(28.65)	(26.19)	(25.02)
BUS_SERV		0.596	0.572		0.567	0.533
		(7.51)	(6.76)		(7.71)	(6.76)
BANKS		0.180	0.166		0.196	0.194
		(4.11)	(3.52)		(5.25)	(4.81)
HML		0.408	0.378		0.302	0.273
		(5.33)	(4.85)		(4.41)	(3.89)
SIZE			0.174			0.197
			(1.27)			(1.54)
MOM			-0.073			-0.046
			(-1.64)			(-1.12)
SMB			0.006			0.033
			(0.08)			(0.52)
p (SIZE=MOM=SMB=0)			0.226			0.243
Adj. R^2	0.301	0.354	0.354	0.290	0.343	0.343
Durbin Watson	2.001	2.027	2.023	2.016	2.037	2.041
Observations	3154	3154	3154	3154	3154	3154

Table 3: 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 Fama-French "Business Services" industry portfolio minus MARKET; BANKS is defined similarly with the Fama-French "Banking" portfolio;. HML is the value factor. 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 six months before default. REVERSAL-LETTER equals REVERSAL if the rating after the second change of the sequence has a different letter rating than the previous rating, zero else. *VARIABLE*-IG equals *VARIABLE* if the rating sequence contains a crossing of the investment grade boundary (Baa3), 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	0.058	0.057	0.062	0.061
	(1.27)	(1.26)	(1.46)	(1.45)
MARKET	0.973	0.972	0.857	0.855
	(22.22)	(22.16)	(26.24)	(26.12)
BUS_SERV	0.602	0.600	0.571	0.570
	(7.57)	(7.53)	(7.77)	(7.74)
BANKS	0.182	0.184	0.198	0.199
	(4.18)	(4.19)	(5.34)	(5.34)
HML	0.401	0.401	0.297	0.296
	(5.25)	(5.24)	(4.33)	(4.32)
REVERSAL	-0.409	-0.441	-0.438	-0.484
	(-2.34)	(-1.90)	(-2.65)	(-2.18)
DRIFT	-0.028	0.003	-0.029	0.000
	(-0.33)	(0.03)	(-0.38)	(0.00)
LARGE	0.024	0.024	0.010	0.012
	(0.32)	(0.28)	(0.14)	(0.16)
DEFAULT	-0.997	-0.993	-0.829	-0.825
	(-2.26)	(-2.25)	(-2.30)	(-2.29)
REVERSAL-		0.071		0.112
LETTER		(0.21)		(0.35)
DRIFT-IG		-0.238		-0.218
		(-1.31)		(-1.25)
LARGE-IG		0.006		-0.002
		(0.05)		(-0.02)
Adj. R^2	0.356	0.356	0.345	0.344
Durbin Watson	2.029	2.028	2.037	2.036
Observations	3154	3154	3154	3154

Table 4: Do eminent reversals and defaults lead to larger stock price reactions?

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 Fama-French "Business Services" industry portfolio minus MARKET; BANKS is defined similar with the Fama-French "Banking" portfolio; HML is the value factor. REVERSAL is one on days with a rating reversal and is split into two types, NON-EMINENT and EMINENT. A reversal is classified as eminent if the issuer is a sovereign or a Fortune 500 company or if there is more than one reversal per day. DEFAULT is one on days on which an issuer rated investment-grade six months before default defaulted. EMINENT and NON-EMINENT DEFAULTS are defined analogously to REVERSALS. CRISIS is one for days in the 07/2007-12/2010 period. T-statistics (in parentheses) are estimated with the White-correction for heteroskedasticity

	Dependent variable			
	Moody's return (in %)		Winsorized Moody's return (in %)	
CONSTANT	0.059	0.067	0.063	0.089
	(1.65)	(1.80)	(1.83)	(2.47)
MARKET	0.972	0.966	0.860	0.856
	(22.11)	(21.94)	(24.85)	(24.66)
BUS_SERV	0.599	0.600	0.565	0.562
	(7.54)	(7.56)	(6.45)	(6.43)
BANKS	0.182	0.185	0.240	0.243
	(4.18)	(4.26)	(5.69)	(5.77)
HML	0.397	0.388	0.197	0.188
	(5.22)	(5.10)	(2.43)	(2.31)
NON-EMINENT	-0.129	-0.122	-0.142	-0.132
REVERSAL	(-0.70)	(-0.66)	(-0.78)	(-0.72)
EMINENT	-1.110	-1.090	-1.118	-1.108
REVERSAL	(-3.03)	(-2.92)	(-3.19)	(-3.12)
NON-EMINENT	-1.026	0.143	-0.967	0.252
DEFAULT	(-1.81)	(0.29)	(-1.95)	(0.45)
EMINENT	-0.837	0.089	-0.887	-0.059
DEFAULT	(-1.25)	(0.18)	(-1.43)	(-0.14)
CRISIS ×		-2.414		-2.152
DEFAULT		(-2.82)		(-2.99)
CRISIS		-0.032		-0.075
		(-0.36)		(-0.93)
R^2	0.357	0.360	0.391	0.394
Durbin Watson	2.036	2.041	2.042	2.048
Observations	3154	3154	2577	2577

Table 5: Empirical reversal frequency for Moody's ratings and hypothetical rating systems based on EDFs

Using the Hodrick-Prescott (HP) filter, Moody's KMV Expected Default frequencies (EDFs) are split into a trend and cycle component. The filter minimizes

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

with $\ln(EDF_{it}) = HPTREND_{it} + HPCYCLE_{it}$, $\tau(i)$ denoting the starting date of series i, and $T(i)$ its end date. λ is the smoothing constant. Both raw EDFs and the Hodrick-Prescott filtered trends for two different choices of λ are then 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%	0.44%
reversals/drifts	3.13	1.61	1.00	0.11

Figure 1: Distribution of rating events over the observation period

Dots mark days on which the respective event dummy variable takes the value one.

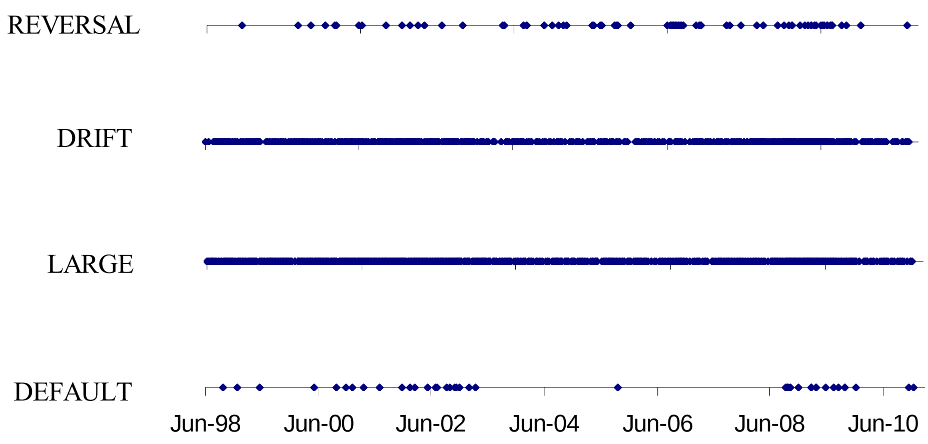


Figure 2: Average abnormal returns surrounding eminent reversals and defaults of financial services firms

The event EMINENT REVERSAL occurs if the rating change of a sovereign or a Fortune 500 company is reversed or if there is more than one reversal per day. The event CRISIS DEFAULT occurs if an issuer defaults that was rated investment grade six months before default and if the day lies within the 07/2007-12/2010 period. Abnormal returns are determined with a multi-factor model (market return, bank and business services industry returns, HML factor) and an 90-day estimation window ending 20 days before the events. Confidence intervals are estimated using the method of Brown and Warner (1980).

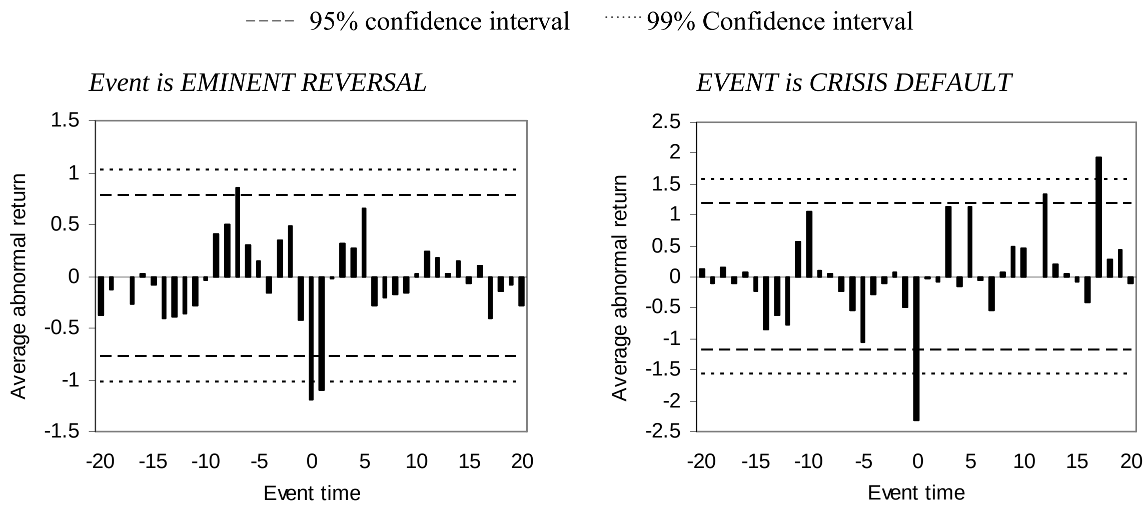
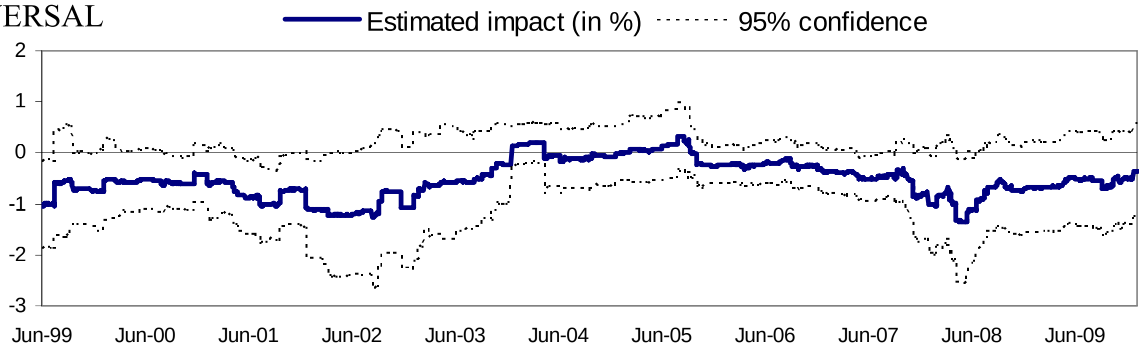


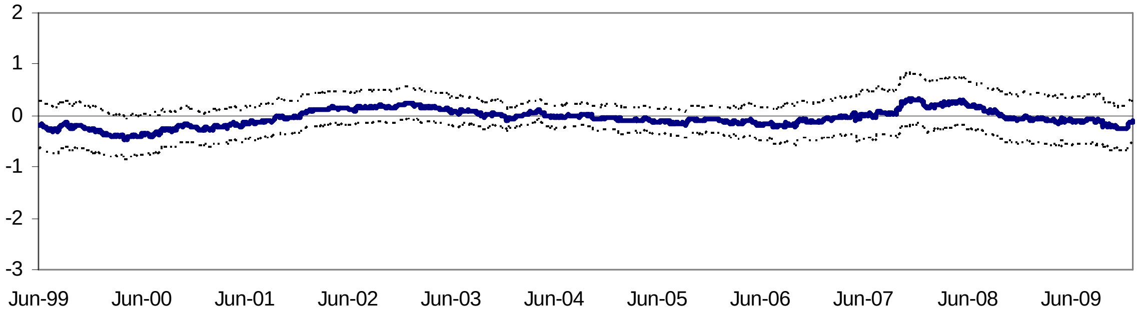
Figure 3: Do market reactions change over time? − Estimated impact of rating actions on Moody's stock price in rolling regressions with a two-year centered window

The estimated impact is the coefficient of the respective rating action dummy in regressions using the specification of Table 3, first column. Rather than using the entire sample as in Table 3, coefficients are estimated with rolling two-year windows centered on the date that is specified on the x-axis.

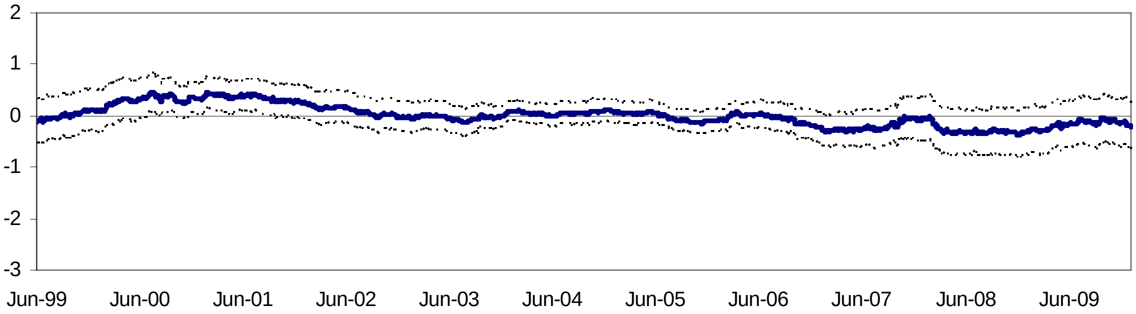
REVERSAL



DRIFT



LARGE



DEFAULT

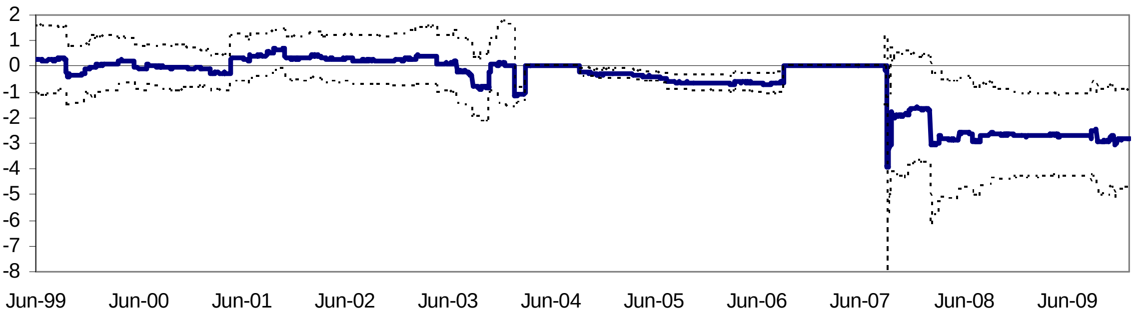


Figure 4: EDF and EDF trends over time for one company

Using the Hodrick-Prescott (HP) filter, Moody's KMV Expected Default frequencies (EDFs) are split into a trend and cycle component. The filter minimizes

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

with $\ln(EDF_{it}) = HPTREND_{it} + HPCYCLE_{it}$, $\tau(i)$ denoting the starting date of series i, and $T(i)$ its end date. λ is the smoothing constant. The figure shows EDFs and the Hodrick-Prescott trends for two different choices of λ. The ordinate is scaled logarithmically.