Rating Watchlists and the Informational Content of Rating Changes

Christian Hirsch *

Christina E. Bannier †

This version: September 2, 2008

Abstract

Credit rating agencies do not only disclose simple ratings but announce watchlists (rating reviews) and outlooks as well. This paper questions whether and, if so, how the review procedure has affected the agencies' information provision. Using Moody's rating data between 1982 and 2004, we find that the information content of rating changes has indeed increased after the introduction of the watchlist instrument. Rating reviews do not appear to "buy time" for an eventual rating action. Rather, they seem to establish implicit contracts between agencies and borrowers similar to Boot, Milbourn, and Schmeits (2006) and as such allow the transmission of a higher quality of information.

Keywords: Credit Rating Agencies; Watchlist; Market Reactions; Event Study

JEL: G14, G29, G33

*Goethe-University Frankfurt, Department of Economics and Business Administration, Mertonstr. 17, Uni-Pf 88, D-60054 Frankfurt am Main, Germany. Phone: +49 69 798 23907, Fax: +49 69 798 28951, E-mail: hirsch@finance.uni-frankfurt.de

†Commerzbank Professor of Corporate Finance / SME Finance, Frankfurt School of Finance and Management, Sonnemannstr. 9-11, 60314 Frankfurt, Germany. Phone: +49 69 154008 755, Fax: +49 69 154008 4755, E-mail: c.bannier@frankfurt-school.de

The authors would like to thank Moody's Investors Service, especially Richard Cantor, for providing the data and for commenting on a first draft of the paper. We are also grateful to Jan Pieter Krahnen, Conrad Zimmer, Karl-Hermann Fischer, Christian Schlag, Nicole Branger, Ralf Elsas, Marcel Tyrell, Gunter Löffler, Paula Hill, Reint Gropp, André Güttler and Isabel Schnabel for valuable comments and suggestions. Any remaining errors are our own responsibility.

Credit rating agencies such as Standard and Poor's (S&P), Moody's Investors Service (Moody's), or Fitch, Inc., provide qualitative statements on the creditworthiness of entities and their financial obligations. Use of credit ratings has expanded in recent years, mostly due to the globalization of financial markets, the growing complexity of financial products, and, generally, an increasing usage of ratings in financial regulation and contracting (Frost, 2006).

The widespread use of credit ratings has been accompanied by a rise in the complexity of the rating information. Most credit rating agencies not only offer a rating for a company issuing securities and for the individual financial products issued, but supplement their service by providing additional information via rating outlooks and rating reviews ("watchlists")[1] that give indications of future credit rating changes. While rating outlooks represent agencies' opinions on the development of a credit rating over the medium term,[2] rating watchlists are stronger statements, as they focus on a much shorter time horizon - three months, on average (Keenan, Fons, and Carty, 1998).[3]

Review listings are usually triggered by sudden corporate events such as the announcement of a merger, a share buy-back, or major changes in management. A rating may be put on review for possible downgrade or upgrade or with direction uncertain. During the watchlist interval, the rating agency collects additional information on the firms it rates, which typically leads to an interaction between rating analysts and firm management. The watchlist is resolved by the announcement of either a rating change or confirmation of the initial rating. The proportion of ratings "on watch," has sharply risen in recent years: until 1998 about 10 percent of bond issuers, on average, were under review at Moody's; between 2000 and 2004, this percentage increased to more than 50 percent (Hamilton and

[1]Moody's reports ratings currently as being under review on their "Watchlist"; S&P refers to its "CreditWatch." In the following, we use the notions of rating watchlists and rating reviews interchangeably.

[2]Rating outlooks are generally terminated after 12 to 18 months.

[3]In the study by Keenan, Fons, and Carty (1998), the 10 (90) percent quantile is 22 (95) days for firms that are placed on watchlist with designation downgrade. For firms entering the watchlist with designation upgrade, the mean is 115 days with 21 (218) as the 10 (90) percent quantile.

Cantor, 2004). Obviously, rating watchlists have grown into heavily used instruments to transmit information to financial markets.

Moody's, as one of the oldest rating agencies, has been publishing a list of ratings on review since 1985.[4] However, it only started to consider watchlist assignments as "formal rating action" on October 1, 1991. Since that time, Moody's has employed a full rating committee to decide whether to place a borrower under review and the subsequent watchlist resolution. Interestingly, Moody's states: "That rating changes for issuers placed on the watchlist are different from issuers not on the watchlist, implies that the watchlist is an important source of information for market participants interested in measuring credit risk." (Keenan, Fons and Carty, 1998). In this paper we use Moody's rating data to try to answer two questions with respect to this statement. First, is it true that there is a difference between watch-preceded rating action and direct, i.e., not review-preceded, rating action? Second, if so, how can we explain this difference?

We employ an event study methodology on the complete history of Moody's estimated senior unsecured ratings between 1982 and 2004 to analyze these questions. In particular, we compare the market reaction to different rating events - measured by the rated companies' cumulative abnormal stock returns.[5] As a first step, we test for a time-series break in the market reaction to rating changes (disregarding any differences between direct and watch-preceded rating changes) due to the institutional implementation of the watchlist on October 1, 1991. In line with earlier work (Cantor, 2004; Vassalou and Xing, 2005), we find a significant reaction following negative rating changes only, but not following upgrades.[6] Comparing the pre-watchlist period (April 26, 1982 - September 1991) with

[4]Standard and Poor's instituted a watchlist in November 1981.

[5]Comparing stock market and bond price effects following rating changes, Wansley and Clauretie (1985) report that the bond market appears to be less efficient in the sense that relative bond prices tend to react as long as seven months after a rating change.

[6]There are exceptions to this generally accepted asymmetry in market reaction: Jorion, Liu, and Shi (2005) find a significant positive abnormal return following upgrades after the introduction of the Regulation Fair Disclosure on October 23, 2000, by the SEC. Second, Goh and Ederington (1993) find a significant negative abnormal return only for downgrades associated with a deterioration of the firm's expected financial performance but not for those attributed to a reorganization or an increase in financial

the post-watchlist period (October 1991 - December 2004) and focusing on rating downgrades, we see that the informational content of ratings significantly increased after the watchlist introduction. This result continues to hold even when we control for additional explanatory factors such as business-cycle, regulatory or sample-composition effects and, consequently, underlines the conjecture implicit in the initial Moody's statement that the watchlist instrument has in some sense influenced rating agencies' traditional role as information providers.

In a second step, we investigate more closely the differences in reactions to direct rating changes vs. watch-preceded rating changes in the post-watchlist period. Very few studies have examined the informational content of watchlist resolutions. Holthausen and Leftwich (1986) use a sample of S&P's Credit Watch data in the period 1981 to 1983 and find tentative evidence that watch-preceded rating downgrades provide less information than rating changes not preceded by a formal review process. However, they point out that the small sample size hampers reliable inferences. Purda (2006) and Hill and Faff (2007), in contrast, conclude that the market does not react any differently to the two types of rating changes. However, Purda (2006) distinguishes between expected and unexpected rating changes, where rating reviews are only one among several ingredients affecting rating change expectations. Hill and Faff (2007) focus solely on sovereign ratings and find that prior to a watch-preceded downgrade, the market seems to anticipate the event by displaying negative returns but has a significantly positive reaction after the downgrade.[7]

Our work differs from the aforementioned studies in several respects. Most important, we question the economic role of the rating review as a particular rating procedure. We test two different lines of argument: first, by creating an additional, more lengthy rating process via the watchlist, agencies may simply "buy time" for assessing the entities in question in order to demonstrate timely action without compromising the long-term char-

leverage. Regarding cross-sectional aspects, stronger market effects are generally found for downgrades to and within the sub-investment-grade rating category (Goh and Ederington, 1999).

[7]This result is supported by Hull, Predescu, and White (2004), who focus on ratings' effects on credit default swaps and find that while additions to the watchlist (with designation downgrade) are informative, the eventual rating downgrades are not.

acter of their rating assessments (Cantor and Mann, 2006). As an alternative, however, it has recently been argued that credit ratings may also be employed as an instrument to coordinate investors' anticipation of credit risk (Carlson and Hale, 2006). As a consequence, an intensive monitoring process (via the watchlist) allows rating agencies to influence firms' risk choices by threatening them with imminent rating changes and subsequent investor reactions. Boot, Milbourn, and Schmeits (2006) have shown that this "implicit contract" is effective at least for specific types of companies, so that watchlists enable credit ratings to convey information of a higher quality to market participants: Whereas a direct downgrade signals a firm's lack of either capability or willingness to uphold a specific credit quality, a watch-preceded downgrade signals a failure in the attempt.

The two arguments (buying time versus implicit contracting) allow the derivation of distinct predictions with respect to the market reaction to direct vs. watch-preceded rating changes. In a univariate analysis – focusing on negative rating changes again – we find that the buying time argument is not confirmed. Rather, rating agencies seem to strategically differentiate between companies that are placed under review and companies that are confronted with direct rating changes. With further scrutinization, we show that a company's management quality and financial flexibility are the main criteria for the watchlist placement decision. Controlling for these selecting factors, a multivariate analysis finds that the informational content of direct downgrades is stronger than that of review-preceded downgrades. Various robustness tests come to the same conclusion. Again, this contradicts the buying time argument, according to which the watchlist simply shifts the decision on a rating change to a later date, so that the market should react identically to both types of rating action.

However, while we do see that agencies discriminate between obligors placed under review and those who are dealt direct rating changes – concurring with the implicit contracting argument –, our results do not fully support Boot, Milbourn, and Schmeits' (2006) interpretation of their theoretical model. Rather, it seems to be the case that their model fails to take into account the endogeneity of information triggered by a watchlist assignment. Several theoretical and experimental studies have shown that the public/private composition of information plays a particularly important role in financial markets due to the strategic considerations of market participants (Morris and Shin, 2002, 2004; Allen, Morris, and Shin, 2006; Angeletos and Pavan, 2007; Heinemann, Nagel, and Ockenfels, 2004, 2008). Since the announcement of a review process implies a sudden loss of accuracy of the initial rating – the public information – investors may be induced to collect supplementary private information on credit risk over the course of the watchlist procedure. This naturally increases the heterogeneity of investors' information base, so that the market reaction to subsequent rating actions will arguably be smaller than in the case of a rating change without a preceding watchlist designation.

The rest of the paper proceeds as follows. Section 1 derives the hypotheses to be tested. Section 2 describes our data set and lays out its main characteristics. Section 3 contains the empirical analyses and provides an interpretation of the results. Section 4 concludes.

1 Derivation of Hypotheses

1.1 Does the Watchlist Instrument Change the Informational Content of Credit Ratings?

Several questions arise with respect to the economic rationale behind the watchlist procedure. As a first step, we analyze whether or not the introduction of the watchlist instrument generally influences the informational content of ratings. Therefore, we test for a time-break in the effects of rating changes on the value of firm equity, i.e., on the cumulative abnormal stock return, at the time of the formal introduction of the watchlist on October 1, 1991. In our analysis, we focus mainly on rating downgrades, since upgrades very rarely deliver significant results. If relevant, however, we also comment on the implications of positive rating changes.[8] Disregarding any differences between direct

[8] If not displayed in the paper, the results from rating upgrades are available from the authors upon request.

downgrades and watch-preceded downgrades (i.e., watchlist resolutions leading to a downgrade) in the post-watchlist period, we seek a larger market effect in the post-watchlist era. This leads to our first hypothesis:

Hypothesis 1 *The effect of downgrade announcements on the market value of firm equity is stronger in the post-watchlist era, as compared to the era before the introduction of the watchlist procedure.*

Apart from the argument mentioned above, a confirmation of Hypothesis 1 could also be due to simple time trends or reasons of sample composition. Blume, Lim, and MacKinlay (1998), for instance, suggest that rating standards applied by credit rating agencies may have hardened over time. If true, this implies that for a given firm quality (measured in terms of observable financial variables and ratios), agencies tend to assign lower rating notches over time. Blume, Lim, and MacKinlay (1998) confirm this hypothesis for the 1980s and early 1990s. Even if rating standards remained the same, however, the average market reaction to rating announcements may have been affected by a sample-composition effect. If the reaction to a downgrade differs across rating notches, and the sample population is not stationary with respect to the distribution of firms across rating notches, the cumulative abnormal return may be affected even if there is no effective change in rating policy.

Therefore, the second hypothesis controls for these additional effects:

Hypothesis 2 *The watchlist effect on a firm's market value of equity is not explained by a change in rating standards, by the sample composition, or by other time trends.*

1.2 What is The Economic Function Underlying Rating Reviews?

With respect to the economic rationale behind the introduction of an institutionalized rating review process, two lines of argument may be distinguished.

First, the introduction of a formal review process may have been the agencies' reaction to a heightened demand for timely credit risk information from financial markets. Agency ratings typically adjust more slowly to new information than market-based measures of corporate default risk such as, e.g., KMV's distance-to-default measure (Löffler, 2004a; Vassalou and Xing, 2005; Robbe and Mahieu, 2005).[9] However, while market prices respond prior to rating events, they tend to react more aggressively than is warranted ex-post. Agency ratings, in contrast, are supposed to reflect changes in credit quality only when they are "unlikely to be reversed within a relatively short period of time" (Cantor, 2001).[10]

According to this argument, watchlists may help to alleviate the traditional conflict between rating stability and accuracy in that they allow agencies to buy time for an eventual rating decision while signalling immediate (monitoring) activity. Consequently, an agency directly changes a rating if it is certain that the change in the borrower's credit quality is sufficiently strong and long-lasting; otherwise, the agency puts a firm under review. Over the watchlist period, the rating agency collects additional information about the borrower's repayment capability and terminates the procedure as soon as a sufficient certainty about the change in the borrower's credit quality is obtained. Ex-post, therefore, a direct downgrade and a watchlist downgrade imply the same informational content, and the market should react in exactly the same way. Adding a rating to the watchlist, in contrast, would be an uninformative event (at least for watchlist entries with direction "uncertain") according to this hypothesis and, hence, should not trigger a market reaction.

Second, following the argument in Boot, Milbourn, and Schmeits (2006), the watchlist may be interpreted as an agency's means of engaging in an implicit contract with the borrowing firm. In their theoretical model, a firm seeks external financing in order to conduct a business project with moderate payoff and variable probability of success. The

[9] Interestingly, the KMV measure of credit risk was introduced in 1989, i.e., only shortly before Moody's released its institutionalized watchlist.

[10] Löffler (2005) provides empirical proof of agency-ratings' stability and analyzes why rating reversals may be harmful. Löffler (2004b) examines the tradeoff between rating timeliness and accuracy against the background of portfolio governance rules.

company may, however, deviate from this viable project and choose instead an inefficient high-risk project with only a low probability of success but a very high payoff. Project choice is unobservable to financiers, so moral hazard in the form of an asset substitution problem arises. Provided that the viable project's probability of success - and, thus, the firm's credit quality - takes on an intermediate value that is common knowledge (i.e., fully precise public information), multiple self-fulfilling equilibria may arise. In these equilibria, the firm's project choice depends critically on investors' repayment requests:[11] If investors believe that the high-risk project will be realized, they will request high coupon payments to compensate for the higher risk. These high costs aggravate the moral hazard problem and induce the company to fulfill investors' beliefs and choose the high-risk project. On the other hand, if financiers believe that the viable project will be selected, they will request only a low repayment, which induces the company to abstain from asset substitution.

Boot, Milbourn, and Schmeits (2006) show that credit ratings may serve as coordination mechanisms in this case, solving the multiple equilibria problem. Provided that enough financiers condition their investment decisions on the rating level,[12] the rating announcement may guide all investors toward either the viable or high-risk project, which will, subsequently, trigger the realization of this project. Credit ratings, hence, deliver information on the investment behavior of a large proportion of the market. This allows agencies to determine the remaining investors' decisions as well, which, in turn, influences firms' risk choice and, subsequently, validates the rating.

Interestingly, this coordination function brings rating agencies in a position to solve further moral hazard problems. Boot, Milbourn, and Schmeits (2006) consider the case where a company is hit by a negative shock that reduces the success probability of the viable project considerably. The firm has the possibility to exert a recovery effort that

[11] Multiple equilibria are possible whenever the viable project's probability of success is not so low that the firm will always succumb to moral hazard and choose the high-risk project, or so high that the firm will always select the viable project.

[12] Many institutional investors are often obliged by specific investment guidelines to engage only in highly-rated (non-speculative grade) investments. See also Hill (2004) for an overview of ratings-based U.S. regulations.

may completely offset the shock's negative effect on credit quality. However, the recovery effort is costly and successful only with a positive probability that is firm-specific and commonly known in the market. As investors cannot observe whether the firm undertook all necessary measures to improve its credit quality, a second recovery-effort moral hazard problem arises.

Boot, Milbourn, and Schmeits (2006) show that the watchlist procedure may solve both types of moral hazard. Firms with low probabilities of a successful recovery effort will be downgraded directly after a negative shock. As a consequence, they fully succumb to the asset substitution problem and realize the high-risk project. Companies with high recovery-effort success probabilities, in contrast, will be placed under review.[13] Afterward, the rating agency will observe whether or not the recovery effort is successful. In the former case it will announce a "good" rating, inducing the choice of the viable project; in the latter case it will assign a "bad" rating, leading to realization of the high-risk project. As a watch-preceded downgrade is due solely to failure of a recovery effort – a random event that cannot be anticipated – the authors conclude that a watchlist downgrade should trigger a strong market reaction. A direct downgrade, in contrast, does not deliver any new information to investors: Each company's probability of a successful recovery effort is common knowledge. Thus, an agency's decision to place a firm on the watchlist, as opposed to a direct downgrade, is fully anticipated by the market and, therefore, should not lead to any reaction. Note that this interpretation rests critically on the assumption that both the firm's credit quality and the success probability of recovery effort are publicly known.

The two lines of argument, buying time vs. implicit contracting, hence lead to opposite projections with respect to the market reaction to direct and watch-preceded rating changes. We phrase Hypothesis 3 such that confirmation would be support for the buying time argument:

Hypothesis 3 *The effect of a watchlist-preceded downgrade on the value of firm equity is not stronger than the effect of a direct downgrade.*

[13] As a necessary condition for watchlist placement, the proportion of institutional investors that condition their decisions on credit ratings must be sufficiently large.

2 Data Selection and Descriptive Statistics

Our data comprise the complete history of Moody's estimated senior unsecured ratings of U.S. issuers.[14] Since Moody's started to add numerical modifiers to its letter ratings on April 26, 1982, we chose to exclude all rating information prior to this date. Consistent with the existing literature, we converted Moody's letter ratings into a numerical scale, where 1 is equivalent to Aaa, 2 is equivalent to Aa1,..., and 21 is equivalent to C.

We made several further refinements to our raw data. First, as we match rating information with firm-specific data later on, we restrict the reported database to include only those firms' ratings for which stock price information is available. Second, we delete all watchlist entries that led to rating reversals (e.g., additions to the watchlist with direction upgrade that were downgraded subsequently). This deletion of data is uncritical, as we lose only six observations altogether. Third, we control for contaminated data, i.e., all rating actions that were tied to obvious events such as corporate mergers, etc. (Jorion, Liu, and Shi, 2005). An observation is considered to be contaminated if any firm-specific, price-relevant information appears in the Wall Street Journal Abstracts within a three-day window surrounding the event day of the rating change. Our final sample consists of 3,180 (direct and watch-preceded) downgrades and 1,609 (direct and watch-preceded) upgrades.

In line with Blume, Lim, and MacKinlay (1998), we find that the mean rating in our sample declined monotonically over time (from 8.07 in 1982 to 10.50 in 2004).[15] However, the rating universe also changed considerably throughout our observation period. The number of rated issuers per year increased almost tenfold from 1982 to 2004. Thus, the

[14]Estimated senior unsecured ratings are usually calculated as issuer ratings, rarely as issue ratings. By using this type of rating, we avoid the problem of multiple ratings for one issuer, which facilitates comparability across firms and also over time. In the case of multiple ratings, the watchlist decision cannot be attributed to a particular issue rating. Therefore, we assume that it affects all outstanding ratings of this firm. For a detailed description of the respective algorithm employed by Moody's to calculate the issuer rating, see Hamilton (2005).

[15]The corresponding tables are not displayed in the paper, but are available upon request.

observed rating deterioration may have been caused by a rise in riskiness of the underlying pool of firms, which supports Hypothesis 2.

With respect to watchlist placements, two subperiods can be clearly distinguished: the watchlist build-up from 1991 to 1997 and the stable period following 1998, where more than 1,000 watchlist assignments per year are observed. Watchlist additions with designation downgrade are roughly twice as frequent as those with designation upgrade.

Regarding direct rating changes, we find that the number of upgrades is fairly stable over time while downgrades seem to display a much stronger dependency on the business cycle,[16] with a peak in the 1999-2001 period. More recently, the number of upgrades exceeds downgrades. Comparing direct and watch-preceded rating changes, we find that between 2000 and 2004, more than 50 percent of overall rating action in our sample is conducted via the watchlist.

Table 1 (2) provides the distribution of the number and the average size of direct and watchlist-preceded downgrades (upgrades) per year. With respect to the time series dimension, we dispose of considerably more data points in the post-watchlist era as compared to the pre-watchlist era (2,150 downgrades altogether versus 1,030, and 1,191 upgrades versus 418). The proportion of direct to watchlist-driven downgrades in the post-watchlist period is roughly 60:40, for upgrades it is 70:30. This again confirms the perception that the watchlist has become an important tool for rating agencies. Comparing the average size of rating changes, we can also see that watchlist-preceded changes tend to be larger than direct rating changes, with the effect being more pronounced for upgrades than for downgrades. Over time, however, the average size of the rating change for both downgrades and upgrades seems to have decreased.

We provide a summary of the size distribution of downgrades (upgrades) in Table 3 (4). During the pre-watchlist era, we find a higher proportion of more-than-one-notch rating downgrades as compared to both the post-watchlist period, in general, and watchlist-downgrades, in particular. In contrast, whereas 48.83 percent of all downgrades in the

[16]Using the NBER classification for recession, we have three recession periods in our sample period: April 1982 to November 1982, July 1990 to March 1991, and March 2001 to November 2001.

pre-watchlist period are a change by one notch, this proportion rises to 58.33 percent in the post-watchlist era. This may at least partly be a consequence of the favorable economic conditions prevailing during most of the 1990s, given that the number of downgrades is positively correlated with recessions. In the post-watchlist period, watch-preceded downgrades seem to be slightly larger than direct downgrades (the proportion of rating changes larger than two notches is a bit higher). Very similar results are obtained with respect to upgrades.

Our empirical tests mainly use a standard event study methodology à la MacKinlay (1997). Generally, we calculate the cumulative abnormal stock return in response to a rating event over a short window surrounding the event date. The cumulative abnormal return (CAR) is computed as the cumulative stock return over the event window minus the return of the market portfolio. The event window spans three days, beginning at -1 and ending at +1, with the event being the direct rating change or review-preceded rating change. Our estimation window spans the time period -120 to -20. We take stock price information from CRSP daily tapes and calculate the market model using the value-weighted index in CRSP.

3 Empirical Analysis

3.1 Informational Content of Rating Action

Hypothesis 1 states that ratings have become more informative since the watchlist instrument has been introduced. In a first univariate test, we analyze the effects of rating changes on cumulative abnormal stock returns, differentiating between market reactions before and after the introduction of the watchlist procedure. The results are displayed in Table 5.

In line with earlier studies, we find statistically significant (negative) CARs only following downgrades. Furthermore, the general market reaction to downgrades (both direct and watch-preceded) seems to be stronger in the post-watchlist era (with a CAR of

-3.26 percent) than in the pre-watchlist period (with only -2.18 percent). The difference is both statistically and economically significant. This result lends support to Hypothesis 1, as it indicates that ratings have indeed become more informative since the introduction of the watchlist, thereby increasing the negative stock price reaction to a rating downgrade. For upgrades, in contrast, we find no significant market reaction.[17]

Hence, univariate results seem to support Hypothesis 1. We now proceed to a test in a multivariate framework. As the univariate analysis indicated insignificant CAR effects from upgrades, we focus solely on downgrades in the following, using model 1,

$$\begin{aligned} CAR_j = {} & \beta_0 + \beta_1\, RCHANGE_j + \beta_2\, IGRADE_j + \beta_3\, DAYS_j \\ & + \beta_4\, POST1991^{*}RCHANGE_j + \beta_5\, POST1991^{*}IGRADE_j \\ & + \beta_6\, POST1991^{*}DAYS_j + \epsilon_j. \end{aligned} \tag{1}$$

In line with Holthausen and Leftwich (1986) and Jorion, Liu, and Shi (2005), we test the influence of the size of the rating change (in number of notches, RCHANGE), the crossing of the investment grade boundary (a dummy variable, IGRADE), and, finally, the number of days since the previous rating action[18] (DAYS) on the cumulative abnormal return of firm j. In order to test Hypothesis 1, we create a dummy variable (POST1991) equal to one if the rating event falls into the post-watchlist era, and zero otherwise. This dummy variable enters our model as an interaction term with the other control variables.

Focusing on the effects of rating downgrades, we expect to find a negative coefficient for RCHANGE. Ratings are proxies for default risk. To the extent that a rating change conveys new information to the market, a downgrade should raise the firm's future debt refinancing costs and, hence, lower the firm's market value. This negative effect should increase in the size of the rating change. Note that the probability of default rises exponentially with decreasing rating notches, so that a downgrade by two notches has an effect on the firm's net worth more than twice as large as a one-notch rating change.

[17] Note that our results do not change if we use different methods of calculating CARs. As an alternative, e.g., we used the method by Boehmer, Masumeci, and Poulsen (1991).

[18] Our results are virtually unchanged if we use the on-watchlist date instead of the off-watchlist date.

The variable IGRADE is expected to display a negative coefficient as well. Large investors, pension funds in particular, are usually not allowed to hold non-investment grade rated products.[19] When bonds pass the boundary to junk status, portfolio managers are often forced to sell. Thus, the market for investment-grade bonds may differ substantially in terms of participants, volume, and risk preferences from the market for junk bonds, leading to a downward jump in CAR due to a crossing of the investment-grade boundary. However, as we use issuer ratings (senior unsecured ratings), this effect may be weaker than for issue ratings.

With respect to regressor DAYS, both a positive and a negative coefficient may be conceivable. On the one hand, the longer the time period between two sequential ratings, the stronger may be the informational novelty of a downgrade, leading to a strongly negative effect on CAR. On the other hand, the more time passes, the more likely it becomes that the market has already updated its belief with respect to the creditworthiness of the borrower based on other pieces of private and public information. In this case, a rating change no longer conveys new information to the market (Jorion, Liu, and Shi, 2005). A downgrade may even lead to a positive market reaction if it is less pronounced than the unconfirmed market pessimism.

Our key variable in model 1 is the interaction of RCHANGE with the POST1991-dummy. If this variable turns out to be significantly negative, this should confirm Hypothesis 1 that the introduction of the watchlist has increased the informational content of rating events. We also include interaction variables with the IGRADE and DAYS variables. However, we cannot offer clear predictions with regard to the signs of these two variables and include them mainly for consistency.

The results of model 1 are presented in Table 6, column 2. While variable RCHANGE displays the expected negative sign and is significant, regressor IGRADE turns out to be economically and statistically significantly positive. One explanation for this could be the high degree of activity in the junk bond market during the 1980s. This may be supported by the negative but insignificant POST1991*IGRADE dummy, indicating that after the

[19]For an overview of rating-related regulation of investment decisions, see Partnoy (2002).

breakdown of the junk bond market, a downgrade from investment grade to speculative grade is perceived as negative information. However, the combined effect is not statistically significant. The variable DAYS turns out to be insignificant. When interacted with the POST1991 dummy, however, the DAYS regressor yields a statistically highly significant and positive coefficient. For our key variable, POST1991*RCHANGE, we find a significantly negative coefficient. Given that the economic effect of this regressor is much stronger than that of the simple RCHANGE variable (-0.017 vs. -0.006), the informational content of rating downgrades seems to have strongly risen due to the introduction of the watchlist. This is consistent with Hypothesis 1.[20]

3.2 Time Trends and Sample Composition

So far, our results tend to confirm Hypothesis 1 that the introduction of the watchlist procedure on October 1, 1991, has increased the informational content of rating changes. However, there are alternative explanations for our findings, which are summarized in Hypothesis 2. This section first addresses the time trend in some of our variables as an explanatory factor for the observed pattern of regression coefficients. We use two alternative specifications for time trend. First, following Blume, Lim, and MacKinlay (1998), we include a set of (n-1) year dummies into the regression equation of model 1 in order to capture a linear time trend. This constitutes model 2. We present the results n Table 6, column 3. Note that the year dummies' coefficients are not displayed.

While all the other variable of the model stay roughly unchanged, also in terms of significance level, the coefficient of the POST1991*RCHANGE variable increases slightly from -0.017 to -0.014, which is statistically significant at the 5 percent-level. The variable IGRADE loses statistical significance weakly. This again strengthens the view that the positive sign in model 1 is time dependent. Note that the POST1991* IGRADE variable is almost unchanged in value but remains statistically insignificant.

[20] Note that our estimation procedure corrects for the fact that companies may enter our database repeatedly. Clustering of standard errors reduces reported significance levels, so that our results are relatively conservative.

In order to allow for the time series of coefficients to follow a macroeconomic cycle, we also included a business cycle dummy, labeled BCYCLE, to constitute model 3. It equals one if the observation is from an NBER recession period, and zero otherwise. According to the NBER criterion, there were three recessions in our sample period: April 1982 to November 1982, July 1990 to March 1991, and March 2001 to November 2001. Results are given in Table 6, column 4. We find the business cycle dummy to have a positive, but statistically insignificant effect. Compared to model 1, the remaining results are unchanged. Overall, the tests performed so far lend support to Hypothesis 2. Although we find evidence of a time dependence in our data, this cannot fully explain the different abnormal returns in the two subperiods.

As the SEC's introduction of the Regulation Fair Disclosure on October 23, 2000 falls into our observation period, we also control for this event by including a dummy variable RegFD in model 4. Regulation Fair Disclosure prohibits U.S. public companies from making selective, non-public disclosures to favored investment professionals. Rating agencies, however, are exempted from this rule, which seems to improve the ratings' informational content: Jorion, Liu, and Shi (2005) even find significant positive abnormal returns following upgrades in the aftermath of this regulatory change. As can be seen from Table 6, column 5, however, this dummy has no explanatory power in our regression and leaves the earlier results unchanged.[21]

One further robustness check concerns the development of corporate financial risk over our sample period. Again, we use two distinct approaches. The first directly addresses the capital structure of the firms in our sample, while the second refers to the sample composition effect, which is relevant here because the mapping of rating notches into the probability space is non-linear.

In order to control for changes in the capital structure of the rated firms, which by itself may explain the increasing response of stock prices to a given rating in the post-watchlist era, we include two measures of leverage to obtain model 5: the ratio of

[21] Note that we use issuer ratings in our empirical analysis, while Jorion, Liu, and Shi (2005) use issue ratings. This may, at least partly, explain the insignificant coefficient.

short-term debt to total assets (SHORT), and the ratio of long-term debt to total assets (LONG).[22] Since the marginal costs of a rating change are directly proportional to the volume of debt financing, in particular short-term debt financing, we expect the coefficients to be negative for both variables.

The results are reported in Table 6, column 6. Again, the inclusion of the additional capital structure variables does not alter the sign nor the significance of the regressors of model 1. The adjusted R^2 of the extended specification rises from 2.52 percent in model 1 to 4.18 percent.[23] The leverage variables have the expected sign and turn out roughly identical in terms of economic significance. Since all other variables of the original model 1 are largely unaffected, we conclude that the increased leverage in the post-watchlist era affects CARs but is unable to explain all variation in the two samples.

A second, alternative test concentrates on the exponential relation between rating notches and probability of default. By using RCHANGE as a dependent variable in the basic model, we have implicitly assumed that the distribution of firms across rating notches is stationary over the entire period. If, however, the composition of our sample shifts over time to lower rating categories, and in these lower rating categories a one-notch rating change implies a larger increase in default probability, then a sheer sample composition effect may just as well yield the results that we have found. From earlier studies, we know that a rating improvement by one notch, say from Baa3 to Ba1, raises the probability of default from 0.52 percent to 0.81 percent. However, a rating change from Ba3 to B1, which is also one notch, raises the default probability from 2.69 percent to 4.04 percent, i.e., four times more than in the first case (Keenan, Hamilton, and Berthault, 2000). The exponential rise in default probability is particularly pronounced in the non-investment grade sector of the rating scale (Jorion and Zhang, 2007). To capture these effects, we include dummy variables in model 1 for each whole letter rating class (i.e., AA, A, BAA, BA, B), where the dummy variable equals one if the firm rating before the event falls into

[22]Our results do not change if we use, e.g., total debt scaled by the market value of the firm.

[23]Note that R^2 levels tend to be very low in these types of regressions, as can also be seen in Jorion, Liu, and Shi (2005) or Hand, Holthausen, and Leftwich (1992).

this rating category, and zero otherwise. This is model 6.

As can be seen from Table 6, column 7, the dummies are both statistically and economically significant. However, their inclusion does not change the overall results obtained in model 1,[24] but the absolute size of the coefficients (the economic significance) is altered. The effect of POST1991*RCHANGE decreases both in statistic and economic significance compared to the basic regression. We interpret this as evidence that there is, indeed, a sample composition effect, which partly explains the increased strength of the announcement effect in the post-watchlist era.[25] However, we are left with an unexplained part that we attribute to the enhanced informational value of the observed rating action. In sum, we find evidence consistent with Hypothesis 2.

3.3 Economic Rationale Behind Rating Reviews

3.3.1 Univariate Results

The test of Hypothesis 3 on the economic rationale behind the introduction of the watchlist procedure starts again with a univariate approach. Table 7 displays the CARs following from direct and watch-preceded rating changes in the post-watchlist period. Direct rating downgrades trigger a much stronger market reaction (-3.99 percent) than watch-preceded downgrades (-2.14 percent). The difference is also highly significant (at the 1 percent-level). For upgrades, no significant effects can be found.

It should be kept in mind, however, that the results so far considered only the "off-watch" effects. This procedure tends to underestimate the true stock market reaction to rating changes, because the anticipatory effect implicit in the price reaction to the announcement of a rating's addition to the watchlist has been neglected. Table 8 gives an indication of the omitted anticipation effects. As can be seen, the market has a strong

[24]One exception is that the variable POST1991*IGRADE becomes statistically significant at the 5 percent level.

[25]This result is strengthened by the fact that R^2 increases strongly from 2.52 percent in model 1 to 5.35 percent in model 5.

negative reaction to a watchlist addition with designation downgrade, and a significantly positive reaction to watchlist additions with direction uncertain and upgrade. However, since firms are usually put under review immediately after credit-relevant corporate events, any "on-watch" market reaction may be expected to be superposed by these corporate events (Keenans, Fons, and Carty, 1998; Linciano, 2004). Furthermore, as there is a strong dependence between the initial watchlist designation and the final resolution,[26] we believe that a simple summing up of on-watch and off-watch CARs is not a sensible approach. Additionally, if issuers are, indeed, affected by the watchlist procedure, their quality will change over the course of the review procedure, so that on- and off-watch effects do not relate to the same corporate entity and, therefore, should not be aggregated (Hirsch and Krahnen, 2007). However, it is interesting to see that even watchlist additions with indication uncertain trigger a significantly positive market reaction. This obviously contradicts the buying time argument, according to which, (at least neutral) watchlist assignment decisions should not be informative events.

In order to take the market reaction to the watchlist addition into account while not simply summing up non-comparable CAR-values, we conduct an additional univariate robustness test, where CARs have been measured using a longer event window, starting one day before the watchlist announcement and ending one day after the watchlist resolution.[27] To facilitate comparability, we use the mean length of the watchlist period in our sample as the length of the event window for direct rating changes as well. Results are displayed in Table 9. As can be seen, our former result is confirmed: The market reacts much more strongly to direct rating downgrades than to watch-preceded downgrades, with a strongly significant difference. Interestingly, this persistent result contradicts both the buying time argument and the implicit contracting hypothesis à la Boot, Milbourn, and Schmeits (2006).

[26]In our sample, for instance, the probability of a downgrade, given the firm is placed on watchlist with designation downgrade, is 0.64.

[27]In our sample, the watchlist spans a time period between 13 and 266 days. The median length is 79 days.

3.3.2 Multivariate Results

When testing Hypothesis 3 in a multivariate approach, we face a potential methodological problem: If the Boot, Milbourn, and Schmeits (2006) argument holds true, rating agencies will preselect firms for addition to the watchlist, so that the difference in effects from direct rating action versus watch-preceded rating action becomes endogenous.

In order to account for this preselection, we split our empirical model into two separate regressions. The first contains the agency's decision to put a firm on the watchlist, while the second captures the relation between the rating change and the market's reaction to it. Note that these two equations constitute a sequential equation model: the watchlist placement influences the observed cumulative abnormal returns but not vice versa. Such recursive equation models can be estimated consistently using OLS, provided that the error terms of the two equations are uncorrelated. Hence, we proceed as follows: First we use a logit regression, modeling the watchlist-addition decision of the rating agency. Second, we test Hypothesis 3 including all variables found significant in the first regression as control variables.

We commence our sample selection process by using all events with either a watchlist assignment with designation downgrade or with a direct downgrade in the post-watchlist period (October 1, 1991, to December 31, 2004). We exclude events with insufficient balance sheet information. In line with earlier work on capital structure (e.g., Flannery and Rangan, 2006), we exclude financial (SIC 6000-6999) and regulated (SIC 4900-4999) firms from the regression, since their capital structure differs markedly from those of other companies. This leads to a total sample of 4,351 observations.

To perform the logit regression, we create a dummy variable WATCHLIST PLACEMENT equal to one if the rating is placed on watchlist with direction downgrade, and zero otherwise. Independent variables referring to the firm's ability to react to the requirements set forth by the rating agency during review procedure are likely determinants of watchlist placement. Obvious candidates in this respect should be measures of financial and technological flexibility. We measure financial flexibility as cashflow divided by total assets (CASHFLOW). The more financial slack the firm has at hand (as measured by CASHFLOW), the more the firm is able to, e.g., retire debt. Therefore, we expect CASHFLOW to have a positive effect on watchlist placement. Following MacKay and Phillips (2005) we employ CAPITAL INTENSITY (fixed assets over number of employees) as a proxy for technological flexibility. The higher the capital intensity of the production technology used by the firm, the more difficult the recovery effort becomes. Thus, we expect a negative impact of this variable on the watchlist placement decision.

The likelihood of being placed on review should also be positively correlated with management quality. As suggested by Boot, Milbourn, and Schmeits (2006), we proxy management quality by SIZE, calculated as the logarithm of book value of assets. Furthermore, we include two variables for capital structure in our regression: LEVERAGE is measured as the book value of total debt over the market value of the firm, and SHORT gives the proportion of debt due within one year to total debt. We include LEVERAGE because we expect firms with higher leverage to be more exposed to an increase in the cost of debt caused by a rating deterioration. This should, in turn, increase the firm's willingness to comply with the agency's request. Thus, we expect a positive sign for the LEVERAGE variable. Likewise, the proportion of short-term debt should have a positive sign, since these liabilities expire within the next year, therefore making the consequences of a rating change more severe.

We also include the degree of competition in the industry (COMPETITION), even though we cannot offer a clear prediction with regard to this regressor's sign. While competition may be positively correlated with the willingness to engage in recovery effort, highly competitive markets may also lead to less financial slack, which reduces a firm's ability to exert recovery effort. The degree of competition is calculated as the number of firms per year from Compustat operating in the industry, where each industry is given by its four-digit SIC code. Finally, we also follow MacKay and Phillips (2005) and consider the variable RISK, calculated as the standard deviation of cashflow to total assets using a minimum of four annual observations. We expect firms with a higher variability of cashflow to be less able to engage in a recovery effort.[28]

Thus, the first regression model with respect to the watchlist decision is given by

$$\begin{aligned} WATCHLIST\,PLACEMENT_j &= \beta_0 + \beta_1\,SIZE_j + \beta_2\,LEVERAGE_j \\ &+\beta_3\,CASHFLOW_j + \beta_4\,SHORT_j + \beta_5\,CAPITAL\,INTENSITY_j \\ &+\beta_6\,COMPETITION_j + \beta_7\,RISK_j + \epsilon_j\ . \end{aligned} \tag{2}$$

All variables are evaluated at, or immediately before, the announcement date.[29] Results are reported in Table 10. All variables have the expected sign. Only the two capital structure variables, LEVERAGE and SHORT, display a negative instead of the expected positive sign. Moreover, financial flexibility (CASHFLOW), management quality (SIZE), and capital structure (LEVERAGE) turn out to be highly significant, while capital intensity, competition, and risk are insignificant. Nevertheless, the regression has a pseudo R^2 of only 7.9 percent, i.e., it is rather low.

We now turn to test Hypothesis 3 using the following model:

$$\begin{aligned} CAR_j &= \beta_0 + \beta_1\,RCHANGE_j + \beta_2\,IGRADE_j + \beta_3\,DAYS_j \\ &+\beta_4\,WATCHLIST^{*}RCHANGE_j + \beta_5\,WATCHLIST^{*}IGRADE_j \\ &+\beta_6\,WATCHLIST^{*}DAYS_j + \beta_7\,SIZE_j + \beta_8\,LEVERAGE_j \\ &\beta_9\,CASHFLOW_j + \epsilon_j\ . \end{aligned} \tag{3}$$

Here, the dependent variable is the cumulative abnormal return for firm j; RCHANGE, IGRADE, and DAYS are the same as in model 1; and the control variables SIZE, LEVERAGE, and CASHFLOW are the same as in the logit regression. We exclude SHORT because the variable only weakly influences the watchlist vs. direct downgrade decision. WATCHLIST is a dummy variable equal to one, if the firm is downgraded after a preceding watchlist, and zero otherwise. Our key variable is the interaction between WATCHLIST

[28] In an earlier version of the paper, we also included the value of rated debt outstanding. Yet, as results stay virtually the same, we decided to abstain from considering this additional regressor.

[29] The announcement date refers to either the date the firm is placed on watchlist with direction downgrade or the date of the direct downgrade.

and RCHANGE. An insignificant coefficient would support Hypothesis 3, that the watchlist was introduced to buy time for investigating firms with uncertain development, while a significantly negative coefficient would support the notion of implicit contracting between the rating agency and the firm during the watchlist episode.

Results are displayed in Table 11, column 2. RCHANGE and DAYS turn out to be statistically significant. This confirms our previous result that these two variables have a strong influence on the cumulative abnormal return. The coefficients of the other control variables correspond to our overall findings in the logit regression. An exception is the SIZE variable, which is insignificant, implying that the difference in abnormal return is not correlated with firm size (respectively with management quality).

Our key variable WATCHLIST*RCHANGE turns out to be positive and statistically significant at the 1 percent level. This implies that watch-preceded downgrades lead to a less negative CAR reaction than direct downgrades. Therefore, we can reject Hypothesis 3 of equivalent stock market reactions for direct and watch-preceded downgrades.

There is, however, a point of severe criticism that could potentially invalidate our result: the assumption of uncorrelated error terms of the two equations may be incorrect. Due to the rather low R^2 in the logit regression, correlation could be simply caused by omitted variables in the watchlist placement equation. If true, such a simultaneous equation problem would have to be estimated by two-stage least squares procedures. Therefore, in order to validate our results, we use an additional instrumental variable (IV) approach. For choosing an instrument, we follow Boot, Milbourn, and Schmeits (2006).[30] They propose the quality of management - approximated by firm size - as an important factor for watchlist placements. Indeed, in the watchlist placement regression displayed in Table 10, we found a significantly positive relationship between review placement and size, implying that size is correlated with the watchlist decision even after controlling for other potential explanatory variables. Additionally, due to the relatively large size of the

[30] A valid instrument *(i)* has to be uncorrelated with the error term of the second equation on the stock market reaction, *(ii)* should be correlated with the watchlist placement decision of the rating agency, and *(iii)* may not be included as an explanatory variable in the CAR equation (Murray, 2006).

rated entities in general, we expect no significant correlation between a specific company's size and the market reaction to rating changes. This is also supported by Faulkender and Petersen (2006) as well as our earlier OLS regression results in Table 11, column 2.

Table 11, column 3, presents the IV-regression results.[31] While the economic significance of variables is reduced, statistical significance is unchanged. The variable WATCHLIST, in particular, turns out to be positive and significant, again implying that watch-preceded downgrades lead to less negative CARs than direct downgrades. Results are unchanged when we use alternative approaches.[32] This corroborates our earlier conclusion that the introduction of the watchlist instrument seems not to be driven by arguments to buy time.

3.3.3 Interpretation

Interestingly, both univariate and multivariate analyses reject the buying-time hypothesis, as the informational content of direct rating changes seems much stronger than that of watch-preceded downgrades. However, our results do not support the interpretation of the implicit-contracting argument à la Boot, Milbourn, and Schmeits (2006) either. Rather, it seems to be the case that rating agencies do preselect borrowers for inclusion in the watchlist procedure, but cannot deliver very much new information via the watchlist resolution. Yet, Boot et al.'s conclusion requires each firm's credit quality and probability of a successful recovery effort to be fully precise pieces of public information (common knowledge). In reality, however, this assumption is more than dubious. In particular, the announcement of a watchlist placement is a clear signal to the market that the rating level

[31] Performing the same regression using the CARs including both the on-watchlist as well as the off-watchlist event does not change the results.

[32] As Jorion, Liu, and Shi (2005) find that SIZE and CARs are correlated after the introduction of regulation FD in October 2000, we re-estimated our initial IV-regression using only observations before the introduction of regulation FD in October 2000. This roughly cuts our sample in half. However, coefficients turn out to be not affected by this procedure. We also employed the IV-approach on the series of CARs calculated with the event window spreading over the total length of the watchlist. Again, results remain unchanged.

is going to be reassessed with a considerable likelihood of rating change. This naturally erodes the perceived precision of the outstanding rating and, hence, of public information about the firm's credit quality. At the same time, it is reasonable to believe that a company's capability of exerting a successful recovery effort definitely falls into the rating agency's area of expertise rather than being common knowledge.

From theoretical, experimental, and empirical research, it is known that the composition of private and public pieces of information strongly influences investors' behavior on financial markets (Morris and Shin, 2002, 2004; Allen, Morris, and Shin, 2006; Bannier, 2006; Angeletos and Pavan, 2007; Heinemann, Nagel, and Ockenfels, 2004, 2008). More recently, it has been found that less precise public information about financial claims induces investors to individually collect private information - particularly in the presence of coordination effects (Bannier and Tyrell, 2006). Translated into the present context, it may be reasonable to assume that the announcement of a watchlist addition triggers investors to collect more precise private information. Provided that this additional information does not consist of pure noise, the market reaction to the subsequent (watch-preceded) rating change should then be muted, as it will have been partly anticipated.

According to this argument, rating reviews do serve as implicit contracts, because ratings continue to coordinate investors' actions. However, over the course of the watchlist procedure, market participants collect additional, private information, so that watch-preceded rating actions are, to some extent, individually expected. In sum, our empirical results may be interpreted as follows: Direct rating downgrades signal that the rating agency is not convinced that this borrower is *capable or willing* to uphold its initial credit quality. A watchlist-placement decision signals exactly the opposite. Both a direct downgrade and a review placement contain new information, as it is within the agency's expertise to assess the impact of a negative shock on a firm's credit quality in relation to the firm's capability to deal with it. The subsequent decision on a rating confirmation, respectively watch-preceded rating downgrade, finally conveys information on whether or not the recovery effort was *truly successful.* As the rating assignment induces investors to collect additional private information on the firm's credit standing and capability to recover from the shock, the watchlist resolution decision will be partly anticipated and, hence, delivers a lower informational content. Note that this interpretation is also supported by our empirical results: As can be seen from Table 11, column 2, for watchlist placements, the DAYS variable does not display any significant effect, implying that the market more or less correctly anticipates the eventual rating action.

4 Conclusion

Our study examined whether Moody's formal introduction of the watchlist procedure in 1991 influenced the informational content of credit ratings and possibly extended the economic role that rating agencies play in financial markets. We find that after the introduction of the review instrument, rating downgrades lead to stronger market reactions than in the pre-watchlist period. Furthermore, our empirical study lends support to the hypothesis that the watchlist procedure allows rating agencies to enter into an implicit contract with the rated firms, as has been suggested by Boot, Milbourn, and Schmeits (2006). Consequently, rating reviews add a finer level of detail to information in financial markets: whereas direct rating downgrades make a statement on issuers' lack of *capability or willingness* to sustain their credit quality, watchlist downgrades inform market participants of borrowers' lack of *success* in doing so.

In this respect, our study confirms the initial statement by Moody's that rating changes for issuers placed on the watchlist are different from those not preceded by a review procedure. The watchlist instrument seems to have developed into an active monitoring device that allows the rating agencies to exert real pressure on the reviewed companies. An interesting question arises from this observation: Was the review procedure introduced with this objective or did it unintentionally develop into such a specific instrument? While our study did not focus on this specific question and, hence, cannot provide an answer, we would like to point out that watchlists seem to have a different impact with respect to sovereign ratings (Hill and Faff, 2007). It is possible that the implicit contracting feature does not operate in an environment where the counterparty consists of a relatively undefined group of politicians and statesmen instead of the much smaller management circle, as in the case of corporate ratings. Although outside the scope of this paper, interesting general conclusions might be drawn from this comparison with respect to the objective function of credit rating agencies.

References

Blume, Marshall E., Felix Lim, and A. Craig MacKinlay, 1998, The declining credit quality of U.S. corporate debt: Myth or reality?, *Journal of Finance* 53, 1389–1413.

Boehmer, Ekkehart, Jim Masumeci, and Anette B. Poulsen, 1991, Event-study methodology under conditions of event-induced variance, *Journal of Financial Economics* 30, 253–272.

Boot, Arnoud W. A., Todd. T. Milbourn, and Anjolein Schmeits, 2006, Credit ratings as coordination mechanism, *Review of Financial Studies* 19, 81–118.

Cantor, Richard, 2004, An introduction to recent research on credit ratings, *Journal of Banking and Finance* 28, 2565–2573.

Faulkender, Michael, and Mitchell A. Petersen, 2006, Does the Source of Capital Affect Capital Structure?, *Review of Financial Studies* 19, 45–79.

Flannery, Mark J., and Kasturi P. Rangan, 2006, Partial adjustment toward target capital structures, *Journal of Financial Economics* 79, 469–506.

Frost, Carol Ann, 2006, Credit Rating Agencies in Capital Markets: A Review of Research Evidence on Selected Criticisms of the Agencies, mimeo.

Goh, J., and H. Ederington, 1999, Cross-Sectional Variation in the Stock Market Reaction to Bond Rating Changes, *Quarterly Review of Economics and Finance* 39, 101–112.

Goh, Jeremy, and Louis Ederington, 1993, Is a bond rating downgrade bad news, good news, or no news for stockholders?, *Journal of Finance* 48, 2001–2008.

Hamilton, David T., 2005, Moody‘s senior rating algorithm and estimated senior ratings, *Moody‘s Investors Service.*

Hamilton, David T., and Richard Cantor, 2004, Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History, *Moody's Investors Service, Special Comment.*

Hand, John R. M., Robert W. Holthausen, and Richard W. Leftwich, 1992, The Effect of Bond Rating Agency Announcements on Bond and Stock Prices, *Journal of Finance* 47, 733–752.

Hill, C., 2004, Regulating the Rating Agencies, *Washington University Law Quarterly* 82, 42–95.

Hill, Paula, and Robert Faff, 2007, Do Credit Watch Procedures Affect the Information Content of Sovereign Credit Rating Changes?, Mimeo.

Hirsch, Christian, and Jan Pieter Krahnen, 2007, A pirmer on rating agencies as monitors: an analysis of the watchlist period, Working Paper.

Holthausen, Robert W., and Richard W. Leftwich, 1986, The effect of bond rating changes on common stock prices, *Journal of Financial Economics* 17, 57–89.

Hull, J., M. Predescu, and A. White, 2004, The Relationship Between Credit Default Swap Spreads, Bond Yields, and Credit Rating Announcements, *Journal of Banking and Finance* 28, 2789–2811.

Jorion, Philippe, Zhu Liu, and Charles Shi, 2005, Informational effects of regulation FD: evidence from rating changes, *Journal of Financial Economics* 76, 309–330.

Jorion, Phillippe, and Gaiyan Zhang, 2007, Information Effects of Bond Rating Changes: The Role of the Rating Prior to the Announcement, *Journal of Fixed Income* 16, 45–59.

Keenan, Sean C., Jerome S. Fons, and Lea V. Carty, 1998, An historical analysis of Moody‘s watchlist, *Moody‘s Investors Service.*

Keenan, Sean C., David T. Hamilton, and Alexandra Berthault, 2000, Historical default rates of corporate bond issuers, 1920-1999, *Moody‘s Investors Service.*

Löffler, Gunter, 2004a, An anatomy of rating through the cycle, *Journal of Banking and Finance* 28, 695–720.

Löffler, Gunter, 2004b, Ratings versus market-based measures of default risk in portfolio governance, *Journal of Banking and Finance* 28, 2715–2746.

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.

Linciano, Nadia, 2004, The Reaction of Stock Prices to Rating Changes, mimeo.

MacKay, Peter, and Gordon M. Phillips, 2005, How does industry affect firm financial structure?, *Review of Financial Studies* 18, 1433–1466.

MacKinlay, A.Craig, 1997, Event studies in Economics and Finance, *Journal of Economic Literature* 35, 13–39.

Murray, Michael P., 2006, Avoiding Invalid Instruments and Coping with Weak Instruments, *Journal of Economic Perspectives* 20, 111–132.

Partnoy, Frank, 2002, The Paradox of Credit Ratings, in Richard M. Levich, Giovanni Majnoni, and Carmen M. Reinhart, eds.: *Ratings, Rating Agencies, and the Global Financial System* (Kluwer Academic Publishers, Norwell, Mass.).

Purda, Lynnette D., 2006, Stock Market Reaction to Anticipated Versus Surprise Rating Changes, mimeo.

Robbe, Paul, and Roland Mahieu, 2005, Are the standards too poor? An empirical analysis of the timeliness and predictability of credit rating changes, Working Paper.

Vassalou, Maria, and Yuhang Xing, 2005, Abnormal equity returns following downgrades, Working Paper.

Wansley, James W., and Terrence M. Clauretie, 1985, The impact of creditwatch placement on equity returns and bond prices, *The Journal of Financial Research* 8, 31–42.

Tables

Table 1: Distribution and size of rating changes by year - downgrades

The table contains number and mean size of rating downgrades for each year of the sample. The sample period after October 1, 1991, includes direct changes as well as watchlist-preceded rating changes. Size reports the mean of all rating changes (in notches) in a given year.

Year	All downgrades		Direct downgrades		Watchlist-preceded downgrades	
	#	Size	#	Size	#	Size
1982	81	1.8	81	1.8	-	-
1983	62	1.46	62	1.46	-	-
1984	66	1.8	66	1.8	-	-
1985	83	1.74	83	1.74	-	-
1986	125	2.13	125	2.13	-	-
1987	96	2.01	96	2.01	-	-
1988	109	2.11	109	2.11	-	-
1989	129	1.86	129	1.86	-	-
1990	191	1.65	191	1.65	-	-
1991	88	1.55	88	1.55	-	-
1992	66	1.4	59	1.44	7	1.14
1993	77	1.53	52	1.53	25	1.52
1994	70	1.37	48	1.33	22	1.45
1995	99	1.53	66	1.57	33	1.45
1996	87	1.5	54	1.46	33	1.57
1997	81	1.45	51	1.45	30	1.46
1998	175	1.57	119	1.53	56	1.64
1999	217	1.64	141	1.69	76	1.55
2000	229	1.65	149	1.63	80	1.67
2001	350	1.74	231	1.74	119	1.73
2002	360	1.66	185	1.57	175	1.75
2003	213	1.59	91	1.59	122	1.59
2004	126	1.34	53	1.39	73	1.3
PRE1991	1030	1.83	1030	1.83	-	-
POST1991	2150	1.59	1299	1.58	851	1.6
Total	3180	1.67	2329	1.69	851	1.6

Table 2: Distribution and size of rating changes by year - upgrades

The table contains number and mean size of rating upgrades for each year of the sample. The sample period after October 1, 1991, includes direct changes as well as watchlist-preceded rating changes. Size reports the mean of all rating changes (in notches) in a given year.

Year	All upgrades		Direct upgrades		Watchlist-preceded upgrades	
	#	Size	#	Size	#	Size
1982	18	1.44	18	1.44	-	-
1983	45	1.68	45	1.68	-	-
1984	42	1.45	42	1.45	-	-
1985	54	1.53	54	1.53	-	-
1986	48	1.47	48	1.47	-	-
1987	46	1.82	46	1.82	-	-
1988	60	1.8	60	1.8	-	-
1989	46	1.39	46	1.39	-	-
1990	34	1.47	34	1.47	-	-
1991	25	1.56	25	1.56	-	-
1992	48	1.41	42	1.38	6	1.66
1993	84	1.51	61	1.54	23	1.43
1994	92	1.32	67	1.23	25	1.56
1995	79	1.32	59	1.22	20	1.65
1996	127	1.25	98	1.26	29	1.2
1997	102	1.18	84	1.15	18	1.33
1998	107	1.37	76	1.27	31	1.61
1999	89	1.22	64	1.17	25	1.36
2000	92	1.67	53	1.22	39	2.28
2001	81	1.23	56	1.14	25	1.56
2002	60	1.23	41	1.21	19	1.26
2003	97	1.17	51	1.15	46	1.19
2004	133	1.26	77	1.18	56	1.37
PRE1991	418	1.58	418	1.58	-	-
POST1991	1191	1.31	829	1.24	362	1.49
Total	1609	1.38	1247	1.35	362	1.49

Table 3: Summary of rating downgrades by absolute magnitude

The table presents the number as well as the proportion of all 3180 rating downgrades in our sample by absolute magnitude of the rating change. The sample is split into two periods: The pre-watchlist period from April 26, 1982, to September 30, 1991 (PRE1991), and the post-watchlist period from October 1, 1991, to December 31, 2004 (POST1991). Rating change is the absolute value of rating change in notches.

Rating change	PRE1991		POST1991			
			All		From watchlist	
	#	%	#	%	#	%
1	503	48.83	1254	58.33	503	59.11
2	336	32.62	623	28.98	233	27.38
3	123	11.94	193	8.98	76	9.93
4	39	3.79	55	2.56	26	3.06
5	11	1.07	17	0.79	10	1.18
6	7	0.68	6	0.28	3	0.35
7	7	0.68	2	0.09	-	-
8	1	0.1	-	-	-	-
>8	3	0.3	-	-	-	-
Total	1030	100	2150	100	851	100

Table 4: Summary of rating upgrades by absolute magnitude

The table presents the number as well as the proportion of all 1609 rating upgrades in our sample by absolute magnitude of the rating change. The sample is split into two periods: The pre-watchlist period from April 26, 198,2 to September 30, 1991 (PRE1991), and the post-watchlist period from October 1, 1991, to December 31, 2004 (POST1991). Rating change is the absolute value of rating change in notches.

Rating change	PRE1991		POST1991			
			All		From watchlist	
	#	%	#	%	#	%
1	259	61.96	952	79.93	263	72.65
2	114	27.27	180	15.11	66	18.23
3	26	6.22	26	2.18	15	4.14
4	10	2.39	14	1.18	9	2.49
5	4	0.96	9	0.76	4	1.1
6	2	0.48	4	0.34	1	0.28
7	2	0.48	2	0.17	1	0.28
8	-	-	-	-	-	-
>8	1	0.24	4	0.33	3	0.84
Total	418	100	1191	100	362	100

Table 5: Stock market response to rating changes: PRE1991/POST1991

The table provides the cumulative abnormal returns for both direct and watch-preceded downgrades and upgrades. The sample consists of only uncontaminated rating events in the period between April 26, 1982, and December 31, 2004. PRE1991 is used with reference to the pre-watchlist period from April 26, 1982, to September 30, 1991, while POST1991 denotes the post-watchlist era from October 1, 1991 to December 31, 2004. Panel A refers to downgrades, Panel B to upgrades. The cumulative abnormal return (CAR) is calculated over a three-day event window (-1,+1) around the date the rating change becomes effective. The CAR is the cumulative abnormal stock return minus the return of the market portfolio, where the market portfolio is given by the value-weighted portfolio from CRSP. Wilcoxon T values are given below the median and t-values below the mean. ***, **, and* indicate significance at the 1%, 5%, and 10% level. Mean and median values are tested using one-sided t-test and Wilcoxon T test, respectively.

	Panel A: Downgrades		
	Mean	Median	$CAR < 0$ (%)
PRE1991	-2.18	-0.72	59.12
	(-6.55)***	(-6.76)***	
POST1991	-3.26	-0.91	58.27
	(-10.84)***	(-10.09)***	
Difference	-1.08	-0.19	-0.85
(POST1991-PRE1991)	(-2.39)***	(-1.26)	
	Panel B: Upgrades		
	Mean	Median	$CAR < 0$ (%)
PRE1991	0.93	-0.83	48.56
	(0.52)	(-0.16)	
POST1991	0.14	-0.53	49.03
	(0.11)	(-0.59)	
Difference	-0.79	-0.3	0.47
(POST1991-PRE1991)	(-0.35)	(-0.21)	

Table 6: The effect of watchlist introduction on the stock market reaction to rating downgrades

The sample consists of 3180 non-contaminated downgrades in the period between April 26, 1982 and December 31, 2004. The dependent variable is the cumulative abnormal return (CAR). RCHANGE is the absolute value of rating change in notches; IGRADE is a dummy variable equal to 1 if the rating downgrade crosses the investment grade boundary, and 0 otherwise; DAYS is the log of the number of days since the last rating change (downgrades as well as upgrades); POST1991 is a dummy variable equal to 1 if the observation is from the watchlist period (October 1, 1991 to December 31, 2004), and 0 otherwise; BCYCLE is a dummy variable equal to 1 if the rating change is from a time period defined as recession by NBER, and 0 otherwise; SHORT is calculated as short-term debt (Compustat item #34)/book value of total assets (#6); LONG is calculated as long-term debt (#9)/book value of total assets (#6). AA, A, BAA, BA, and B are dummy variables equal to 1 if the rating of the observation before the rating change is within the respective rating class, and 0 otherwise. Note, that AA also includes rating changes coming from AAA. Rating categories CAA and below serve as the reference category. ***, **, and * indicate significance at the 1%, 5%, and 10% level. t-values are given in parenthesis. All results are obtained accounting for clustering in the sample.

Explanatory variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
INTERCEPT	-0.009*	0.034	-0.01*	-0.008	0.019**	-0.068***
	(-1.67)	(0.28)	(-1.68)	(-1.33)	(2.36)	(-4.30)
RCHANGE	-0.006*	-0.008*	-0.006*	-0.006*	-0.004	-0.008**
	(-1.82)	(-2.08)	(-1.81)	(-1.95)	(-1.58)	(-2.48)
IGRADE	0.017**	0.017*	0.017*	0.018**	0.015	0.014
	(1.97)	(1.93)	(1.96)	(2.00)	(1.45)	(1.51)
DAYS	-0.001	0	0	-0.001	-0.003	0
	(-1.02)	(-0.59)	(-1.05)	(-1.16)	(-0.32)	(-0.77)
POST1991*RCHANGE	-0.017***	-0.014**	-0.017***	-0.016***	-0.017***	-0.010**
	(-3.97)	(-2.08)	(-3.98)	(3.63)	(-3.66)	(-2.29)
POST1991*IGRADE	-0.009	-0.008	-0.009	-0.008	-0.014	-0.021
	(-0.69)	(-0.62)	(-0.68)	(-0.68)	(-0.79)	(-1.62)
POST1991*DAYS	0.004***	0.004***	0.004***	0.005***	0.003***	0.004***
	(3.73)	(3.39)	(3.75)	(3.89)	(2.78)	(3.32)
BCYCLE			0.013			
			(0.73)			
RegFD				-0.006		
				(-1.18)		
SHORT					-0.051***	
					(-4.48)	
LONG					-0.057***	
					(-3.73)	
AA (or above)						0.082***
						(5.68)
A						0.077
						(5.43)
BAA						0.065***
						(4.55)
BA						0.06***
						(4.718)
B						0.033**
						(2.26)
Year dummies	no	yes	no	no	no	no
$Adj.R^2(\%)$	2.52	3.48	2.53	2.52	4.18	5.35
F	5.64***	2.89***	4.92***	5.37***	12.61***	11.35***
Observations	3180	3180	3180	3180	2849	3180
Clusters	1532	1532	1532	1532	1442	1532

Table 7: Stock market response to rating changes: direct / watchlist ratings

The table provides the cumulative abnormal returns following direct and watch-preceded downgrades and upgrades. The sample consists of only uncontaminated rating events in the post-watchlist period from October 1, 1991, to December 31, 2004. The cumulative abnormal return (CAR) is calculated over a three-day event window (-1,+1) around the date the rating change becomes effective. The CAR is the cumulative abnormal stock return minus the return of the market portfolio, where the market portfolio is given by the value-weighted portfolio from CRSP. Wilcoxon T values are given below the median and t-values below the mean. ***, **, and* indicate significance at the 1%, 5%, and 10% level. Mean and median values are tested using one-sided t-test and Wilcoxon T test, respectively.

Panel A: Downgrades			
	Mean	Median	$CAR < 0$ (%)
Direct	-3.99	-1.49	59.76
	(-9.44)***	(-9.05)***	
From watchlist	-2.14	-0.4	55.22
	(-5.38)***	(-4.57)***	
Difference	1.85	1.09	-4.54
(From watchlist-Direct)	(3.18)***	(3.06)***	
Panel B: Upgrades			
	Mean	Median	$CAR < 0$ (%)
Direct	-0.06	0	49.32
	(-0.37)	(-0.53)	
From watchlist	0.18	-0.09	47.51
	(0.76)	(-0.19)	
Difference	0.24	-0.09	-1.80
(From watchlist-Direct)	(0.84)	(-0.15)	

Table 8: Stock market response to watchlist additions

The table provides the cumulative abnormal returns following watchlist additions. We differentiate between watchlist additions with direction downgrade, uncertain and upgrade. The cumulative abnormal return (CAR) is calculated over a three-day event window (-1,+1) around the date the watchlist addition is announced. The CAR is the cumulative abnormal stock return minus the return of the market portfolio, where the market portfolio is given by the value-weighted portfolio from CRSP. P-values are given below the mean. ***, **, and* indicate significance at the 1%, 5%, and 10% level.

	#	Mean	Median	$CAR < 0$ (%)
Down	1083	-3.41	-0.86	58.26
		(-8.18)***	(-7.65)***	
Uncertain	56	1.88	0.48	58.78
		(2.40)**	(0.32)	
Up	560	1.62	0.21	53.93
		(4.43)***	(2.20)**	

Table 9: Stock market response to rating changes: direct / watchlist ratings

The table provides the cumulative abnormal returns following downgrades. The sample consists of only uncontaminated rating events in the post-watchlist period from October 1, 1991, to December 31, 2004. The cumulative abnormal return (CAR) is calculated over a event window beginning one day before watchlist placement and ending one day after watchlist resolution for firms coming from watchlist. For direct downgrades the event window is set as the the median length of the watchlist period in our sample. The CAR is the cumulative abnormal stock return minus the return of the market portfolio, where the market portfolio is given by the value-weighted portfolio from CRSP. Wilcoxon T values are given below the median and t-values below the mean. ***, **, and* indicate significance at the 1%, 5%, and 10% level. Mean and median values are tested using one-sided t-test and Wilcoxon T test, respectively.

	Mean	Median	$CAR < 0$ (%)
Direct	-13.96	-9.36	60.82
	(-8.71)***	(-7.97)***	
From watchlist	-0.96	-1.65	60.1
	(-0.56)	(-2.22)**	
Difference	13	7.71	0.72
	(-5.57)***	(-4.29)***	
(From watchlist-direct)			

Table 10: Which firms are placed on watchlist?

The sample consists of 4351 direct downgrades and watchlist placements with direction downgrade in the watchlist period between October 1, 1991 and December 31 2004, respectively. Ratings are issuer ratings provided by Moody's. The dependent variable is a dummy variable equal to 1 if the observation is placed on watchlist with designation downgrade, and 0 otherwise. SIZE is calculated as log of book value of total assets (Compustat item #6); LEVERAGE is calculated as total debt (#9 + #34)/(total debt (#9 + #34) + market value of equity (#199)); CASHFLOW is calculated as earnings before depreciation (#18) / book value of total assets (#6); SHORT is calculated as short-term debt(#34)/total debt (#9 + #34); CAPITAL INTENSITY is calculated as property, plant, and equipment (#8)/ number of employees(#29); COMPETITION is the number of firms in a given industry in a given year reported by Compustat, where industry is defined by the 4-digit SIC code; finally, RISK is calculated as the standard deviation of the CASHFLOW variable defined above. *** ,** , and * indicate significance at the 1%, 5%, and 10% level. z-values are given in parenthesis.

Explanatory variables	Coefficient (z-statistic)
INTERCEPT	-2.252***
	(-10.16)
SIZE	0.254***
	(10.52)
LEVERAGE	-1.477***
	(-9.46)
CASHFLOW	1.989***
	(5.74)
SHORT	-0.334*
	(-1.84)
CAPITAL INTENSITY	-0.016
	(-0.43)
COMPETITION	0
	(0.17)
RISK	-0.273
	(-0.92)
$PseudoR^2(\%)$	7.9
LR χ^2	417.71***
Observations	4351

Table 11: CAR for firms coming from watchlist vs. direct downgrades

The sample consists of 2150 non-contaminated downgrades in the period between October 1, 1991 and December 31, 2004. Ratings are issuer ratings provided by Moody's. The sample includes direct downgrades as well as downgrades following watchlist placements. The results in colum 2 (3) are obtained using the OLS (IV) estimation methods, where the IV approach uses SIZE as the instrument. The dependent variable is the cumulative abnormal return. RCHANGE is the absolute value of rating change in notches; IGRADE is a dummy variable equal to 1 if the rating downgrade crosses the investment grade boundary, and 0 otherwise; DAYS is the log of the number of days since the last rating change (downgrades as well as upgrades); WATCHLIST is a dummy variable equal to 1 if the rating change follows a watchlist placement, and 0 otherwise; SIZE is calculated as log book value of total assets (Compustat item #6); LEVERAGE is calculated as total debt (#9 + #34)/(total debt (#9 + #34) + market value of equity (#199)); finally CASHFLOW is calculated as earnings before depreciation (#18) / book value of total assets (#6). ***, **, and * indicate significance at the 1%, 5%, and 10% level. t-values are given in parenthesis.

Explanatory variables	Model 1	Model 2
INTERCEPT	-0.006	-0.031***
	(-0.34)	(-2.93)
RCHANGE	-0.024***	-0.024***
	(-6.12)	(-7.16)
IGRADE	0	0.010
	(-0.02)	(0.99)
DAYS	0.004***	0.003***
	(3.38)	(3.00)
WATCHLIST		0.063***
		(2.87)
WATCHLIST*RCHANGE	0.010**	
	(2.22)	
WATCHLIST*IGRADE	0.004	
	(0.20)	
WATCHLIST*DAYS	-0.001	
	(-1.13)	
SIZE	0.001	
	(0.98)	
LEVERAGE	-0.045***	
	(-3.60)	
CASHFLOW	0.091***	
	(4.21)	
$Adj.R^2(\%)$	5.29	0.006
F	14.11***	18.58***
Method	OLS	IV
Observations	2111	2125