The added value of Rating Outlooks and Rating Reviews to corporate bond ratings

Edward I. Altman [1] and Herbert A. Rijken [2]

May 2007

JEL categories: G20, G33

[1] NYU Salomon Center, Leonard N. Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012, USA. email: ealtman@stern.nyu.edu
[2] Free University, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands. email: hrijken@feweb.vu.nl

Acknowledgement

We thank Moody's for funding this study and for providing an extended version of data on outlooks and rating Reviews for Moody's corporate issuer credit ratings**.**

Abstract

A prime objective set by credit rating agencies is to achieve stability of their corporate bond ratings by a rating through-the-cycle methodology. However rating stability is at the expense of rating timeliness and default prediction performance. To mitigate this tension between stability and accuracy, agencies publish outlooks – rating Outlooks and rating Reviews (Watchlist) – in addition to corporate bond ratings. Outlooks provide an indication of the direction and timing of likely rating changes in the future.

In this study we quantify the added value of outlooks to corporate bond ratings and investigate to what extent outlooks are able to compensate for the disadvantages of rating stability. Results show that outlooks do indeed partially close the gap between the agencies' through-the-cycle perspective and the investor's point-in-time perspective. After adjusting ratings by their outlooks default prediction performance does improve, but only slightly, especially for short prediction horizons. Default prediction performance and point-in-time characteristics of adjusted ratings could be enhanced even further if outlooks become more accurate measures of credit risk.

We conclude that accuracy in credit risk information signaled by outlooks can be improved most likely by standardizing credit risk information in the outlook scale. Credit risk dispersion in the outlook scale could be enhanced by a factor two. Perhaps, as outlooks are not intended to quantify credit risk information explicitly, agencies have not standardized credit risk information in the outlook assignment process. We also considered the impact of the explicit timing objective of outlooks which to some degree overrides credit risk information signaled by outlooks. From a pure credit risk perspective the timing objective of outlooks shortens durations for rating Reviews, lengthens durations for rating Outlooks and partially circumvents the use of rating Outlooks as "intermediate" states between Stable Outlooks and rating Reviews. However this specific outlook migration policy has little effect on credit risk accuracy in outlooks.

1 Introduction

In addition to their corporate issuer credit ratings, agencies provide rating Outlooks and rating Reviews (Watchlist). Outlooks signal the likely direction of a rating migration in one to two years' time. In response to an event or an abrupt break in a trend, a corporate-issuer credit rating is placed on a Watchlist by Moody's, on CreditWatch by Standard & Poor's or on a rating Watch by Fitch. In these cases ratings are said to be under review and the outcome is disclosed typically within 90 days. In this paper, outlooks refer both to rating Outlooks and rating Reviews. Ratings refer to corporate issuer credit ratings.

Outlooks have become an important source of credit risk information to investors in addition to corporate bond ratings. Ratings are the outcome of a through-the-cycle methodology which makes them relative stable, insensitive to temporary credit risk fluctuations and oriented towards the long-term. A drawback of this methodology is a lower timeliness compared to a one year point-in-time investor's perspective, which most investors have. Investors monitor outlooks to compensate for this loss in timely credit risk information. According to a survey conducted by Moody's in 2002, "*Investors agree with the goal of more timely rating actions including shorter review periods. However they use and appreciate the rating Review and rating Outlook signaling process; they derive substantial information from them.*" In their response to this survey Moody's intended to improve rating timeliness by more frequent (internal) review of ratings and to retain the provision of outlooks. Without having to change their through-the-cycle methodology, agencies can fulfill the investors' desire for more timely credit risk information by issuing outlooks in addition to ratings.

Strictly according to their definition, outlooks are indications of the likely direction of a rating migration in the short or medium term. Agencies do not pretend to specify their expectations for size and probability of a potential rating migration. However, from historical data one can compute the average rating migration at the resolution of outlooks. The outcome of a Moody's DOWN Review is on average a rating migration of -1.0 notch steps. Moody's NEG, STA, POS Outlooks and UP Reviews result on average in a rating migration of -0.4, -0.1, +0.2 and +1.0 notch steps. Keeman et al. (1998) record the informational value of rating Reviews by constructing rating migration matrices conditional to whether issuers are held on a rating review.

Although outlooks are not meant to be a correction for ratings in the first place, one can use them as a secondary credit risk measure on top of the rating scale and adjust ratings by their outlooks. The adjustment of ratings is done by adding or subtracting one or two notch steps depending on the type and sign of outlooks. A study by Hamilton and Cantor (2004) reveals that adjustment of ratings significantly improves default prediction performance. The accuracy ratio increases by 4% for a one year prediction horizon. Another interesting aspect of adjusted ratings is the absence of serial correlation, which is profoundly present in ratings (see Altman and Kao, 1992). In a follow up study Hamilton and Cantor (2005) show that outlooks are able to explain differences between actual ratings and ratings implied by CDS spreads. This finding indicates that adjusted ratings offer a more timely reflection of the investor's short-term point-in-time perspective on credit risk.

So far most studies on outlooks are event studies, testing whether changes in outlooks signal new information to the market. As outlooks are secondary to ratings these studies also test the information value of rating migrations. Recent surveys reveal that investors are not satisfied with the timeliness of ratings (Association for Financial Professionals (2002), Ellis (1998) and Baker and Mansi (2002)). So rating migrations might not disclose new information to the financial markets. Norden and Weber (2004) indeed find no response of stock and CDS markets to rating downgrades. Hull, Predescu and White (2004) confirm these results for the CDS market. However, studies with more dated stock market datasets - ending before 1998 - show a negative impact of rating downgrades on pricing (for a recent literature overview see Steiner and Heinke (2001)).

For outlook announcements literature largely agrees on the negative impact of DOWN Review announcements on stock pricing and the absence of a significant response to NEG and POS Outlook announcements. However, for UP Review announcements results are mixed. For both UP and DOWN Review announcements Hand, Holthausen and Leftwich (1992) find significant excess bond returns, while more recently Steiner and Heinke (2001) and Hull et al. (2004) find only a significant response to DOWN Review announcements.

In this paper we report the results of an empirical study on the added value of outlooks to corporate bond ratings. Key in our study is the estimation of credit scoring models for outlook prediction, rating prediction and default prediction and the simulation of ratings and outlooks by equivalent credit model ratings and outlooks. Insight in the added value of outlooks is obtained by comparing parameters of various credit scoring models, quantifying credit risk dispersion in the outlook scale by variations in credit scores, benchmarking actual outlook dynamics with simulated outlook dynamics and comparing properties of adjusted ratings vs. unadjusted ratings.

First, in order to reveal the credit risk nature of outlooks we compare parameters of rating and outlook prediction models with parameters of default prediction models for various prediction horizons. In contrast to ratings, which are intended to be insensitive to temporary fluctuations in credit quality, outlooks appear to be sensitive to more volatile credit risk variables – even more sensitive than would be expected from a one-year point-in-time perspective, which most investors have. Outlooks, especially rating Reviews, are mainly driven by events and breaks in trends.

Second, we quantify credit risk information signaled by outlooks. For this purpose we link the outlook scale with the (notch) rating scale. In terms of notch steps the creditworthiness of issuers with a DOWN Review is on average positioned 3 notch steps below the average creditworthiness of all issuers in a given rating class. For NEG, STA and POS Outlooks and UP Reviews these figures are respectively -1.5, 0.5, 1.5 and 2 notch steps. These figures are not consistent in time and are conditional to a rating migration event in the near future or the near past. Moreover, the credit risk dispersion in the outlook scale is low compared to the credit risk dispersion in the outlook scale of simulated outlooks. Simulated outlooks are obtained by ranking credit scores of outlook prediction models and converting these scores to outlooks in such a way that the distributions of actual and simulated outlooks match. The lower credit risk dispersion in the actual outlook scale suggests that the accuracy of credit risk information signaled by outlook ratings could be improved. Credit risk dispersion in the outlook scale could be enhanced by a factor two. Perhaps, as outlooks are not intended to quantify credit risk information explicitly, agencies do not standardize credit risk information signaled by outlooks.

Third, we investigate the agencies' outlook migration policy. Therefore we compare dynamic properties of simulated outlooks with actual outlooks. In contrast to actual outlooks, the dynamics of simulated outlooks is not driven by an explicit timing objective and only reflects the dynamics of credit scores. In the benchmark we show that outlook migrations are heavily concentrated before and at the rating migration event – more than one would expect from a pure credit risk perspective. In addition the outlook migration policy shortens durations for rating Reviews, lengthens durations for rating Outlooks and partially circumvents the use of rating Outlooks as "intermediate" states between Stable Outlooks and rating Reviews.

Fourth, we test the ability of outlooks to compensate for the disadvantages of rating stability and the ability to add more timely credit risk information to ratings. For this purpose we adjust ratings by their actual outlooks and their simulated outlooks. The adjustment is based on the linkage of the outlook scale to the (notch) rating scale. We test to what extent these adjusted ratings have a more point-in-time character and a better default prediction performance. Ratings adjusted by their actual outlooks have a better default prediction performance and their character is halfway between through-the-cycle and point-in-time. However, ratings adjusted by their (benchmark) simulated outlooks show a better default prediction performance and do have a point-in-time character. We suggest that a lack of credit risk standardization in the actual outlook scale confines the default prediction performance and the point-in-time characteristics of adjusted ratings. The outlook migration policy has much less impact.

The study is carried out with Moody's Outlook and Watchlist data. We are not aware of a reason why empirical results and conclusions presented here should not apply for outlooks of Standard & Poor's and Fitch as these agencies have disclosed a similar policy on rating Outlooks and rating Reviews. Discussions and conclusions can therefore be generalized to all outlooks for ratings of Moody's, Standard & Poor's and Fitch.

This paper is organized as follows. Chapter 2 elaborates on outlook definitions provided by agencies. Chapter 3 describes the benchmark credit scoring models and the procedure to construct simulated ratings and outlooks. Chapter 4 examines the credit risk nature of outlooks in terms of through-the-cycle vs. point-in-time. Chapter 5 links the outlook scale to the (notch) rating scale.

Chapter 6 explores the agencies' migration policy for outlooks. Chapters 7 and 8 report on default prediction performance and point-in-time characteristics of adjusted ratings. Chapter 9 draws conclusions.

2 Agencies' definition of outlooks – rating Outlooks and rating Reviews

In their guide to ratings, rating process and rating practices, Moody's describes the meaning of rating Outlooks and rating Reviews as follows: "*A Moody's rating Outlook is an opinion regarding the likely direction of a rating over the medium term, typically 18 to 36 months.* […] *An RUR (rating(s) under review) designation indicates that the issuer has one or more ratings under review for a possible change and thus overrides the outlook designation.* […] *Moody's uses the Watchlist to indicate that a rating is under review for possible change in the short term, usually within 90 days.* […] *A credit is removed from Watchlist when the rating is upgraded, downgraded or confirmed*" (Moody's, 2004).

In a similar way Standard & Poor's explains the meaning of their rating Outlooks and rating Reviews on CreditWatch (2005): "*A Standard & Poor's rating Outlook assesses the potential direction of a long-term credit rating over the intermediate term (typically six months to two years). In determining a rating Outlook, consideration is given to any changes in economic and/or fundamental business conditions.* […] *Credit Watch highlights the potential direction of a short- or long-term rating. It focuses on identifiable events and short-term trends that cause ratings to be placed under special surveillance by Standard & Poor's analytical staff. These may include mergers, recapitalizations, voter referendums, regulatory action or anticipated operating developments. Ratings appear on CreditWatch when an event or a deviation from an expected trend occurs and additional information is necessary to evaluate the current rating.* […] *Such rating Reviews are normally completed within 90 days, unless the outcome of a specific event is pending.*" (Standard & Poor's, 2005).

Fitch rating Outlooks indicate the direction a rating is likely to move over a one- to two-year period. Fitch place their long-term credit ratings on a Rating Watch "*to notify investors that there is a reasonable probability of a rating change and the likely direction of such change. These are designated as Rating Watch and are typically resolved over a relatively short period.*" (Fitch, 2005).

Rating Outlooks and rating Reviews indicate different time horizons at which a rating migration might occur. Rating Outlooks signal likely rating migrations in the medium term. In their definitions agencies appear to have slightly different notions of what is meant by medium term. For example Moody's notion of medium term is one year longer than that of Standard & Poor's. However these horizons are only indications. A fine average is one to two years. For rating Reviews Standard & Poor's and Moody's are more explicit on the time horizon: 90 days; Fitch refers to a "relatively short period".

Although not mentioned explicitly in the definitions, rating Outlooks and rating Reviews differ in the likelihood of a rating migration to occur. Descriptions to define rating migration probabilities signaled by rating Reviews are stronger than those for rating Outlooks. For example Fitch uses "likely" for rating Outlooks and "reasonable probability" for rating Reviews. Moody's and Standard & Poor's emphasize the urgency of the situation when a rating is placed under review. Historical data on Moody's outlooks shows that two thirds of the rating Reviews have been followed by a rating migration with the indicated sign, while one third of the rating Outlooks have ultimately resulted in such a migration. In terms of expected rating migration probability and urgency, rating Reviews can be interpreted as stronger versions of rating Outlooks.

Only Standard & Poor's (2005) provides some insight into the criteria for rating Outlooks and rating Reviews. According to Standard & Poor's, rating Reviews are triggered by events or sudden changes in expected trends that require a formal review procedure in the short term. Standard & Poor's defines rating Outlooks as a response to changes in economic and fundamental business conditions. Our interpretation of these criteria for rating Outlooks is that agencies are aware of developing changes in the medium term, but do not judge them severe enough yet to consider a rating migration.

3 Definition benchmark credit scoring models and credit model ratings/outlooks

3.1 Outlook data and statistics

Data on Moody's ratings is obtained from the July 2005 version of the Moody's DRS database, which includes all corporate issuer credit rating revisions and default events in the period January 1971-May 2005. An extended version of the outlook dataset – which has been made available to us by Moody's – includes all outlooks provided by Moody's for their ratings in the period September 1991-February 2005. In 1991 Moody's started to provide information on rating Reviews (DOWN, UP). In 1995 Moody's began to publish rating Outlook information (positive Outlook POS, stable Outlook STA and negative Outlook NEG) as well. All five outlook categories have been available since 1995. Therefore our analysis covers the January 1995-December 2004 period. NOA Outlooks (No Outlook Available) are also included in the database. For issuers with a NOA Outlook the provision of outlook information is (mostly temporarily) suspended by agencies.

Not all issuers rated by Moody's are included in our analysis. For benchmarking purposes, ratings and outlooks are linked with accounting and market data from COMPUSTAT. In order to ensure consistency in accountancy information we selected only non-financial US issuers, and issuers with sufficient accounting and market data available in COMPUSTAT. This selection of issuers reduces the number of issuer-monthly observations from 507,824 to 71,962, including the NOA Outlooks. When the NOA Outlooks are excluded, 52,595 observations are left. The large reduction in observations changes the outlook distribution. Along the different selection steps, as outlined in table I, the fraction of DOWN Reviews and POS Outlooks increase by almost a factor of 2 and the fraction of NOA Outlooks is reduced by a factor of 1.5. As no dramatic changes appear in the outlook distribution, we believe the conclusions of this study will largely hold for non-US, private and financial issuers.

In the first years after the introduction of outlooks by Moody’s the percentage of NEG, STA and POS Outlooks steadily increases (see table I). After 1998 the outlook distribution becomes relative stable up to 2003. In 2004 almost all NOA Outlooks seem to be converted to STA Outlooks.

Outlook distributions vary among rating categories. Most issuers in the Caa category and below have a NEG Outlook. Aaa rated issuers have obviously no POS Outlooks and UP Reviews. Notable is the relative small fraction of DOWN Reviews and NEG Outlooks in the Aaa rating category. DOWN Reviews are more likely to appear in investment grade categories while POS Outlooks and UP Reviews are more present in speculative grade categories.

3.2 Specification of credit scoring models

As outlined in the introduction key in our study is the estimation of credit scoring models for default prediction, rating prediction and outlook prediction.

All default prediction models are estimated by the following logit regression model in a panel data setting

$$\begin{aligned} \mathrm{CS}_{\mathrm{i,t}} = \alpha + \beta_1 \frac{\mathrm{WK}_{\mathrm{i,t}}}{\mathrm{TA}_{\mathrm{i,t}}} + \beta_2 \ln(1 - \frac{\mathrm{RE}_{\mathrm{i,t}}}{\mathrm{TA}_{\mathrm{i,t}}}) + \beta_3 (1 - \frac{\mathrm{EBIT}_{\mathrm{i,t}}}{\mathrm{TA}_{\mathrm{i,t}}}) + \beta_4 (1 + \ln \frac{\mathrm{ME}_{\mathrm{i,t}}}{\mathrm{BL}_{\mathrm{i,t}}}) \\ + \beta_5 \mathrm{Size}_{\mathrm{i,t}} + \beta_6 \mathrm{SD(AR)}_{\mathrm{i,t}} + \beta_7 \mathrm{AR}_{\mathrm{i,t}} + \varepsilon_{\mathrm{i,t}} \end{aligned} \tag{3.1}$$

$$E(p_{i,t}) = \frac{1}{1 + \exp(CS_{i,t})} \tag{3.2}$$

$\mathrm{CS}_{i,t}$ is the credit score of issuer i at time t, $E(p_{i,t})$ is the expected probability of default of issuer i at time t, WK is net working capital, RE is retained earnings, TA is total assets, EBIT is earnings before interest and taxes, ME is the market value of equity, and BL is the book value of total liabilities. Size is the log-transformation of total liabilities normalized by the total value of the US equity market Mkt: ln(BL/Mkt). The abnormal stock return AR_t is the stock return relative to equal weighted market return in the twelve months preceding t. $\mathrm{SD(AR)}_t$ is the standard deviation in monthly abnormal returns in the twelve months preceding t.

The parameters of the logit regression model α and β are estimated by a standard maximum likelihood procedure. This estimation procedure seeks an optimal match between the actual outcome $p_{i,t}$ and the expected outcome of the model $E(p_{i,t})$. $p_{i,t} = 0$ when issuer i defaults before t + T and $p_{i,t} = 1$ when issuer i survives beyond t + T. Default prediction models are estimated for various time horizons T. An SDP model is estimated for T = 6 months, a DP1 model for T = 12 months and an LDP model for T = 6 years. Credit scores of these models are point-in-time measures of credit risk, giving weight to both the permanent and temporary credit risk component. SDP model gives maximum weight to the temporary credit risk component and the LDP model gives a substantial lower weight to the temporary credit risk component.

The sensitivity to the temporary credit risk component can be even further suppressed when the default prediction focuses exclusively on default probability in a specific future period. In this way credit scores are forced to be relative insensitive to temporary fluctuations in credit risk and focus as much as possible on the permanent credit risk component. The binary variable $p_{i,t}$ is set to 0 only for issuers defaulting in a future period ($t + T_1$, $t + T_1 + \Delta T$). Default events in the near future (t, $t + T_1$) are disregarded by excluding observations of issuers defaulting in this period from the model estimation.[1] A MDP model is estimated for $T_1 = 3$ years and $\Delta T = 3$ years. The parameter estimates of the MDP model do not change substantially when T_1 is varied between 3 and 6 years and ΔT is allowed to vary between one and three years.

The rating prediction model (RP model) models the discrete rating scale N with an ordered logit regression model in a panel data setting. In this model, the credit score $RP_{i,t}$ is an unobservable variable

$$\begin{aligned} RP_{i,t} &= \alpha + \beta_1 \frac{WK_{i,t}}{TA_{i,t}} + \beta_2 \ln(1 - \frac{RE_{i,t}}{TA_{i,t}}) + \beta_3 (1 - \frac{EBIT_{i,t}}{TA_{i,t}}) + \beta_4 (1 + \ln \frac{ME_{i,t}}{BL_{i,t}}) \\ &+ \beta_5 Size_{i,t} + \beta_6 SD(AR)_{i,t} + \beta_7 AR_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{3.3}$$

The $RP_{i,t}$ score is related to the rating R as follows

$$N_{i,t} = R \quad if \quad B_{R-1} < RP_{i,t} \leq B_R \tag{3.4}$$

where R is one of the rating categories, $N_{i,t}$ is the actual rating of issuer i at time t, B_R is the upper boundary for the RP score in rating category R, $B_0 = -\infty$ and $B_{18} = \infty$. We consider the following 18 rating categories N: Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3, Ba1, Ba2, Ba3, B1, B2, B3, Caa and Ca. In order to have a reasonable number of observations in each rating category, the rating categories Caa3, Caa2 and Caa1 are combined into a single rating category Caa and the rating categories C and Ca are combined into Ca. In the ordered logit model, the probability that $N_{i,t}$ equals R is specified by

$$P(N_{i,t} = R) = F(B_R - RP_{i,t}) - F(B_{R-1} - RP_{i,t}) \quad (3.5)$$

where F is the cumulative logistic function. Parameters α, β, and B_R are estimated with a maximum likelihood procedure. This estimation procedure seeks an optimal match between the actual rating $y_{i,t}$ and the expected outcome of the model $P(N_{i,t} = R)$.

The outlook prediction model (OP model) is estimated following the same ordered logit regression methodology. Instead of 18 rating categories N, five outlook categories O are modeled: NEG, STA and POS Outlooks and UP and DOWN Reviews. The outlook scale is a secondary scale on top of the rating scale. Outlooks are *relative* credit risk measures *within* a rating category N. The seven model variables X in equation 3.3 are therefore replaced by their differentials ΔX relative to their mean values $X_{N,t}$ of all issuers in rating category N at time t

$$\Delta X_{N,i,t} = \frac{X_{i,t} - X_{N,t}}{SD(X)_{N,t}} \text{ with } i \in N \quad (3.6)$$

$SD(X)_{N,t}$ is the cross-sectional standard deviation of model variable X for all issuers in rating category N at time t. Normalization of ΔX by SD(X) makes ΔX better comparable among different rating categories N. This allows a robust and single estimate of the outlook prediction model including outlooks of all issuers along the entire rating scale N.

3.3 Parameter estimates credit scoring models

Table II reports the estimated parameters α and β_i of four default prediction models (SDP, DP1, LDP, MDP) and five versions of a rating prediction model. These models are estimated with a dataset covering the period April 1982-December 2004. The starting date is motivated by the fact that Moody's started in April 1982 to refine their ratings by adding 1, 2 and 3 modifiers. At the end of each month Moody's (corporate issuer credit) ratings are linked to stock price data and accounting data from COMPUSTAT. Accounting data is assumed to be publicly available three months after the end of the fiscal year. The resulting panel dataset includes the time series of 2239 issuers with durations between 1 and 273 months and an average duration of 85 months. The total number of issuer-monthly observations is 189,248. Depending on the time horizon the estimation period for default prediction models is restricted to April 1982-December 2004 *minus* the prediction horizon, thereby avoiding an overweight of short-term defaults in the estimation.

Table III reports the parameter estimates of various outlook prediction models. As outlined in section 3.1 52,595 issuer-monthly observations are available in the 1995-2004 period to estimate these models.

All credit scoring models employ the same model variables. This allows a fair comparison of the relative weights RW_k of model variables k

$$RW_k = \frac{\beta_j \sigma_k}{\sum_{j=1}^{7} |\beta_j| \sigma_j} \tag{3.7}$$

b_j is the parameter estimate for model variable j, and σ_j is the standard deviation of model variable j in the pooled sample of all observations. Table II and III show the RW values for all credit scoring models.

3.4 *Rating and outlook simulation*

Credit scores of default prediction models and rating prediction models are converted to credit model ratings equivalent to actual ratings. This enables to compare the dynamics of agency ratings unambiguously with the dynamics of credit scores. At the end of each month all companies are ranked by their credit score. On the basis of this ranking, credit model ratings, Aaa, Aa1, …. , B3, Caa and C/Ca, equivalent to agency ratings, are assigned to individual issuers. So at the end of each month the number of issuers in each agency rating class N equals the number of firms in the equivalent credit model rating class.

The influence of agency rating migration policies is included in credit model ratings by adjusting credit scores following a particular policy model. We model the rating migration policy by two parameters: a threshold parameter and an adjustment parameter. The threshold parameter TH specifies the size of a credit risk interval [-TH,+TH], in which credit risk is allowed to fluctuate without triggering a rating migration. If a rating migration is triggered, ratings are not fully adjusted to the actual credit risk level. The adjustment fraction AF specifies the partial adjustment of ratings. Partial adjustment of ratings (i.e. the spreading of a target rating adjustment over time) generates drift in ratings. After modifying credit scores following this migration model the modified scores are converted to simulated ratings. Appendix A describes the details of the rating simulation procedure.

Based on RP model scores, - reflecting the agencies' through-the-cycle perspective - we have constructed simulated RP(TH, AF) ratings for a range of TH and AF values. A best match in rating migration probability and rating drift properties between simulated RP(TH,AF) ratings and actual ratings N is obtained for a threshold TH of 1.8 notch steps and an adjustment fraction AF of 0.7 on the downside and 0.6 on the upside. Actual ratings are best reproduced by RP(1.8,0.7/0.6) ratings. These implied migration policy parameters for Moody's ratings equal to those derived for Standard & Poor's ratings (see also Altman and Rijken, 2004). [2]

For outlook simulation we follow exactly the same procedure, except we leave out the adjustment parameter AF because outlooks exhibit no drift. For OP model scores, on average, a best match between the dynamics of actual and simulated outlooks is obtained for a threshold TH of 1.5 – 2.0 notch steps (see chapter 6).

Appendix A discusses the influence of marginal high-frequency non-informative "noise" in credit scores on the dynamics of credit model ratings/outlooks. The most important conclusion is that the influence of this marginal "noise" in credit scores is suppressed by threshold levels of 0.5 notch steps and above. These threshold levels avoid large numbers of reversal migrations in subsequent months. The implied TH levels for ratings and outlooks are safely above the "noise" level of 0.5 notch steps, so the marginal "noise" in credit scores is of no concern in the benchmark of actual rating/outlook dynamics with credit model rating/outlook dynamics.

4 Credit risk nature of outlooks

4.1 Through-the-cycle vs. Point-in-time perspective

According to the survey of Moody's in 2002, investors regard outlook information as an additional source of information to compensate for the inadequate timeliness of (corporate issuer credit) ratings. A well-accepted explanation for the perceived delays in rating migrations is the through-the-cycle methodology that agencies apply in their rating assessment. This methodology has two aspects: first, a focus on the permanent credit risk component of default risk and, second, a prudent migration policy.

The first aspect of the through-the-cycle rating methodology is the disregard of short-term fluctuations in default risk. By filtering out the temporary credit risk component, ratings measure exclusively the permanent, long-term and structural credit risk component. According to Cantor and Mann (2003) the through-the-cycle methodology aims to avoid excessive rating reversals, while holding the timeliness of ratings at an acceptable level: "*If over time new information reveals a potential change in an issuer's relative creditworthiness, Moody's considers whether or not to adjust the rating. It manages the tension between its dual objectives – accuracy and stability – by changing ratings only when it believes an issuer has experienced what is likely to be an enduring change in fundamental creditworthiness. For this reason, ratings are said to 'look through-the-cycle'*." Standard and Poor's (2003) is convinced that "…*the value of its rating products is greatest when its ratings focus on the long term and do not fluctuate with near term performance*."

The second aspect of the through-the-cycle methodology is the enhancement of rating stability by a prudent migration policy. Only substantial changes in the permanent credit risk component result in rating migrations and, if triggered, ratings are partially adjusted to the actual level in the permanent credit risk component. Although not officially disclosed by agencies, practical evidence of such a prudent migration policy exists. In their announcement on the reconsideration of their migration policy, in January 2002, Moody's provides some insight into their migration policy: *"Under consideration are more aggressive ratings changes – such as downgrading a rating by several notches immediately in reaction to adverse news rather than slowly reducing the rating over a period of time – as well as shortening the rating review cycle to a period of weeks from the current period of months"*. [3] The spread of rating adjustments over time is the source of rating drift. The relative long review cycle and high threshold for rating adjustments are the sources of rating stability.

How rating agencies put their through-the-cycle methodology exactly into practice is not clear. Treacy and Carey (2000) describe the through-the-cycle rating methodology as a rating assessment in a worst case scenario, at the bottom of a presumed credit risk cycle. Löffler (2004) explores the through-the-cycle effects on rating stability and default-prediction performance in a quantitative manner by modeling the separation of permanent and temporary credit risk components of default risk in a Kalman filter approach. As outlined in section 3.4 we have taken a different approach to gain insight into the agencies' through-the-cycle methodology by benchmarking agency rating dynamics with credit model scores. We confirm the exclusive focus of ratings on the permanent credit risk component. We suggest that a rating migration is triggered if the actual credit risk, as indicated by the permanent credit risk component, exceeds a threshold of 1.8 notch steps relative to the average credit risk level in a rating category. If triggered, ratings are partially adjusted to the actual credit risk level, 60% at the upside and 70% at the downside.

Rather than the through-the-cycle methodology, bankers have a point-in-time perspective on corporate credit risk with a time horizon of between one and seven years (see Basle Committee, 2000). Most bankers have a one-year point-in-time perspective on credit risk. It is reasonable to assume that this perspective applies to most other investors as well. The point-in-time perspective looks at the current credit risk situation without attempting to suppress the temporary credit risk component. It weights both the temporary and permanent credit risk component. The relative weight of these two components depends on the horizon in the point-in-time perspective. For a one year horizon the temporary credit risk component has a larger weight than it does for a longer horizon.

4.2 Parameter estimates of default and rating prediction models

According to the parameter estimates of default prediction models (see table II) market equity information has a large share in the explanation of short-term defaults. The ME/BL variable dominates the SDP model with a RW value of 37.2 %. Short-term dynamics of abnormal stock returns measured by AR and SD(AR) have a relative weight of respectively 16.5% and 10.6%. Although equity information is dominant, accounting information and Size add substantially to the explanation of the default incidence.

Prediction horizon has a significant impact on the relative weight of the model variables. Especially for RW values of the RE/TA, ME/BL, and Size variable, a clear shift is observed from the SDP, DP1, LDP to the MDP model, in that order of sequence. Not surprisingly, the short-term oriented SDP model depends heavily on the more volatile variables, like ME/BL and the trend variable AR, while the MDP model is more driven by less volatile variables like RE/TA and Size. The MDP model is relative insensitive to the most volatile variables EBIT/TA, SD(AR) and AR, which is consistent with the objective of this model to suppress the temporary credit risk component as much as possible.

Of all default prediction models the RW values of the MDP model most closely matches the RW values of the rating prediction model RP. The MDP model focuses as much as possible on the permanent credit risk component while the RP model reflects the agencies' through-the-cycle perspective. However, these two perspectives do not fully overlap. As opposed to the MDP perspective the agencies' through-the-cycle perspective puts less weight on WK/TA and ME/LIB but more on Size.

RP model parameters have been re-estimated for various parts of the rating scale. For issuers rated B1 and above the RP model parameters are robust. For example, the RW values for investment graded issuers are about equal to RW values of B1, Ba3, Ba2, and Ba1 speculative graded issuers (see table II). At the bottom end of the rating scale - for issuers with ratings below B1 the RW values differ. In this range the RW value of Size is much lower and the RW value of EBIT/TA is higher, comparable to the SDP model. Similar RW values are observed for issuers approaching a default event. Apparently, ratings of distressed issuers have a point-in-time perspective instead of a through-the-cycle perspective. In order to model the agencies through-the-cycle perspective as close as possible the agency rating prediction model is re-estimated excluding issuers rated below B1 (TTC model).

Robustness checks have been carried out to test for sector influence, time period and default event definition. Parameter estimates do not differ substantially between two periods, 1982-1994 and 1995-2004. We compared the impact of three default event definitions on the default prediction model estimation: bankruptcy filing, Moody's default event definition and Standard & Poor's definition. Default prediction model parameters appear to be only moderately sensitive to the definition of a default event. [4]

4.3 Parameter estimates of outlook prediction models

RW values of the OP model (see table III) closely match the RW values of the SDP model (see table II). [5] This finding implies that the outlook scale is *on average* a measure of short-term point-in-time credit risk. A more diverse picture shows up when the outlook prediction model is estimated for four subscales, each of them including STA Outlooks with either NEG Outlooks, POS Outlooks, DOWN Reviews or UP Reviews. Parameter estimates vary considerably between these four outlook subscales, so the outlook scale appears not to have a uniform credit risk nature. At the downside outlooks depend more on volatile variables SD(AR) and ME/LIB, while at the upside outlooks are related to less volatile variables Size and RE/TA (with a negative sign).

In the OP model estimation, rating Reviews have a relatively low weight because of their low share in the outlook distribution and more outlooks appear at the downside than at the upside of the outlook scale (see table I). To control for this unequal outlook distribution a weighted ordered regression model (OPW model) is estimated which gives equal weight to all five outlook categories in the model estimation. Between the OP and OPW model RW values differ by 7.5% at most.

4.4 Weighting the permanent and temporary credit risk components in credit scoring models

In a first stage, credit scoring models are estimated with seven model variables. In a second stage they are re-estimated with only two model variables X_P and X_T, proxies for the permanent and temporary credit risk component. The permanent credit risk component X_P is proxied by TTC scores. The TTC model represents the agencies' through-the-cycle perspective (see section 4.2). In contrast to TTC scores, SDP scores are sensitive to temporary fluctuations in credit risk because of their short six month prediction horizon. SDP scores are strongly correlated with TTC scores (correlation coefficient of 0.49). SDP scores follow both the permanent and temporary credit risk component. To obtain a pure proxy for the temporary credit risk component $X_{T,}$ TTC scores are subtracted from SDP scores. [6]

Default and rating prediction models are re-estimated using X_P and X_T. Outlook prediction models are re-estimated after converting X_P and X_T to their differentials following equation 3.6. Results are presented in table IV. In the SDP model 41.4% of the variations in credit risk are explained by the temporary credit risk component X_T. Increasing the horizon in the default prediction models from six months to six years reduces the weight to X_T from 41.4% to 29.4%. Focusing exclusively on default prediction in a future period (MDP model) suppresses the weight to X_T further down to 22.6%. The weight of X_T in the TTC model – the agencies' through-the-cycle perspective for issuers rated B1 and above - is by definition 0%. At the low end of the rating scale – below B1, X_T has a 31.9% weight in the rating scale, comparable to the point-in-time LDP model. The difference in weight to X_T between the MDP model and the TTC model reflects the difference between two long-term views on credit risk.

The outlook scale is *on average* more sensitive to the temporary credit risk component than the SDP model is. The weight of X_T varies between the four outlook subscales. The DOWN Review – driven by events according to the Standard & Poor's definition – is most sensitive to temporary fluctuations in credit risk (see table IV). Down Reviews and NEG Outlooks are even more sensitive to the temporary credit risk component than would be expected from a one-year point-in-time perspective as proxied by the SDP model. The upside of the outlook scale is more moderately sensitive to short-term credit risk fluctuations. The nature of POS outlooks matches the SDP perspective while the nature of UP Reviews matches the LDP perspective. Apparently, DOWN Reviews are driven mainly by suddenly changing circumstances (events) while UP Reviews – at the other side of the spectrum - are more driven by more moderate changes in economic and fundamental business conditions. This is probably not a matter choice in setting the outlook migration policy but due to an asymmetry in credit risk dynamics which is far more volatile at the downside than at the upside.

Comparing pseudo R^2 values between first stage credit scoring models (table III) and second stage credit scoring models (table IV) gives insight into the extent to which information is lost by replacing the seven model variables by X_P and X_T. Pseudo R^2 indicates the goodness of fit by the (ordered) logit regression model. As expected the pseudo R^2 of the default prediction models and agency rating prediction models are hardly affected. For the O-DOWN model R^2 is lowered by about 15%. For the O-NEG, O-POS and O-UP models the reduction in pseudo R^2 is larger, respectively 27%, 26% and 51%. From this large reduction in pseudo R^2 we conclude that outlooks are also driven by other risk factors, not captured by SDP and TTC scores. Interestingly, the pseudo R^2 of the OP model – which models the common factors for all five outlooks - is reduced only by 8%. This implies that the common factors in the *entire* outlook scale are largely captured by X_P and X_T.

5 Linking the outlook scale to the rating scale

Credit risk variations among issuers within a rating category N can be measured by variations in credit scores ΔCS

$$\Delta CS_{N,i,t} = \frac{CS_{i,t} - CS_{N,t}}{\gamma_{N,t}} \tag{5.1}$$

$CS_{N,t}$ is the average credit score for all issuers in rating category N at time t. ΔCS is converted to a notch rating scale by $\gamma_{N,t}$. This scaling factor $\gamma_{N,t}$ is obtained as follows: For default and rating prediction model scores the scaling factor $\gamma_{N,t}$ reflects the slope between the numerical rating scale N and the average credit scores CS in rating categories N. The numerical rating scale runs from Ca/C = 1, Caa = 2, B3 = 3, B2 = 4, up to Aaa = 18. This numerical rating scale is an arbitrary but quite intuitive choice that is commonly found in the mapping of bank internal-rating models to agency ratings. Roughly three groups of rating categories can be distinguished with a close to linear relationship between CS and N: $N \in [1 .. 4]$, $N \in [5 .. 10]$ and $N \in [11 .. 18]$. For each of these groups and for each month $\gamma_{N,t}$ is derived. For outlook prediction model scores, the scaling factor $\gamma_{N,t}$ equals the change in OP-score when transforming the differential variables $\Delta X_{N,i,t}$ in the outlook prediction model by replacing $X_{N,t}$ by $X_{N+1,t}$ or $X_{N-1,t}$ (see equation 3.6). The α and β parameters of the outlook prediction model are kept fixed. Note that this scaling factor

$\gamma_{N,t}$ does not convert credit scores to a rating scale in a sense that it removes the temporary credit risk component. The scaling factor converts credit scores to a *notch* rating scale and defines a common ground to compare variations in credit scores of various prediction models.

Table V reports the average ΔCS values for five outlook categories. By definition the average ΔCS values for all outlooks is zero. Based on ΔOP scores, the credit risk of issuers with DOWN Reviews is centered 2.7 notch steps lower than is indicated by their rating class N. Issuers with NEG, STA, POS Outlooks and UP Reviews are centered at respectively -1.4, +0.4, +1.7 and +1.8. Average ΔOP scores do not differ much between POS Outlooks and UP Reviews. This is caused by the relative low number of observations with POS Outlooks and UP Reviews in the OP model estimation. When all outlooks are equally weighted in the estimation procedure, the different nature of the outlook scale at the upside gets more weight. So for ΔOPW scores a more clear distinction appears between POS Outlooks and UP Reviews: 1.7 vs. 2.8 notch steps. As indicated by ΔOPW scores, rating Outlooks diverge 1.5 notch steps from the centre and rating Reviews diverge 3 notch steps from the centre. These numbers are consistent with the adjustments Hamilton and Cantor (2004) apply in their search for an optimal adjustment of ratings by their outlooks.

Credit risk variation in the outlook scale as indicated by the ΔOP and ΔOPW scores is best captured by ΔSDP scores and ΔDP1 scores. This empirical finding emphasizes once more the short-term point-in-time character of outlooks. Point-in-time measures with longer time horizons (ΔLDP scores) and through-the-cycle measures (ΔRP scores) detect less credit risk dispersion in the outlook scale.

Credit risk information signaled by the outlook scale is not constant in time. In periods with no upcoming or recent rating migration event only little credit risk dispersion shows up in the outlook scale. Also, little credit risk dispersion shows up between DOWN Reviews and NEG Outlooks just before and just after a rating *downgrade* (see Table V). Similar findings hold for STA Outlooks, POS Outlooks and UP Reviews just before and just after a rating *upgrade*. Credit risk information of -1.6 notch steps and 2.5 notch steps signaled by STA outlooks just before respectively a rating downgrade and upgrade suggests that not all outlooks are set appropriately. In contrast, the credit risk information signaled by simulated OP(0) outlooks is constant in time and unconditional to rating migration events. Furthermore, the credit risk dispersion in the OP(0) outlook scale is much larger than the actual outlook scale.

The time inconsistencies and relative low credit risk dispersion in the actual outlook scale are most likely due to a lack of credit risk standardization in the assignment of outlooks. This result is not surprising in view of the purpose of outlooks. First, as outlooks are intended to signal likely upcoming rating migrations, outlooks are set more accurately when rating analysts are considering rating migrations and extra alert after rating migrations. Second, as outlooks are not intended to quantify credit risk information explicitly, agencies do not standardize credit risk information signaled by outlooks. Instead, changes in outlooks are triggered by sudden events, breaks and shifts in trends, not by measuring short-term point-in-time credit risk along a standardized scale.

6 Agency migration policy for outlooks

The prime objective of outlooks is to signal upcoming rating migrations. This timing information is disclosed by the timing of outlook migrations following a specific outlook migration policy. Insight into the agencies' outlook migration policy is obtained by benchmarking the dynamics of actual outlooks with the dynamics of simulated OP(TH) outlooks. In contrast to actual outlooks, simulated outlooks have no explicit timing objective and their dynamics are only driven by a variation in credit risk within a rating category N – as measured by OP scores. This benchmark aims to investigate to what extent the explicit timing objective of outlooks overrides credit risk information.

6.1 Outlook distribution conditional to a rating migration event

Table VI reports the distribution of actual outlooks and simulated OP(0) outlooks conditional to a rating downgrade event or a rating upgrade event. Just before a rating downgrade at $t = -1$, DOWN Reviews occur a factor of 3 more frequently than NEG and STA Outlooks do. In the months before a rating downgrade a large part of the actual STA and NEG Outlooks are changed to DOWN Reviews – excessively large from a pure credit risk perspective as indicated by simulated outlooks. Moreover, just before a rating downgrade the credit risk level as indicated by ΔOP and ΔOPW scores hardly distinguish between NEG Outlooks and DOWN Reviews (see table V). So prior to a rating migration event, NEG Outlooks and DOWN Reviews mostly differ in timing information on an upcoming rating downgrade rather than differ in credit risk information. Credit risk information is (partly) overridden by the timing objective when putting ratings on a DOWN Review to signal a likely upcoming rating migration.

A strong increase in rating Reviews before a rating migration also appears at the upside. Just before a rating upgrade the fraction of UP Reviews raises sharply – again much larger than would be expected from a pure credit risk perspective. In contrast to actual outlooks the distribution of simulated outlooks does not change significantly conditionally to a rating upgrade, except for a reduction in UP Reviews by half just after the rating upgrade. The precise timing of rating upgrades is difficult to establish from a pure credit quality perspective. At the upside changes in credit quality happen less abrupt. Therefore rating upgrades are not clearly preceded by significant changes in credit risk, which hampers a good timing match between upward dynamics of OP scores and upgrades of actual outlooks.

6.2 Outlook migration matrix

Table VII presents the unconditional outlook migration matrix on a monthly basis. “Exit” observations – observations at the end of an issuer time series in the database – are not included in the construction of the matrix because a database exit is not initiated by agencies and no credit scores are available to benchmark these exit migrations. No bias is introduced as the distribution of outlooks for these “exit” observations is comparable to the outlook distribution for all observations. [7]

At first NOA Outlooks (no outlook available) are treated as a separate outlook category in the computation of the outlook migration matrix. The migration probabilities for DOWN and UP Reviews are 23.2% and 16.1%, which implies an average duration of 4.3 months and 6.2 months. These duration figures are somewhat higher than reported by Cantor and Hamilton (2004) as we do not include “exit” observations. The migration probabilities for NEG, STA, POS and NOA Outlooks are roughly a factor of 4 lower. The duration for these outlooks is about 20 months.

In order to simulate outlooks – including NOA Outlooks – by ranking of credit scores, NOA Outlooks need to be positioned on the outlook scale. We assume NOA Outlooks to be comparable to STA outlooks. We justify this assumption as follows.

1. Most migrations to NOA Outlooks happen at a rating migration event. Almost all of these migrations take place from DOWN and UP Reviews. After a rating migration event, DOWN and UP Reviews are no longer of any use to signal upcoming rating events. Apart from NOA Outlooks the most likely outlook after a rating migration event is a STA Outlook.
2. In 2004 Moody’s converted almost all NOA Outlooks (27% of total outlooks) to STA Outlooks in their dataset. Apparently Moody’s regard NOA Outlooks to be comparable to STA Outlooks.
3. Migration probabilities for NOA and STA Outlooks are comparable. The both have a probability of about 0.5% to migrate to POS Outlooks and UP Reviews and a probability of about 1% to migrate to NEG Outlooks and DOWN Reviews.

For these reasons we convert all NOA Outlooks to STA outlooks. The outlook simulation procedure is repeated to match simulated outlooks with the new outlook distribution.

6.3 Agency migration policy for outlooks

Table VII compares the migration matrices of actual and simulated outlooks. Compared to the dynamics of simulated OP(1.5) outlooks - a pure and consistent credit risk perspective -, the dynamics of actual outlooks deviates in a number of ways:

1. Migration probabilities are relative high for DOWN(UP) Reviews and relative low for NEG(POS) Outlooks. In other words durations for DOWN and UP Reviews are about 4 – 5 months while durations for NEG and POS Outlooks are about 20 months. These durations are consistent with the time horizons of expected rating changes set in the outlook definitions by the agencies.

For simulated outlooks, from a pure credit risk perspective, when outlooks are only aimed to signal credit risk information and DOWN (UP) Reviews are positioned at the extreme side of NEG (POS) Outlooks, the durations of Reviews and Outlooks do not differ much and are about 7 – 8 months. The timing objective for actual outlooks clearly influences the duration of outlooks.

2. The frequency a rating is put on a DOWN Review is twice as high compared to the frequency a rating is put on a NEG Outlook. A large number of two-step migrations between STA Outlooks and DOWN Reviews circumvent intermediate NEG Outlooks, thereby reducing the number of one-step migrations. Similar findings apply to the upside of the outlook scale. For simulated outlooks two-step migrations occur a factor 3 – 4 less, which illustrates the effect of timing objective for actual outlooks.

From a pure credit risk perspective the timing objective of outlooks shortens durations for rating Reviews, lengthens durations for rating Outlooks and partially circumvents the use of rating Outlooks as "intermediate" states between Stable Outlooks and rating Reviews. These deviations are consistent with the official agencies' policy for Reviews (see chapter 2).

The outlook migration policy has not changed much in the period 1995 - 2004. The migration probabilities of actual DOWN Reviews vary between 20 and 25%, while migration probabilities of NEG, POS and STA Outlooks vary between 2% and 5%. Extraordinary changes in migration probabilities are only observed for UP Reviews. Since 2002 the duration of UP Reviews is reduced by a factor of two compared to the average level of 7 months in the years 1996-2001. Apparently Moody's brought the durations of UP Reviews more in line with DOWN Reviews. This significant change in migration policy demonstrates that the timing of rating upgrades is not as clear as the timing of rating downgrades. Upgrade dynamics is more moderate than upward dynamics.

6.4 The implied threshold for the outlook migration policy

The implied threshold of the agencies outlook migration policy is derived by searching for a best match between the dynamics of simulated outlooks OP(TH) and actual outlooks. In order to avoid the interference of the overriding timing objective as much as possible, we constructed outlook migration matrices for a restricted outlook scale with three outlook categories: a merged DOWN/NEG category, STA outlooks and a merged UP/POS category. This restriction controls for the circumvention of NEG and POS categories by two step migrations between STA Outlooks and DOWN Reviews.

For all observations, unconditionally to a rating migration event, a best fit in migration probabilities between actual outlooks and simulated OP(TH) outlooks is obtained for a threshold TH of 1.5 notch steps (see table VII, panel C), which seems slightly lower the 1.8 notch threshold in the *rating* migration model. However, this implied 1.5 notch threshold can be considered less prudent since ratings are only sensitive to the permanent credit risk component and outlooks are sensitive to the temporary credit risk component as well. The standard deviation of OP scores in the time series of individual issuers is a factor of two higher than that of RP scores. So the outlook migration policy can be considered at least a factor of two less prudent than the rating migration policy.

The prudence of the outlook migration policy varies conditionally to a rating migration event. Six months after a rating migration event a best match between actual outlooks and simulated OP(TH) outlooks is obtained for a threshold TH of 3.0 notch steps. In periods with no rating migration event in the near past and near future the outlook migration policy is slightly less restrictive, with an implied threshold level TH of 2.5 notch steps. Six months before a rating migration event takes place, NEG/DOWN and POS/UP categories hardly move and the implied TH level is at least 3.5 notch steps. In contrast, STA Outlooks are less stable before a migration event with an implied TH level of 0.5 notch steps.

7 Default prediction performance of adjusted ratings

7.1 Definition of adjusted ratings

On top of through-the-cycle credit risk information signaled by ratings, outlooks provide point-in-time credit risk information and in addition, especially DOWN and UP Reviews, provide timing information on upcoming rating migrations. Rating information and outlook information is combined by adding outlooks O to (notch) ratings N, resulting in adjusted ratings. For this purpose the outlook scale is converted to a notch scale based on the ΔCS values as reported in table V. Following the (unconditional) credit risk dispersion in the actual outlook scale, measured by ΔOP scores, ratings with a DOWN Review are adjusted by subtracting 2.5 notch steps (a numerical scale of for ratings N is assumed: Ca/C = 1, Caa = 2, up to Aaa = 18). Ratings with a NEG Outlook, a STA Outlook, a POS Outlook and an UP Review are adjusted respectively by -1.5 notch steps, +0.5 notch steps, +1.5 notch steps and +2.0 notch steps. We round the adjustments to 0.5 notch steps. Unless stated otherwise, ratings are adjusted following this adjustment scheme.

7.2 Default prediction performance of adjusted ratings

Default prediction performance of ratings is expected to improve when point-in-time credit risk information and timing information on upcoming rating migrations are added. Default prediction performance of unadjusted ratings and adjusted ratings is measured by an accuracy ratio for different prediction horizons T (for details ACR computation see appendix B). The accuracy ratio measures the overall default prediction performance of a rating scale, weighting type I and type II errors equally in distinguishing defaulters and non-defaulters. It varies between 0% (random scale) and 100% (perfect prediction scale). We define ΔACR as the difference in ACR value between an adjusted rating scale R and the unadjusted rating scale N. The standard error in ACR is 1.5% for a 6 month prediction horizon and 2% for a 3 year horizon. The standard error in ΔACR is a factor of two lower (see appendix B).

Table VIII reports the ACR values of unadjusted and ΔACR values of adjusted ratings for various default prediction horizons, ranging from 6 months to 3 years. For the unadjusted rating scale N ACR is 69.1% and 57.2% for a prediction horizon of respectively one year and three years. Adjusting ratings by their actual outlooks increases ACR by 3.7% and 1.5%. The lower improvement for longer time horizons is consistent with the point-in-time credit risk nature of the outlook scale. Hamilton and Cantor (2004) report similar figures in their study covering the 1999-2003 period: respectively 4.3% and 2.0%.

When adjusting actual ratings by their simulated OP(0) outlooks the ACR jumps by +9.1% and +5.7% for a prediction horizon of respectively one year and three years. This increase in ACR can be largely ascribed to the downside of the outlook scale. When adjustments are only made for simulated NEG Outlooks and DOWN Reviews ACR improves by 8.3%, while adjustments at the upside improve ACR by only 2.1%. It might be that at the downside changes in credit quality happen more abruptly and faster and are therefore easier to detect.

For simulated outlooks ΔACR values are moderately sensitive to the adjustment scheme (see table VIII). Adjustment scheme 1 and 4 are based on the (unconditional) credit risk dispersion in respectively the actual outlook scale and simulated outlook scale (see table V). As opposed to the moderate adjustment scheme 1, adjustment scheme 4 stretches the outlooks scale to -6.5 notch steps for DOWN Reviews and +7.5 notch steps for UP reviews. Adjustment scheme 3 - in between adjustment scheme 1 and 4 - with adjustments ranging from - 4.5 notch steps for DOWN Reviews to + 4.5 notch steps for UP Reviews offers most improvement in default prediction performance: ΔACR = 10.2% for a prediction horizon of 12 months. Apparently the credit risk dispersion within a rating class is substantial and exceeds standard deviations beyond 3 notch steps, which is at least a factor two larger than indicated by actual outlooks.

Large variations in rating adjustments have only moderate impact on ΔACR values. Apparently outlooks have most value at the bottom of the rating scale. Regardless whether the adjustment scheme is moderate or extreme, all troubled issuers – with low ratings and outlooks at the downside - are pushed to the bottom of the adjusted rating scale. The robustness of default prediction performance to various adjustment schemes is also demonstrated by Hamilton and Cantor (2004).

7.3 Explaining the differences in default performance between actual and simulated outlooks

For simulated outlooks the ΔACR values are a factor 2 – 4 larger than for actual outlooks. Most obvious explanation is the lack of credit risk standardization in the assignment of outlooks.

Chapter 5 documents the relative low credit risk dispersion in the actual outlook scale. Also credit risk information signaled by actual outlooks is not constant in time and conditional to the rating migration event. In contrast, credit risk information signaled by simulated outlooks is more dispersed and does not depend on rating migration events. Simulated outlooks are only based on the ranking of credit scores each month. Therefore, as pure and consistent measures of credit risk, simulated outlooks perform better than actual outlooks. Direct empirical evidence for the better credit quality of simulated outlooks is found in the outlook migration matrix. As indicated by simulated outlooks 89% of the defaulted issuers had a NEG Outlook and a DOWN Review just before default (see Table VII). For actual outlooks this percentage is 62%. The simulated outlook scale is more effective in separating defaulters from non-defaulters.

A few alternatives are considered to explain the difference in ΔACR values between actual and simulated outlooks.

1. The timing objective overrides credit risk information of the outlook scale. It shortens durations for rating Reviews, lengthens durations for rating Outlooks and partially circumvents the use of the "intermediate states" of rating Outlooks. In order to control for these migration policy effects we merged the DOWN and NEG outlooks categories, and the POS and UP outlook categories. After the removal of "intermediate states" the dynamics of actual outlooks closely matches the dynamics of simulated outlooks and the migration policy effects are largely suppressed. The distribution of outlook durations for the two merged categories and STA outlook category largely overlaps between actual and simulated outlooks. An exception is the larger number of DOWN/NEG outlooks with short durations up to 4 months for actual outlooks, but the number of outlooks with short durations has already been reduced substantially by a factor 2 by the merger of DOWN and NEG outlooks categories. For this restricted outlook scale ratings are adjusted as follows: DOWN/NEG: -1.5, STA: + 0.5, POS/UP: +1.5. The restriction to three outlook categories hardly changes ΔACR values for both actual and simulated outlooks (see table VIII). From this analysis we conclude that the outlook migration policy does not severely override credit risk information in the outlook scale.
2. Dynamics and timely adjustment of actual outlooks are restricted by a migration threshold TH. The implied threshold is on average 1.5 not steps (see section 6.4). However, the impact of a threshold TH on ΔACR is low. For simulated OP(3.0) outlooks ΔACR values decreases by only a half percentage point.
3. At the downside actual outlooks are more sensitive to the temporary credit risk component than simulated outlooks are. The weight to the temporary credit risk component in the O-DOWN model is higher than it is in the OP model, which represents a best average for all five outlook categories (see table IV). An (over)sensitivity to temporary credit risk fluctuations could worsen default prediction performance, especially for longer time horizons. However, ΔACR values only marginally decrease when outlooks are simulated based on O-DOWN model scores instead of OP model scores.

These explanations can only explain 1.5% of the 5.5% gap in ΔACR values between actual and simulated outlooks. The remainder of this gap is highly likely due to a lack of credit risk standardization in the assignment of outlooks.

8 Point-in-time characteristics of adjusted ratings

To what extent does the lack of credit risk standardization in the outlook scale hamper the ability of outlooks to bridge the agencies' through-the-cycle perspective and the investor's one-year point-in-time perspective? Or to what extent do adjusted ratings have point-in-time characteristics? To answer these questions we compare rating properties of adjusted ratings with credit model ratings based on RP(TH) and LDP(TH) scores. For these ratings, table IX presents the rating migration probabilities, rating drift properties and sensitivity to the temporary and permanent credit risk component.

Adjustment of ratings N by their actual outlooks has a significant impact on rating dynamics. Outlooks add point-in-time credit risk information to the through-the-cycle ratings. As a result migration probabilities increase from 1.8% to 3.0% for downgrades and from 0.7% to 1.9% for upgrades. Rating drift disappears and even reversal effects show up at the downside. The weight to X_T increases from 10.2% to 18.7% for all issuers and from 0% to 15.2% for issuers rated B1 and above. [8] A 30% weight to X_T relates to the point-in-time LDP model (see table IV), so a level of 15% can be interpreted as halfway across the bridge between the agencies' through-the-cycle perspective and a long-term point-in-time perspective.

In contrast to actual outlooks, simulated outlooks are capable to adjust ratings to point-in-time measures, although the implied prediction horizon is long, about 6 years. Adjusting ratings N by their simulated OP(1.5) outlooks, resulting in N-OP(1.5) ratings, increases the weight to X_T from 10.2% to 27.3%. Dynamic properties and default prediction performance of N-OP(1.5) ratings are most in line with LDP(1) and LDP(1.5) ratings. The implied threshold level of 1.25 notch steps is a compromise between the implied threshold for ratings (1.8 notch steps) and the implied threshold for outlooks (1.5 notch steps). [9]

9 Conclusions

Corporate bond ratings become sensitive to short-term fluctuations in credit risk when they are adjusted by their outlooks. As a result default prediction improves – especially for short prediction horizons. However, adjusted ratings reflect a more moderate version of a through-the-cycle perspective.

Credit risk standardization in the outlook scale could improve the default prediction performance of adjusted ratings even further and could enable to bridge the gap between the agencies' through-the-cycle perspective and the investor's point-in-time perspective. Potentially, adjusted ratings have a long-term point-in-time perspective when agencies standardize credit risk information in the outlook scale.

We suggest changing the outlook assignment process as follows. Set short-term point-in-time credit risk standards for NEG, STA and POS Outlooks. Review these outlooks at least on a quarterly basis. Preserve the timing objective by migrations to DOWN and UP Reviews when rating migrations are considered on the short term. These suggestions are by no means a fundamental revision of current practice. It is an adequate response to the critique that investors have on rating timeliness. Depending on their rating stability preferences, investors can use either ratings without any additional outlook information when they prefer rating stability or ratings in combination with outlook information when they prefer rating timelines and maximum accuracy. To serve the last group best, accuracy and timing should be of equal importance in the assignment of outlooks and meet the same quality standards as agencies apply to ratings.

Literature

Altman E.I. en H.A. Rijken, 2004, "How rating agencies achieve rating stability", *Journal of Banking & Finance* 28, 2679-2714

Altman E. I. and D. L. Kao, 1992, "The implications of corporate bond ratings drift", *Financial Analysts Journal*, June, 64-75

Association for Financial Professionals, 2002, www.afponline.org

Basel Committee on Banking Supervision, 2000, "Range of practice in banks' internal rating systems", discussion article

Baker H. K. and S. A. Mansi, 2002, "Assessing credit agencies by bond issuers and institutional investors", *Journal of Business Finance & Accounting* 29 (9,10), 1367-1398

Cantor R. and D. Hamilton, 2005, "Rating Transition and Default Rates Conditioned on Outlooks Revisited: 1995-2005", *Special Comment*, Moody's Investor Services, December

Cantor R. and D. Hamilton, 2004, "Rating Transition and Default Rates Conditioned on Outlooks", *Journal of Fixed Income*, September, 54-70

Cantor R. and C. Mann, 2003, "Are Corporate Bond Ratings Procyclical?", *Special Comment*, Moody's Investor Services, October

Ellis D., 1998, "Different sides of the same story: investors' and issuers' views of rating agencies", *The Journal of Fixed Income* 7(4), 35 – 45

Fitch, 2005, "Fitch Ratings Definitions, www.fitchratings.com

Hand, J.R.M., R.W. Holthausen and R.W Leftwich, 1992, "The effect of bond rating announcements on bond and stock prices", *The Journal of Finance* 47, 733-752

Hull J., M. Predescu and A. White, "The relationship between credit default swap spreads, bonds yields, and credit rating announcements", *Journal of Banking & Finance* 28

Keeman S.C., J.S. Fons and L.V. Carty, "An historical analysis of Moody's Watchlist", *Special Comment*, Moody's Investor Services, October

Löffler G., 2004, "An anatomy of rating through-the-cycle", *Journal of Banking and Finance* 28, 695-720

Moody's Investor Service, 2004, "Guide to Moody's ratings, rating process and rating practices", by J.S. Fons and J. Viswanathan, www.moodys.com, June

Moody's Investor Service, 2002 "Understanding Moody's corporate bond ratings and rating process" by J.S. Fons, R. Cantor and C. Mahoney, Special Comment, May

Rogers, W.H. 1993. "Regression Standard Errors in Clustered Samples." State Technical Bulletin 13:19–23.

Standard & Poor's, 2005, "Corporate Ratings Criteria", www.standardandpoors.com

Standard & Poor's, 2003, "Corporate Ratings Criteria"

Steiner M. and V.G. Heinke, 2001, "Event study concerning international bond price effects of credit rating actions", *International Journal of Finance and Economics* 6, 139-157

Treacy W.F. and M. Carey, 2000, "Credit rating systems at large US banks", *Journal of Banking & Finance* 24, 167-201

Appendix A Definition of credit model ratings including a migration policy

The computation of credit model ratings following a particular migration policy involves two-steps. In the first step credit model scores, CM-scores, are modified to CM^M scores, reflecting a particular migration policy. In the second step, the modified CM^M scores are converted to credit model ratings CM(TH,AF).

The migration policy model has two parameters: a threshold parameter and an adjustment parameter. The threshold parameter TH specifies the size of a credit risk interval [-TH,+TH], in which credit risk is allowed to fluctuate without triggering a rating migration.[10] If a rating migration is triggered, ratings are not fully adjusted to the actual credit risk level. The adjustment fraction AF specifies the partial adjustment of ratings.

Step 1: Modification of CM scores

For each observation, the CM score is converted to a modified score CM^M in such a way that it reflects a specific migration policy, characterized by a threshold TH and an adjustment fraction AF. When following the time-series of the CM_t scores for a particular issuer, modified CM^M_t scores are computed. At the beginning of the time-series of each issuer, CM^M_0 is set equal to CM_0. The CM^M_t score is held constant as long as the CM_t score stays within the threshold interval $(CM^M_{t-1} - \gamma\times TH, CM^M_{t-1} + \gamma\times TH)$:

$$CM^M{}_t = CM^M{}_{t-1} \quad , \quad \text{if} \quad \frac{\left|CM_t - CM^M{}_{t-1}\right|}{\gamma_{N,t}} < TH \tag{A1}$$

where $t \in (0,t^{max})$ and t^{max} is the period of unbroken stay of a particular issuer in the dataset. TH is expressed in notch steps, the scaling factor $\gamma_{N,t}$ converts CM scores to a notch scale. As soon as the CM_t score exceeds the threshold interval, the CM^M_t score is adjusted. If AF = 1, the CM^M_t score is fully adjusted to the current CM score. If AF < 1, the CM^M_t score is partially adjusted to the current CM score as follows:

$$CM^M{}_t = AF\times(CM_t - CM^M{}_{t-1}) + CM^M{}_{t-1} \quad \text{if} \quad \frac{\left|CM_t - CM^M{}_{t-1}\right|}{\gamma_{N.,t}} \geq TH \tag{A2}$$

Step 2: Conversion of CM^M scores to CM(TH,AF) ratings

CM^M scores are converted to CM(TH,AF) ratings, equivalent to ratings, as follows. At the end of each month all issuers are ranked by their CM^M score. On the basis of this ranking, eighteen credit score ratings, Aaa/Aa1, Aa2, Aa3,…., B3, Caa/Ca, equivalent to agency ratings, are assigned to individual issuers. So at the end of each month the number of issuers in each rating category N equals the number of issuers in the equivalent CM rating category. Eighteen rating categories are defined on a "notch" scale level. Rating categories are separated from their neighbors by one notch step.

The time-series of CM^M scores is an irregular pattern of upward and downward jumps. The time period between these jumps varies between 1 and t^{max} years. An unambiguous conversion of these jumps to CM(TH,AF) migrations is crucial to reflect correctly the influence of the migration policy on rating dynamics. This unambiguous conversion is checked and safeguarded as follows. The minimum size of the jump in CM^M scores is $\gamma \times AF \times TH$, which is sufficient to convert nearly all jumps in the modified CM^M score to CM(TH,AF) migrations. The conversion procedure, however, does not prevent a CM(TH,AF) migration from happening, when no jump occurs in the CM^M score. To prevent these non-intended migrations, CM(TH,AF) ratings are replaced by lagged ratings, when the CM^M_t score equals its one-year lagged CM^M_{t-1} score. As a consequence, the distribution of the CM(TH,AF) ratings is slightly altered. The number of observations in each rating category, before and after this correction, differs by 10% at most. This change in rating distribution only marginally affects the comparability of CM(TH,AF) ratings with ratings.

Dynamic properties of credit model ratings as a function of threshold parameter TH

In the absence of a threshold TH, marginal high-frequency (1-2 months) “noise” in credit scores triggers a large number of CM(0,AF) migrations, almost all followed by reversals in subsequent months. To investigate the influence of this noise on rating dynamics, reversal probabilities are derived for RP(TH) ratings as a function of TH (AF = 1) and period P. Reversal probabilities are computed for a period P after the upgrade or downgrade events.

In order to focus as much as possible on reversals due to high-frequency noise in credit scores, this examination is carried out with RP ratings. RP scores are insensitive to the temporary credit risk component since they represent the agencies’ through-the-cycle perspective. As a result, rating reversals caused by temporary changes in credit risk are suppressed. This allows a clearer cut between the marginal high-frequency noise in credit scores and the more moderate and significant dynamics in the permanent credit risk component.

Table A presents the reversal probabilities of RP(TH) ratings as a function of TH and time period P. For RP(0) ratings the reversal probability in the month following a downgrade(upgrade) is 10.3%(13.3%), on a monthly basis. In this case rating dynamics is dominated by (marginal) high-frequency noise in credit scores. The number rating reversals decreases linearly by the length of period P if the number of reversals is dominated by high-frequency noise in credit scores. This is indeed the case for TH = 0. This linear relationship disappears when the TH level is raised to 0.4 notch steps. For TH = 0.6 and above reversal probabilities are comparable for different periods P, which means that the influence of high-frequency noise in credit scores is suppressed and rating dynamics have a flatter frequency spectrum.

The reduction of rating dynamics by an increase in threshold level has little impact on the informational value of credit model ratings, measured by the default prediction performance. For example, the accuracy ratio ACR for RP(TH) ratings decreases by 3.3% when the threshold level is increased from zero to 2.2 notch steps (one year prediction horizon).

Appendix B Measurement of default prediction performance

A well-accepted methodology to measure the overall default prediction performance of a rating scale, weighting type I and type II errors equally in distinguishing defaulters and non-defaulters, is to construct a "cumulative accuracy profile" curve. This CAP curve is obtained by plotting, for each rating category R, the proportion of default observations in the *same and lower* rating category $F_D(R)$ (Y-axis), against the proportion of all survival and default observations in the same and lower rating category $F_A(R)$ (X-axis).

$$F_A(R,T) = \frac{\sum_{C=1}^{R} \sum_{t=1982}^{2002-T} \sum_{i=1}^{N_{C,t}} (S_{C,T,i,t} + D_{C,T,i,t})}{N_{DS}(T)} \tag{B1}$$

where $F_A(0,T) = 0$. $N_{C,i}$ is the total number of observations in rating category C at time t. $S_{C,T,i,t}$ indicates whether an issuer i rated in category C at time t survives at least until t + T (in other words, is the particular observation at least present in the database until t + T). $D_{C,T,i,t}$ indicates whether an issuer i rated in category C at time t defaults within the period (t, t+T). $N_{DS}(T)$ is the total number of default observations ($D_{C,T,i,t} = 1$) and survival observations ($S_{C,T,i,t} = 1$) with a time horizon T in the dataset.

A similar definition holds for $F_D(R,T)$ summing up only the number of default observations.

$$F_D(R,T) = \frac{\sum_{C=1}^{R} \sum_{t=1982}^{2002-T} \sum_{i=1}^{N_{C,t}} D_{C,T,i,t}}{N_D(T)} \tag{B2}$$

where $F_D(0,T) = 0$ and $N_D(T)$ is the total number of default observations with a time horizon T in the dataset.

The higher the proportion of default events happening in the lower categories – in other words the higher the surface below the CAP curve – the better the rating scale performs. The accuracy ratio ACR measures the surface below the CAP curve relative to the surface below the CAP curve for a random rating scale (=½). Based on cumulative default rates ACR is given by

$$ACR(T) = \frac{\sum_{R=1}^{16} \left(\begin{array}{l} [F_A(R,T) - F_A(R-1,T)] \times \\ \qquad [F_D(R-1,T) + \tfrac{1}{2}(F_D(R,T) - F_D(R-1,T))] \end{array} \right) - \frac{1}{2}}{1/2} \tag{B3}$$

ACR varies between 0% (random scale) and 100% (perfect prediction scale). The standard error in ACR is 1.5%, 2% respectively for time horizons T of one year and 3 years. [11] When comparing ACR values of two different scales the standard errors are a factor two lower. [12]

Table I Outlook statistics

Data on Moody's outlooks is obtained from an extended version of the Moody's DRS database. It includes all outlooks provided by Moody's for their ratings in the period September 1991-February 2005. This study covers the January 1995-December 2004 period.
For benchmarking purposes, outlooks are linked with accounting and market data from COMPUSTAT. In order to ensure consistency in accountancy information only non-financial US issuers are selected of which sufficient accounting and market data is available in COMPUSTAT. This selection reduces the number of issuer-monthly observations from 507,824 to 71,962, still including the NOA Outlooks. When the NOA Outlooks are excluded, 52,595 observations are left. The table presents the outlook distribution following a few selection steps. For the final selected 71,962 observations the outlook distribution is broken down to annual periods and major rating categories.

	number of monthly observations	no outlook available	watch down DOWN	negative outlook NEG	stable outlook STA	positive outlook POS	watch up UP
dataset selection, 1995-2004							
all Moody's rated issuers	507,824	40.1%	3.6%	12.7%	35.4%	6.1%	2.1%
+ US issuers	329,866	44.2%	3.7%	12.0%	32.1%	5.9%	2.1%
+ US listed issuers	112,475	33.0%	5.0%	14.4%	35.9%	8.7%	2.9%
+ accounting info available	87,146	32.2%	5.2%	15.3%	35.6%	9.2%	2.5%
+ excluding financial sector	71,962	26.9%	6.0%	15.9%	38.5%	10.0%	2.8%
non-financial US listed issuers with COMPUSTAT accounting information available							
1995	2,188	69.5%	10.5%	4.2%	5.2%	5.1%	5.6%
1996	3,629	51.9%	9.0%	7.1%	12.3%	13.7%	6.1%
1997	5,019	47.2%	5.9%	8.0%	18.6%	15.4%	4.8%
1998	6,655	34.5%	6.0%	9.5%	33.4%	12.9%	3.8%
1999	8,390	25.1%	5.6%	14.7%	41.6%	9.3%	3.8%
2000	8,931	23.7%	5.8%	14.9%	42.9%	9.9%	2.8%
2001	9,199	22.2%	6.4%	19.2%	40.3%	9.4%	2.5%
2002	9,339	26.0%	6.8%	21.1%	35.8%	9.3%	1.0%
2003	9,278	27.2%	5.2%	19.5%	38.8%	8.5%	0.9%
2004	9,334	1.0%	4.0%	20.5%	64.2%	8.0%	2.2%
non-financial US listed issuers with COMPUSTAT accounting information available, 1995 - 2004							
Caa	4,827	17.9%	4.4%	44.2%	23.6%	8.3%	1.6%
B	20,281	16.8%	4.1%	18.2%	44.7%	13.5%	2.6%
Ba	15,440	23.0%	6.1%	10.5%	40.7%	15.4%	4.4%
Baa	17,889	33.3%	6.8%	13.7%	36.3%	7.0%	2.9%
A	11,389	42.9%	8.2%	11.4%	32.6%	3.0%	1.8%
Aa	1,757	34.9%	9.4%	10.0%	40.3%	4.3%	1.0%
Aaa	379	22.2%	2.9%	5.5%	69.4%	-	-

Table II Parameter estimates of default prediction models and rating prediction models

The table presents parameter estimates α and β_i of four default prediction models for various prediction horizons: six months, one-year, six-years and a three year period starting three years in the future. Five versions of a rating prediction model are estimated for various groups of issuers as specified in the table. Standard errors in the logit regression estimation are a generalized version of the Huber and White standard errors, which relaxes the assumptions on the distribution of error terms and independence among observations of the same issuer. [13] z-statistics are given in brackets. Pseudo R^2 is a measure for the goodness of the fit. The last rows of the table give the relative weight of the parameters (see equation 3.7).

	default prediction models				agency rating prediction model				
model	SDP	DP1	LDP	MDP	RP	TTC	RP spec1	RP spec2	RP def
	default prediction horizon				issuers included in estimation				
	half year [0,0.5]	one year [0,1]	six year [0,6]	future period [3,6]	all issuers	all issuers, except Ca, Caa,B3,B2	B1, Ba3, Ba2, Ba1 issuers	Ca, Caa, B3, B2 issuers	issuers 6 months before def
estimation period	82/4-04/6	82/4-03/12	82/4-98/12	82/4-98/12	82/4-04/12	82/4-04/12	82/4-04/12	82/4-04/12	82/4-04/12
parameters regression									
α constant	7.09 (15.5)	6.32 (13.9)	5.33 (10.2)	5.58 (10.2)	ordered logit [1]				
β_1 WK/TA	1.22 (3.3)	0.78 (2.2)	-0.98 (2.6)	-1.53 (3.9)	-1.01 (4.5)	-1.56 (5.8)	0.07 (0.3)	0.57 (1.7)	0.74 (1.9)
β_2 RE/TA	0.44 (2.0)	0.77 (3.7)	1.32 (4.8)	1.16 (4.2)	3.10 (20.2)	3.21 (16.5)	2.53 (10.0)	1.48 (6.5)	1.55 (4.6)
β_3 EBIT/TA	4.97 (5.9)	4.85 (6.0)	0.96 (1.1)	0.03 (0.0)	2.61 (6.2)	1.04 (2.0)	0.47 (0.8)	4.14 (6.1)	2.68 (3.0)
β_4 ME/BL	1.09 (16.1)	0.98 (14.5)	0.85 (10.5)	0.67 (7.9)	1.02 (23.4)	1.06 (19.3)	0.75 (12.0)	0.44 (7.1)	0.84 (8.2)
β_5 Size	0.24 (5.2)	0.27 (5.7)	0.38 (6.8)	0.33 (5.8)	1.03 (25.9)	1.10 (22.7)	0.85 (17.1)	0.09 (1.6)	0.33 (4.6)
β_6 SD(AR)	-4.38 (6.1)	-4.77 (6.7)	-5.81 (7.8)	-3.07 (3.8)	-7.83 (16.7)	-8.06 (12.3)	-4.39 (7.2)	-3.14 (6.6)	-5.96 (7.6)
β_7 AR	12.81 (10.4)	12.76 (10.9)	5.50 (4.7)	1.25 (0.9)	-4.53 (8.7)	-6.06 (9.5)	-4.36 (6.1)	-0.82 (1.1)	2.51 (1.7)
pseudo R^2	0.446	0.415	0.263	0.141	0.240	0.220	0.131	0.104	0.161
# obs.	184,050	178,164	114,535	107,445	189,248	161,663	66,861	27,585	2,130
# default obs.	2108	4262	13,304	6,214	-	-	-	-	-
relative weight model variables									
WK/TA	6.8%	4.4%	-6.7%	-13.4%	-3.6%	-5.7%	0.4%	5.4%	4.6%
RE/TA	4.9%	8.6%	16.2%	18.4%	22.0%	21.6%	21.7%	21.9%	15.8%
EBIT/TA	12.4%	12.2%	2.8%	0.1%	4.2%	3.1%	1.2%	19.8%	8.4%
MV/LIB	37.2%	33.8%	31.6%	32.1%	22.4%	21.2%	25.2%	27.8%	28.4%
Size	11.7%	12.8%	21.7%	24.5%	31.7%	34.0%	35.6%	5.6%	14.2%
SD(AR)	-10.6%	-11.6%	-13.7%	-9.3%	-12.2%	-10.0%	-9.6%	-16.8%	-22.9%
AR	16.5%	16.6%	7.4%	2.2%	-3.8%	-4.4%	-6.3%	-2.8%	5.7%

[1] Due to space considerations the 15 boundary parameters B_R in the ordered logit model are not shown.

Table III Parameter estimates of outlook prediction models

The table presents parameter estimates α and β_i of various outlook prediction models. All models, except the OPW model, are estimated by weighting the five outlook categories by their occurrence. The OPW model is estimated by equally weighting the five outlook categories.

Standard errors in the (ordered) logit regression estimation are a generalized version of the Huber and White standard errors, which relaxes the assumptions on the distribution of error terms and independence among observations of the same issuer. z-statistics are given in brackets. Pseudo R^2 is a measure for the goodness of the fit. The last rows of the table give the relative weight of the parameters (see equation 3.7).

outlooks in regression analysis	DOWN NEG STA POS UP	DOWN STA	NEG STA	STA POS	STA UP	DOWN NEG STA POS UP
model	OP	O-DOWN	O-NEG	O-POS	O-UP	OPW
parameters regression						
α constant	ordered logit[1]	2.11 (36.3)	0.931 (16.7)	-1.44 (20.0)	-2.97 (36.3)	ordered logit[1]
β_1 ΔWK/TA	-0.004 (0.1)	0.187 (3.3)	-0.034 (0.6)	0.024 (0.3)	-0.332 (4.4)	-0.080 (1.8)
β_2 ΔRE/TA	-0.080 (2.4)	-0.082 (1.5)	0.063 (1.2)	-0.175 (2.9)	-0.207 (3.6)	-0.120 (3.6)
β_3 ΔEBIT/TA	0.180 (5.1)	0.255 (4.1)	0.108 (2.0)	0.192 (3.1)	0.043 (0.6)	0.152 (3.5)
β_4 ΔME/BL	0.409 (10.7)	0.332 (5.8)	0.416 (6.8)	0.238 (3.3)	0.349 (4.4)	0.481 (10.4)
β_5 ΔSize	0.031 (0.9)	-0.158 (3.0)	-0.053 (1.0)	0.136 (1.9)	0.423 (6.2)	0.151 (3.5)
β_6 ΔSD(AR)	-0.127 (5.4)	-0.291 (8.9)	-0.079 (2.4)	-0.010 (0.2)	0.063 (1.3)	-0.129 (3.9)
β_7 ΔAR	0.172 (6.6)	0.336 (7.6)	0.031 (0.8)	0.028 (0.7)	0.302 (4.8)	0.350 (9.1)
pseudo R^2	0.039	0.099	0.045	0.023	0.083	0.058
# obs.	52,595	31,996	39,085	34,862	29,689	52,595
relative weight model variables						
ΔWK/TA	-0.4%	10.1%	-3.9%	2.7%	-17.7%	-4.9%
ΔRE/TA	-7.8%	-5.0%	7.9%	-22.1%	-12.2%	-8.1%
ΔEBIT/TA	16.9%	14.6%	13.1%	23.2%	2.4%	9.8%
ΔMV/LIB	40.3%	20.1%	53.0%	30.3%	20.8%	32.8%
ΔSize	3.0%	-9.2%	-6.6%	16.7%	24.3%	9.9%
ΔSD(AR)	-14.5%	-20.3%	-11.6%	-1.5%	4.3%	-10.2%
ΔAR	17.2%	20.6%	4.0%	3.6%	18.2%	24.2%

[1] Due to space considerations the boundary parameters B_R in the ordered logit model are not shown. These are available on request.

Table IV Sensitivity of credit scoring models to the permanent and temporary credit risk component in credit risk

Credit scoring models are re-estimated with two model variables: proxies for the permanent and temporary credit risk component. The permanent credit risk component is proxied by TTC scores and the temporary credit risk component is proxied by the difference in SDP scores and TTC scores. TTC and SDP credit scoring models are estimated in a first stage estimation, see Table II.

The table presents parameter estimates α and β_i of various credit scoring prediction models. Standard errors in the logit regression estimation are a generalized version of the Huber and White standard errors, which relaxes the assumptions on the distribution of error terms and independence among observations of the same issuer. z-statistics are given in brackets. Pseudo R^2 is a measure for the goodness of the fit. The last rows of the table give the relative weight of the parameters (see equation 3.7).

	default prediction models			agency rating prediction models					outlook prediction models				
model	SDP	LDP	MDP	TTC	RP spec1	RP spec2	RP def	model	OP	O-DOWN	O-NEG	O-POS	O-UP
	default prediction horizon			issuers included in estimation					outlooks in regression analysis				
	half year horizon [0,0.5]	six year horizon [0,6]	future period horizon [3,6]	all issuers except Ca, Caa, B3 issuers	B1, Ba3, Ba2 Ba1 issuers	Ca, Caa, B3 B2 issuers	Ca, Caa, B3 B2 issuers		DOWN NEG STA POS UP	DOWN STA	NEG STA	STA POS	STA UP
estimation period	82/4-04/6	82/4-98/12	82/4-98/12	82/4-98/12	82/4-04/12	82/4-04/12	82/4-04/12	estimation period	95/1-04/12	95/1-04/12	95/1-04/12	95/1-04/12	95/1-04/12
parameter estimates regression													
α constant	0	2.56	4.36	ordered logit [1]				α constant	o-logit [1]	2.069	0.942	-1.427	-2.806
	(0.0)	(5.0)	(8.2)							(37.0)	(17.4)	(20.4)	(34.8)
β_1 temporary	1	0.431	0.206	0	0.042	0.215	0.468	β_1 Δ temporary	0.637	0.969	0.514	0.301	0.310
	(25.9)	(10.9)	(4.5)	(0.0)	(1.5)	(8.2)	(9.1)		(18.5)	(16.6)	(9.2)	(4.8)	(3.9)
β_2 permanent	1	0.753	0.513	1	0.718	0.437	0.798	β_2 Δ permanent	0.637	0.563	0.475	0.380	0.696
	(35.5)	(20.8)	(12.7)	(29.4)	(22.7)	(14.1)	(14.6)		(14.9)	(9.0)	(8.1)	(5.2)	(7.7)
pseudo R^2	0.446	0.259	0.134	0.220	0.125	0.077	0.154	pseudo R^2	0.036	0.085	0.033	0.017	0.041
# observations	184,050	114,535	107,445	161,663	66,861	27,585	2,130	# observations	52,595	31,996	39,085	34,862	29,689
relative weight model variables													
temporary	41.4%	29.4%	22.6%	0.0%	5.4%	31.9%	32.7%	Δ temporary	51.9%	65.7%	53.1%	42.0%	28.9%
permanent	58.6%	70.6%	77.4%	100.0%	94.6%	68.1%	67.3%	Δ permanent	48.1%	34.3%	46.9%	58.0%	71.1%

[1] Due to space considerations the boundary parameters B_R in the ordered logit model are not shown. These are available on request.

Table V Credit risk dispersion indicated by the actual and simulated outlook scale

Credit risk variations within a rating category N are measured by variations in credit scores $CS_{i,t}$ compared to the average credit model score $CS_{N,t}$ for all issuers in a rating category N at time t: $\Delta CS = (CS_{i,t} - CS_{N,t}) / \gamma_{N,t}$. ΔCS are converted to a notch rating scale by the scaling factor $\gamma_{N,t}$. For various credit scoring models the table presents the average ΔCS values unconditionally and conditionally to rating migration events as indicated in the table. ΔCS values are given for actual outlooks (panel A) and simulated outlooks (panel B).

outlook	number of observations	average ΔCS-score (in notch steps)					
		outlook prediction models		default prediction models and rating prediction model (RP)			
		ΔOP	ΔOPW	ΔSDP	ΔDP1	ΔLDP	ΔRP
Panel A: actual outlooks							
all observations							
DOWN	4,317	-2.7	-2.7	-2.8	-2.5	-1.4	-0.5
NEG	11,406	-1.4	-1.5	-1.2	-1.1	-0.8	-0.6
STA	27,679	0.4	0.2	0.5	0.5	0.2	0.1
POS	7183	1.7	1.7	1.6	1.4	0.8	0.5
UP	2010	1.8	2.8	1.9	1.7	1.4	0.9
NOA	19,367	-0.4	0.0	-0.4	-0.3	0.1	0.3
observations with no rating migration event ΔN in the past 12 months and future 12 months							
DOWN	1,149	-0.6	-0.3	-0.8	-0.7	-0.3	-0.2
NEG	5,462	-0.4	-0.3	-0.6	-0.5	-0.4	-0.5
STA	17,346	0.6	0.4	0.8	0.7	0.4	0.1
POS	3,915	1.6	1.6	1.6	1.4	0.9	0.5
UP	887	1.4	2.2	1.2	1.0	0.9	0.5
NOA	9,607	0.1	0.5	0.1	0.2	0.3	0.4
observations in a half yearly period before and after a downgrade event ΔN							
DOWN	2,663	-3.8	-4.1	-3.9	-3.5	-2.0	-0.7
NEG	3,368	-3.2	-3.7	-2.4	-2.3	-1.5	-0.6
STA	3,605	-1.6	-2.0	-1.6	-1.4	-0.9	-0.3
observations in a half yearly period before and after an upgrade event ΔN							
STA	2,045	2.5	3.2	2.5	2.2	1.2	0.4
POS	1,220	3.2	3.9	2.7	2.4	1.5	0.7
UP	763	2.5	3.9	2.8	2.7	2.0	1.5
Panel B: simulated OP(0) outlooks							
all observations							
DOWN	4,317	-6.2	-6.6	-6.4	-5.9	-3.6	-1.8
NEG	11,406	-3.3	-3.5	-2.9	-2.7	-1.7	-0.9
STA	27,679	0.6	0.5	0.6	0.6	0.4	0.2
POS	7,183	4.6	4.9	4.4	3.9	2.3	1.0
UP	2,010	7.5	7.9	7.6	6.7	3.5	1.5
observations with no rating migration event ΔN in the past 12 months and future 12 months							
DOWN	1,335	-5.6	-5.6	-6.2	-5.6	-3.5	-2.1
NEG	5,194	-3.0	-3.0	-3.1	-2.8	-1.7	-1.0
STA	17,101	0.7	0.5	0.8	0.8	0.5	0.2
POS	4,079	4.7	4.7	4.6	4.1	2.4	1.1
UP	1,050	7.5	7.5	7.6	6.7	3.5	1.6
observations in a half yearly period before and after a downgrade event ΔN							
DOWN	2,182	-6.7	-7.3	-6.7	-6.1	-3.7	-1.6
NEG	3,706	-3.8	-4.3	-2.9	-2.7	-1.8	-0.7
STA	3,909	-0.2	-0.4	-0.2	-0.2	0.0	0.2
observations in a half yearly period before and after an upgrade event ΔN							
STA	2,086	1.2	2.0	1.0	1.0	0.8	0.4
POS	1,365	4.6	5.4	4.2	3.8	2.2	0.9
UP	528	7.6	8.6	8.1	7.2	3.8	1.7

Table VI Outlook distribution conditional to a rating migration event

The table presents the distribution of actual outlooks and simulated OP(0) outlooks as a function of time relative to a rating migration event ΔN in period (-1,0). In the first column the average outlook distribution is given for observations with no ΔN happened in the past 12 months and future 12 months.

outlook		timing t relative to an agency migration in period (-1,0)							
	t > 11, t < -12	t = -12	t = -6	t = -3	t = -1	t = 0	t = 2	t = 5	t = 11
actual outlook distribution, conditional to a rating downgrade in (-1,0)									
DOWN	4.5%	3.9%	14.1%	38.3%	56.7%	18.9%	12.6%	8.3%	5.4%
NEG	25.1%	23.2%	27.6%	22.9%	17.1%	42.3%	44.1%	42.5%	39.1%
STA	58.1%	59.0%	48.3%	31.5%	20.9%	33.6%	36.9%	41.2%	44.5%
POS	10.6%	12.0%	9.4%	6.6%	4.9%	4.5%	5.4%	6.8%	9.0%
UP	1.7%	1.8%	0.6%	0.7%	0.4%	0.7%	1.0%	1.1%	2.0%
DOWN+NEG	29.6%	27.1%	41.7%	61.2%	73.8%	61.2%	56.7%	50.8%	44.5%
UP+POS	12.3%	13.8%	10.0%	7.3%	5.3%	5.2%	6.4%	7.9%	11.0%
simulated OP(0) outlook distribution, conditional to a rating downgrade in (-1,0)									
DOWN	5.8%	9.4%	15.8%	25.8%	35.4%	18.2%	16.1%	12.7%	6.1%
NEG	22.9%	22.5%	33.1%	33.1%	32.2%	41.0%	38.3%	36.6%	31.8%
STA	57.2%	54.8%	44.1%	35.2%	28.0%	35.1%	40.4%	46.2%	56.0%
POS	11.7%	11.5%	6.0%	5.2%	3.2%	5.0%	4.7%	4.1%	5.9%
UP	2.5%	1.8%	0.9%	0.7%	1.2%	0.8%	0.5%	0.3%	0.2%
DOWN+NEG	28.7%	31.9%	48.9%	58.9%	67.6%	59.2%	54.4%	49.3%	37.9%
UP+POS	14.2%	13.3%	6.9%	5.9%	4.4%	5.8%	5.2%	4.4%	6.1%
actual outlook distribution, conditional to a rating upgrade in (-1,0)									
DOWN	3.5%	1.2%	1.0%	1.3%	1.3%	1.0%	1.4%	1.7%	2.6%
NEG	9.8%	11.4%	9.8%	7.1%	4.8%	4.6%	3.9%	5.0%	4.2%
STA	62.7%	44.6%	40.6%	29.8%	23.0%	63.2%	63.8%	66.2%	61.9%
POS	20.1%	33.7%	34.3%	29.8%	24.5%	27.4%	27.6%	24.1%	28.0%
UP	3.9%	9.0%	14.3%	31.9%	46.4%	3.9%	3.3%	3.0%	3.2%
DOWN+NEG	13.3%	12.6%	10.8%	8.4%	6.1%	5.6%	5.3%	6.7%	6.8%
UP+POS	24.0%	42.7%	48.6%	61.7%	70.9%	31.3%	30.9%	27.1%	31.2%
simulated OP(0) outlook distribution, conditional to a rating upgrade in (-1,0)									
DOWN	3.6%	1.8%	2.4%	2.1%	1.7%	3.4%	1.7%	3.3%	4.2%
NEG	11.4%	10.8%	5.9%	7.1%	6.5%	4.8%	4.7%	6.0%	10.6%
STA	61.8%	48.8%	52.4%	45.3%	45.8%	49.2%	50.0%	47.8%	48.1%
POS	17.6%	30.7%	27.6%	29.1%	28.4%	32.7%	35.1%	34.1%	29.1%
UP	5.6%	7.8%	11.5%	16.5%	17.6%	9.9%	8.6%	8.7%	7.9%
DOWN+NEG	15.0%	12.6%	8.3%	9.2%	8.2%	8.2%	6.4%	9.3%	14.8%
UP+POS	23.2%	38.5%	39.1%	45.6%	46.0%	42.6%	43.7%	42.8%	37.0%

Table VII Outlook migration matrix

Panel A of the table presents the monthly outlook migration probabilities for six outlooks, including the NOA Outlooks (NOA = no outlook available). Panel B of the table presents the outlook migration matrix after converting all NOA Outlooks to STA Outlooks. This allows the migration matrix of actual outlooks to be compared with the migration matrix of simulated OP(1.5) outlooks. Panel C of the table presents the outlook migration matrix after merging the DOWN and NEG categories and merging the UP and POS categories.

Observations with no succeeding observation available in the dataset – "exit" observations – are excluded from the computation of the outlook migration matrices. The numbers of default observations and "exit" observations – observations at the end of an issuer time series in the database – are given in last column.

panel A: NOA Outlooks are treated separately in the migration matrix

	initial outlook at t = 0	# obs.	end outlook at t = 1, one month later								# exit
			DOWN	NEG	STA	POS	UP	NOA	default (% and number)		
actual outlook	DOWN	4,235	76.8%	4.1%	4.0%	0.3%	0.3%	14.2%	0.40%	17	82
	NEG	11,181	2.0%	95.0%	1.5%	0.2%	0.3%	0.2%	0.89%	100	225
	STA	26,997	1.1%	0.9%	96.9%	0.5%	0.5%	0.1%	0.08%	21	682
	POS	7,065	0.6%	0.6%	2.0%	95.8%	0.8%	0.1%	0.03%	2	118
	UP	1,873	0.2%	0.2%	4.2%	1.4%	83.9%	10.1%	0.05%	1	137
	NOA	19,200	1.4%	0.8%	1.4%	0.3%	0.5%	95.2%	0.26%	49	167

panel B: NOA Outlooks are converted to STA outlooks

	initial outlook at t = 0	# obs.	end outlook at t = 1, one month later							# exit
			DOWN	NEG	STA	POS	UP	default (% and number)		
actual outlook	DOWN	4,235	76.8%	4.1%	18.2%	0.3%	0.3%	0.40%	17	82
	NEG	11,181	2.0%	95.0%	1.6%	0.2%	0.3%	0.89%	100	225
	STA	46,197	1.2%	0.9%	96.9%	0.4%	0.5%	0.15%	70	849
	POS	7,065	0.6%	0.6%	2.2%	95.8%	0.8%	0.03%	2	118
	UP	1,873	0.2%	0.2%	14.4%	1.4%	83.9%	0.05%	1	137
OP(1.5) outlook	DOWN	4,087	85.9%	9.0%	3.9%	0.0%	0.0%	1.08%	44	65
	NEG	10,495	3.3%	88.6%	6.9%	0.0%	0.0%	1.20%	126	216
	STA	46,906	0.4%	1.9%	96.5%	1.0%	0.1%	0.04%	20	917
	POS	7,028	0.1%	0.1%	8.6%	89.3%	2.0%	0.00%	0	155
	UP	2,035	0.0%	0.0%	5.2%	8.0%	86.9%	0.00%	0	58

panel C: DOWN and NEG categories are merged, UP and POS categories are merged

	initial outlook at t = 0	# obs.	end outlook at t = 1, one month later					# exit
			DOWN/ NEG	STA	POS/ UP	default (% and number)		
actual outlook	DOWN/NEG	15,416	92.6%	6.2%	0.4%	0.76%	117	307
	STA	46,197	2.1%	96.9%	0.9%	0.15%	70	849
	POS/UP	8,938	1.0%	4.7%	94.2%	0.03%	3	255
OP(1.5) outlook	DOWN/NEG	14,582	92.8%	6.0%	0.0%	1.17%	170	281
	STA	46,906	2.3%	96.5%	1.2%	0.04%	20	917
	POS/UP	9,063	0.1%	7.8%	92.1%	0.00%	0	213
OP(2.0) outlook	DOWN/NEG	14,414	93.9%	4.9%	0.0%	1.17%	169	288
	STA	46,764	1.9%	97.1%	0.9%	0.04%	21	909
	POS/UP	9,373	0.1%	6.4%	93.5%	0.00%	0	214

Table VIII Default prediction performance of adjusted ratings

The table presents the default prediction performance of adjusted and unadjusted rating scales. Default prediction performance of a rating scale is measured by an accuracy ratio ACR (see Appendix B). ACR weights type I and type II errors equally in distinguishing defaulters and non-defaulters. It varies between 0% (random scale) and 100% (perfect prediction scale).
The table reports the ACR values for the unadjusted rating scale N, ACR(N), and the difference in ACR values between an adjusted rating scale R and the rating scale N, ΔACR(R). ACR(N) and ΔACR(R) values are given for various prediction horizons, ranging from 6 months to 36 months. Standard error in ACR(N) is 1.5% for a 6 month prediction horizon and 2% for a 3 year horizon. The standard error in ΔACR(R) is a factor of two lower.
ΔACR(R) values are computed for adjustments based on actual outlooks and simulated OP(TH) outlooks. The applied adjustment schemes are given in the table. Adjustment scheme 1neg and 1pos restrict the adjustment of ratings to respectively the downside of the outlook scale (DOWN: -2.5, NEG: -1.5, STA: +0.5, POS: + 0.5 and UP: + 0.5) and the upside of the outlook scale (DOWN: +0.5, NEG: +0.5, STA: +0.5, POS: + 1.5 and UP: + 2).

	basis of outlook adjustment	adjustment scheme	prediction horizon (months) 6	12	18	24	36
adjustment scheme 1, DOWN: -2.5 , NEG: -1.5, STA, +0.5, POS: +1.5, UP:+2							
ACR(N)	no adjustment	-	76.8%	69.1%	65.3%	62.4%	57.2%
ΔACR(R)	actual outlook	1	4.1%	3.7%	2.2%	1.5%	1.5%
ACR(N)	actual outlook	1neg	3.7%	2.9%	1.2%	0.6%	0.7%
ΔACR(R)	actual outlook	1pos	1.3%	1.7%	1.8%	1.6%	1.4%
ΔACR(R)	OP(TH = 0)	1	8.7%	9.1%	8.0%	6.9%	5.7%
ΔACR(R)	OP(TH = 0)	1neg	8.1%	8.3%	7.1%	6.1%	5.0%
ΔACR(R)	OP(TH = 0)	1pos	1.8%	2.1%	2.1%	1.9%	1.7%
adjustment scheme 2, DOWN/NEG: -1.5 , STA + 0.5, POS/UP:+1.5							
ΔACR(R)	actual outlook	2	3.9%	3.5%	2.3%	1.7%	1.7%
ΔACR(R)	OP(TH = 0)	2	7.5%	8.0%	7.1%	6.2%	5.3%
ΔACR(R)	OP(TH = 3.0)	2	6.9%	7.5%	6.5%	5.4%	4.5%
adjustment scheme 3, DOWN: -4.5 , NEG: -2.5, STA, +0.5, POS: +3, UP:+4.5							
ΔACR(R)	actual outlook	3	3.4%	3.1%	1.0%	-0.1%	-0.5%
ΔACR(R)	OP(TH = 0)	3	9.7%	10.2%	8.4%	6.7%	4.9%
ΔACR(R)	OP(TH = 3.0)	3	8.5%	8.8%	6.6%	4.6%	3.1%
adjustment scheme 4, DOWN: -6.5 , NEG: -3.5, STA, +0.5, POS: +4.5, UP:+7.5							
ΔACR(R)	actual outlook	4	1.5%	1.1%	-1.7%	-3.1%	-4.0%
ΔACR(R)	OP(TH = 0)	4	9.2%	9.5%	7.1%	4.9%	2.3%
ΔACR(R)	OP(TH = 3.0)	4	7.7%	7.4%	4.4%	1.8%	-0.5%

Table IX Properties of adjusted ratings

The table presents the properties of unadjusted ratings, ratings adjusted by their actual outlooks and simulated OP(TH) outlooks and credit model ratings RP(TH,AF) and LDP(TH). Adjustment by outlooks follows adjustment scheme 1 (DOWN: -2.5 , NEG: -1.5, STA: +0.5, POS: +1.5, UP:+2.0). Credit model ratings are computed following the procedure described in appendix A. Simulated RP(1.98,0.7/0.6) ratings are based on RP scores and a migration policy with a threshold TH of 1.8 notch steps and an adjustment fraction of 0.7 at the downside and 0.6 at the upside. RP(TH) and LDP(TH) ratings are based on respectively RP scores and LDP scores and a migration policy with threshold TH and an adjustment fraction of 1.

The following properties are reported: rating stability, rating drift, sensitivity to temporary credit risk component and default prediction performance. Rating stability is measured in terms of migration probabilities per month. Rating drift is measured by the average migration two years after a rating migration event ΔR. Sensitivity to the temporary credit risk component is measured by the relative weight to the temporary credit risk variable in the logit regression analysis, as described in section 4.4. Default prediction performance is measured by the difference in ACR between an adjusted rating scale R and the unadjusted rating scale N, ΔACR(R). The prediction horizon is one year.

rating	outlook adjustment	downgrade			upgrade			sensitivity to temporary credit risk component			ΔACR(R) one year horizon
		monthly downgrade probability	average migration conditional to a down-grade in period (-1,0)		monthly upgrade probability	average migration conditional to an up-grade in period (-1,0)		all issuers	all, except Ca, Caa B3, B2 issuers	only Ca,Caa B3, B2 issuers	
			(-1,0)	(0,24)		(-1,0)	(0,24)				
actual N	no adjustment	1.8%	-1.40	-0.33	0.7%	1.22	0.48	10.2%	0.0%	27.8%	69.1%
RP(1.8,0.7/0.6)	no adjustment	1.7%	-1.40	-0.28	0.9%	1.26	0.45	1.5%	1.4%	0.2%	-2.5%
RP(1.8,1)	no adjustment	1.7%	-1.86	0.09	1.0%	1.69	0.11	3.4%	4.2%	1.0%	-1.7%
RP(0,1)	no adjustment	16.1%	-0.68	0.08	12.7%	0.66	-0.13	12.2%	12.9%	11.0%	1.8%
actual N	actual outlook	3.0%	-1.93	0.41	1.9%	1.59	0.00	18.7%	15.2%	35.0%	3.7%
actual N	OP (TH=1.5)	5.1%	-1.45	0.33	3.4%	1.40	-0.31	27.3%	27.0%	40.3%	9.1%
LDP(TH=1)	no adjustment	6.3%	-1.41	0.38	5.5%	1.15	-0.35	29.3%	30.3%	28.9%	10.2%
LDP(TH=1.5)	no adjustment	4.0%	-1.80	0.44	3.3%	1.46	-0.40	28.3%	29.5%	27.6%	9.1%

Table A Influence of the threshold parameter TH on rating dynamics of RP(TH) ratings

The table presents the properties of credit model ratings RP(TH). Credit model ratings RP(TH) are computed following the procedure described in appendix A. RP(TH) ratings are based on respectively RP scores and a migration policy with threshold TH and an adjustment fraction of 1.The following properties are reported: reversal probabilities - on a monthly basis - for various periods P after a rating migration in (-1,0) and the default prediction performance by accuracy ratios ACR.

period P after a rating migration in (-1,0)	RP(TH,AF = 1), various TH (notch steps) 0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	N	RP (1.8,0.7/0.6)
reversal probability conditional to a downgrade in month (-1,0)														
one month	10.3%	3.3%	1.2%	0.7%	0.4%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%
six month	2.5%	2.0%	1.4%	1.0%	0.8%	0.6%	0.5%	0.4%	0.3%	0.3%	0.3%	0.2%	0.1%	0.1%
one year	1.4%	1.3%	1.1%	1.0%	0.9%	0.8%	0.7%	0.6%	0.6%	0.5%	0.5%	0.5%	0.2%	0.3%
two year	0.7%	0.7%	0.6%	0.6%	0.6%	0.6%	0.5%	0.5%	0.5%	0.5%	0.4%	0.4%	0.2%	0.3%
reversal probability conditional to an upgrade in month (-1,0)														
one month	13.3%	5.3%	2.3%	1.4%	0.8%	0.6%	0.4%	0.4%	0.2%	0.2%	0.1%	0.0%	0.2%	0.1%
six month	3.5%	3.0%	2.3%	1.7%	1.3%	1.0%	0.8%	0.7%	0.6%	0.5%	0.5%	0.4%	0.2%	0.2%
one year	1.9%	1.8%	1.5%	1.2%	1.1%	0.9%	0.7%	0.6%	0.5%	0.5%	0.5%	0.4%	0.2%	0.3%
two year	1.1%	1.0%	0.9%	0.8%	0.7%	0.7%	0.6%	0.5%	0.5%	0.4%	0.4%	0.4%	0.3%	0.3%

prediction time horizon	RP(TH,AF = 1), various TH (notch steps) 0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	N	RP (1.8,0.7/0.6)
accuracy ratio ACR (%)														
one year	77.1%	77.0%	76.8%	76.7%	76.5%	76.3%	75.7%	75.8%	75.6%	74.8%	74.6%	73.8%	76.8%	74.0%
three years	67.2%	67.2%	67.1%	66.9%	66.9%	66.7%	66.5%	66.5%	66.3%	65.7%	65.5%	64.9%	68.3%	65.3%
six years	59.6%	59.6%	59.5%	59.5%	59.4%	59.3%	59.1%	59.0%	59.2%	58.7%	58.6%	58.2%	61.4%	58.4%

Endnotes

1 An alternative, including observations of issuers defaulting in period (t, t+T_1) in the analysis by setting $p_{i,t}$ = 1 for these observations, does not change the model estimation significantly, since the number of defaulting observations is relatively small compared to the number of surviving observations.

2 We have not revealed the parameters of the Fitch through-the-cycle methodology.

3 see The Financial Times, 19 January 2002, "Moody's mulls changes to its ratings process".

4 Default prediction models estimated with the Moody's default dataset are fairly similar to those estimated with the Standard & Poor's definition of default (see Altman and Rijken, 2004). Although the Moody's default definition differs from the Standard & Poor's default definition, the default prediction models estimated with Moody's default events are as good as equal to default prediction models estimated with Standard & Poor's default events. In contrast to Standard & Poor's, Moody's counts delayed payments made within a grace period and explicitly counts issuer files for bankruptcy (Chapter 11 and Chapter 7) and legal receivership. Furthermore, the rating prediction model estimated with Moody's ratings is almost an exact replication of the rating prediction model estimated with Standard & Poor's ratings. This is not surprising as Moody's ratings differ only by 1 to 2 notch steps at most from those of Standard & Poor's.

5 Robustness tests show that outlook prediction model parameters do not change much along the entire rating scale N. Even in the extreme case of Caa category, RW values and model parameters are comparable to investment graded categories and other speculative graded categories. Other robustness tests shows that just before and just after a rating migration event the outlook scale relies more on short-term trends in stock prices at the expense of the ME/LIB variable. Similar differences are observed between time period 1995-1999 to 2000-2004. More recently the outlook scale depends less on equity trends and more on market leverage.

6 In a first attempt to characterize the dynamic properties of X_T and X_P we decomposed these scores into a permanent component and a cyclical component. A significant cyclical component showed up for X_T with a cycle of 3-4 years. However this cyclical component contains no credit risk information. It has no added value in explaining one-year default probabilities.

7 In the distribution for "exit" observations STA Outlooks and UP Reviews are slightly outnumbered compared to the distribution for all observations by respectively 48.3% vs. 38.3% and 9.7% vs. 2.7%, while NOA Outlooks are underrepresented by 11.8% vs. 27.2%.

8 Sensitivity of ratings to the temporary credit risk component is measured by the weight to the temporary credit risk component X_T in the logit regression analysis as described in section 4.4. The weight to X_T is by definition 0% for the unadjusted ratings of all issuers rated B1 and above. For highly distressed Ca, Caa, B3 and B2 rated issuers the rating scale has a 27.8% weight to X_T which comparable to a long-term point-in-time measure (see table IV). Simulated RP(1.8,0.7/0.6) ratings are not sensitive to X_T for the entire rating scale as they reproduce the agency rating dynamics. Relaxing the migration policy by threshold removal (TH = 0) and full adjustment (AF = 1) increases the sensitivity to X_T from 1.5% to 12.2%. As part of the through-the-cycle methodology, the prudent rating migration policy substantially reduces the sensitivity to X_T.

[9] In order to compare threshold levels for different credit model ratings the threshold levels have to be rescaled to the volatility level of the underlying credit scores. The 1.8 notch steps threshold for ratings and the 1.5 threshold for OP(1.5) ratings translate into respectively an equivalent threshold of 2.1 and 0.7 for LDP ratings. The average of 1.4 notch steps is close to the implied threshold TH of 1.25 notch steps.

[10] The minimum threshold level imposed by a discrete agency rating scale is 0.5 notch steps.

[11] The stochastic defaulting process can be modeled by the following exponential distribution function $\alpha \times \exp(-\alpha F_A)$. With this distribution function the CAP curve can be modeled by $1 - \exp(-\alpha F_A)$ with $F_A < 1$. The surface below the CAP curve is $1 - 1/\alpha$, when approximating $\exp(-\alpha) \approx 0$. In that case ACR is $1 - 2/\alpha$. In a sampling experiment with n defaulting events the expected average F_A for the exponential distribution is $1/\alpha$ and the variance in F_A $VAR(F_A)$ is $1/(n\,\alpha^2)$. In that case the standard error in ACP is $2/(\alpha\sqrt{n})$. For a time horizon of three years, a best fit with the actual CAP curve is obtained for $\alpha = 10$, so the standard error is 0.020 ($n = 162$). For a one year horizon the standard error is 0.015.

[12] The standard errors in ΔACR are 0.75 percent for T = one year, 1.0 percent for T = three years, and 1.25 percent for T = six years. The standard errors in comparing differences between accuracy ratios of agency ratings and credit-model ratings, $\sigma(\Delta ACR)$, are lower than the standard error of ACR itself because the underlying stochastic defaulting process (same dataset and same defaulting events) is the same for all rating scales. Because the CAP curves of agency ratings and CM ratings are comparable, variation in this stochastic process are expected to have a comparable impact on the ACRs of these ratings. However, a standard error $\sigma(\Delta ACR)$ still exists. An approximation of $\sigma(\Delta ACR)$ for the pooled sample was obtained from a time-series analysis of the ACR and ΔACR. The standard deviation in annual times series of the ACR for agency ratings and CM ratings is roughly 2 percent higher than the standard deviation in annual time series of ΔACR for these ratings. So, based on the pooled sample's standard errors for ACR, the pooled sample standard error for $\sigma(\Delta ACR)$ is approximately 0.75 percent for a time horizon of one year and goes up to 1.25 percent for a time horizon of six years.

[13] The standard errors in the logit regression estimation are a generalized version of the Huber and White standard errors. In a standard logit model setting, the error terms, ε_i, are assumed to be identically and independently distributed [$var(\varepsilon_i) = \sigma^2$, $cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$]. In reality, these conditions are violated. To obtain the correct statistics, Huber–White standard errors are used to relax the assumption of homoscedasticity. A generalization of Huber–White standard errors (Rogers 1993) also relaxes the assumption of independence among all observations. Instead, only independence between observations of different companies is assumed. "Pseudo R^2" is a measure of the goodness of the fit.