

The effects of rating through the cycle on rating stability, rating timeliness and default prediction performance

Edward I. Altman¹ and Herbert A. Rijken²

December 2004

JEL classification: G20, G33

Keywords: Rating agencies, through-the-cycle rating methodology, migration policy, credit scoring models, default prediction

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 Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. email: hrijken@feweb.vu.nl

Abstract

The role and performance of credit rating agencies are currently under debate. Several surveys conducted in the United States reveal that most investors believe that rating agencies are too slow in adjusting their ratings to changes in corporate creditworthiness. Well known is that agencies achieve rating stability by their through-the-cycle methodology. This study aims to provide quantitative insight in this methodology and to quantify the effects of this methodology on rating stability, rating timeliness and default prediction performance, from an investor's point-in-time perspective. We believe that our results can guide the search for an optimal balance between rating stability, rating timeliness and default prediction performance.

I Introduction

The quality of corporate credit ratings issued by the major rating agencies has come under closer scrutiny, especially because there is a perception that the adjustment of these ratings is too slow. Recent surveys reveal that investors are not satisfied with the timeliness of ratings, e.g. the Association for Financial Professionals (AFP, 2003).¹ A similar case had been made earlier by Ellis (1998) and Baker and Mansi (2001). In the survey conducted by Ellis, 70% of investors believed that ratings should reflect recent changes in default risk, even if they are likely to be reversed within a year.

At the same time, investors want to keep their portfolio rebalancing as low as possible and desire some level of rating stability. They do not want ratings to be changed to reflect small changes in financial condition. This is the argument put forward by rating agencies. Standard & Poor's (2003) is convinced that stable ratings are of most value to investors and Moody's observe a fervent support for the current level of rating stability among investors. From a regulatory perspective, rating stability is desirable to prevent procyclicality effects. A prompt and full response to changes in current creditworthiness could deepen a financial crisis. Linkages of portfolio strategies and portfolio mandates with NRSRO ratings and, in the future, linkages of bank capital requirements with NRSRO ratings, can force banks and investors to liquidate their positions hurriedly as ratings decline, which might ultimately result in a credit crunch. A third argument for rating stability is to maintain the reputation of agencies. Rating reversals within a short period have a negative impact on an agency's reputation, even when they reflect true changes in creditworthiness. In a sense, it is better to be late and right, than fast and wrong. A strong reputation, which underlies the recognition of ratings in financial markets, is in the interest of agencies, regulatory authorities, investors and bond issuing companies.

Apparently, investors want both stable and timely ratings, which are likely to be two conflicting objectives. Moody's tries to find a compromise: "*Moody's analysts attempt to balance the market's need for timely updates on issuer risk profiles, with its conflicting expectation for stable ratings*" (Cantor, 2001). In response to criticism on rating timeliness, in January 2002 Moody's considered changing ratings more aggressively and updating them more frequently. However, the Moody's then renounced their intention after broad consultation with investors, companies and financial authorities. In their meetings, Moody's repeatedly heard that investors value the current level of rating stability and do not want ratings simply to follow market prices. Moody's therefore decided not to change their rating policy, and they continue to produce stable ratings (see Fons et al., 2002). Curiously, the APF-survey, conducted in the same year, reveals the opposite conclusion by investors. It is a dilemma for investors and prompts our continued interest, see Altman and Rijken (2004).

In the discussion on rating timeliness a rigorous discussion of the pros and cons of rating stability in quantitative terms is lacking as yet. How stable are ratings? How do rating agencies achieve rating stability? What are the costs of rating stability in terms of rating timeliness and default prediction performance? In this article we attempt to provide answers to these questions.

Chapter 2 Through-the-cycle methodology

A widely accepted explanation for the sometimes inadequate timeliness of rating changes is the through-the-cycle methodology that agencies apply in their rating assessment. This methodology has two aspects: first, a focus on the permanent component of default risk, and second, a prudent migration (rating change) policy.

Based on the first aspect of the through-the-cycle rating methodology, agency ratings² disregard short-term fluctuations in default risk. By filtering out the temporary component of default risk, they measure only the permanent, long-term and structural component. According to Cantor and Mann (2003), the through-the-cycle methodology aims to avoid excessive rating reversals, while holding the timeliness of agency 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) states 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 component of default risk lead to rating migrations and, if triggered, ratings are partially adjusted to the actual level in the permanent component of default risk. Although not officially disclosed by agencies, practical evidence of such a prudent migration policy exists. Moody's provided some insight into their migration policy in their announcement of January 2002, stating that it was to be reconsidered: *"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

In contrast the through-the-cycle methodology bankers have a point-in-time perspective on corporate credit quality with a time horizon in between one and seven years (see Basle

Committee, 2000). It is reasonable to assume that this perspective applies to other investors as well. The point-in-time perspective looks at the current default risk of a counterparty without attempting to suppress the temporary component of default risk. It weights both the temporary and permanent component of credit quality. The relative weight between these two components depends on the time horizon. For a one year horizon the temporary component weights more heavy than for a longer time horizon.

Precisely how rating agencies put into practice their through-the-cycle methodology 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 quality cycle. Löffler (2004) explores the through-the-cycle effects on rating stability and default prediction performance quantitatively, by modeling the separation of permanent and temporary components of default risk in a Kalman filter approach. We take a different approach to understand the impact of the through-the-cycle methodology, by benchmarking the agency rating dynamics with credit scoring models, which serve as proxies for the point-in-time investor's perspective. In order to connect as close as possible with the investor's perception of rating timeliness, it is important to formulate credible and accurate proxies (benchmarks) for the investor's perspective on credit quality – with no desire for rating stability. For this purpose, credit scoring models for default prediction with various time horizons are estimated.

An earlier article, Altman and Rijken (2004), focuses mainly on the *modeling* of the through-the-cycle methodology, especially the prudent migration policy. In this article we emphasize the *quantitative consequences* of the through-the-cycle methodology on rating stability, rating timeliness and default prediction performance. This study examines corporate-issuer credit ratings, which are relative measures of default probability. Only corporate bond ratings are

investigated in this study; the additional Rating Outlook and Watchlist information is not taken into account.

The following chapter discusses the definition of benchmark credit scoring models. Chapter 4 presents the benchmark research setup. Chapter 5 and chapter 6 reports the results on rating stability and rating timeliness. Chapter 7 considers benchmark results on default prediction performance. Chapter 8 gives a summary of the main results and draws final conclusions.

Chapter 3 Benchmark credit scoring models

3.1 Default prediction models

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

$$CS_{i,t} = \alpha + \beta_1 \frac{WK_{i,t}}{TA_{i,t}} + \beta_2 \ln\left(1 - \frac{RE_{i,t}}{TA_{i,t}}\right) + \beta_3 \left(1 - \frac{EBIT_{i,t}}{TA_{i,t}}\right) + \beta_4 \left(1 + \ln \frac{ME_{i,t}}{BL_{i,t}}\right) + \beta_5 Size_{i,t} + \beta_6 Age_{i,t} + \varepsilon_{i,t} \quad (3.1)$$

$$\log\left(\frac{p_{i,t}}{1 - p_{i,t}}\right) = CS_{i,t} \quad (3.2)$$

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: $\ln(BL/Mkt)$. Age is the number of years since a firm was first rated by an agency.⁴

$p_{i,t} = 0$ if firm i defaults before $t + T$ and $p_{i,t} = 1$ if firm i survives beyond $t + T$. Default prediction models are estimated for various time horizons T .

In addition, *marginal* default prediction models are estimated. These models focus exclusively on default probability in a specific future period, i.e. the permanent component of default risk, and the binary variable $p_{i,t}$ is set to 0 only for firms defaulting in this future period ($t + T_1, t + T_2$), with $0 < T_1 < T_2$. Default events in the near future ($t, t + T_1$) are ignored by setting $p_{i,t} = 1$ for firms surviving and defaulting in this period. An alternative, to leave out the observations of firms defaulting in the near future, does not change the estimate significantly, since the number of defaults is relative small.

3.2 Agency rating prediction model

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

$$AR_{i,t} = \alpha + \beta_1 \frac{WK_{i,t}}{TA_{i,t}} + \beta_2 \ln\left(1 - \frac{RE_{i,t}}{TA_{i,t}}\right) + \beta_3 \left(1 - \frac{EBIT_{i,t}}{TA_{i,t}}\right) + \beta_4 \left(1 + \ln \frac{ME_{i,t}}{BL_{i,t}}\right) + \beta_5 Size_{i,t} + \beta_6 Age_{i,t} + \varepsilon_{i,t} \quad (3.3)$$

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

$$y_{i,t} = R \quad \text{if} \quad B_{R-1} < AR_{i,t} \leq B_R \quad (3.4)$$

where k is one of the agency rating classes,⁵ $y_{i,t}$ is the actual agency rating, B_R is the upper boundary for the AR score in rating class R , $B_0 = -\infty$ and $B_{16} = \infty$. In the ordered logit model, the probability that $y_{i,t}$ equals k is specified by

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

where F is the cumulative logistic function. The parameters α , β , and B_k are estimated with a maximum likelihood procedure.

3.3 Estimation benchmark credit scoring models

Data on agency ratings is obtained from the July 2002 version of the Standard & Poor's CREDITPRO database, which includes all S&P corporate credit ratings in the period January 1981-July 2002. At the end of each calendar quarter – March, June, September and December – corporate ratings are linked to stock price data and accounting data. Accounting data is assumed to be widely publicly available three months after the end of the fiscal year. The resulting panel dataset covers the period 1981-2001 and includes time series of 1629 obligors with period lengths in between 1 and on average 27.0 quarters. In addition, the dataset contains 9253 firm-quarter observations of firms with non-rated S&P status. These observations maximize the number of default events in the default prediction model estimation.⁶

Strictly speaking, all empirical results presented here refer to ratings by Standard and Poor's. However, we are not aware of any reason why empirical results and conclusions presented here for Standard and Poor's ratings should not apply for ratings by Moody's and Fitch. The discussions and conclusions in this article are therefore generalized to ratings by Standard and Poor's and Moody's and Fitch.

Table I reports the estimated parameters of three default prediction models, the agency rating prediction model and two bankruptcy prediction models. Three default prediction models are estimated: a short-term default prediction model with a time horizon of one year (SDP model), a long-term default prediction model with a time horizon of six years (LDP model) and a marginal default prediction model for a one year period, starting 5 years in the future (MDP model). The estimation period is 1981 - 1995. In this period each firm-year observation has a horizon up to six years, which allows a fair comparison of the parameters between the three default prediction models (see next section). The time horizon of six years in the LDP model and MDP model is a result of a compromise. A six year horizon is just beyond the length of the temporary credit

quality cycles of about 4 years (see figure II) while the length of the estimation period, 14 years, is kept at an acceptable level. Robustness checks for time horizons between 5 and 8 years show no substantial changes in the parameters of the MDP model.

Two bankruptcy prediction models are estimated to ensure that the parameters of the default prediction models are not uniquely related either to the particular S&P corporate bond dataset or the Standard and Poor's definition of default. A new dataset is constructed including all bankruptcies reported by COMPUSTAT.⁷ The estimation methodology of the bankruptcy prediction models is identical to the estimation of the default prediction models, as described above, apart from the omission of the Age variable. Bankruptcy prediction models are estimated for two periods: 1970-1995 (BP model) and 1970-1980 (BPO model). The time horizon T is one year. In the remainder of the article, the BPO model is considered to be an *out-of-sample* model when testing the default prediction in the period 1981-2001.

The parameters of the credit scoring models are robust over time in the period 1981-1999. No substantial differences are observed in parameter estimates between two sub-periods, 1981-1990 and 1991-1999. The period 2000-2001 is an exception. Most notable is the absence of the too-big-to-fail default protection in this period. When controlling for industry sector differences, the model parameters vary only slightly.⁸ A specific test for the AR model shows the robustness of the estimated parameters to a split of observations into non-investment graded (BB+ and below) firms and investment graded (BBB- and above) firms.⁹ The AR-parameters do not vary substantially with agency rating level, which enables to model the entire agency rating scale with one single parameter set. These robustness tests demonstrate the universal character of the credit scoring models that makes them a suitable benchmark for agency ratings.

3.4 Comparison credit scoring models

All credit scoring models employ the same model variables, apart from the BPO model which leaves out the Age variable. This allows a fair comparison of the relative weights RW_k of the model variables k

$$RW_k = \frac{|\beta_k| \sigma_k}{\sum_{j=1}^6 |\beta_j| \sigma_j} \quad (3.6)$$

β_k is the parameter estimate for model variable k , and σ_k is the standard deviation in the pooled sample distribution of model variable k in period 1981 – 1995. Table I shows the RW values for the estimated credit scoring models.

The ME/BL variable dominates in the SDP model with a RW value of 41.7 %. This is consistent with Moody's KMV structural model, in which market equity and total liabilities play a key role as well. Although the ME/BL variable is most important, accounting information – particularly the obligor characteristics Size and Age – add substantially to the explanation of the default incidence. The WK/TA, RE/TA, and EBIT/TA variables play a minor role. The relative weight of the model variables appear to be robust to dataset choice, estimation period and default event definition. The weight of the model variables in the BP model, BPO model and SDP model, all with a one-year time horizon, are similar.

Time horizon has a significant impact on the relative weight of the model variables. Especially for the RE/TA, ME/BL, and Size variable, a clear shift in relative weight is observed in the SDP, LDP, MDP and AR model, in that order of sequence. Not surprisingly, the short-term oriented SDP model depends heavily on variables which follow most closely the credit / business cycle, like ME/BL, while the AR model and MDP model place relatively more weight on variables which are less sensitive to credit cycles, like RE/TA and Size. The relative weight of the model

variables in the AR model matches most closely the MDP model, which suggests that agency ratings only weight the long-term permanent component of default risk. This finding is consistent with the aim of rating agencies to filter out the temporary component of default risk. In contrast, the LDP model and SDP model weights *both* temporary and permanent component of default risk.

Chapter 4 Benchmark setup

4.1 Conversion of credit scores to ratings

SDP scores, LDP scores, AR scores and BPO scores (CM scores) are converted to credit score ratings (CM ratings), equivalent to agency ratings. This enables to compare the dynamics of agency ratings unambiguously with the dynamics of credit scores. At the end of each quarter all companies are ranked by their credit score. On the basis of this ranking, sixteen credit score ratings, AAA, AA, ..., B, CCC/CC, equivalent to agency ratings, are assigned to individual companies. So at the end of each quarter the number of firms in each agency rating class R equals the number of firms in the equivalent CM rating class R.

4.2 Influence of migration policy on rating dynamics

SDP, LDP, AR and BPO ratings are point-in-time ratings, in the sense that they reflect the most recently available credit quality information, without any delay imposed by a migration policy. According to this definition, AR ratings are in fact point-in-time measures of the long-term default risk view of agencies. AR ratings represent only one aspect of the through-the-cycle methodology: a focus on the permanent component of default risk after filtering out the cyclical component. The dynamic influence of the prudent migration policy, the second aspect of the through-the-cycle methodology, is not picked up by the static ordered logit-regression

methodology.¹⁰ In order to study the influence of the prudent migration policy on rating dynamics, the AR score is adjusted, following a particular migration policy model.

We model the migration policy of agencies by two parameters: a threshold parameter and an adjustment parameter. The threshold parameter TH specifies the size of a credit quality interval $[-TH, +TH]$, in which credit quality is allowed to fluctuate without triggering a rating migration.¹¹ This threshold prevents small credit quality fluctuations from triggering a rating migration thereby reducing the rating migration probability. If a rating migration is triggered, the ratings are not fully adjusted to the actual credit quality level. The adjustment fraction AF specifies the partial adjustment of agency ratings. Partial adjustment of agency ratings (i.e. the spreading of the target rating adjustment over time) is responsible for the observed drift in agency ratings.

After adjusting the AR score in line with a particular migration policy, the resulting adjusted AR scores are converted to adjusted AR ratings. By varying the threshold parameter TH and the adjustment fraction AF parameter at the upside and the downside, migration probabilities and rating drift properties of adjusted AR ratings are varied. A best match with the dynamics of agency ratings is found for a threshold of 1.8 notch steps and an adjustment fraction of 0.6 at the upside and one of 0.7 at the downside. Apparently, the agencies migration policy is slightly more conservative on the upside. It is beyond the scope of this article to describe the details of this simulation experiment; for more information we refer to Altman and Rijken (2004).

The adjusted AR ratings that best match the dynamic properties of agency ratings are labeled as ARS ratings in the remainder of this article. These ratings represent both aspects of the through-the-cycle methodology. The (static) level of ARS ratings reflects *on average* the long-term perspective of agencies on default risk. The dynamics of the ARS ratings are influenced by both aspects of the through-the-cycle methodology.

4.3 Benchmark setup

Figure I shows the benchmark setup. Stability, timeliness and default prediction performance of agency ratings, point-in-time CM ratings and ARS ratings are compared.

- Differences between SDP ratings and LDP ratings illustrate the effect of the investor's horizon on the dynamics of point-in-time credit quality.
- Differences between LDP ratings and AR ratings shed more light on the interpretation of the long-term perspective of agency ratings.
- Differences between AR ratings and ARS ratings quantify the influence of the prudent migration policy on rating dynamics.
- Differences between ARS ratings and agency ratings are an indication of the information gap between agency ratings and credit scores. Dynamics of simulated ARS ratings and agency ratings are identical as both ratings are a result of the same through-the-cycle methodology. After controlling for the through-the-cycle methodology, the only difference is the accurateness and timeliness of credit quality information underlying the ratings. Agencies rate firms on the basis of in-depth analysis of private and public information, while credit scores are based on six model variables, available to the public.
- Differences between SDP ratings and BPO ratings illustrate whether, in this study, out-of-sample effects are relevant in measuring default prediction performance.

Chapter 5 Rating stability and rating drift

5.1 Unconditional rating migration probability

Rating stability is quantified by migration probabilities. Table II (panel A) reports the migration probability in a quarterly period for agency ratings and CM ratings. For agency ratings and ARS

ratings the migration probability is 5.6% and 5.4% respectively. Elimination of the prudent migration policy (→ AR ratings) increases the migration probability to 27%. Reducing the time horizon to one year and giving full weight to short-term default risk fluctuations (→ SDP ratings) increases the migration probability by another 12.6%. Similar results are achieved when migration policy elimination and time horizon reduction are carried out in reverse order: first a shift to a short-term focus (→ "adjusted" SDP ratings, similar to the way AR ratings are adjusted to ARS ratings) and subsequently the elimination of the prudent migration policy (→ SDP ratings). The prudent migration policy has more impact on migration probability than the disregard of the temporary component of default risk, the other aspect of the through-the-cycle methodology.

The rating migration probability of LDP-ratings is in between the rating migration probability of AR ratings and SDP ratings. This finding suggests that agencies put heavy weight on the permanent component of default risk compared to the LDP model, which only moderately weights the temporary component. Putting emphasis on the temporary component, like SDP ratings do, results in higher rating migration probabilities. Like the results in section 3.4, this empirical result is consistent with the exclusive focus of agencies on the permanent component of default risk and the disregard of credit quality cycles.

5.2 Mean rating migration figures

In order to calculate mean migration figures, a numerical scale is assigned to the ordinary notch scale of agency ratings and equivalent CM ratings: D = 0, CCC/CC/C = 1, B- = 2, B = 3, AA- = 14, AA = 15, and AA+/AAA = 16. 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.

Mean rating migration figures are computed for upgrades and downgrades in the period 1981-2001 (see Table II, panel B).¹² Without a prudent migration policy the average rating migration is just above one notch step, which is expected when necessary changes are made immediately and in full. The threshold TH of 1.8 notch steps, in combination with the moderating influence of the adjustment fraction AF, increases the average migration step to about 1.4 at the upside and 1.5 at the downside.

The unconditional mean rating migration $\Delta R(u)$ in each quarter is about -0.02 for agency ratings and -0.01 for CM ratings. Technically, this unconditional migration is equal to the difference in rating level between firms entering the dataset and firms exiting the dataset, divided by the number of quarters of unbroken stay in the dataset (= on average 27 quarters). Defaulting firms are mainly responsible for the unconditional downward drift in ratings.

5.3 Rating drift properties

Conditional on an upgrade, downgrade or no migration ΔR in quarterly period Q0, the mean rating migration figures $\Delta R(+)$, $\Delta R(-)$ and $\Delta R(0)$ are computed for subsequent quarters Q1, Q2 until Q8¹³. These conditional rating migration figures are corrected for the unconditional $\Delta R(u)$: $\Delta R(+)$ - $\Delta R(u) \rightarrow \Delta R(+)$ and $\Delta R(-)$ - $\Delta R(u) \rightarrow \Delta R(-)$. Only $\Delta R(+)$ and $\Delta R(-)$ are of interest, since $\Delta R(0) \approx \Delta R(u)$.

Table II reports $\Delta R(+)$ and $\Delta R(-)$ for Q1, the quarter immediately following Q0 (panel C), for Q2, the second quarter following Q0 (panel D) and for the seven quarter period Q2 until Q8 (panel E). For point-in-time CM ratings a short term reversal behavior shows up. The duration of this reversal effect is very short. It disappears in subsequent quarter Q2. In quarters Q2 to Q8 rating drift is absent, which suggests a random behavior of point-in-time corporate credit quality beyond

a quarterly cycle. The origin of this short-term reversal effect needs further study. A likely cause are the seasonal patterns in accounting figures.

Given the random behavior of the underlying credit risk fundamentals, rating drift is expected when ratings are partially adjusted to actual credit quality ($AF = 0.7/0.6$, see section 4.2).¹⁴ In the eight quarters after a downgrade or upgrade, agency ratings and ARS ratings drift with a steady rate up to about -0.30 at the downside and +0.30 at the upside. Drift at both sides is expected as the underlying source of rating drift is effective in both directions. In quarter Q1, shortly after the migration event in Q0, agency ratings and ARS ratings drift most strongly at the downside, perhaps because downturns in credit quality happen more rigorously than upturns.

Chapter 6 Rating timeliness

Conditional on an *agency* rating upgrade or downgrade, average changes in CM ratings surrounding this migration event are investigated. The magnitude of conditional changes in CM ratings before the agency migration event is an indication of the timeliness of agency ratings.

The conditional migration $\Delta R(+)$ and $\Delta R(-)$ figures are recomputed by following exactly the same procedure as described in previous section 5.3, except for one difference: all $\Delta R(+)$ and $\Delta R(-)$ figures are conditional on an *agency rating migration event* ΔN in Q0, instead of ΔR itself. The cumulative rating change ΔR^C_t since $t = -4.25$, conditional on an agency rating migration event in Q0, is given by

$$\Delta R^C(v)_t = \sum_{k=-4}^t \Delta R(v)_{k-0.25,k} \quad (6.1)$$

where $v = "+"$ or $"-"$.

Figure II shows the cumulative rating changes ΔR^C for agency ratings (ΔN^C), ARS ratings (ΔARS^C), AR ratings (ΔAR^C), LDP ratings (ΔLDP^C), and SDP ratings (ΔSDP^C). In order to compare these figures on a comparable scale, in terms of agency rating notch steps, the ΔR^C of all CM ratings are scaled by a factor of $1/\kappa_R$; κ_R equals the slope in the regression equation: $CM = \kappa_R N + \text{constant}$. For ARS ratings, AR ratings, LDP ratings, SDP ratings and BPO ratings, κ_R equals 0.856, 0.841, 0.807, 0.744 and 0.756, respectively. Because of the strong variation in credit scores within an agency rating class, with a "standard error" up to 3 notch steps, and the boundaries of the discrete rating scale, agency rating migrations do not always show up in CM rating changes, even if all agency rating migrations are correctly picked up by changes in credit scores.¹⁵

On average, all point-in-time CM rating changes clearly anticipate agency rating migrations, except for ARS ratings, as expected. Among CM ratings, SDP ratings anticipate an agency rating migration event most strongly. In the two-year period surrounding the agency rating migration event, ΔSDP^C and to a lesser extent ΔLDP^C show "overshooting" behavior (see Figure II). Just after the agency-migration date, ΔSDP^C clearly exceeds the change in the permanent component of default risk, as proxied by ΔAR^C and ΔN^C at $t = 4$. This overshooting behavior is due to the sensitivity of SDP ratings to changes in the temporary component of default risk. As expected these temporary changes in SDP ratings are reversed. The overshooting behavior is less pronounced for LDP ratings, as they are only moderately sensitive to the temporary component of default risk. The absence of overshooting behavior for AR ratings is consistent with the disregard of short-term fluctuations in default risk by agencies.

The quantification of rating timeliness is based on the cumulative rating changes ΔR^C starting at $t = -2.25$ and ending at $t = 2$. The choice of this time interval is of course arbitrary, but in this period most of the rating changes do take place. Longer time intervals do not change the results substantially, but are at a cost of statistical significance. Table III reports the total cumulative rating change ΔR^C_{TOT} in this time interval and the percentages of these cumulative rating change happening in the 2 years before Q0 (-2.25,-0.25), in Q0 (-0.25,0) and in the two years after Q0 (0,2). These figures are not scaled by the factor of $1/\kappa_R$. As expected, ARS ratings do not show a timing advantage. In this case an equal fraction of total cumulative rating change ΔR^C_{TOT} happens before and after the agency rating migration event. For point-in-time CM ratings the majority of the conditional rating changes (60-102%) happen before Q0. After Q0 the cumulative rating change ΔR^C_{TOT} are relative low, and are even negative for SDP^C and BPO^C, due to the "overshooting" behavior.

The timeliness of agency ratings relative to CM ratings is defined as follows. The maximum cumulative rating change in interval (-2,2) ΔR^C_{MAX} is a proxy for the total change in both the permanent and temporary component of default risk, conditional on an agency rating migration event. The timing when half of ΔR^C_{MAX} is reached, is a proxy for the *average* timing of these permanent and temporary changes in CM ratings, conditional on the agency rating migration event at $t = -0.125$. The difference between these two timing moments is an indication for the timeliness of agency ratings relative to CM ratings. Table III reports the results. The timeliness disadvantage of agency ratings compared to LDP ratings, SDP ratings and BPO ratings is about 0.75 years at the upside and 0.5 years at the downside. This is consistent with the evidence that agencies are more conservative at the upside (for example the lower threshold level in the migration policy at the upside, see section 4.2). After controlling for the through the cycle effects the timing of agency rating migrations and corresponding changes in ARS ratings, based on

public information, is negligible at the upside (-0.03 year), while agency ratings are slightly more responsive than ARS ratings at the downside (+0.18 year).

At the downside both aspects of the though-the-cycle methodology equally affect the timeliness. Due to insensitivity to the temporary component of credit quality, the AR ratings are delayed by about 0.4 years compared to LDP, SDP and BPO ratings, while ARS ratings are delayed by another 0.3 years, due to the prudent migration policy. At the upside, the prudent migration policy has the most impact. Perhaps changes in credit quality at the upside have a more permanent and less abrupt character, which makes the time horizon in default risk less relevant.

Chapter 7 Default prediction performance

7.1 Cumulative default rates

The probability for a firm, in a particular rating class R , to default *within* T years is measured from historical data by

$$\text{Cumulative default rate } (R, T) = \frac{\sum_{t=1981}^{2002-T} \sum_{i=1}^{N_{R,T,t}} D_{R,T,i,t}}{\sum_{t=1981}^{2002-T} \sum_{i=1}^{N_{C,T,t}} (S_{R,T,i,t} + D_{R,T,i,t})} \quad (7.1)$$

where $D_{R,T,i,t}$ and $S_{R,T,i,t}$ are binary variables identifying default observations ($p_{i,t} = 0$) and survival observations ($p_{i,t} = 1$) in rating class R with a time horizon T (see section 3.1). Firms are characterized as surviving firms if they survive beyond T years. They are characterized as defaulting firms if they default within T years. Firms exiting the dataset by other means than default within T years (for example mergers and migrations to a non-rated credit rating status) are excluded from the default rate calculation. $N_{R,T,t}$ is the total number of defaulting and survival observations in rating class R , in year t with a time horizon T . This cumulative default rate

definition for pooled samples is similar to the Static Pools Cumulative Averaged Default Rates as reported by Standard and Poor's.

Table IV shows the *three-year* cumulative default rates for all 16 classes of agency ratings and CM ratings. In general, point-in-time CM ratings (AR, LDP, SDP and BPO ratings) perform slightly better than agency ratings in the non-investment regime (below BB+). In this regime the default rates of CM ratings are higher in bottom rating classes CCC/CC and B- and lower in B+, BB-, BB rating classes. So the type I errors and type II errors are lower for CM ratings if firms in bottom rating classes are classified as defaulters and firms in higher rating classes are classified as non-defaulters. In the investment regime (above BB+) the number of type I errors is lower for agency ratings. On a three-year horizon the average investment-grade default rate is 0.31% for agency ratings and 0.41 - 0.58% for CM ratings.

7.2 Accuracy ratios

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 class R, the proportion of default observations in the *same and lower* rating class $F_D(R)$ (Y-axis), against the proportion of all survival and default observations in the same and lower rating class $F_A(R)$ (X-axis).

$$F_A(R, T) = \frac{\sum_{C=1}^R \sum_{t=1981}^{2002-T} \sum_{i=1}^{N_{C,T,t}} (S_{C,T,i,t} + D_{C,T,i,t})}{N_A(T)} \quad (7.2)$$

where $F_A(0, T) = 0$ and $N_A(T)$ is the total number of default and survival observations 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.

The higher the proportion of default events happening in the lower classes – in other words the higher the surface below the CAP curve – the better the performance of the rating scale. The accuracy ratio ACR measures the surface below the CAP curve, relative to the surface below the CAP curve for a random rating scale ($=1/2$). Based on cumulative default rates ACR is given by

$$ACR(T) = \frac{\sum_{R=1}^{16} \left(\frac{[F_A(R, T) - F_A(R-1, T)] \times [F_D(R-1, T) + \frac{1}{2}(F_D(R, T) - F_D(R-1, T))]}{1/2} \right)}{1/2} \quad (7.3)$$

ACR varies between 0% (random scale) and 100% (perfect prediction scale). Table IV shows the ACR of agency ratings, ARS ratings and point-in-time CM ratings¹⁶. The standard error in ACR is 1.5%, 2% and 2.5% for respectively time horizons T of one year, 3 years and 6 years¹⁷.

In general, point-in-time CM ratings perform better than agency ratings up to a time horizon of 3-4 years. On a three-year horizon the accuracy ratios ACR(3) do not vary much among agency ratings and CM ratings. Based on a *pooled* sample (equations 7.2 and 7.3) the ACR is 73.0% for agency ratings and between 71.3% and 74.3% for CM ratings. Beyond a time horizon T of 3-4 years agency ratings show a slightly better default prediction performance. Watchlist and Outlook information has not been included in the analysis. Interesting follow-up research would be to see whether Outlook and Watchlist information on agency ratings is as timely as credit scores. This would make the analysis complete. First tests have been carried out by Cantor and Hamilton (2004). They show that the ACR for agency ratings improves significantly when Outlook and Watchlist information is added.

Differences in ACR between agency ratings and various CM ratings, $\Delta ACR = ACR(\text{CM ratings}) - ACR(\text{agency ratings})$ enables to reveal the impact of the through-the-cycle methodology and the information advantage of agencies on default prediction performance. Figure III shows ΔACR as

a function of time horizon T. The standard error in Δ ACR are 0.75%, 1.0 and 1.25% for respectively time horizons T of one year, 3 years and 6 years ¹⁸.

On a one year horizon, the ACR of agency ratings is 6% lower than the ACR of LDP ratings, which weights both the temporary and permanent component of default risk (see Figure III). Filtering out the temporary component of credit quality measurement reduces ACR by 4%, as proxied by the difference in ACR between LDP ratings and AR ratings. The prudent migration policy reduces ACR by another 4%, as proxied by the difference in ACR between AR ratings and ARS ratings. The information advantage of agency ratings improves the ACR of agency ratings by 2% relative to the point-in-time CM ratings, as proxied by the difference in ACR between ARS ratings and agency ratings. So main conclusion is that the negative impact of the through-the-cycle methodology (8%) fully overshadows the information advantage of agency ratings (2%), resulting in a ACR reduction of 6%. In a comparable analysis Fons and Viswanathan (2004) found a 1.7% advantage in ACR for ratings based on solely accounting data compared to the ACR of actual Moody's ratings (analysis period is 1999 - 2003).

As expected, for a six-year horizon negative impact of the through-the-cycle methodology is less severe. For longer horizons the weighting the temporary component of default risk has almost no impact on the ACR. The prudent migration policy lowers the ACR by only 2%. In this case the information advantage of 3.5% is only partly masked by the through-the-cycle methodology, resulting in an ACR advantage of about 1% for agency ratings.

Superior performance is expected when CM ratings, which have superior timeliness properties, are combined with agency ratings, which have superior accuracy in default risk assessments. To prove this, a combined rating scale (COMBI rating) is constructed. Firms are ranked on the basis of the mean value of agency ratings and BPO ratings and, subsequently, by fine-tuning on the

basis of BPO scores. After this ranking procedure, COMBI ratings – equivalent to agency ratings – are assigned to firms each quarter by following the same procedure as outlined in section 4.1. The superior performance of COMBI ratings, as shown in Figure III, proves that credit-model scores are complementary to agency ratings. This result is consistent with the common practice of using credit scores and EDF scores in addition to agency ratings, not only as a second opinion, but also because of the superior timeliness of credit scores.

The estimation period of credit scoring models largely overlaps the default performance analysis period, except for the "out-of-sample" BPO model. This could raise questions as to whether measuring the default prediction performance of CM ratings "in-sample" is appropriate. However, the underlying credit scoring models are robust to dataset choice, default definition and robust in time (see section 3.3), so a distinction between "in-sample" and "out-of sample" is not very relevant. For example, the ACR(T) of SDP ratings and BPO ratings, both proxies for the one-year default probability, practically overlap each other (see Figure III).

Chapter 8 Summary and conclusions

From the benchmark results we confirm the exclusive focus of agency ratings on the permanent component of credit quality and disregard of the temporary component. In addition we modeled the prudent migration policy. A rating migration is triggered if the actual credit quality – permanent credit quality component - exceeds a threshold of 1.8 notch steps relative to the average credit quality level in a rating class. If triggered, ratings are partially adjusted to the actual credit quality level, 60% at the upside and 70% at the downside.

This rating through the cycle lowers the rating migration probability in a quarterly period by a factor of 7, from an investor's one-year point-in-time perspective, weighting both permanent and

temporary component of default risk (as proxied by LDP ratings). At the same time it delays the timing of rating migrations by 0.5 years at the downside and 0.75 years at the upside. The agencies turn out to be slightly more responsive at the downside than at the upside.

Rating stability affects default prediction performance, especially in the short term. The accuracy ratio in predicting one-year defaults drops by 8%. The disregard of the temporary component of default risk and the prudent migration policy – both aspects of the through-the-cycle methodology – have an equal share in this reduction. Controlling for the through-the-cycle effects shows that agencies have an information advantage over credit scoring models. On an accuracy ratio scale the information gap is 3%. This information advantage is fully overshadowed by the negative impact of the through-the-cycle methodology up to time horizons of 3-4 years. A follow-up research question is whether the information advantage of agencies reappears when Rating Outlook and Watchlist information is taken into account.

Main purpose of this study is to quantify the effects of the through-the-cycle methodology on rating stability, rating timeliness and default prediction performance. It is up to investors and authorities to judge whether this balance between rating stability and timeliness (i.e. default prediction performance) matches best the investor's interest.

Literature

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

APF, 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, 2004, "Rating Transitions and Defaults Conditional on Outlooks, Watchlists, and Rating History", *Special Comment*, Moody's Investor Services, February
- Cantor R. and C. Mann, 2003, "Are Corporate Bond Ratings Procyclical?", *Special Comment*, Moody's Investor Services, October
- Cantor R., 2001, "Moody's investor service response to the consultative article issued by the Basel Committee on Bank Supervision "A new capital adequacy framework", *Journal of Banking & Finance* 25, 171 - 185
- 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
- Fons J.S., and J. Viswanathan, 2004, "A user's guide to Moody's Default predictor model: an accounting ration approach", *Special Comment*, Moody's Investor Services, December
- Fons J.S., R. Cantor and C. Mahoney, 2002, "Understanding Moody's corporate bond ratings and rating process", *Special Comment*, Moody's Investor Services, May
- Moody's Investor Service, 2002 "Understanding Moody's corporate bond ratings and rating process", *Special Comment*, May
- Löffler G., 2004, "An anatomy of rating through-the-cycle", *Journal of Banking and Finance* 28, 695 - 720
- Standard & Poor's, 2003, "Corporate ratings criteria", www.standardandpoors.com
- Treacy W.F. and M. Carey, 2000, "Credit rating systems at large US banks", *Journal of Banking & Finance* 24, 167 - 201

Figure 1 The benchmark study setup

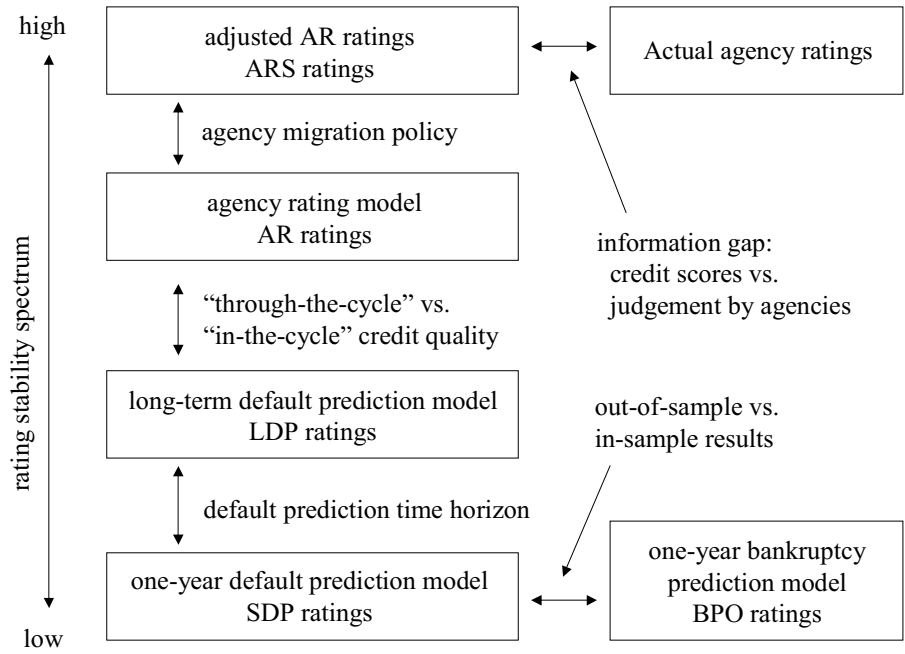


Table I Estimation default prediction models and agency rating prediction model

The table presents parameter estimates of (1) three default prediction models for various prediction time horizons: a one-year period, a six-year period and an annual period starting five years in the future, (2) the agency rating prediction model and (3) two bankruptcy prediction models for two estimation periods, 1970-1980 and 1970-1995. The 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 firm¹⁹. Z-statistics are given in brackets. Pseudo R² is a measure for the goodness of the fit. The last rows of the table give the relative weight (see equation 3.6) of the parameters in a particular model.

prediction time horizon estimation period	default prediction models			AR model	bankruptcy prediction models	
	t ∈ (1,4)	t ∈ (1,24)	t ∈ (21,24)	not applicable	t ∈ (1,4)	
	'81-'95	'81-'95	'81-'95	'81-'95	'70-'95	'70-'80
	SDP	LDP	MDP	AR	BP	BPO
regression results						
constant	8.12 (7.83)	5.44 (6.81)	7.00 (8.19)	-5.97 * (0.61)	7.61 (31.18)	10.15 (14.30)
WK/TA	1.09 (1.54)	0.19 (0.36)	-0.30 (0.56)	-2.25 (6.32)	0.60 (3.51)	0.07 (0.15)
RE/TA	0.05 (0.09)	1.02 (2.92)	1.07 (3.42)	3.59 (3.59)	0.09 (1.05)	0.49 (1.78)
EBIT/TA	5.39 (3.51)	2.81 (2.41)	-1.31 (1.12)	4.87 (7.88)	2.83 (10.75)	3.55 (4.29)
ME/BL	1.44 (9.86)	0.96 (9.51)	0.42 (3.85)	0.97 (14.38)	0.88 (23.28)	0.92 (8.95)
Size	0.53 (5.10)	0.50 (6.08)	0.35 (4.19)	0.91 (13.36)	0.29 (13.29)	0.50 (7.65)
Age	0.16 (4.14)	0.13 (4.69)	0.05 (1.80)	0.10 (6.92)	-	-
pseudo R ²	0.381	0.288	0.081	0.217	0.195	0.162
# observations	31829	24656	24656	28333	111510	33242
# default obs.	278	1677	343	-	720	119
relative weight model variables						
WK/TA	5.8%	1.2%	3.3%	8.7%	5.7%	0.5%
RE/TA	0.5%	12.0%	21.3%	25.0%	2.0%	6.7%
EBIT/TA	12.5%	8.1%	6.3%	8.2%	18.8%	15.1%
ME/BL	40.7%	33.5%	24.6%	20.1%	47.7%	42.4%
Size	24.0%	28.2%	33.4%	30.0%	25.7%	35.3%
Age	16.5%	17.1%	11.2%	7.9%	-	-

* Due to space considerations only the estimated boundary between the rating category BB+ and BBB- is shown. In this particular case the standard error of the cutoff is given in the brackets, not the z-statistics.

Table II Rating stability and rating drift properties

Panel A presents the rating migration probability in a quarterly period for three migration events ME: an upgrade, a downgrade and no migration.¹² Panel B presents average rating migration figures for upgrade and downgrade events. Panel C, D and E present rating drift properties. Panel C shows the average rating migration in quarter Q1, conditional on an upgrade or downgrade in previous quarter Q0. Panels D and E show these statistics for respectively Q2 and the seven quarter period Q2 to Q8. These rating drift figures are corrected for unconditional rating migration figures. Standard errors are given in brackets.

unconditional results						
migration event ME	agency ratings	ARS ratings	AR ratings	LDP ratings	SDP ratings	BPO ratings
A rating migration event ME in a quarterly period						
upgrade	2.3%	2.2%	13.1%	16.3%	19.2%	16.5%
no migration	94.4%	94.6%	73.0%	67.0%	60.4%	62.7%
downgrade	3.3%	3.2%	13.9%	16.7%	20.4%	20.8%
B average rating migration for a rating migration event ME						
upgrade	1.36 (0.03)	1.44 (0.02)	1.06 (0.00)	1.09 (0.00)	1.17 (0.01)	1.14 (0.01)
downgrade	-1.56 (0.02)	-1.51 (0.02)	-1.14 (0.01)	-1.17 (0.01)	-1.26 (0.01)	-1.24 (0.01)
mean migration figures conditional on a rating migration event ME in Q0						
migration event ME	agency ratings	ARS ratings	AR ratings	LDP ratings	SDP ratings	BPO ratings
C average rating migration in Q1						
upgrade	0.03 (0.01)	0.02 (0.01)	-0.22 (0.01)	-0.19 (0.01)	-0.18 (0.01)	-0.20 (0.01)
downgrade	-0.08 (0.01)	-0.10 (0.01)	0.18 (0.01)	0.17 (0.01)	0.12 (0.01)	0.12 (0.01)
D average rating migration in Q2						
upgrade	0.04 (0.01)	0.08 (0.01)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)
downgrade	-0.04 (0.02)	-0.10 (0.01)	0.00 (0.01)	-0.02 (0.01)	-0.04 (0.01)	-0.05 (0.01)
E average rating migration in period Q2-Q8						
upgrade	0.27 (0.04)	0.29 (0.05)	0.06 (0.02)	0.05 (0.03)	-0.04 (0.03)	-0.03 (0.03)
downgrade	-0.30 (0.05)	-0.27 (0.04)	0.01 (0.02)	-0.02 (0.03)	-0.06 (0.03)	-0.02 (0.03)

Figure II Changes in ARS and DP ratings conditional on an agency rating migration event

The figure shows the cumulative rating migration ΔR^C , for agency ratings N, ARS ratings, AR ratings, LDP ratings and SDP ratings conditional on an agency rating migration event. Starting time for the accumulation of rating migrations is -4.25. The expressions $\Delta N^C(+)$ and $\Delta N^C(-)$ refer to the cumulative agency rating migration conditional on respectively an upgrade in period $(-0.25,0)$ and a downgrade in period $(-0.25,0)$. Comparable definitions hold for the CM ratings. For all ratings the cumulative rating migration $\Delta R^C(+)$ and $\Delta R^C(-)$ are corrected for unconditional rating migration figures and scaled by a factor $1/\kappa_R$ (see text).

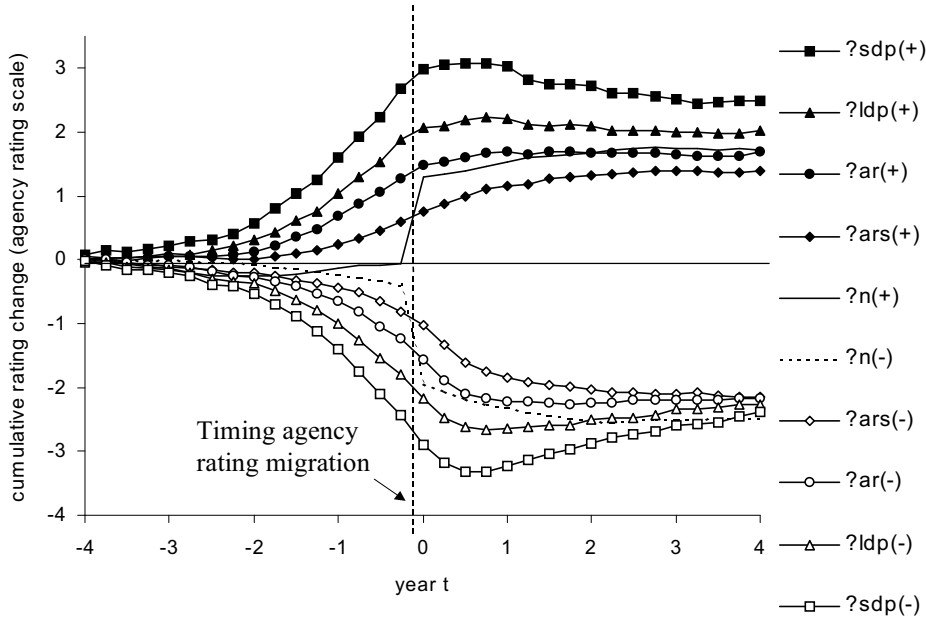


Table III Rating timeliness

Proxies of agency rating timeliness are based on the cumulative rating migration ΔR^C of CM ratings, conditional on an agency rating upgrade or downgrade in $(-0.25,0)$. Two ΔR^C -figures are computed: the total ΔR^C over a period starting at $t = -2.25$ and ending at $t = 2$ (ΔR^C_{TOT}) and the maximum ΔR^C in this period (ΔR^C_{MAX}). Two proxies for rating timeliness are the percentage of ΔR^C_{TOT} happening before $t = -0.25$ and the timing when half of ΔR^C_{MAX} is reached. The timeliness of CM ratings relative to agency ratings N is the difference in timing when half of ΔR^C_{MAX} is reached.

	agency ratings	CM ratings				
		ARS ratings	AR ratings	LDP ratings	SDP ratings	BPO ratings
<i>Conditional on an agency rating upgrade</i>						
ΔR^C_{TOT}	1.93 (0.07)	1.29 (0.07)	1.60 (0.11)	1.87 (0.13)	2.33 (0.17)	2.18 (0.16)
% $\Delta R^C_{TOT} (-2,-0.25)^*$	10%	46%	77%	89%	98%	102%
% $\Delta R^C_{TOT} (-0.25,0)^*$	74%	15%	14%	11%	14%	16%
% $\Delta R^C_{TOT} (0,2)^*$	16%	40%	9%	0%	-12%	-18%
ΔR^C_{MAX}	1.93 (0.07)	1.29 (0.07)	1.61 (0.10)	2.02 (0.11)	2.68 (0.14)	2.44 (0.13)
timing $\frac{1}{2} \Delta R^C_{MAX}$	-0.11 (0.01)	-0.14 (0.08)	-0.72 (0.09)	-0.82 (0.07)	-0.90 (0.08)	-0.85 (0.08)
timeliness CM vs. N	-	-0.03	-0.61	-0.71	-0.79	-0.74
<i>Conditional on an agency rating downgrade</i>						
ΔR^C_{TOT}	-2.49 (0.07)	-1.85 (0.07)	-2.01 (0.10)	-2.17 (0.11)	-2.47 (0.15)	-2.44 (0.13)
% $\Delta R^C_{TOT} (-2,-0.25)^*$	16%	42%	60%	77%	92%	93%
% $\Delta R^C_{TOT} (-0.25,0)^*$	68%	15%	20%	20%	23%	22%
% $\Delta R^C_{TOT} (0,2)^*$	16%	43%	20%	3%	-15%	-15%
ΔR^C_{MAX}	-2.49 (0.07)	-1.85 (0.07)	-2.01 (0.09)	-2.33 (0.09)	-2.91 (0.11)	-2.80 (0.11)
timing $\frac{1}{2} \Delta R^C_{MAX}$	-0.11 (0.01)	0.07 (0.04)	-0.25 (0.05)	-0.53 (0.06)	-0.67 (0.06)	-0.64 (0.06)
timeliness CM vs. N	-	0.18	-0.14	-0.42	-0.56	-0.53

* the percentages of ΔR^C_{TOT} happening in the sub-periods $(-2,-0.25)$, $(-0.25,0)$ and $(0,2)$.

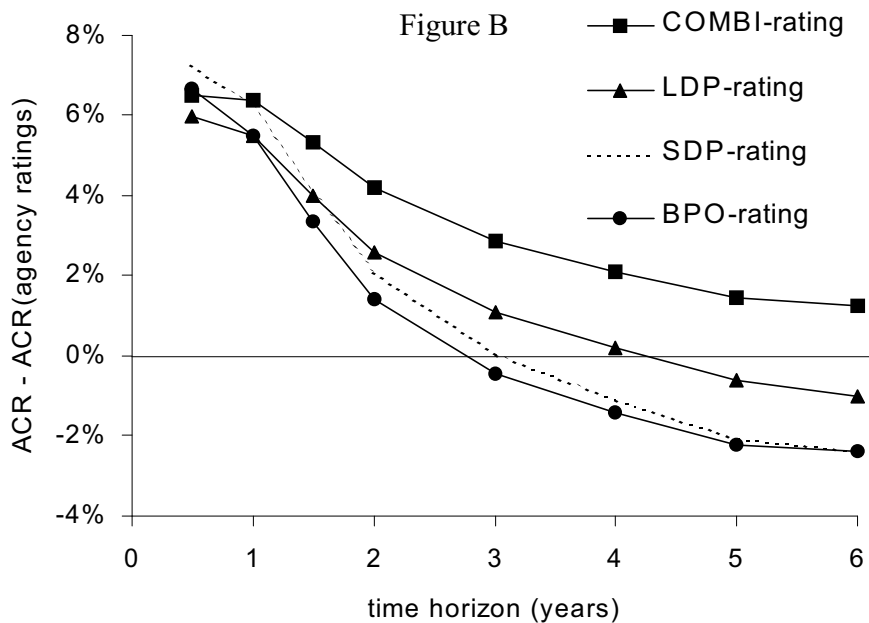
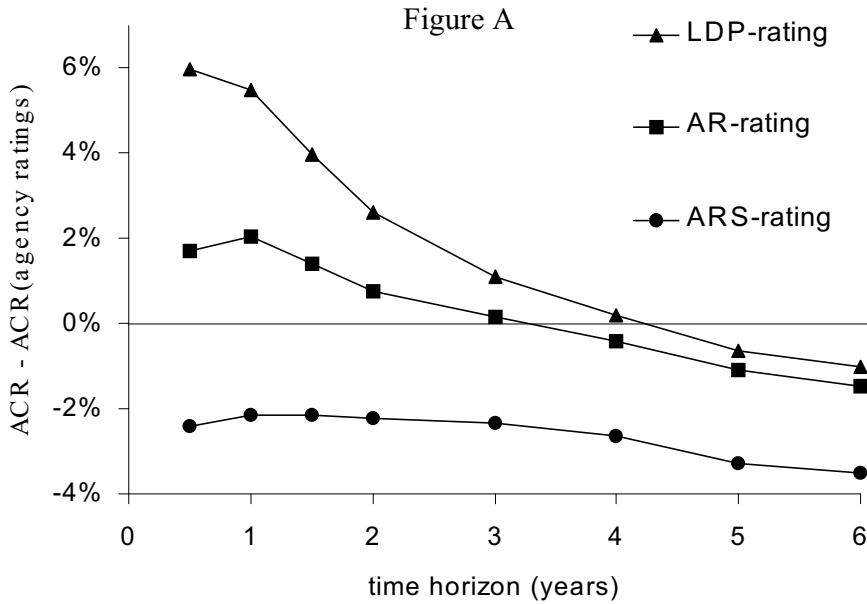
Table IV Cumulative default rates for a three year horizon

The table presents the probabilities of default within three years for each of the 16 agency rating classes CCC/CC, B-, ... AA, AA+/AAA, for each of the equivalent CM rating and COMBI rating classes. In the bottom three rows the default prediction performance of the different rating scales are compared by (1) the average default rate in the investment regime, which measures the type I error in this regime, by (2) the accuracy ratio ACR, which weights both type I errors and type II errors equally (= 0% for a random scale = 100% for a perfect scale) and by (3) the mean rating of firms defaulting within three years. Standard errors are given in brackets.

Rating class (or equivalent)	Agency ratings	"in-sample" results				"out-of-sample" results		# obser- vations
		ARS ratings	AR ratings	LDP ratings	SDP ratings	BPO ratings	COMBI ratings	
1 CCC/CC	53.7%	56.1%	61.6%	71.9%	69.1%	73.5%	70.1%	708
2 B -	39.3%	37.0%	41.6%	41.0%	41.6%	40.4%	43.0%	1098
3 B	32.1%	26.7%	31.1%	30.9%	31.3%	32.3%	35.1%	2192
4 B +	17.1%	15.9%	16.4%	15.8%	15.3%	13.8%	14.8%	5456
5 BB -	11.9%	8.8%	9.1%	7.9%	7.8%	7.9%	7.9%	4003
6 BB	5.3%	6.0%	4.8%	5.0%	4.7%	4.7%	5.0%	3163
7 BB +	2.7%	2.9%	3.6%	3.4%	2.8%	2.9%	2.7%	2103
8 BBB -	1.9%	1.9%	2.3%	2.0%	2.2%	2.7%	1.5%	2949
9 BBB	0.5%	0.9%	0.9%	1.6%	1.7%	1.9%	1.0%	3671
10 BBB +	1.2%	2.0%	1.5%	1.0%	1.4%	1.3%	0.9%	3126
11 A -	0.4%	0.8%	0.7%	0.4%	0.8%	1.1%	0.7%	2948
12 A	0.3%	0.5%	0.3%	0.4%	0.6%	0.7%	0.2%	4628
13 A +	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2746
14 AA -	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1550
15 AA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2351
16 AA+/AAA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1332
# default obs.	2214	2214	2214	2214	2214	2214	2214	
# default events	151	151	151	151	151	151	151	
default rate investment grade	0.31%	0.46%	0.42%	0.41%	0.51%	0.58%	0.32%	
ACR pooled sample	73.0% (2.0%)	71.3% (2.0%)	73.4% (2.0%)	74.3% (2.0%)	73.0% (2.0%)	72.6% (2.0%)	75.9% (2.0%)	
mean rating default obs.	3.72 (1.90)	3.73 (2.11)	3.67 (2.07)	3.58 (2.09)	3.65 (2.22)	3.68 (2.31)	3.46 (1.97)	

Figure III Default prediction performance of CM ratings relative to agency ratings

Figures A and B present differences in ACR between agency ratings and CM ratings, $ACR(\text{CM ratings}) - ACR(\text{agency ratings})$, as a function of time horizon T . The ACR values are calculated on the basis of cumulative default rates. Standard errors are approximated by 0.75% for a time horizon of 1 year up to 1.25% for a time horizon of 6 years.



Endnotes

- ¹ The critique on rating agencies is mainly focused on the timeliness properties of agency ratings and not on the accuracy level itself. The AFP survey reveals that 83% of the investors believe that most of the time agency ratings accurately reflect the issuer's creditworthiness.
- ² In the remainder of this article, *agency ratings* refer to the corporate issuer credit ratings of Standard and Poor's, Moody's, and Fitch.
- ³ see The Financial Times, 19 January 2002, "Moody's mulls changes to its ratings process".
- ⁴ The Age variable is set to 10 for observations with Age values above 10 and for all observations of firms already rated at the start of the dataset in 1981.
- ⁵ In order to have a reasonable number of observations in each rating class, the agency rating classes C, CC, CCC-, CCC and CCC+ are combined into a single rating class CCC/CC, and the agency rating classes AA+ and AAA are combined into a single rating class AA+/AAA. A numerical scale is assigned to the agency ratings from 1 to 16 referring to the rating classes CCC/CC up to AA+/AAA.
- ⁶ Firms with an NR status keep on being monitored for default events. If they default, the NR status changes to a D status.
- ⁷ The bankruptcy dataset covers the 1970-1995 period and contains 111,510 survival observations and 720 bankruptcy observations, which are defined in a similar manner as survival and default observations in the S&P corporate bond dataset. Only a small fraction of these bankruptcy observations overlap the default observations in the Standard and Poor's corporate bond dataset.
- ⁸ WK/TA is an exception.
- ⁹ The only significant difference is the absence of a significant parameter for the Age variable for non-investment grade firms.
- ¹⁰ Ratings predicted by the AR model are slightly overstated as a result of a prudent migration policy. This overstatement is explained as follows. Temporarily, ratings may in fact either be understated or overstated due to a prudent migration policy. If the number of overstated and understated ratings are equal over the sample period – neutralizing the variation in overstated and understated ratings due to the prudent migration policy and business cycles – the migration policy will not affect the parameter estimates. In that case it will only widen the distribution of the error term ϵ in the logit regression. However, the number of downgrades is 30% higher than the number of upgrades and the average agency rating migration shows a downward trend, so the number of overstated ratings is expected to be slightly higher. As a consequence the ratings predicted by the AR score are expected to be slightly overstated due to the prudent migration policy. However when AR scores are converted to AR ratings, the shift in AR scores due to overstatement is not relevant; only relative AR scores matter. Consequently the AR rating dynamics are insensitive to the migration policy.
- ¹¹ The minimum threshold level imposed by a discrete agency rating scale is 0.5 notch steps.
- ¹² With quarterly data, the sign of a migration event in Q0 is strictly defined by the net rating migration of all actual rating migration events in Q0. However, more than one migration event happening in one quarter is rare, so it is appropriate to designate the net migration in quarters as single events.
- ¹³ In detail, the conditional mean rating migration figures are computed as follows. For each firm-quarter observation (firm i and quarter Q0) the net rating change $\Delta R_{i,0.25,0}$ in Q0 and the net rating change $\Delta R_{i,0.25,t}$ in the 32 quarters surrounding Q0 (Q-16, ..., Q-1, Q0, Q1, ..., Q16, $t \in (-4, -3.75, \dots, 3.75, 4)$) are computed. Due to dataset boundaries, defaulting firms, new firms entering the dataset etc., the time series $\Delta R_{i,0.25,t}$ is not complete for 50% of the 40,440 firm-quarter observations. The mean $\Delta R_{i,0.25,t}$ for all observations, ignoring missing data, is the *unconditional* average rating migration $\Delta R(u)_{i,0.25,t}$. In addition, the *conditional* average rating migration $\Delta R_{i,0.25,t}$ is calculated for observations with an upgrade in Q0 ($\Delta R(+)_i,0.25,t$), for observations with a downgrade in Q0 ($\Delta R(-)_i,0.25,t$) and for observations with a zero migration in Q0 ($\Delta R(0)_i,0.25,t$). This procedure is carried out separately for agency ratings and CM ratings.
- ¹⁴ An alternative to measure the non-random rating dynamics of agency ratings and ARS ratings are the rating reversal probabilities. Given an unconditional migration probability level, determined by the threshold level TH and volatility in credit quality dynamics, the rating reversal probability is significantly reduced by the adjustment fraction AF in the short term. For example, the upgrade probability of agency ratings in Q1 and Q2, following a downgrade in Q0 is 1.4%, a factor 3.5 lower

- than the unconditional probability of 4.8% for upgrades in a semi-annual period. The same numbers apply to ARS-ratings. The reversal probability of agency ratings, following an upgrade in Q0, is 1.4% in Q1 and Q2, which is a factor of 5.5 lower than the unconditional probability of 6.6% for downgrades in a semi-annual period. In this case the reversal probability is higher for ARS ratings: 2.8%, which suggests agencies try to avoid downgrades shortly after an upgrade.
- ¹⁵ A large fraction of agency rating migrations is picked up by changes in ARS ratings. After scaling ΔARS^C by $1/\kappa_R$ the ΔN^C and ΔARS^C converge in the years after the agency migration event. At $t = 4$ $\Delta ARS(+)^C/\kappa_R$, $\Delta N(+)^C$, $\Delta ARS(-)^C/\kappa_R$, $\Delta N(-)^C$ are respectively 1.40, 1.72, -2.16 and -2.48 notch steps. So, at the upside 19% of the agency rating migrations is not picked up by changes in ARS ratings and at the downside only 11% is missing.
- ¹⁶ An alternative to the ACR methodology is to measure the average rating of firms defaulting within T years. This average rating methodology weights type I and II errors proportionally to the numerical rating scale R, while the ACR methodology weights these errors proportionally to $\bar{F}_A(R)$. The lower the average rating figure is, the better ratings anticipate a possible default event. The average agency rating is 3.72 for firms defaulting within three years, and varies between 3.58 and 3.73 for CM ratings (see Table IV). As with the ACR methodology, differences in default performance between agency ratings and CM ratings are small.
- ¹⁷ 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 \bar{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 ($n = 151$). For a time horizon of six years this standard error is slightly higher: 0.025 ($n = 130$ and $\alpha = 7$). For a one year horizon the standard error is 0.015. To verify the theoretical standard error for ACR(1) in the pooled sample, ACR(1) is computed for each annual period between 1981 and 2001. The standard deviation in ACR(1) in the resulting time series is 9.5%. This experimental standard error is close to a theoretical expected standard error of 6.7% (= standard error of 0.015 for a one year horizon multiplied by the square root of the total number of defaults divided by the number of years in the time series = $1.5\% \times \sqrt{200/21}$). This theoretical error of 6.7% is a lower estimate assuming a constant number of defaults in each annual period, while in reality the number of defaults in each annual period varies in between 3 and 56.
- ¹⁸ The standard errors in comparing *differences* between ACRs of agency ratings and CM ratings $\sigma(\Delta ACR)$ are lower than the standard error in ACR itself, since the underlying stochastic defaulting process (same dataset and same defaulting events) is the same for all rating scales. As the CAP curves of agency ratings and CM ratings are comparable, a variation in this stochastic process is expected to have 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 is obtained from a time series analysis of ACR and ΔACR figures. The standard deviation in annual times series of ACR, for agency ratings and CM ratings, is roughly a factor two higher than the standard deviation in annual times series of ΔACR among these ratings. So based on the pooled-sample standard errors for ACR, the pooled-sample standard error $\sigma(\Delta ACR)$ is approximately 0.75% for a time horizon of 1 year and up to 1.25% for a time horizon of 6 years.
- ¹⁹ In a standard logit model setting, the error terms ε_i are assumed to be identically distributed and independently distributed ($Var(\varepsilon_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$). In reality, these error term conditions are violated. To obtain the correct statistics, Huber-White standard errors are used to relax the assumption of homoskedasticity. A generalization of these Huber-White standard errors (see W.H. Rogers, 1993, Regression standard errors in clustered samples. State Technical Bulletin 13, 19-23) relaxes the assumption of independency among all observations as well. Instead, only independency between observations of different firms is assumed.