Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Banking & Finance 28 (2004) 2679–2714
Journal of BANKING & FINANCE
www.elsevier.com/locate/econbase

How rating agencies achieve rating stability

Edward I. Altman [a,*], Herbert A. Rijken [b]

[a] NYU Salomon Center, Leonard N. Stern School of Business, New York University, 44 West 4th Street, New York, NY 10012, USA
[b] Vrije Universiteit Amsterdam, Faculty of Economics and Business Administration, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Available online 26 August 2004

Abstract

Surveys on the use of agency credit ratings reveal that some investors believe that rating agencies are relatively slow in adjusting their ratings. A well-accepted explanation for this perception on the timeliness of ratings is the through-the-cycle methodology that agencies use. According to Moody's, through-the-cycle ratings are stable because they are intended to measure default risk over long investment horizons, and because they are changed only when agencies are confident that observed changes in a company's risk profile are likely to be permanent. To verify this explanation, we quantify the impact of the long-term default horizon and the prudent migration policy on rating stability from the perspective of an investor – with no desire for rating stability. This is done by benchmarking agency ratings with a financial ratio-based (credit-scoring) agency-rating prediction model and (credit-scoring) default-prediction models of various time horizons. We also examine rating-migration practices. The final result is a better quantitative understanding of the through-the-cycle methodology.

By varying the time horizon in the estimation of default-prediction models, we search for a best match with the agency-rating prediction model. Consistent with the agencies' stated objectives, we conclude that agency ratings are focused on the long term. In contrast to one-year default prediction models, agency ratings place less weight on short-term indicators of credit quality.

We also demonstrate that the focus of agencies on long investment horizons explains only part of the relative stability of agency ratings. The other aspect of through-the-cycle methodology – agency-rating migration policy – is an even more important factor underlying the

* Corresponding author. Tel.: +1 212 998 0709; fax: +1 212 995 4220.
E-mail addresses: ealtman@stern.nyu.edu (E.I. Altman), hrijken@feweb.vu.nl (H.A. Rijken).

doi:10.1016/j.jbankfin.2004.06.006

stability of agency ratings. We find that rating migrations are triggered when the difference between the actual agency rating and the model predicted rating exceeds a certain threshold level. When rating migrations are triggered, agencies adjust their ratings only partially, consistent with the known serial dependency of agency-rating migrations.

JEL classification: G20; G33
Keywords: Rating agencies; Through-the-cycle rating methodology; Migration policy; Credit-scoring models; Default prediction

1. Introduction

The credit ratings of Moody's, Standard and Poor's, and Fitch play a key role in the pricing of credit risk and in the delineation of investment strategies. The future role of these agency ratings will be further expanded with the implementation of the Basle II accord, which establishes rating criteria for the capital allocations of banks. Given the rather sudden meltdown in Asian countries and corporations in 1998 and the large increase in defaults in the 2001–2002 period, the timeliness of agency ratings has come under closer scrutiny and criticism.

A recent survey conducted by the Association for Financial Professionals (2002) reveals that most participants believe that agency ratings are slow in responding to changes in corporate credit quality. [1] Surveys by Ellis (1998) and Baker and Mansi (2002) report the same finding. The slowness in rating adjustments is well recognized by investors. Indeed, it seems that investors anticipate the well documented serial correlation in downgrades. [2] In a survey conducted by Ellis (1998), 70% of investors believe that ratings should reflect recent changes in credit quality, 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. On the issue of two conflicting goals – rating timeliness and rating stability – investors appear to have ambiguous opinions. In their meetings with the institutional buyside in 2002, Moody's repeatedly heard that investors value the current rating stability level and do not want ratings simply to follow market prices (see Fons et al., 2002).

The objective of agencies is to provide an accurate *relative* (i.e., ordinal) ranking of credit risk at each point in time, without reference to an explicit time horizon (see Cantor and Mann, 2003). In order to achieve rating stability, agencies take an undefined long-term perspective, which lowers the sensitivity of

[1] The critique of rating agencies focuses mainly on the timeliness of agency ratings, and not on the accuracy of agency ratings. The AFP survey reveals that 83% of the investors surveyed believe that, most of the time, agency ratings accurately reflect the issuers' creditworthiness.

[2] This view has been echoed in a large number of conversations and interviews with market practitioners.

agency ratings to short-term fluctuations in credit quality. In their corporate ratings criteria document, Standard and Poor's (2003) takes the position that "the value of its products is greatest when its ratings focus on the long term and do not fluctuate with near term performance." [3] Agencies aim to respond only to the perceived permanent (long-term) component of credit-quality changes. In addition, agencies follow a prudent migration policy. Only significant changes in credit quality result in rating migrations and, if triggered, ratings are partially adjusted.

The through-the-cycle rating methodology of agencies is designed to achieve an optimal balance between rating timeliness and rating stability. [4] The methodology has two key aspects: first, a long-term default horizon and, second, a prudent migration policy. These two standpoints are aimed at avoiding excessive rating reversals, while holding the timeliness of agency ratings at an acceptable level. [5] It is unclear so far, which aspect of the through-the-cycle approach makes the primary contribution to rating stability.

So far details on how the through-the-cycle methodology is put into practice by agencies and quantitative details on its impact on rating stability are largely unknown to the outside world. [6] The prime objective of this article is to shed some light in this black box. First we quantify the impact of the two aspects of the through-the-cycle methodology to rating stability from an investor's

[3] In their disclosure on corporate ratings criteria, Standard and Poor's explains how to interpret their credit ratings (Standard and Poor's (2003), Corporate Ratings Criteria): *"Standard and Poor's credit ratings are meant to be forward looking; that is, their time horizon extends as far as is analytically foreseeable. Accordingly, the anticipated ups and downs of business cycles – whether industry specific or related to the general economy – should be factored in the credit rating all along. This approach is in keeping with Standard's and Poor's belief that the value of its rating products is greatest when it's rating does not fluctuate with near term performance. Ratings should never be a mere snapshot of the present situation. There are two models for how cyclicality is incorporated in credit ratings. Sometimes, ratings are held constant throughout the cycle. Alternatively, the rating does vary – but within a narrow band".*

[4] According to Moody's, through-the-cycle methodology manages the tension between rating timeliness and rating stability: "*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.'*" (Cantor and Mann, 2003).

[5] According to Moody's, the optimal balance between rating stability and rating timeliness results from a close interaction between agencies and market participants: "*In response to persistent market feedback, Moody's manages its ratings with an eye towards minimizing abrupt changes in rating levels.*" (Cantor, 2001).

[6] There is no consensus on the details of the implementation of the through-the-cycle methodology. Carey and Hrycay (2001) describe through-the-cycle methodology as a rating assignment based on a stress scenario. When firms are consequently rated in the bottom of the credit-quality cycle, agency ratings are insensitive to the credit-quality cycle and focus on the long term. An alternative interpretation of the through-the-cycle methodology is to extract the permanent component from changes in the observed credit quality, on the basis of a forecasting analysis: "*Even though an issuer might experience a change in its financial performance as a result of an adjustment in the macroeconomic environment, its rating may nonetheless remain unchanged if it is likely that its previous financial condition will be restored during the next phase of the cycle.*" (Cantor and Mann, 2003).

perspective – with no desire for rating stability, and second, we provide a further understanding of the through-the-cycle methodology by modeling the prudent migration policy. In order to measure rating stability we formulate credit-scoring models, reflecting the investor's perspective on credit quality. In a benchmark analysis, we compare the agency-rating dynamics with credit-model score dynamics. The conclusions of this study are useful to formulate policies to achieve an optimal balance between rating stability and rating timeliness, to provide guidelines how to use agency ratings in the Basel II framework and to define conditions when it is acceptable to use agency-rating migration matrices as input to rating-based credit risk models.

Of crucial importance to this benchmark study is the formulation of a credible and accurate proxy for the investor's perspective on credit quality – with no desire for rating stability. For this purpose, credit-scoring default prediction models of various time horizons are estimated. We assume investors to have a point-in-time perspective on credit quality, comparable to the well documented perspective of banks. As opposed to through-the-cycle methodology, banks state that they base their internal ratings on the borrower's current condition, i.e., current position in the credit quality cycle (see the Basel Committee on Banking Supervision, 2000; Treacy and Carey, 1998). Their measures of the point-in-time credit quality reflect the current, possibly transient, market perception on credit quality. As a consequence, banks examine both the permanent and transitory components of credit-quality changes. The extent to which point-in-time credit quality measures include temporary fluctuations in credit quality depends on the default horizon. A large number of banks assess the credit quality with a one-year horizon, but nearly as many banks apply horizons of 3–7 years (see Basel Committee on Banking Supervision, 2000). In contrast to rating agencies, banks have no stated objective for rating stability or, more specifically, for avoiding rating reversals that could be caused by overreactions to temporary shocks. Agency ratings are aimed at ignoring temporary shocks and are therefore less likely to be reversed within a short period of time.

From the benchmark study we confirm that agency ratings focus on predicting relative default risk over long horizons. We obtain this empirical finding by modeling the agency-rating scale with an ordered logit regression model and by modeling the default probability with a logit regression model for various time horizons. The agency-rating prediction model parameters closely match the parameters of a default-prediction model with a default horizon of 6 years.

A prudent migration policy is the second aspect of through-the-cycle methodology. The key issue of the migration policy is a reliable detection of the permanent (long-term) component in credit-quality changes and avoidance of rating reversals. Few details are known about the identification of the permanent component by agencies. No straightforward method exists to forecast whether the nature of a credit-quality change is permanent or transitory. A combination of thorough analysis and expert judgment is needed to separate the permanent and transitory components. Because of the uncertainty inherent in forecasting credit quality, agencies follow a prudent migration policy. We characterize the prudent migration policy by two parameters – a threshold parameter and an adjustment fraction parameter. First, a rating change is only considered by agencies if the actual rating differs significantly (by a specific threshold) from an estimated rating based on the most up-to-date credit quality information. Second, if triggered, the ratings are partially adjusted to the estimated rating. This partial adjustment is the source of serial correlation in agency-rating migrations, as reported by Altman and Kao (1991, 1992). Both the direction from which a rating class is reached and the time period of a stay in a particular class, are correlated with the following downgrade or upgrade intensity (see also Lando and Skødeberg, 2002).

In a simulation experiment we quantify the two migration policy parameters. A rating migration is triggered when the rating predicted by a (credit-scoring) agency-rating prediction model differs by at least a threshold level of 1.25 notch steps from the actual agency rating. If triggered, ratings are only partly adjusted, by about 75%. The rating adjustments are split and executed at different times. Agencies appear to follow a moderate "wait-and-see" policy.

In the same spirit, Löffler (2002) examines a rating-migration policy model based on the idea that agencies try to avoid a rating bounce. In this model, agencies set different thresholds for each rating-migration step. The level of these thresholds is determined by a target rating bounce probability, which is set by the agencies and ideally kept as low as possible. Although the notions behind the modeling of agency-rating dynamics are similar to our model, the technical construction of our model differs. Instead of multiple thresholds, we include one threshold level at the upside and one at the downside. We further assume the ratings to be adjusted by a fraction to their predicted rating. [7] In addition, we apply a different simulation approach to test the validity of the rating-migration policy model. Instead of modeling credit-quality dynamics, we proxy the credit-quality dynamics by the dynamics of credit-model scores.

This paper proceeds as follows. In Section 2, the benchmark credit-scoring models are described. In this section the (credit-scoring) agency-rating prediction model is compared with (credit-scoring) default-prediction models for various time horizons. Extensive attention is paid to the credibility of the estimated credit-scoring models to serve as a benchmark for agency ratings. Section 3 outlines the benchmark study. Section 4 benchmarks agency rating dynamics and timeliness of agency-rating migrations. Section 5 describes the simulation experiment in which the migration policy parameters are quantified. Section 6 summarizes the consequences of the long-term default horizon and the prudent migration policy on agency-rating dynamics. Section 7 draws conclusions.

[7] This adjustment fraction explains the serial correlation in agency-rating migration. Löffler's (2002) multiple threshold model explains rating drift as well. In a closely related paper, Löffler (2004) examines alternative explanations for rating drift. The partial rating adjustment hypothesis seems to be most convincing, however.

2. Benchmark credit-scoring models

2.1. Formulation of credit-scoring models

In our benchmark study, we examine the corporate-issuer credit ratings of Standard and Poor's. [8] These ratings reflect "the obligor's ability and willingness to meet its financial commitments on a timely basis" (see Standard and Poor's, 2003). According to this definition, the corporate-issuer credit ratings of agencies are measures of default probability.

We formulate two benchmark credit-scoring models: a default-prediction model (DP model) and an agency-rating prediction model (AR model). Both the DP model and AR model employ the same model variables. This allows an unambiguous comparison of the dynamics of AR-scores and DP-scores. The estimation of these models differs in the use of data (default events versus agency ratings) and statistical methodology.

At first, the DP model is estimated for a short horizon of 1 year. The one-year default probability p_i is modeled as follows:

$$\mathrm{DP}_i = \alpha + \beta X_i + \varepsilon_i, \tag{2.1}$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathrm{DP}_i, \tag{2.2}$$

where X_i is the set of model variables for firm-year observation i. In a standard logit model setting, the error terms ε_i are assumed to be identically distributed and independently distributed ($\mathrm{Var}(\varepsilon_i) = \sigma^2$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ 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 Rogers, 1993) relaxes the assumption of independency among all observations as well. Instead, only independency among clusters of observations – a cluster includes all observations of the same firm – is assumed. The parameters α, β are estimated by a maximum likelihood procedure. The DP-score is directly related to the one-year default probability p_i.

The agency-rating prediction model (AR model) models the discrete agency-rating scale N with an ordered logit regression model. [9] In this model, the AR-score (AR_i) is an unobservable variable:

[8] The empirical analysis is conducted using data on Standard and Poor's corporate-issuer credit ratings. So, strictly speaking, the empirical results refer only to the ratings of Standard and Poor's. We are not aware, however, of a reason why the empirical results and the conclusions presented here for Standard and Poor's ratings should not apply for the ratings of Moody's and Fitch. The discussions and conclusions in this article are therefore generalized to the agency ratings of Standard and Poor's and Moody's and Fitch.

[9] Bond ratings are modeled mainly for the purpose of forecasting agency-rating migrations (see for example Ederington, 1985; Kaplan and Urwitz, 1979; Blume et al., 1998; Kamstra et al., 2001). Applied statistical methodologies are typically either ordinary least square analysis, ordered probit regression analysis, or discriminant analysis. In order to be consistent with the logit regression methodology of the default-prediction model, we model the agency ratings by an ordered logit model.

$$\mathrm{AR}_i = \alpha + \beta X_i + \varepsilon_i, \tag{2.3}$$

where X_i is the set of model variables for firm-year observation i. The AR_i score is related to the agency rating k as follows:

$$y_i = k \quad \text{if } B_{k-1} < \mathrm{AR}_i \leqslant B_k, \tag{2.4}$$

where k is one of the agency-rating classes {CCC/CC, B−, B, . . ., AA, AA+/AAA}, [10] y_i is the actual agency rating, B_k is the upper boundary for the AR-score in rating class k, $B_0 = -\infty$, and $B_{16} = \infty$. In the ordered logit model, the probability that y_i equals k is specified by:

$$P(y_i = k) = F(B_k - \mathrm{AR}_i) - F(B_{k-1} - \mathrm{AR}_i), \tag{2.5}$$

where F is the cumulative logistic function. The parameters α, β, and B_k are estimated with a maximum likelihood procedure. As for the DP model, the generalized Huber–White standard errors are computed, thus relaxing the homoskedasticity assumption and the assumption of independency among observations of the same firm.

The AR-score is in fact a point-in-time measure of the long-term default risk view of agencies. It represents primarily one aspect of the through-the-cycle methodology, the long-term default horizon focus. The migration policy – the second aspect of the through-the-cycle methodology – has little impact on the estimation of the AR-model. The AR-scores might be slightly overstated but this does not effect the benchmarking of rating *dynamics*.

The slight overstatement of AR-scores is explained as follows. Due to a prudent migration policy, the agency ratings may be temporarily understated or overstated. If the number of overstated ratings and the number of understated ratings – due to the prudent migration policy – are equal over the sample period, the migration policy will not affect the AR-model estimate. In that case it will only widen the distribution of the error term ε. However, the number of downgrades is 30% higher than the number of upgrades and the agency-rating migration shows a downward trend. The number of overstated ratings is expected to be slightly higher and, as a consequence, the predicted ratings by the AR model are expected to be slightly higher than in absence of a prudent migration policy. This small shift in predicted rating level does not affect the dynamic properties of these ratings.

2.2. Model variables in the credit-scoring models

The DP-score (Eq. (2.1)) and the AR-score (Eq. (2.3)) are calculated on the basis of the following set of six model variables:

[10] 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 to a single rating class CCC/CC, and the agency-rating classes AA+ and AAA are combined into a single rating class AA+/AAA.

$$\text{DP-, AR-score} = \alpha + \beta_1 \frac{\text{WK}}{\text{TA}} + \beta_2 \frac{\text{RE}}{\text{TA}} + \beta_3 \frac{\text{EBIT}}{\text{TA}} + \beta_4 \frac{\text{ME}}{\text{BL}} + \beta_5 \text{ Size} + \beta_6 \text{ Age}, \tag{2.6}$$

where 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 equals total liabilities normalized by the total value of the US equity market (Mkt) and log-transformed: ln(BL/Mkt). Age is the number of years since a firm was first rated by an agency. [11] In order to increase the effectiveness of the RE/TA, EBIT/TA and ME/BL variables in the logit model estimate, these variables are log-transformed as follows: RE/TA $\rightarrow -\ln(1-\text{RE/TA})$, EBIT/TA $\rightarrow -\ln(1-\text{EBIT/TA})$ and ME/BL $\rightarrow 1+\ln(\text{ME/BL})$. [12]

The choice of the six model variables is inspired by the Z-score model (Altman, 1968). [13,14] The WK/TA variable is a proxy for the short-term liquidity of a firm. The RE/TA, EBIT/TA, and ME/BL variables proxy for historic, current, and future profitability, respectively. The ME/BL variable also proxies for market leverage, which can be interpreted as the willingness of the stock market to invest in a particular firm. Multiple interpretations are possible for the ME/BL variable, as the market value of equity is a catchall variable of actual information regarding future earnings, confidence of investors et cetera. Empirical evidence of a ‘‘too-big-to-fail’’

[11] The Age variable is set to 10 for observations with Age values above 10 and for firms already rated at the start of the dataset in 1981.

[12] The distribution of the ME/BL variable is positively skewed. To a lesser extent, the distributions of the RE/TA variable and EBIT/TA variable are negatively skewed. The information content in the fat tails of the distributions is relatively low. For example, the difference between a ME/BL value of 50 and 25 is far less informative than a difference between a ME/BL value of 1 and 0.5, which might distinguish a healthy firm from a firm approaching default. The effectiveness of the ME/BL variable in the logit regression model estimate can be improved by a log-transformation of the ME/BL variable: $\rightarrow 1+\ln(\text{ME/BL})$. This log-transformation stretches the informative part at the lower side of the ME/BL scale and compresses the non-informative part at the upper side of the ME/BL scale. For the same reason, the RE/TA and EBIT/TA variables are log-transformed: $-\ln(1-\text{RE/TA})$ and $-\ln(1-\text{EBIT/TA})$. The log-transformation reduces the skewness in the distribution of these variables. The average value of these distributions is hardly affected, as the log-transformation centers around 1 for the ME/BL variable and around 0 for the EBIT/TA and RE/TA variables.

[13] The sales-to-asset ratio is not included in the default-prediction model. This variable adds little additional value to a default-prediction model when estimated for a sample of firms covering a wide range of industries.

[14] In a report on their rating methodology, Standard and Poor's (2003) describes a set of 8 key ratios. These ratios include two interest coverage ratios, two cash flow ratios, two earnings profitability ratios and two leverage ratios. In numerous empirical studies on credit-scoring models, different sets of variables are proposed to proxy for these four groups of credit-risk fundamentals. In general, interest coverage ratios and cash flow ratios appear to add surprisingly little to the explanation of default. The strong correlation of these variables with earnings profitability and leverage presumably prevents a significant marginal contribution. Moreover, interest coverage ratios often suffer from ambiguity problems, as both the denominator values (interest) and numerator values (EBIT) are centered close to 0. Only the profitability and leverage ratios, therefore, are included in the benchmark credit-scoring models.

default protection, [15] and empirical evidence of a strong negative relationship between Age and the default rate (for Age values below 10) [16] motivate the inclusion of Size and Age variables in the credit-scoring models.

We do not try to find an optimal set of model variables in the logit model. First, it would be beyond the scope of this article to replicate the numerous studies on finding an optimal set of model variables. [17] Second, the Z-score variables have a good track record. Third, we believe that variation in proxies for profitability and leverage could improve the effectiveness of the credit-scoring models only marginally.

2.3. Parameter estimates of the credit-scoring models

Data on agency ratings is obtained from the Standard and Poor's CREDITPRO database, the July 2002 version, which includes all S&P corporate credit ratings in the period January 1981–July 2002. Less than half of the data in CREDITPRO can be linked with COMPUSTAT data. [18] The requirement of stock price data availability restricts the sample to public firms. In addition, only non-financial US firms are selected.

The panel dataset covers the 1981–2001 period and includes the time series of 1772 obligors with period lengths between 1 and 21 years. It contains 11,890 firm-year observations with known S&P ratings and 1828 firm-year observations with

[15] Although the size of total market equity is often included in default-prediction models, this variable strongly correlates with the ME/BL variable. We include the size of total liabilities instead. Apart from this technical reason, the size of total liabilities is more directly related to the too-big-to-fail protection. One explanation for the too-big-to-fail protection is that credit holders might shy away from the large potential losses in case of a default or bankruptcy, hoping that the problems will be solved by time. The potential loss of large loans, potential damage to bank reputations, and the number of credit holders involved may all slow down the decision process, thereby allowing more time for companies with larger loans to restructure themselves.

[16] A strong negative relationship exists between the Age variable and the default rate for Age values below 10. An exception forms the low default rate in the first year after being rated for the first time (Age = 1); that is, the so-called aging or mortality factor documented in Altman and Bana (2003). This suggests a need for a dummy variable. In a multivariate logit model estimate, however, the parameter of this dummy variable is not statistically significant. The lower default rate in the first year is probably captured by the healthier financial ratios. New ratings often coincide with bond issues, which enhance the financial condition of the issuing firms, at least temporarily.

[17] Most of the literature on credit-scoring models was written in the seventies and the eighties. Research on credit-scoring models has recently gained renewed interest for two reasons. First, the record high default rates in the years 2001 and 2002 (e.g., Altman and Bana, 2003) stimulated a further improvement and refinement of these models. Second, the expected implementation of the Basle II accord has triggered efforts to upgrade internal rating systems of banking institutions.

[18] Apart from minor deviations, the distribution in S&P ratings is not affected by this data reduction and selection of public firms. The percentage of defaulting observations at the beginning of the nineties shrinks, however, while the percentage of defaulting observations in the years 2000 and 2001 increases. Presumably, the credit quality of public firms is less affected than that of private firms by the recession in the beginning of the nineties. In the years 2000 and 2001, the opposite occurred.

non-rated S&P status. [19] Each firm-year observation consists of the S&P rating at the end of June and the values of the model-variables known to the public at that date. Market equity values are based on stock price and total shares outstanding at the end of June. To ensure that the accounting information is publicly available, all balance sheet data refer to the latest fiscal quarter in the previous calendar year. The income statement data refer to the four fiscal quarters in the previous calendar year. [20] This six-month lagging condition for accounting information may be somewhat conservative, as most accounting data become available in the first months after the end of a fiscal year/quarter. For troubled firms, however, financial information is, in general, slower in reaching the financial community.

The panel observations are split into surviving observations and defaulting observations. For the 13,447 surviving observations, stock price data are available both at the end of June in the current year and at the end of June in the subsequent year. [21] For the 271 defaulting observations, the default event happens in the subsequent year. [22] The dependent binary variable p_i in the logit-regression model estimation (see Eq. (2.2)) is equal to 1 for surviving observations and 0 for defaulting observations.

Table 1 provides mean and median values for the model variables, after truncation of their most extreme values. [23] The log-transformations of the ME/BL, RE/TA and EBIT/TA variables reduce the skewness in their panel distributions considerably. The panel distribution of the log transformed ME/BL variable approaches a normal distribution.

[19] The reason to include panel observations of firms with a S&P non-rated status in the estimation of the DP model is to maximize the number of observations in the default-prediction model estimate. Firms with a non-rated status are monitored for default events as well. When defaulting, the rating status of firms with a non-rated status changes to D status.

[20] For companies whose fiscal years end in December, the accounting information refers to the previous fiscal year which equals the previous calendar year. For about 30% of the companies, the fiscal year does not end in December. For these companies the (approximately) six-month lagging accounting information at the end of June is derived as follows: the income statement data are averaged for the four fiscal quarters ending in the previous calendar year. In addition, the balance sheet data are taken from the latest-ending fiscal quarter in the previous calendar year.

[21] The surviving observations are observations of firms at the end of June in year X that also have stock exchange listings at the end of June in year $X+1$. This imposes a survivorship bias. Robustness tests show that this bias does not significantly affect the parameter estimates of the benchmark models.

[22] The defaulting observations are observations of firms at the end of June in year X that default between the end of June in year X and the end of June in year $X+1$.

[23] The raw COMPUSTAT data produce some extreme values for the model variables that contain little relevant information. In order to reduce the impact of these observations the 0.5% highest values and the 0.5% lowest values are truncated for each model variable. These values are replaced by values ranked at 99.5% and 0.5%, respectively. Even though defaults are extreme events, a little amount of defaulting observations is affected by this truncation procedure.

Table 1
Descriptive statistics for the model variables included in the credit-scoring models

Agency rating	N	WK/TA (4–5)/6	RE/TA −ln(1–36/6)	EBIT/TA −ln(1–178/6)	ME/BL 1 + ln(ME/181)	Size ln(181/Mkt)	Age
Mean statistic per rating class							
AAA	317	0.15	0.64	0.17	1.98	−6.53	9.14
AA	1198	0.12	0.52	0.14	1.67	−7.51	9.07
A	2885	0.14	0.39	0.12	1.35	−7.97	8.72
BBB	2603	0.14	0.27	0.10	1.05	−8.48	7.77
BB	2396	0.18	0.11	0.09	0.82	−9.37	6.01
B	2323	0.22	−0.03	0.04	0.56	−10.05	5.29
CCC/CC	168	0.08	−0.21	−0.05	−0.66	−10.08	5.45
NR	1828	0.26	0.21	0.08	1.28	−10.55	9.06
Mean of observations 1 year preceding default (1 Y before D) and at default (D)							
1 Y before D	271	0.11	−0.17	−0.02	−0.83	−10.02	5.33
D	167	−0.13	−0.45	−0.07	−2.44	−9.86	5.88
Statistics for all 13,718 panel observations (excluding D ratings)							
Mean	13,718	0.17	0.23	0.092	1.07	−8.95	7.53
Median	13,718	0.15	0.20	0.092	1.08	−9.05	10
Std. dev.	13,718	0.19	0.35	0.087	1.11	1.66	3.37
Min	13,718	−0.51	−1.22	−0.33	−3.78	−15.04	1
Max	13,718	0.73	1.56	0.40	4.16	−3.18	10
Kurtosis	13,718	3.36	5.68	7.13	3.83	2.89	2.04
Skewness	13,718	0.42	0.03	−0.42	−0.22	0.20	−0.86

The table presents descriptive panel data statistics for the model variables in the credit-scoring models. The dataset consists of 13,718 observations from the period 1981–2001, including 271 observations of firms less than 1 year before default (1 Y before D). For the period after the default event, sufficient data on the model variables are available for 167 of the defaulted firms (D). 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 equals total liabilities, normalized by the total value of the US equity market (Mkt). Age is the number of years since a firm was first rated by an agency. The numbers in the first row refer to COMPUSTAT data codes. The variables RE/TA, EBIT/TA, ME/BL, and Size are log transformed as indicated in the table.

The DP model parameters are estimated for the period 1981–1999 (see the first column of Table 2). The signs of all estimated parameters match expectations. The ME/BL variable turns out to be the dominant variable in the DP model. [24] This is consistent with the success of Moody's KMV structural model, in which market equity and total liabilities play a key role. Although the ME/BL variable is the most important variable, accounting and firm-descriptive information – particularly the obligor characteristics of Size and Age – add substantially to the explanation of the incidence of default.

[24] A logit model that excludes the ME/BL variable is less effective in explaining default rates than is a logit model that includes only the ME/BL variable. Including the ME/BL variable in the logit model reduces the weights of the EBIT/TA variable and the RE/TA variable considerably.

Table 2
Parameter estimates for the DP and the AR model

	Default-prediction model (logit model)		Agency-rating prediction model (ordered logit model)	
	1981–1999	2000–2001	1981–1999	2000–2001
Const	7.06	0.68	–	–
	(9.35)	(0.76)		
WK/TA	0.41	−1.37	−1.80	−1.71
	(0.75)	(−2.11)	(−6.27)	(−4.39)
RE/TA	0.20	0.89	3.33	2.84
	(0.58)	(1.94)	(16.48)	(10.59)
EBIT/TA	3.78	8.36	4.70	8.11
	(3.79)	(5.76)	(8.70)	(9.80)
ME/BL	1.28	1.01	0.85	0.52
	(12.83)	(9.73)	(15.89)	(11.15)
Size	0.47	−0.24	0.95	0.93
	(6.40)	(−2.84)	(17.59)	(14.78)
Age	0.20	0.14	0.082	0.082
	(6.83)	(3.63)	(6.76)	(4.02)
Boundaries B_k				
AAA/AA+	–	–	−0.24	0.26
AA	–	–	−1.61	−0.69
AA−	–	–	−2.25	−1.51
A+	–	–	−3.06	−2.53
A	–	–	−4.14	−3.66
A−	–	–	−4.73	−4.31
BBB+	–	–	−5.30	−5.04
BBB	–	–	−5.99	−5.79
BBB−	–	–	−6.55	−6.53
BB+	–	–	−7.01	−7.00
BB	–	–	−7.64	−7.73
BB−	–	–	−8.56	−8.78
B+	–	–	−10.31	−10.15
B	–	–	−11.65	−11.52
B−	–	–	−13.05	−12.93
CCC/CC	–	–	$-\infty$	$-\infty$
Pseudo-R^2	0.355	0.374	0.214	0.231
N observations	11990	1728	10345	1545
N obs. 1 year preceding default	150	121	–	–

The table presents the parameter estimates for the DP and the AR model. The dependent binary variable in the logit regression model estimation is 0 for the defaulting observations (firms defaulting within 1 year) and 1 for all surviving observations (firms surviving in subsequent year). The dependent variable in the ordered logit regression model estimation is the agency-rating scale. The standard errors in the logit regression estimation are a generalized version of the Huber and White standard errors, which relaxes the assumptions concerning the distribution of error terms and independence among observations of the same firm. The z-statistics are given in brackets. The pseudo-R^2 is a measure for the goodness of the fit.

Because of the arbitrary nature of the DP model, the robustness of the estimated DP parameters is extensively tested.

- For two sub-periods, 1981–1990 and 1991–1999, the DP parameters largely agree with each other, thus demonstrating the time stability of the DP model for the entire 1981–1999 period. Observations in the period 2000–2001 are excluded from the estimation of the DP model. In this period, the DP parameters differ significantly from the period 1981–1999 (see Table 2). The EBIT/TA variable has become more informative on credit quality. Furthermore, the too-big-to-fail default protection has disappeared. Instead, firms with large Size values experienced a higher default rate; 90 firms with liabilities greater than $1 billion defaulted over the 30-month period between January 2001 and June 2003 (see Altman and Bana, 2003). We must wait to determine whether these abrupt changes in DP parameters represent a regime change or should be ascribed to temporally exceptional circumstances. Notice that a large number of large liability failures have occurred in the telecommunications sector.
- When controlling for industry differences, the DP parameters change only slightly, 20% at most. By exception, the estimated parameter of the WK/TA variable increases to a significant value of 0.94. When estimating the DP model separately by industry, the DP parameters are largely comparable. [25]
- To ensure that the DP parameters are not related either to this particular S&P corporate bond dataset or the Standard and Poor's definition of default, the DP model is re-estimated for all bankruptcies, reported by COMPUSTAT. [26] The Age variable is omitted in this re-estimation. Due to space considerations, these results are not presented in this article. [27] The relative weights of the DP parameters appear to be robust to the choice of dataset and the definition of the default event. When using the bankruptcy dataset, the relative weight of the DP parameters is stable over time, varying at the most by 20% between the two sub-periods 1970–1980 and 1981–1998. This allows the DP model to be considered as an out-of-sample model for the entire period 1981–2001.

In summary, for the period 1981–1999, the DP model is stable over time and robust to the definition of default and to dataset choice. It is applicable for different industry sectors and obligors of different sizes. This emphasizes the universal character that makes the DP model a suitable benchmark for agency ratings (excluding the financial sector).

The AR parameters are estimated for the period 1981–1999 (see third column of Table 2). All parameter estimates have the expected sign. [28] As are the DP parameters, the AR parameters are robust to a split in sample period: 1981–1990

[25] The DP model is estimated separately for six industry sectors, defined by the first digit of the SIC code. The sign of the estimated parameters does not vary; the magnitude of the parameters varies within a factor two among these six industry sectors. The parameter for the WK/TA variable is an exception to this finding. It appears to be significantly positive for the infrastructure services sector.

[26] The bankruptcy dataset covers the 1970–1998 period and contains 118,154 surviving observations and 755 bankruptcy observations. Only a small percentage of these bankruptcy observations overlap the defaulting observations in the Standard and Poor's corporate bond dataset.

[27] Results are available on request.

[28] The WK/TA variable is an exception.

and 1991–1999. Observations from 2000 and 2001 are excluded from this model estimation as well, as the AR parameters for this period differ for the EBIT/TA and ME/BL variables (see Table 2). The AR parameters are robust to a split of observations into non-investment graded (BB+ and below) firms and investment graded (BBB– and above) firms. [29] This allows to model the entire agency-rating scale with one single parameter set.

2.4. Identifying the time horizon of the agency-rating prediction model

While the DP model has a known one-year horizon by construction, the AR model has no immediately identifiable time horizon. In order to measure the implicit time horizon of the AR model, we compare the AR parameters with those of the long-term default-prediction models (LDP models).

As for the DP model, the LDP models are estimated with a logit-regression model (see Eqs. (2.1) and (2.2)). The only difference between the DP model estimation and the LDP model estimation is the definition of surviving observations ($p_i=1$) and defaulting observations ($p_i=0$). For a given time horizon T, surviving observations are observations of firms surviving beyond T years, and defaulting observations are observations of firms defaulting within T years. The LDP score represents the probability of default in the coming T years.

LDP models are estimated for a four-year and a six-year horizon. The parameters of these models are estimated for the period 1981–1995 (see Table 3). [30] The generalized version of the Huber and White standard errors accounts for the overlapping periods in the estimation of the LDP model. [31]

The relative weights of the model variables in the AR, DP, and LDP models are compared using the following formula:

$$\mathrm{RW}_i = \frac{|\beta_i|\sigma_i}{\sum_{j=1}^{6}|\beta_j|\sigma_j}, \tag{2.7}$$

where RW_i is the relative weight of model variable i, β_i is the parameter estimate for model variable i, and σ_i is the standard deviation in the panel distribution of model variable i in the period 1981–1995.

---

[29] The only major difference is the absence of a significant parameter for the Age variable for non-investment grade firms.

[30] For comparability reasons, all models presented in Table 3 are estimated for the sample period 1981–1995. In this case, each firm-year observation can "look" 6 years ahead. Including observations in years after 1995 would lower the effective time horizon.

[31] Unlike the DP model, the LDP model models multiyear cumulative default rates. Observations of the same firm are not only correlated because of the relative stable credit quality position over time, but also because of the overlapping multiyear periods in the definition of defaulting observations and surviving observations. Because of their time robustness, the estimated DP parameters and AR parameters hardly differ between an estimation period 1981–1999 (Table 2) and 1981–1995 (Table 3).

Table 3
Comparison of the DP, LDP, and AR models

	DP model	LDP model		AR model
Default-prediction time horizon:	1 year	4 years	6 years	–
Regression results				
Constant	7.72	7.06	6.84	−6.77[a]
	(8.36)	(7.74)	(8.06)	(−0.56)
WK/TA	0.00	0.56	0.82	−1.85
	(0.00)	(1.01)	(1.53)	(−5.36)
RE/TA	0.52	0.70	1.12	3.58
	(1.04)	(1.91)	(3.21)	(15.68)
EBIT/TA	3.61	3.31	1.57	4.49
	(2.46)	(2.83)	(1.30)	(7.48)
ME/BL	1.34	1.09	0.93	0.94
	(8.71)	(9.68)	(8.54)	(14.23)
Size	0.51	0.64	0.66	1.00
	(5.44)	(6.89)	(7.44)	(15.20)
Age	0.183	0.179	0.151	0.080
	(4.89)	(5.31)	(4.82)	(5.57)
Pseudo-R^2	0.347	0.326	0.304	0.213
# surviving obs.	8639	7424	6782	7419
# default obs.	83	293	400	–
Relative weight model variables RW				
WK/TA	0.0%	3.1%	4.8%	7.3%
RE/TA	5.2%	6.9%	11.8%	24.7%
EBIT/TA	8.9%	8.0%	4.0%	7.6%
ME/BL	41.7%	32.9%	29.8%	20.1%
Size	25.7%	31.5%	33.9%	34.7%
Age	18.5%	17.6%	15.7%	5.6%

The table presents the parameter estimates α and β (see Eqs. (2.1) and (2.3)) and the relative weight RW of the model variables (see Eq. (2.7)) for the DP model, the LDP models, and the AR model. In case of the LDP models, the dependent binary variable in the logit regression model estimation is 0 for the defaulting observations (firms defaulting within the default prediction time horizon) and 1 for all surviving observations (firms surviving within the default prediction time horizon). The parameters and RW values are estimated for the period 1981–1995. The standard errors in the logit regression estimation are a generalized version of the Huber and White standard errors, which relaxes the assumptions concerning the distribution of error terms and independence among observations of the same firm. The z-statistics are given in brackets. The pseudo-R^2 is a measure for the goodness of the fit.

[a] Due to space considerations, only the estimated boundary between the rating category BB+ and BBB− (B_7, see Eq. (2.4)) is shown. In this particular case, the standard error of this boundary value is given in the brackets.

The ME/BL variable dominates in the DP model with a RW value of 41.7%. The Size and Age variables have substantial RW values in the DP model as well. These three variables account for most of the variation in the DP-score. The WK/TA, RE/TA, and EBIT/TA variables play only a minor role. The AR model gives the most weight to the Size, RE/TA and ME/BL variables. The RW values of the AR model match most closely with the RW values of the six-year LDP model. This confirms the long term perspective of agency ratings.

Especially for the RE/TA and ME/BL variables, a clear shift in relative weight is observed in the DP, LDP, and AR models, in that order (see Table 3). [32] Not surprisingly, the short-term oriented DP model depends heavily on variables which follow most closely the business cycle, like ME/BL. The AR model and LDP model place relatively more weight on variables which are less sensitive to business cycles, like historical earnings and Size. The traditional measures of "fundamental" risk dominate in long-term measures of credit risk. In the short term, however, if a firm is valued poorly in the marketplace and needs cash to avoid default, it will default.

In the remainder of this article the AR model will refer to the model estimate of the agency-rating scale in the period 1981–1999 (see Table 2), the DP model will refer to the model estimate of the one-year default probability in period 1981–1999 (see Table 2), and the LDP model will refer to the model estimate of the six-year default probability in period 1981–1995 (see Table 3). CM-scores refer in general to AR-scores, LDP-scores and DP-scores. The benchmark analysis itself covers the period 1981–2001.

2.5. Matching CM-scores with agency ratings

In order to examine the credibility of AR-scores and DP-scores to serve as a benchmark for agency ratings, the CM-scores are matched with agency ratings. Average AR-scores and average DP-scores are computed for each agency-rating class. In Fig. 1, these average values are plotted against a numerical agency-rating scale N: CCC/CC/C = 1, B− = 2, B = 3, B+ = 4,..., 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.

The relationship between average DP-score and agency rating N is close to linear. Apparently, the DP-scores are sufficiently and nearly equally dispersed over the entire agency-rating scale. On a more detailed level, two groups of rating classes with an almost perfect linear relationship can be distinguished ($\mathrm{DP} = \alpha + \gamma N + \varepsilon$). For non-investment rating classes 2–7, the slope γ equals 0.405. For investment rating classes 8–15, the slope γ equals 0.307 (see Fig. 1). [33] Not surprisingly a comparable picture appears for AR-scores as well, with γ-values of 0.690 for $N \in [2,\ldots,7]$ and 0.471 for $N \in [8,\ldots,15]$.

[32] No clear shift is observed in the RW values of the WK/TA, EBIT/TA and Size variables.

[33] The distinction between the non-investment grades (rating numbers 2–7) and investment grades (rating numbers 8–15) is determined by eye. The breaking point could equally well be chosen one notch below or above.

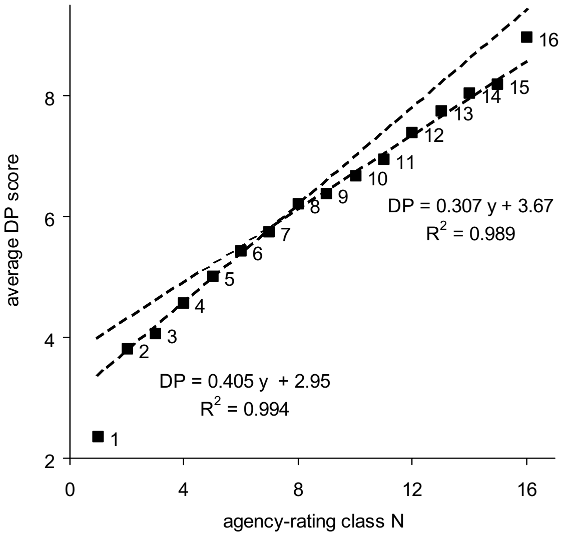

Fig. 1. Average DP-scores for all panel observations in a particular agency-rating class N.

The slope γ depends both on agency-rating class N and time. [34] The time dependency of γ is illustrated in Table 4. This table presents the mean DP-scores and mean AR-scores in each rating class N for the periods 1981–1990 and 1991–2001. The results suggest a slight increase in credit-quality dispersion within the agency-rating scale. On the upper side of the agency-rating scale (rating classes A and above) the mean CM-scores have increased over time (see also Lucas and Lonski, 1992). Blume et al. (1998), who reveal the same findings, ascribe this increase in credit quality to the more stringent rating standards set by agencies in their rating assessment. This explanation is consistent with the decrease in the number of obligors in the upper side of the agency-rating scale. On the lower side of the agency-rating scale (rating classes BBB and below), the mean CM-scores have decreased over time. If CM-scores are time-robust measures of absolute credit quality, this should imply a deterioration in credit quality for the lower rating classes. The increase in default rates for the lower rating classes in the last three decades, as reported by Zhou (2001), supports this suggestion.

[34] The $\gamma(N, t)$ is computed as follows. For each year t, the average CM(N, t)-scores are computed for 16 rating classes N. For $N \in [2,\ldots,7]$, $\gamma(N, t)$ results from the regression equation $\mathrm{CM} = \alpha + \gamma N$, with $N \in [2,\ldots,7]$. For $N \in [8,\ldots,15]$, $\gamma(N, t)$ results from the regression equation $\mathrm{CM} = \alpha + \gamma N$, with $N \in [8,\ldots,15]$. $\gamma(1, t)$ equals $\mathrm{CM}(2, t) - \mathrm{CM}(1, t)$. $\gamma(16, t)$ equals $\mathrm{CM}(16, t) - \mathrm{CM}(15, t)$. In order to reduce noise, the average γ figure is averaged over the current and two previous years: $\gamma(N, t) \rightarrow \{\gamma(N, t) + \gamma(N, t-1) + \gamma(N, t-2)\}/3$. Exceptions are made for $t = 1982$: $\gamma(N, t) \rightarrow \{\gamma(N, t) + \gamma(N, t-1)\}/2$ and $t = 1981$: $\gamma(N, t)$ is not replaced.

Table 4
Descriptive statistics for the AR- and DP-scores within 16 agency-rating classes

		Mean		Median		Standard. dev.		N	
		81–90	91–01	81–90	91–01	81–90	91–01	81–90	91-01
AR-score									
16	AAA/AA+	11.51	11.55	11.98	11.66	1.92	2.42	236	206
15	AA	9.96	11.01	10.31	11.07	1.68	2.40	379	229
14	AA−	9.91	10.18	10.06	10.39	1.40	1.83	222	243
13	A+	9.39	9.73	9.52	9.65	1.35	1.54	409	359
12	A	8.84	9.11	8.85	9.13	1.21	1.55	637	676
11	A−	8.51	8.27	8.52	8.31	1.46	1.55	320	484
10	BBB+	8.36	7.92	8.33	7.93	1.06	1.35	302	541
9	BBB	7.74	7.45	7.71	7.50	1.26	1.36	405	580
8	BBB−	7.58	7.02	7.73	7.02	1.25	1.30	236	539
7	BB+	6.77	6.41	6.83	6.41	1.26	1.37	181	397
6	BB	6.51	5.67	6.65	5.71	1.33	1.45	229	547
5	BB−	5.93	4.95	5.86	4.97	1.31	1.37	324	718
4	B+	5.02	4.15	4.92	4.22	1.23	1.63	571	844
3	B	4.45	3.39	4.50	3.37	1.68	1.77	202	409
2	B−	3.89	2.72	3.80	2.78	1.74	1.73	109	188
1	CCC/CC	3.26	1.82	3.59	1.91	2.37	2.54	54	114
DP-score									
16	AAA/AA+	8.88	9.09	8.92	9.17	1.39	1.68	236	206
15	AA	7.88	8.69	8.02	8.80	1.19	1.48	379	229
14	AA−	7.80	8.26	7.87	8.47	1.13	1.24	222	243
13	A+	7.61	7.91	7.74	8.00	1.06	1.23	409	359
12	A	7.20	7.59	7.28	7.68	1.04	1.21	637	676
11	A−	6.86	7.02	7.03	7.06	1.08	1.18	320	484
10	BBB+	6.66	6.67	6.74	6.75	0.96	1.13	302	541
9	BBB	6.34	6.43	6.36	6.45	1.09	1.06	405	580
8	BBB−	6.27	6.18	6.29	6.21	1.19	1.21	236	539
7	BB+	5.77	5.72	5.80	5.78	0.98	1.36	181	397
6	BB	5.75	5.30	5.75	5.32	1.24	1.50	229	547
5	BB−	5.46	4.81	5.39	4.82	1.25	1.52	324	718
4	B+	4.86	4.38	4.87	4.41	1.33	1.69	571	844
3	B	4.46	3.88	4.56	3.79	1.59	1.82	202	409
2	B−	3.96	3.72	4.11	3.69	1.50	1.89	109	188
1	CCC/CC	3.32	1.90	3.63	2.00	1.52	2.22	54	114

The table presents descriptive statistics for the AR-scores and the DP-scores within 16 agency-rating classes for the periods 1981–1990 and 1991–2001. The AR-scores are scaled to the lower boundary of the B-rating class.

The tractable linear relationship between CM-scores and the numerical agency-rating scale provides further support for the ability of CM-scores to benchmark the entire agency-rating scale consistently. Moreover, the accuracy of CM-scores is comparable for the lower and upper parts of the agency-rating scale. An indication of the accuracy of the CM-scores, relative to the agency-rating scale, is the standard deviation in CM-scores within a particular rating class N (see Table 4). After controlling for γ, the standard deviation in CM-scores varies only up to 25%, leaving aside the agency-rating classes 1 and 16.

Dividing the standard deviation in CM-scores by γ gives an indication of the standard deviation in CM-scores, in terms of notch steps. For example, credit quality as predicted by the AR model varies by about 2 notch steps within a particular rating class. The large variation in CM-scores can be ascribed neither to the supposed noisy character of CM-scores nor to the limited credit-quality information incorporated in the model variables, as compared to the information available to agencies. In that case, the performance of CM-scores in predicting default events is expected to be inferior to that of agency ratings. The default-prediction performance of CM-scores is actually comparable to agency ratings: slightly better in the short term, but slightly worse in the long term. [35] CM-scores and agency ratings are equally informative on credit quality. The credit quality of firms apparently diverges considerably within an agency-rating class and overlaps significantly with neighboring agency-rating classes.

3. Benchmark analysis

3.1. Benchmark research setup

Credit-model scores (CM-scores) are point-in-time measures of short and long term default risk, in the sense that they make maximum use of the most recent data on the model variables to make an optimal estimate. Regardless of whether changes in model variables (i.e., credit fundamentals) have a permanent or temporary character, they all lead to changes in CM-scores. The extent to which these scores incorporate information on temporary fluctuations in credit quality depends on the nature of the credit-scoring model and the time horizon in the default-prediction model. DP model scores are expected to be most sensitive to temporary changes in credit quality, due to their short, one-year default horizon. Both the LDP model and the AR model represent a view on long-term default risk. The sensitivity to short-term fluctuations in credit quality might differ between these two models. The LDP model suppresses sensitivity to temporary credit-quality changes by extending the default horizon. The AR model is less sensitive to temporary credit-quality changes, because it models agency ratings that are intended to be insensitive to short-term credit-quality fluctuations.

Technically, the sensitivity of CM-scores to temporary fluctuations varies because of the differences in weight for (relatively short-term oriented) market information and differences in weight for traditional measures of fundamental risk (see Table 3). Compared to the AR model, the LDP model is slightly more responsive to market equity information and less responsive to historical earnings (see Table 3).

[35] Due to space considerations, these results are not presented in this article. Results are available on request.

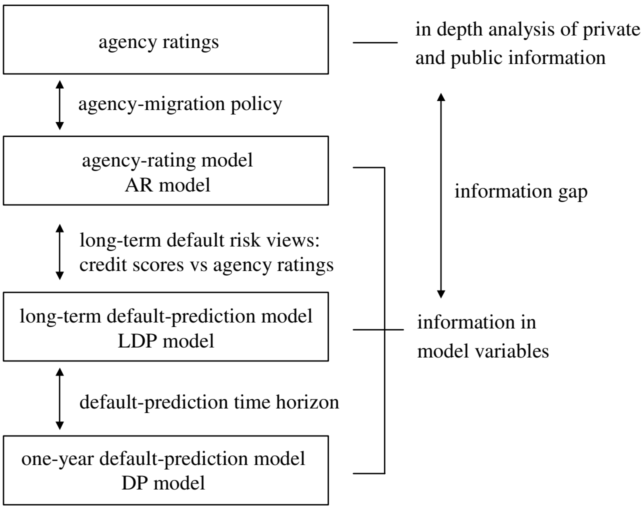

Fig. 2. Concept of the benchmark analysis.

The higher weight for market equity information makes LDP-scores more sensitive to temporary changes in credit quality, as compared to AR-scores. Although an extension of the time horizon in the LDP model could probably close the gap between LDP-scores and AR-scores, the sample period limits a sensible analysis to about six years.

Fig. 2 shows the basic concept of the benchmark analysis. Four measures of default probability, with different stability properties, are compared to each other. Agency ratings and the DP model are placed at opposite ends of the stability spectrum. In between, the AR model bridges the gap between agency ratings and (L)DP model scores, and enables the unambiguous exploration of the two aspects of through-the-cycle methodology (i.e., long-term horizon and prudent migration policy) with regard to rating stability. As explained in Section 2.1, the AR model represents only one aspect of the through-the-cycle methodology: the long-term horizon. Differences in dynamics between agency ratings and AR-scores are to be ascribed exclusively to the prudent migration policy. Differences in dynamics between AR-scores and DP-scores reflect the influence of the default horizon. Differences between AR-scores and LDP-scores reflect only the different weighting on fundamental risk drivers between agency-ratings and credit-scoring models.

The three point-in-time credit-scoring models are compatible with the objectives of internal rating models in banks and with the investors' perspective on credit quality: a point-in-time estimate of the expected default rate in the next 1–7 years. A true proxy for the investor's point-in-time perception is difficult to achieve, of course, as neither a precise reference nor a theoretical framework exists for this perspective. Ultimately, a default-prediction model with the best long-term default-prediction performance in recent history gives the best estimate of point-in-time credit quality.

3.2. Conversion of CM-scores to CM-ratings

Agency-rating dynamics are benchmarked against the dynamics of CM-scores. For a proper comparison, CM-scores are converted to CM-ratings, which are equivalent to agency ratings. Each year, at the end of June, all firms are ranked by CM-scores. On the basis of this ranking, sixteen CM-ratings (AAA/AA+, AA, ..., B−, CCC/CC), equivalent to agency ratings, are assigned to all firms. Each year, the number of firms within each CM-rating class is exactly equal to the number of firms in the equivalent agency-rating class. In order to calculate average migration figures, numbers are assigned to agency ratings and CM-ratings: CCC/CC/C = 1, B− = 2, B = 3,..., AA− = 14, AA = 15, and AA+/AAA = 16 (see also Section 2.5).

4. Benchmarking agency-rating dynamics

Agencies' long-term default horizon and prudent migration policy both reduce the rating migration probability. The extent to which agency-rating migration probability is reduced and the relative importance of these two sources, as observed from the investor's point-in-time perspective, is the aim of the benchmark study. This study is carried out as follows. First the dynamic properties of agency ratings, DP-ratings, and AR-ratings are investigated in detail (see Sections 4.1 and 4.2). A first notion of the impact of the two aspects of the through-the-cycle methodology on agency-rating dynamics is obtained by quantifying the timeliness of agency ratings. This is done by comparing the dynamics of agency ratings and point-in-time CM-ratings – conditional on an agency-rating migration event (Section 4.3). The results underpin the formulation of a migration policy model (Section 5.1). After quantification of the migration policy model parameters in a simulation experiment (Section 5.2) we are in a position to reconstruct rating-migration matrices from a point-in-time perspective with different time horizons, with and without migration policy effects. The rating-migration probabilities uncovered by these matrices quantify the impact of the two aspects of the through-the-cycle methodology on agency-rating dynamics (Section 6).

4.1. Unconditional rating migration

The average rating migration for all observations is the unconditional rating migration. The unconditional rating migration is evaluated for agency ratings and CM-ratings to make sure that differences in rating dynamics are not imposed by the boundaries of the dataset. In the sample period 1981–2001 the annual unconditional rating migration is equal to −0.15 for agency ratings and AR-ratings and −0.13 for DP-ratings. This unconditional downward migration is equal to the difference in rating level between firms entering the dataset and firms exiting the dataset, divided by the number of years of unbroken stay in the dataset (=on average 6.35 years). Firms enter the dataset (1) at the beginning of the dataset in 1981, (2) when they are newly rated, (3) when their non-rated agency-rating status is lifted, or (4) when COMPUSTAT data become available. Firms exit the dataset (1) at the end

Table 5
Unconditional rating migration

	Reasons for firms to enter or exit the dataset	*N*	Weight	Average agency rating	Average AR-rating	Average DP-rating
Enter	Start dataset in 1981	442	26.8%	9.98	10.00	9.94
	Newly rated after 1981	942	57.2%	6.05	5.82	5.65
	NR status lifted up	89	10.7%	7.29	7.13	6.29
	COMP. data become available	176	5.4%	6.70	7.31	8.53
Exit	End of the dataset in 2001	621	37.7%	7.80	7.65	7.47
	Default	240	14.5%	0	0	0
	Rating changed to NR status	340	20.6%	5.55	5.72	6.69
	COMP. data becomes unavailable	448	27.2%	8.10	7.89	7.90
No exit no enter		8621	–	9.22	9.23	9.15
Total enter		1649	100.0%	7.24	7.17	7.14
Total exit (excl default)		1409	85.5%	7.35	7.26	7.42
Total exit		1649	100.0%	6.28	6.20	6.34
Total exit–total enter				−0.95	−0.96	−0.80
Unconditional annual migration				−0.15	−0.15	−0.13

The table presents the calculation of the unconditional migration of agency ratings, AR-ratings and DP-ratings. The average annual rating migration is the difference between the average rating level of firms entering the dataset and the average rating level of firms exiting the dataset, divided by the average number of years of unbroken stay in the dataset (=6.35 years).

of the dataset in 2001, (2) in case of a default event, (3) when a rating changes to a non-rated status, or (4) when COMPUSTAT data become unavailable. [36]

Table 5 presents the analysis of the unconditional migration of agency ratings and CM-ratings. These ratings agree on the rating level at which firms enter the dataset R_{EN} and on the rating level at which firms exit the dataset R_{EX}. R_{EN} is about 7 (BB+) and R_{EX} is about 6 (BB), resulting in an unconditional downward migration of about 1 notch for all firms during their stay in the dataset. Dividing this figure by the average time period of 6.35 years of unbroken stay in the dataset gives the unconditional annual rating-migration of −0.15. Grouping by reason to exit and enter the dataset reveals that 80% of this unconditional rating migration is due to financially troubled firms in the 2–3 years approaching the default event. When these firm-year observations are eliminated from the dataset, R_{EX} is about equal to R_{EN}.

4.2. Rating-migration probability and rating drift

Table 6 shows the one-year rating-migration distribution for agency ratings, DP-ratings, and AR-ratings. The symmetric properties of this distribution are examined with the average rating migration figure and the ratio between the number of down-

[36] We have not examined a possible relationship between rating level and reasons for the absence of COMPUSTAT data and the reasons for a switch to a non-rated status.

Table 6
Panel rating-migration distributions for agency ratings, AR-ratings and DP-ratings

	N	Rating migration ΔR (notch steps) (%)								D/U ratio	Sign test z-value	Rating migration	
		<−3	−3	−2	−1	0	1	2	>2			<Mean	t-stat.
Agency rating													
All	10257	1.5	1.6	3.6	7.9	76.6	6.5	1.8	0.5	1.65	−12.7	−0.15	−16.1
Downgrade previous year	1076	3.2	3.8	6.4	13.8	68.4	3.1	0.9	0.5	6.08	4.8	−0.47	−12.6
No change previous year	6738	1.4	1.4	3.5	7.9	76.8	6.8	1.8	0.4	1.57	−9.4	−0.14	−12.2
Upgrade previous year	806	0.6	0.9	0.9	4.2	78.9	10.5	3.4	0.6	0.46	−13.4	0.08	2.9
Investment grade	6304	0.9	0.9	3.4	7.8	80.4	5.4	1.0	0.2	1.95	−11.7	−0.14	−13.2
Non-investment grade	3953	2.5	2.7	3.9	8.0	70.7	8.4	2.9	0.9	1.39	−6.3	−0.17	−9.6
1981–1990	4336	1.2	1.5	4.1	6.7	78.2	5.7	2.0	0.7	1.61	−7.6	−0.13	−9.3
1991–2001	5921	1.7	1.7	3.2	8.8	75.5	7.2	1.6	0.4	1.67	−10.7	−0.17	−13.1
AR-rating													
All	10257	1.4	1.6	4.5	20.0	51.0	17.9	3.0	0.5	1.29	−10.2	−0.15	−13.9
Downgrade previous year	1854	1.1	1.3	5.2	23.7	46.9	17.3	3.9	0.6	1.44	−5.8	−0.16	−6.4
No change previous year	4588	0.9	0.9	2.9	18.7	58.7	15.7	1.9	0.3	1.31	−6.3	−0.11	−8.1
Upgrade previous year	2178	2.5	2.7	6.3	19.2	43.4	21.7	3.7	0.6	1.18	−4.5	−0.21	−7.5
Investment grade	6290	1.0	1.5	4.1	19.6	55.9	16.0	1.7	0.2	1.46	−10.9	−0.18	−13.5
Non-investment grade	3967	2.1	1.8	5.2	20.7	43.2	21.0	5.1	1.0	1.10	−3.2	−0.12	−6.1
1981–1990	4336	1.0	1.6	4.5	18.7	52.0	18.6	3.0	0.6	1.16	−4.2	−0.11	−6.7
1991–2001	5921	1.7	1.7	4.5	21.0	50.3	17.4	3.0	0.4	1.39	−9.8	−0.19	−12.5
DP-rating													
All	10257	2.7	2.8	6.9	19.9	39.3	18.6	6.6	3.1	1.14	−6.1	−0.13	−8.6
Downgrade previous year	2615	3.4	3.4	8.0	19.5	34.8	20.4	7.5	3.1	1.11	−2.5	−0.16	−5.2
No change previous year	3531	1.2	1.7	5.2	19.2	47.9	18.4	4.8	1.5	1.10	−2.4	−0.07	−3.6
Upgrade previous year	2474	2.5	2.6	7.2	21.3	36.8	18.6	7.0	4.0	1.14	−3.2	−0.09	−2.8
Investment grade	6206	3.2	3.1	7.2	20.0	42.1	17.8	5.0	1.6	1.37	−11.1	−0.27	−14.4
Non-investment grade	4051	2.0	2.4	6.4	19.8	35.1	19.9	9.0	5.4	0.89	3.6	0.08	3.4
1981–1990	4336	2.6	2.5	5.9	17.8	39.6	19.8	7.9	4.0	0.91	2.1	0.00	−0.2
1991–2001	5921	2.8	3.0	7.6	21.5	39.2	17.8	5.6	2.4	1.35	−9.9	−0.22	−11.5

The table presents, for agency ratings, AR-ratings and DP-ratings, the panel distributions of the rating migration ΔR, (1) unconditionally for all observations, (2) conditionally on the rating migration in previous year, (3) for a sub-sample of non-investment graded firms and for a sub-sample of investment graded firms, and (4) for two periods 1981–1990 and 1991–2001. In the last 4 columns, the symmetric properties of these rating-migration distributions are assessed in terms of the average rating-migration and the ratio in number of downgrades and the number of upgrades (D/U ratio). The two-sided sign test determines, whether the number of downgrades differs significantly from the number of upgrades.

grades D and upgrades U (D/U). The two-sided sign test determines whether the number of downgrades deviates significantly from the number of upgrades.

The one-year migration probability is 23.4% for agency ratings, 49.0% for AR-ratings, and 60.7% for DP-ratings, including migrations to default. These significant differences imply that both long-term default horizon and migration policy have a significant impact on the agency-rating stability. These unconditional migration probability figures are robust to splits in both investment graded and non-investment graded firms and to splits in the sample period.

Conditional on the absence of rating migration in the previous annual period, the average agency-rating migration is −0.14, about equal to the unconditional rating migration. Conditional on a downgrade in the previous annual period, the average agency-rating migration is −0.47. Conditional on an upgrade in the previous annual period, the average agency-rating migration is +0.08. The upward drift in agency ratings is much smaller than the downward drift in agency ratings. This asymmetric behavior in rating drift is reported by Altman and Kao (1992). When controlling for unconditional rating migration of −0.15, the rating drift is of equal magnitude at both the downside and the upside. This implies that the underlying source of rating drift is equally effective in both directions.

In contradiction to agency ratings, no significant rating drift is observed for DP-ratings and AR-ratings. Conditional on the rating migration in the previous year, the average AR-rating migration and the average DP-rating migration do not differ significantly from the unconditional rating migrations of −0.15 and −0.13, respectively. Rating drift in agency ratings is apparently a migration policy effect and not the result of a drift in the credit-quality fundamentals, as reflected by the model variables in credit-scoring models. Cantor and Hamilton (2004) have shown that serial dependence in ratings largely disappears once ratings are conditioned on outlook or watchlist information. So, on the basis of timely (point-in-time) credit quality information, credit-quality changes are not (or are at least very little) predictable from historic changes in credit quality. This finding is consistent with absence of drift in CM-ratings. In the first instance, agencies partially update their ratings in response to changes in credit quality and take time to determine whether the credit-quality changes are temporary or permanent. Meanwhile, outlook and watchlist information provides investors more up-to-date information on changes in credit quality.

4.3. Timeliness of agency ratings

Conditional on the agency-rating migration event, average changes in CM-ratings surrounding the migration event are investigated. The magnitude of the conditional changes in CM-ratings just before the agency-migration event is an indication of the timeliness of agency ratings.

For each year T, at the end of June, firms are classified into three samples, conditional on the agency-rating migration $\Delta N_{T-1,T}$ in the previous annual period: one sample of firms with an upgrade ($\Delta N_{T-1,T} > 0$), one sample of firms with a downgrade ($\Delta N_{T-1,T} < 0$), and one sample of firms with no migration ($\Delta N_{T-1,T} = 0$). For each of these three samples, the average rating change $\Delta R_{T+t-1,T+t}$ (agency ratings or

CM-ratings) in eight annual periods surrounding the agency-rating migration event are computed, with $t\in(-4, 4)$. These rating-migration figures are subsequently averaged over the sample period $T\in(1981, 2001)$, resulting in $\Delta R_{t-1,t}$. These $\Delta R_{t-1,t}$ figures are cumulated, starting at $t = -5$, resulting in the cumulative rating change figure ΔR_t^{C}:

$$\Delta R^{\mathrm{C}}(i)_t = \sum_{k=-4}^{t} \Delta R(i)_{k-1,k}, \tag{4.1}$$

where i refers to the three samples, conditional on the agency-rating migration in period $(-1, 0)$: "+" refers to an upgrade, "0" refers to no migration, and "−" refers to a downgrade. The ΔR^{C} figures are computed for agency ratings (ΔN^{C}), AR-ratings ($\Delta \mathrm{AR}^{\mathrm{C}}$), LDP-ratings ($\Delta \mathrm{LDP}^{\mathrm{C}}$), and DP-ratings ($\Delta \mathrm{DP}^{\mathrm{C}}$). The average $\Delta R^{\mathrm{C}}(+)$ and $\Delta R^{\mathrm{C}}(-)$ figures for CM-ratings are measures of the timeliness of agency ratings from a point-in-time perspective.

The cumulative rating changes $\Delta \mathrm{R}^{\mathrm{C}}(+)$ and $\Delta \mathrm{R}^{\mathrm{C}}(-)$ are subtracted by $\Delta \mathrm{R}^{\mathrm{C}}(0)$ and scaled by a factor $1/\kappa_R$:

$$\frac{\Delta R^{\mathrm{C}}(+)_t - \Delta R^{\mathrm{C}}(0)_t}{\kappa_R} \to \Delta R^{\mathrm{C}}(+)_t, \tag{4.2}$$

$$\frac{\Delta R^{\mathrm{C}}(-)_t - \Delta R^{\mathrm{C}}(0)_t}{\kappa_R} \to \Delta R^{\mathrm{C}}(-)_t. \tag{4.3}$$

This conversion of $\Delta R^{\mathrm{C}}(+)$ and $\Delta R^{\mathrm{C}}(-)$ is motivated as follows:

1. The subtraction of $\Delta R^{\mathrm{C}}(0)$ controls for the downward rating trend and potential bias due to missing observations. Due to defaults, mergers, de-listings, et cetera, the composition of the three conditional samples varies for $t\in(-4, 4)$. For example, firms defaulting in the period $(-1, 0)$ are missing in the mean calculations of $\Delta R_{0,1}$, $\Delta R_{1,2}$, $\Delta R_{2,3}$, and $\Delta R_{3,4}$, all of which include the relatively longer-surviving, healthier firms.
2. The scaling factor $1/\kappa_R$ allows a comparison of ΔN^{C}, $\Delta \mathrm{AR}^{\mathrm{C}}$, and $\Delta \mathrm{DP}^{\mathrm{C}}$ in terms of agency-rating notch steps. Because of the disagreement between agency ratings (N) and CM-ratings (CM) the slope κ_R in the regression equation, $\mathrm{CM} = \kappa_R N + \text{constant}$, does not equal 1. In order to compare the $\Delta \mathrm{CM}^{\mathrm{C}}$ figures correctly with ΔN^{C} figures, in terms of agency-rating notch steps, the $\Delta \mathrm{CM}^{\mathrm{C}}$ figures are scaled by κ_R. For AR-ratings, LDP-ratings, and DP-ratings κ_R equals 0.83, 0.79 and 0.74, respectively.

Fig. 3 presents the time series of the converted cumulative rating changes $\Delta R_t^{\mathrm{C}}(+)$ and $\Delta R_t^{\mathrm{C}}(-)$. Conditional on an agency-rating downgrade, the total cumulative rating change $\Delta R_4^{\mathrm{C}}(-)$ is −2.2 notch steps. Just before the downgrade – at $t = -1$, $\Delta \mathrm{AR}_{-1}^{\mathrm{C}}(-)$ predicts a decrease by 0.9 notch steps, while $\Delta N_{-1}^{\mathrm{C}}(-)$ equals −0.3 notch steps. Similar, but in absolute terms, slightly lower numbers are found for an agency-rating upgrade ($\Delta R_4^{\mathrm{C}}(+)$ equals +1.6 notch steps, $\Delta \mathrm{AR}_{-1}^{\mathrm{C}}(+)$ equals 0.7 notch steps and $\Delta N_{-1}^{\mathrm{C}}(+)$ equals −0.1 notch steps).

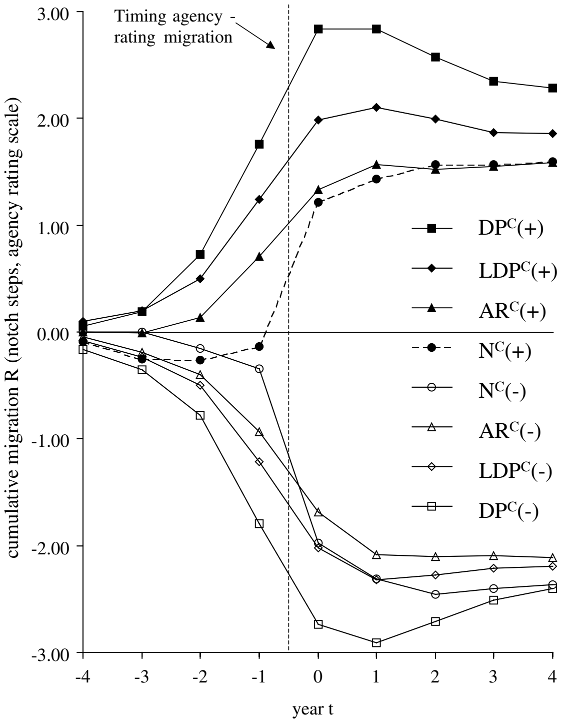

Fig. 3. The figure shows the cumulative rating-migration ΔR^{C}, for agency ratings N, AR-ratings, LDP-ratings and DP-ratings conditional on an agency-rating upgrade and downgrade. Base year for the accumulation of rating migrations is $t = -5$, on average 4.5 years before the agency-rating migration event. The expressions $\Delta N^{\mathrm{C}}(+)$ and $\Delta N^{\mathrm{C}}(-)$ refer to the cumulative agency-rating migration conditional on respectively an upgrade in period $(-1, 0)$ and a downgrade in period $(-1, 0)$. Comparable definitions apply to the AR-ratings, LDP-ratings and DP-ratings. The cumulative rating-migration figures $\Delta R^{\mathrm{C}}(+)$ and $\Delta R^{\mathrm{C}}(-)$ are subtracted by $\Delta R^{\mathrm{C}}(0)$ scales by κ_R – the slope in the regression equation CM = $\kappa_R N$ + constant (see Eqs. (4.1) and (4.2)) and scaled.

On average, CM-rating changes clearly anticipate agency-rating migrations. Among CM-ratings, the DP-ratings anticipate most strongly an agency-rating migration event. In the two-year period surrounding the agency-rating migration date, $\Delta\mathrm{DP}^{\mathrm{C}}$ and (to a lesser extent) $\Delta\mathrm{LDP}^{\mathrm{C}}$ show some "overshooting" behavior (see Fig. 3). [37] Just after the agency-migration date, $\Delta\mathrm{DP}^{\mathrm{C}}$ clearly exceeds the permanent component in the credit-quality change, as proxied by $\Delta\mathrm{AR}^{\mathrm{C}}$ at $t = 4$ and reflected by ΔN^{C} at $t = 4$. This overshooting behavior suggests that DP-scores are highly sensitive to the temporary component in credit-quality changes, and that LDP-scores are moderately sensitive. The absence of overshooting behavior for

[37] From a technical perspective, the greater anticipation evidence in DP-ratings is driven by a higher sensitivity to changes in the ME/BL variable (see Section 2.4).

AR-ratings implies that AR-scores are not sensitive to short-term fluctuations in credit quality. This is consistent with the agencies' objective to suppress the influence of temporary changes in credit quality on their ratings.

The results and conclusions presented above are robust with regard to rating level and sample period. [38] Only some modest differences appear between investment-grade firms and non-investment grade firms. The robustness of the conclusions has two important implications. First, no major differences in migration policy appear between the high and low credit-quality range, and no major change in migration policy shows up between the sample periods 1981–1990 and 1991–2001. Second, CM-scores are just as effective in detecting credit-quality changes in the high credit-quality range as they are in the low credit-quality range.

To summarize, if investors monitor changes in credit quality based on a point-in-time perspective, as proxied by DP-ratings or LDP-ratings, they may be dissatisfied with the timeliness of agency ratings for two reasons:

1. Migration policy masks the dynamics of point-in-time credit quality. Agency ratings do not respond immediately to changes in point-in-time credit quality.
2. Investors with short horizons are sensitive to temporary changes in credit quality, while agency ratings are intended to ignore these changes.

5. Characterization of migration policy

5.1. Migration policy parameters

Differences between the dynamics of AR-ratings and actual agency ratings are ascribed to the agencies' migration policy. In order to understand the impact of migration policy on rating dynamics, we propose a simple model representing the migration policy of agencies. This model is characterized by two parameters:

- 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. [39] This threshold prevents small credit-quality fluctuations from triggering a rating migration thereby reducing the probability of rating migration.
- In case a rating migration is triggered, the ratings are not fully adjusted to the current credit-quality level. The adjustment fraction AF specifies the partial adjustment of agency ratings. The partial adjustment of ratings (i.e., the spreading of the target rating adjustment over time) is responsible for the observed drift in agency ratings.

[38] Due to space considerations, the robustness test results for the two periods are not presented in this article. Results are available on request.

[39] The minimum threshold level, imposed by the discrete agency-rating scale, is 0.5 notch steps.

The threshold level TH can be estimated from the time-series of ΔAR^{C} (see Fig. 3). The ΔAR^{C}-level at which an agency-rating migration is triggered is about 1 notch step for an upgrade and about -1.25 notch steps for a downgrade (an agency-rating migration occurs on average at $t=-0.5$). The threshold level TH is therefore likely to be about 1–1.25 notch steps. A best guess for the adjustment fraction AF can be made as follows. If agencies *do not* spread the intended rating adjustments over more years, the average agency-rating migration in the periods $(-5, -1)$ and $(0, 4)$, surrounding the agency-rating migration event, is expected to be close to zero, given the unpredictability of changes in credit quality (see Section 4.2). However it seems that agencies do spread rating adjustments over time. The total agency-rating migration ΔN_4^{C}, conditional on a downgrade and an upgrade in annual period $(-1, 0)$, equals -2.2 and 1.6 notch steps, respectively. On average, two-thirds of this total migration ΔN_4^{C} occurs in period $(-1, 0)$. The other part occurs in the annual periods surrounding $(-1, 0)$. When triggered, ratings appear, on average, adjusted only by a fraction 2/3 of the target rating level. The remainder of the intended rating adjustment is executed on a later date.

5.2. Simulation of agency-rating dynamics

A simulation experiment is conducted as an alternative way of estimating the two migration policy parameters TH and AF. The simulation experiment involves three steps. In the first step, AR-scores are modified to AR^{M}-scores, reflecting a particular migration policy. In the second step, the AR^{M}-scores are converted to AR(TH, AF)-ratings. In the third step, the migration policy parameters TH and AF are determined by searching for matches in rating-migration distributions and rating-drift properties between agency ratings and AR(TH, AF)-ratings.

Step 1: Modification of AR-scores

For each observation, the AR-score is converted to a modified score AR^{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 AR_t-scores for a particular firm, the modified AR_t^{M}-scores are computed. At the beginning of the time-series of each firm, AR_0^{M} is set equal to AR_0. The AR_t^{M}-score is held constant as long as the AR_t-score stays within the threshold interval $(AR_{t-1}^{M} - \gamma \times TH, AR_{t-1}^{M} + \gamma \times TH)$:

$$AR_t^{M} = AR_{t-1}^{M}, \quad \text{if } \frac{|AR_t - AR_{t-1}^{M}|}{\gamma} < TH, \tag{5.1}$$

where $t \in (0, t^{max})$ and t^{max} is the period of unbroken stay of a particular firm in the dataset. TH is expressed in notch steps, γ converts the AR-score to a notch scale (see

Section 2.5). As soon as the AR_t-score exceeds the threshold interval, the AR_t^M-score is adjusted as follows:

$$AR_t^M = AF \times (AR_t - AR_{t-1}^M) + AR_{t-1}^M \quad \text{if } \frac{|AR_t - AR_{t-1}^M|}{\gamma} \geqslant TH. \tag{5.2}$$

If AF = 1, the AR_t^M-score is fully adjusted to the AR_t-score. If AF < 1, the AR_t^M-score is partially adjusted to the AR_t-score.

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

AR^M-scores are converted to AR(TH, AF)-ratings, equivalent to agency ratings, by following the procedure as described in Section 3.2. The time-series of AR^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 AR(TH, AF)-rating migrations with the correct sign is crucial to the simulation experiment. This unambiguous conversion is checked and safeguarded as follows:

- The minimum size of the jump in AR^M-scores is $\gamma \times AF \times TH$, which is sufficient to convert nearly all jumps in the modified AR^M-score to AR(TH, AF)-rating migrations.
- This conversion procedure, however, does not prevent an AR(TH, AF)-rating migration to happen, when no jump occurs in the AR^M-score. To prevent these non-intended migrations, AR(TH, AF)-ratings are replaced by lagged ratings, when AR_t^M equals the one-year lagged value AR_{t-1}^M. As a consequence, the distribution of the AR(TH, AF)-ratings is slightly altered. The number of observations in each rating class, before and after this correction, differ by 10% at most. This change in rating distribution does not seriously affect the comparability of the AR(TH, AF)-ratings with agency ratings.

Step 3: Determination of migration policy parameters

For a range of TH and AF parameter values the rating-migration distributions and rating-drift properties of AR(TH, AF)-ratings are determined (see Table 7). The actual migration policy parameters are determined by finding a best match in rating-migration distributions and rating-drift properties between agency ratings and AR(TH, AF) ratings. In this analysis, the migrations to default are excluded from the simulation experiment, as these migrations are obviously not initiated by agencies.

The rating-migration probabilities of agency ratings and AR(TH, AF)-ratings are best matched for a TH value of 1.25 notch steps. Practically, the rating-migration probability appears to be insensitive to the adjustment fraction AF. The two migration policy parameters influence rating dynamics nearly independently from each other. The threshold level influences the rating-migration probability, the adjustment fraction influences the strength of the rating-drift and the distribution in rating-migration size (in number of notch steps).

Table 7
Rating-migration distribution and rating-drift properties for agency ratings and simulated AR(TH, AF)-ratings

Rating	All	Rating migration in the previous year (notch steps)				
		<−1	−1	0	1	>1
Average rating migration (notch steps)						
Agency rating	−0.09	−0.30	−0.30	−0.08	0.06	0.18
AR-rating	−0.07	0.09	−0.06	−0.07	−0.15	−0.06
AR-rating, TH = 1.25, AF = 1	−0.06	−0.10	−0.08	−0.07	0.00	−0.01
AR-rating, TH = 1.25, AF = 0.83	−0.06	−0.25	−0.15	−0.07	0.14	0.19
AR-rating, TH = 1.25, AF = 0.66	−0.06	−0.38	−0.31	−0.05	0.21	0.33
AR-rating, TH = 1.25, AF = 0.50	−0.06	−0.59	−0.33	−0.05	0.29	0.50
Standard deviation of rating migration (notch steps)						
Agency rating	0.82	1.16	0.92	0.79	0.70	0.85
AR-rating	0.93	1.21	0.99	0.82	0.98	1.28
AR-rating, TH = 1.25, AF = 1	0.78	0.84	0.70	0.78	0.70	0.86
AR-rating, TH = 1.25, AF = 0.83	0.78	0.84	0.70	0.78	0.70	0.86
AR-rating, TH = 1.25, AF = 0.66	0.68	0.83	0.71	0.65	0.69	0.80
AR-rating, TH = 1.25, AF = 0.50	0.60	0.80	0.63	0.55	0.63	0.75
Rating-migration distribution						
Agency rating	8416[a]	4.4%	7.4	78.7	7.1%	2.4%
AR-rating	8416[a]	5.2%	19.0%	53.8%	18.3%	3.6%
AR-rating, TH = 1.25, AF = 1	8416[a]	6.3%	5.5%	79.4%	4.6%	4.2%
AR-rating, TH = 1.25, AF = 0.83	8416[a]	5.2%	6.9%	79.4%	5.3%	3.3%
AR-rating, TH = 1.25, AF = 0.66	8416[a]	3.8%	8.4%	79.5%	6.0%	2.3%
AR-rating, TH = 1.25, AF = 0.50	8416[a]	2.7%	9.1%	80.5%	6.5%	1.2%

The table presents the rating-migration distribution, mean and standard deviation of the rating migration for agency ratings, AR ratings and simulated AR(TH, AF)-ratings. These figures are given unconditionally for all observations ("all") and conditionally on the migration in the previous year.
[a] Number of observations employed in the analysis.

In case of a full rating adjustment (AF = 1), no significant rating drift is observed, regardless of the threshold level. [40] For AR-ratings – specifically AR-ratings and AR(1.25, 1)-ratings – no significant positive relationship shows up between ΔR and ΔR_{t-1}. [41] In a further refinement of the simulation experiment, the adjustment fraction AF is varied in between 0.5 and 1. As expected, the rating drift appears as soon as ratings are partly adjusted, and the magnitude of the rating drift increases with lower adjustment factors. The best match in rating-drift properties with agency ratings is obtained for AR(1.25, 0.66)-ratings and AR(1.25, 0.83)-ratings. From this finding we conclude that, on average, agencies apply a threshold of 1.25 and partially

[40] The insensitivity of rating-drift properties to the threshold level TH dispels a concern that the absence of rating drift in CM-ratings is due to a countervailing effect of continuously reverting noisy CM-scores. Were this effect to exist, the rating-drift properties should depend on the threshold level, and this is not the case.

[41] Only conditional on $\Delta R_{t-1} < -1$ and $\Delta R_{t-1} = 1$, slight differences in the average rating migration ΔR_t are observed when the threshold level TH is increased to 1.25 notch steps.

adjust their ratings by a factor of about 0.75. These migration policy parameters are not extreme; they point towards a reasonably prudent migration policy. Notice that the 1.25 notch step threshold level is a minimum estimate. In the simulation experiment, the rating-migration trigger is set, on average, at 0.5 year before the agency-rating migration event. The actual threshold level applied by the agencies just before the actual agency-rating migration event is likely to be somewhat higher.

6. Benchmarking the migration matrices of agency ratings

Up to this point, the academic literature on agency-rating migration matrices has focused primarily on the influence of the business cycle, bond rating age, and industry (see Altman, 1998; Nickel et al., 2000; Hu et al., 2002; Bangia et al., 2002). This article aims to draw attention to the volatility of the agency-rating migration matrix. As shown in previous chapters, the stability of agency ratings is significantly enhanced by a prudent agency-migration policy and long-term default horizon. The benchmark CM-ratings and the migration policy model enables to proxy how the agency-rating migration matrix would look like if agencies were to relax their prudent migration policies and if they were to focus on a one-year (instead of a longer term) horizon.

Table 8 presents the one-year rating-migration matrices $T(N_t, N_{t+1})$ for agency ratings, AR(1.25, 0.66)-ratings, AR-ratings, LDP-ratings and DP-ratings on a major rating level. [42] Eliminating the impact of the migration policy, as reflected by AR-ratings, increases the agency-rating migration probabilities by a factor of 2. From a one-year point-in-time perspective (DP-ratings) the agency-rating migration probability is a factor 3 too low. From this short-term perspective, the two aspects of the through-the-cycle approach – prudent migration policy and long-term default horizon – contribute equally to the stability of agency ratings.

We believe that a rating-migration matrix, based on a default-prediction model with a six-year horizon, is a good proxy for the true rating-migration matrix. In this setting, ‘‘true’’ refers to the investor’s point-in-time perspective with no desire for rating stability. Alternative rating-migration matrices were proposed by Kealhofer et al. (1998) and Carey and Hrycay (2001), based on EDF-scores (KMV model) and sole accounting information, respectively. The rating-migration probabilities are a factor 1.5–2 higher than our proposal for a true rating-migration matrix. Ultimately, a default-prediction model with the best default-prediction performance in recent history and estimated with the appropriate investor’s time horizon gives the best estimate of the true rating-migration matrix.

The dynamic properties of the one-year rating-migration matrix are analyzed in greater detail per rating class. The average rating migration $\Delta R(N)$ within 1 year equals:

[42] On a major rating level, rating numbers 1–7 refer to the following agency-rating rating classes: CCC/CC, B, BB, BBB, A, AA and AAA. The rating class CCC/CC is a combination of the rating classes C, CC and CCC.

Table 8
One-year rating-migration matrix

	AAA	AA	A	BBB	BB	B	CCC	Default
Agency rating								
AAA	0.92	0.07	0.01	0.00	0.00	0.00	0.00	0.00
AA	0.01	0.92	0.07	0.01	0.00	0.00	0.00	0.00
A	0.00	0.02	0.91	0.07	0.00	0.00	0.00	0.00
BBB	0.00	0.00	0.05	0.88	0.06	0.01	0.00	0.00
BB	0.00	0.00	0.00	0.06	0.84	0.07	0.00	0.02
B	0.00	0.00	0.00	0.00	0.06	0.82	0.04	0.08
CCC/CC	0.00	0.00	0.00	0.00	0.01	0.14	0.48	0.38
AR(1.25, 0.66)-rating								
AAA	0.92	0.07	0.01	0.00	0.00	0.00	0.00	0.00
AA	0.03	0.88	0.09	0.00	0.00	0.00	0.00	0.00
A	0.00	0.02	0.92	0.06	0.00	0.00	0.00	0.00
BBB	0.00	0.00	0.04	0.89	0.05	0.00	0.00	0.01
BB	0.00	0.00	0.00	0.05	0.85	0.08	0.00	0.02
B	0.00	0.00	0.00	0.00	0.06	0.84	0.03	0.07
CCC/CC	0.00	0.00	0.00	0.00	0.00	0.16	0.50	0.34
AR-rating								
AAA	0.82	0.17	0.01	0.00	0.00	0.00	0.00	0.00
AA	0.04	0.81	0.15	0.00	0.00	0.00	0.00	0.00
A	0.00	0.05	0.84	0.11	0.00	0.00	0.00	0.00
BBB	0.00	0.00	0.10	0.77	0.12	0.01	0.00	0.00
BB	0.00	0.00	0.00	0.13	0.71	0.14	0.00	0.02
B	0.00	0.00	0.00	0.00	0.13	0.75	0.04	0.08
CCC/CC	0.00	0.00	0.00	0.00	0.00	0.25	0.38	0.38
LDP-rating								
AAA	0.80	0.18	0.02	0.00	0.00	0.00	0.00	0.00
AA	0.05	0.80	0.15	0.00	0.00	0.00	0.00	0.00
A	0.00	0.06	0.82	0.11	0.00	0.00	0.00	0.00
BBB	0.00	0.00	0.13	0.74	0.12	0.01	0.00	0.00
BB	0.00	0.00	0.00	0.16	0.68	0.14	0.00	0.01
B	0.00	0.00	0.00	0.00	0.16	0.73	0.05	0.07
CCC/CC	0.00	0.00	0.00	0.00	0.00	0.22	0.27	0.51
DP-rating								
AAA	0.78	0.21	0.00	0.00	0.00	0.00	0.00	0.00
AA	0.05	0.74	0.20	0.01	0.00	0.00	0.00	0.00
A	0.00	0.08	0.76	0.14	0.02	0.00	0.00	0.00
BBB	0.00	0.00	0.16	0.64	0.17	0.02	0.00	0.00
BB	0.00	0.00	0.01	0.22	0.58	0.18	0.00	0.01
B	0.00	0.00	0.00	0.01	0.18	0.69	0.05	0.07
CCC/CC	0.00	0.00	0.00	0.00	0.01	0.26	0.22	0.51

The table presents the one-year rating-migration matrix of agency ratings, simulated AR(1.25, 0.66)-ratings, AR-ratings, LDP-ratings and DP-ratings. The elements in the rating-migration matrix (N_t, N_{t+1}) represent the one-year probability of a rating migration from rating class N_t to rating class N_{t+1}.

$$\Delta R(N) = \sum_{k=1}^{7} T(N,k)(k-N), \tag{6.1}$$

where $T(N_t,\ N_{t+1})$ are the elements of the one-year rating-migration matrices. The rating-migration probability $P(N)$ equals:

$$P(N) = \sum_{k=1}^{7} T(N,k)|k \neq N, \tag{6.2}$$

$\Delta R(N)$ and $P(N)$ are given in Table 9. In order to highlight the rating-reversion rates, default migrations are excluded from the analysis (the default migration trigger is assumed to follow a completely different stochastic process). In the absence of default migration events, ratings tend to migrate towards a mean investment grade (see also Altman, 1998). Part of this rating-reverting behavior could be ascribed to the restricted number of possible upgrades and downgrades, especially for the highest and lowest rating classes. Another part of the reverting behavior is more fundamental. Corporate credit quality, as measured by the CM-score, tends to revert toward mean credit-quality values, with a mean reversion rate depending on the current credit-quality level. The precise characterization of this stochastic process requires further study.

On a more detailed level, per rating class, the same conclusions are derived for the relative contribution of the long-term default horizon and migration policy to the agency-rating stability. For all rating classes, the rating-migration probabilities of

Table 9
Rating migration and rating-migration probability in a one-year rating-migration matrix

Rating class N	AAA	AA	A	BBB	BB	B	CCC	All
Rating migration $\Delta R(N)$								
Agency rating	−0.086	−0.073	−0.060	−0.024	−0.012	0.030	0.247	−0.027
AR(1.25, 0.66)-rating	−0.088	−0.062	−0.048	−0.016	−0.029	0.033	0.244	−0.024
AR(1.25, 1)-rating	−0.110	−0.067	−0.062	−0.016	−0.020	0.046	0.276	−0.025
AR-rating	−0.193	−0.118	−0.070	−0.037	−0.013	0.102	0.398	−0.027
LDP(1.25, 0.66)-rating	−0.084	−0.063	−0.038	−0.010	−0.005	0.054	0.250	−0.011
LDP(1.25, 1)-rating	−0.101	−0.076	−0.053	−0.010	0.004	0.088	0.288	−0.009
LDP-rating	−0.218	−0.107	−0.072	−0.020	0.019	0.126	0.457	−0.013
DP-rating	−0.219	−0.163	−0.111	−0.047	0.052	0.159	0.565	−0.021
Rating migration probability $P(N)$								
Agency rating	8.0%	8.1%	9.0%	11.5%	14.3%	10.5%	23.5%	9.5%
AR(1.25, 0.66)-rating	7.8%	12.1%	8.2%	10.2%	13.7%	9.0%	24.4%	9.0%
AR(1.25, 1)-rating	8.5%	13.4%	10.6%	13.7%	16.2%	10.7%	27.6%	11.5%
AR-rating	17.9%	19.2%	15.7%	22.2%	27.9%	18.6%	39.8%	20.0%
LDP(1.25, 0.66)-rating	8.1%	11.2%	9.0%	14.1%	19.6%	12.0%	25.0%	12.0%
LDP(1.25, 1)-rating	9.0%	12.8%	12.1%	17.6%	22.5%	15.3%	28.8%	15.0%
LDP-rating	20.1%	20.2%	17.6%	26.2%	31.5%	21.9%	45.7%	23.6%
DP-rating	21.6%	25.9%	23.8%	35.9%	40.8%	25.9%	55.1%	33.0%

On a major rating class level N, the table presents the average rating migration $\Delta R(N)$ (see Eq. (6.1)) and rating-migration probability $P(N)$ (see Eq. (6.2)) in a one-year migration matrix for agency ratings, AR(TH, AF)-ratings, LDP(TH, AF)-ratings and DP-ratings. LDP(TH, AF)-ratings are defined in a similar way as AR(TH, AF)-ratings, by modifying LDP-scores instead of AR-scores. Migrations to default are excluded from the computation of these figures.

AR-ratings are higher than those of agency ratings by a factor of approximately 2.5, while the DP-ratings are higher by a factor of approximately 3 (see Table 9).

A key input for many credit-risk pricing models is the credit-risk migration matrix. In general, when using the agency-rating migration matrix as a proxy for the credit-quality dynamics, the corporate bond spreads predicted by these models are too low. Elton et al. (2001) suggest that a higher expected default rate (compensation for default loss), a higher variation in the unexpected default rate (credit risk premium), or a combination of both should account for a relatively large part of the credit risk spread. In other words, the volatility in the agency-rating migration matrix might be too low to explain the credit spreads for corporate bonds. Whether the more volatile rating-migration matrices – as proposed in this article – can explain a more substantial part of corporate bond spreads is an interesting topic for follow-up research.

7. Conclusions

Both aspects of the through-the-cycle methodology – prudent migration policy and default-prediction time horizon – are responsible for the investors' perception of rigidity in agency ratings. From a six-year point-in-time perspective, when no rating stability is desired, the agency-rating migration probability is lower than expected by a factor of 2.5, primarily due to the prudent migration policy. An investor with one-year perspective should apply a slightly higher correction factor: 3.

These empirical results are obtained by benchmarking agency-rating dynamics against rather well established credit-model scores, which proxy for the point-in-time perspective on credit quality. In addition, by varying the default horizon in estimating default-prediction credit-scoring models, we demonstrate that agencies focus on long-term default rates, and not on one-year investment horizons.

In a simulation experiment, the migration policy is characterized by two parameters. An agency-rating migration is triggered when the point-in-time rating prediction differs from the actual agency rating by at least 1.25 notch steps. If triggered, the agency-rating migration closes 75% of the gap between the actual agency-rating level and the predicted rating level. Although these parameters do not suggest that the migration policy lags excessively, they are sufficient to enhance agency-rating stability significantly.

Acknowledgments

The authors wish to thank Richard Cantor, Adrian Buckley, an anonymous referee, seminar participants at the University of Ulm and at the Vrije Universiteit Amsterdam and participants at the FMA-conference in Denver (2003) for their comments and suggestions on a working paper version of this article.

References

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance 35, 1001–1016.

Altman, E.I., Kao, D.L., 1991. Corporate Bond Rating Drift: An Examination of Agency Credit-Quality Changes. AIMR, Charlottesville VA.

Altman, E.I., Kao, D.L., 1992. The implications of corporate bond ratings drift. Financial Analysts Journal (June), 64–75.

Altman, E.I., 1998. The importance and subtlety of credit rating migration. Journal of Banking and Finance 22, 1231–1247.

Altman, E.I., Bana, G., 2003. Default and returns on high yield bonds: The year 2002 in review and the market outlook, Also Report of same title through 2003, (Altman and Fanjul, 2004), NYU Salomon Center, Stern School of Business.

Association for Financial Professionals, 2002. Ratings Agencies Survey: Accuracy, Timeliness, and Regulation. Available from <www.afponline.org>.

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

Bangia, A.F., Diebold, X., Kronimus, A., Schagen, C., Schuermann, T., 2002. Ratings migration and the business cycle with application to credit portfolio stress testing. Journal of Banking and Finance 26, 445–474.

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

Blume, M.E., Lim, F., MacKinlay, A.C., 1998. The declining credit quality of US Corporate Debt: Myth or Reality? Journal of Finance 53, 1389–1413.

Cantor, R., 2001. Moody's investors service response to the consultative paper issued by the Basel Committee on Bank Supervision 'A New Capital Adequacy Framework'. Journal of Banking and Finance, 25.

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

Cantor, R., Hamilton, D., 2004. Rating Transitions and Defaults Conditional on Outlooks, Watchlists, and Rating History. Special Comment, Moody's Investor Services, February.

Carey, M., Hrycay, M., 2001. Parameterizing credit risk models with rating data. Journal of Banking and Finance 25, 197–270.

Ederington, L.H., 1985. Classification models and bond ratings. The Financial Review 20 (4), 237–262.

Ellis, D., 1998. Different sides of the same story: Investors' and issuers' views of rating agencies. The Journal of Fixed Income 9, 35–45.

Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C., 2001. Explaining the rate spread on corporate bonds. The Journal of Finance 56, 247–277.

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

Hu, Y., Kiesel, R., Perraudin, W., 2002. The estimation of migration matrices for sovereign credit ratings. Journal of Banking and Finance 26, 1383–1406.

Kaplan, R.S., Urwitz, G., 1979. Statistical models of bond ratings: A methodology inquiry. Journal of Business 52, 231–261.

Kamstra, M., Kennedy, P., Suan, T., 2001. Combining bond rating forecasts using logit. The Financial Review 36, 75–96.

Kealhofer, S., Kwok, S., Weng, S., 1998. Uses and abuses of default rates. CreditMetrics Monitor, 1st Quarter, pp. 37–55.

Lando, D., Skødeberg, T.M., 2002. Analyzing rating migrations and rating drift with continuous observations. Journal of Banking and Finance 26, 423–444.

Löffler, G., 2002. Avoiding the rating bounce: Why agencies are slow to react to new information, Working paper.

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

Lucas, D., Lonski, J., 1992. Changes in Corporate Credit Quality: 1970–1990. Journal of Fixed Income (March), 7–14.

Nickel, P., Perraudin, W., Varotto, S., 2000. Stability of rating migrations. Journal of Banking and Finance 24, 203–227.

Rogers, W.H., 1993. Regression standard errors in clustered samples. Stata Technical Bulletin 13, 19–23 (Reprinted in Stata Technical Bulletin Reprints 3, 88–94).

Standard and Poor's, 2003. Corporate ratings criteria. Available from <www.standardandpoors.com>.

Treacy, W.F., Carey, M., 1998. Credit risk rating at large US banks. Federal Reserve Bulletin 84, November, pp. 897–921.

Zhou, C., 2001. Credit rating and corporate defaults. The Journal of Fixed Income 11, 30–40.