A Point-in-Time Perspective on Through-the-Cycle Ratings

Author(s): Edward I. Altman and Herbert A. Rijken

Source: *Financial Analysts Journal*, Jan. - Feb., 2006, Vol. 62, No. 1 (Jan. - Feb., 2006), pp. 54-70

Published by: Taylor & Francis, Ltd.

Stable URL: https://www.jstor.org/stable/4480730

FAJ | Financial Analysts Journal
Volume 62 • Number 1
©2006, CFA Institute

A Point-in-Time Perspective on Through-the-Cycle Ratings

Edward I. Altman and Herbert A. Rijken

The role and performance of credit-rating agencies are currently under debate. Several surveys conducted in the United States reveal that most investors believe rating agencies are too slow in adjusting their ratings to changes in corporate creditworthiness. It is well known that agencies achieve rating stability by their through-the-cycle methodology. This study provides quantitative insight into this methodology from an investor's point-in-time perspective and quantifies the effects of the methodology on three, somewhat conflicting, objectives: rating stability, rating timeliness, and performance in predicting defaults. The results can guide the search for an optimal balance among these three objectives.

The quality of corporate credit ratings issued by the major rating agencies has come under close scrutiny because the adjustments of these ratings are perceived to be too slow. A survey by the Association for Financial Professionals in 2002 revealed that investors are not satisfied with the timeliness of ratings.[1] A similar case was made earlier by Ellis (1998) and Baker and Mansi (2002). In the survey conducted by Ellis, 70 percent of investors believed that ratings should reflect recent changes in default risk, even if ratings are likely to be reversed within a year.

At the same time, investors want to keep their portfolio rebalancing as low as possible and desire some level of rating stability. They do not want ratings to be changed to reflect small changes in financial condition. This argument has been put forward by the rating agencies. Standard & Poor's Corporation (2003) is convinced that stable ratings are of most value to investors, and Moody's Investors Service (Fons, Cantor, and Mahoney 2002) has picked up fervent support among investors for the current level of rating stability. From a bank regulatory perspective, rating stability is desirable to prevent procyclical effects. For example, a prompt and full response to changes in current creditworthiness could deepen a financial crisis. Linking portfolio strategies and portfolio mandates to ratings of nationally recognized statistical rating organizations (NRSROs) and, in the future, linking bank capital requirements to NRSRO ratings could force banks and investors to liquidate their positions hurriedly as ratings decline, which could ultimately result in a credit crunch. Another argument for rating stability is that it maintains the reputation of the agencies. Rating reversals within a short period have a negative impact on an agency's reputation, even when the reversals reflect true changes in creditworthiness. In a sense, for the agencies, it is better to be slow and right than fast and wrong. A strong reputation, which underlies the recognition of ratings in financial markets, is in the interest of agencies, regulatory authorities, investors, and bond-issuing companies.

Edward I. Altman is the Max L. Heine Professor of Finance at the Stern School of Business, New York University. Herbert A. Rijken is professor of corporate finance at the Vrije University, Amsterdam, the Netherlands.

Apparently, investors want both stable and timely ratings, which are likely to be conflicting objectives. Moody's tries to find a compromise:

> Moody's analysts attempt to balance the market's need for timely updates on issuer risk profiles, with its conflicting expectation for stable ratings. (Cantor 2001, p. 175)

In response to criticism of rating timeliness, in January 2002, Moody's considered changing ratings more aggressively and updating them more frequently. The company renounced its intention, however, after broad consultation with investors, companies, and financial authorities. In meetings, Moody's repeatedly heard that investors value the current level of rating stability and do not want ratings to simply follow market prices. Moody's, therefore, decided to continue to produce stable ratings (see Fons et al.). Because the Association for Financial Professionals survey reported the opposite conclusion by investors, we believe that timeliness versus stability is a dilemma for investors, and it is one that prompts our continued interest (see Altman and Rijken 2004).

In the discussions of rating timeliness so far, a rigorous analysis of the pros and cons of rating stability in quantitative terms has been lacking. How stable are ratings? How do rating agencies achieve rating stability? What are the costs of rating stability in terms of timeliness and default prediction? We attempt to provide answers to these questions.

Through-the-Cycle Methodology

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

Based on the first aspect of the through-the-cycle rating methodology, the agencies disregard short-term fluctuations in default risk.[2] By filtering out the temporary credit risk component, agency ratings reflect only the permanent, long-term, structural component. According to Cantor and Mann (2003), the through-the-cycle methodology aims to avoid excessive rating reversals while holding the timeliness of agency ratings at an acceptable level:

> If over time new information reveals a potential change in an issuer's relative creditworthiness, Moody's considers whether or not to adjust the rating. It manages the tension between its dual objectives—accuracy and stability—by changing ratings only when it believes an issuer has experienced what is likely to be an enduring change in fundamental creditworthiness. For this reason, ratings are said to "look through-the-cycle." (p. 4)

Standard & Poor's (2003) believes that

> the value of its rating products is greatest when its ratings focus on the long term and do not fluctuate with near term performance.

The second aspect of the through-the-cycle methodology is the enhancement of rating stability by a prudent migration policy. After filtering out the temporary credit risk component, agencies monitor the remaining (permanent) credit risk component for substantial changes. If a migration is triggered, ratings are, *on average*, only partially adjusted to the actual level in the permanent credit risk component. Although this policy has not been officially disclosed by the agencies, practical evidence of such a prudent migration policy exists. Moody's provided some insight into its migration policy in an announcement in January 2002 stating that it was reconsidering its migration policy:

> Under consideration are more aggressive ratings changes—such as downgrading a rating by several notches immediately in reaction to adverse news rather than slowly reducing the rating over a period of time—as well as shortening the rating review cycle to a period of weeks from the current period of months. ("Moody's Mulls . . ." 2002)

In contrast to the through-the-cycle methodology, bankers have a point-in-time perspective on corporate credit quality with a time horizon between one and seven years (Basel Committee on Banking Supervision 2000). A reasonable assumption is that this perspective applies also to other investors. The point-in-time perspective views the current default risk of a counterparty without attempting to suppress the temporary credit risk component. It weights both the temporary and permanent components of credit quality. The relative weights of these two components depend on the time horizon. For a one-year horizon, the temporary credit risk component is weighted more heavily than it would be for a longer time horizon.

Precisely how rating agencies put into practice their through-the-cycle methodology is not clear. Treacy and Carey (2000) described the through-the-cycle rating methodology as a rating assessment in a worst-case scenario, at the bottom of a presumed credit-quality cycle. Löffler (2004) explored the through-the-cycle effects on rating stability and default prediction performance quantitatively by modeling the separation of the permanent and temporary components of default risk in a Kalman filter approach.

We take a different approach to investigate the impact of the through-the-cycle methodology. We benchmark the dynamics of agency ratings with equivalent ratings based on credit scores, which serve as proxies for the point-in-time investor's perspective. To connect with the investor's perception as closely as possible, we use credit-scoring models with a high default prediction accuracy and vary the prediction time horizon.

An earlier article (Altman and Rijken 2004) focused on the *modeling* of the through-the-cycle methodology, especially the prudent migration policy. In this article, we emphasize the *quantitative consequences* of the through-the-cycle methodology for rating stability, rating timeliness, and default prediction performance from a point-in-time perspective.

For this study, we examined corporate issuers' credit ratings of Standard & Poor's. (Additional information from the "Rating Outlook" and "CreditWatch" sources is not included.) Strictly speaking, all the empirical results presented in this article refer to ratings of Standard & Poor's. However, we are not aware of any reason that empirical results and conclusions presented here for Standard & Poor's ratings should not apply to ratings of Moody's and Fitch Ratings. The discussions and conclusions in this article are, therefore, generalizable to ratings of Moody's and Fitch.

Benchmark Credit-Scoring Models

We describe in this section default prediction models and an agency rating prediction model. Scores of these benchmark credit-scoring models represent a range of point-in-time perspectives with different time horizons and different sensitivities to the temporary credit risk component. After converting these scores to credit-model ratings, equivalent to agency ratings, we compare rating stability, rating timeliness, and default prediction performance of credit-model ratings with those of agency ratings.

Default Prediction Models. We estimated all default prediction models by the following logit regression models in a panel data setting:

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

The $CS_{i,t}$ score is related to the expected probability of default as follows:

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

where

- $CS_{i,t}$ = credit score of company i at time t
- WK = net working capital
- RE = retained earnings
- TA = total assets
- $EBIT$ = earnings before interest and taxes
- ME = market value of equity
- BL = book value of total liabilities
- $Size$ = the log transformation of total liabilities normalized by the total value of the U.S. equity market, $\ln(BL/Mkt)$
- Age = number of years since the company was first rated by an agency[3]
- $E(p_{i,t})$ = expected probability of default of company i at time t

The first four variables, before their transformations, are found in the familiar Z-score model (Altman 1968), supplemented by the size and age of the company. The parameters of the logit regression model, α and β, were estimated by a standard maximum-likelihood procedure. This estimation procedure seeks for an optimal match between the actual outcome, $p_{i,t}$, and the expected outcome of the model, $E(p_{i,t})$; $p_{i,t} = 0$ when company i defaults before $t + T$, and $p_{i,t} = 1$ when company i survives beyond $t + T$. We estimated the default prediction models for various time horizons T.

In addition, we estimated *marginal* default prediction models. These models focus exclusively on default probability in a specific future period (i.e., the permanent credit risk component), and the binary variable $p_{i,t}$ was set to 0 only for companies defaulting in this future period $(t + T_1, t + T_2)$, with $0 < T_1 < T_2$. Default events in the near future $(t, t + T_1)$ were ignored by setting $p_{i,t} = 1$ for companies defaulting in this period. (An alternative procedure of leaving out the observations of companies defaulting in the near future did not change the estimate significantly because the number of near-term defaults is relatively small.) We set $p_{i,t}$ to 1 when company i survived beyond $t + T_2$.

Agency Rating Prediction Model. The discrete agency rating (AR) scale, N, is modeled by an ordered-logit regression model in a panel data setting:

$$\begin{aligned} AR_{i,t} &= \alpha + \beta_1 \frac{WK_{i,t}}{TA_{i,t}} - \beta_2 \ln\left(1 - \frac{RE_{i,t}}{TA_{i,t}}\right) \\ &\quad - \beta_3 \ln\left(1 - \frac{EBIT_{i,t}}{TA_{i,t}}\right) + \beta_4\left(1 + \ln\frac{ME_{i,t}}{BL_{i,t}}\right) \\ &\quad + \beta_5 Size_{i,t} + \beta_6 Age_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{2a}$$

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

$$N_{i,t} = R \text{ if } B_{R-1} < AR_{i,t} \leq B_R, \tag{2b}$$

where

- R = one of the agency rating classes
- $N_{i,t}$ = agency rating of company i at time t
- B_R = upper boundary for the AR score in rating class R
- B_0 = $-\infty$
- B_{16} = ∞

In the ordered-logit model, the probability that $N_{i,t}$ equals R is specified by

$$P(N_{i,t} = R) = F(B_R - AR_{i,t}) - F(B_{R-1} - AR_{i,t}), \tag{2c}$$

where F is the cumulative logistic function. The parameters α, β, and B_R were estimated by a maximum-likelihood procedure. This procedure seeks for an optimal match between the actual agency rating and the most likely rating class predicted by the model.

In the estimate, we took into account the following 16 agency rating classes: AAA/AA+, AA, AA–, A+, A, A–, BBB+, BBB, BBB–, BB+, BB, BB–, B+, B, B–, CCC/CC. (To have a reasonable number of observations in each rating class, we combined classes C, CC, CCC–, CCC, and CCC+ into a single rating class, CCC/CC and we combined classes AA+ and AAA into a single rating class, AAA/AA+.)

Estimation Benchmark for Credit-Scoring Models. We obtained data on agency ratings from the July 2002 version of Standard & Poor's CreditPro Database, which includes all Standard & Poor's corporate credit ratings for the January 1981–July 2002 period. We linked corporate ratings at the end of each calendar quarter—March, June, September, and December—to stock price data and accounting data from Compustat. We assumed that accounting data are publicly available three months after the end of the fiscal year. The resulting panel dataset includes time series of 1,629 corporate issuers. The average period in the time series is 27.0 quarters. In addition, the dataset contains 9,253 company-quarter observations of companies with a "non-rated" Standard & Poor's status. These observations maximize the number of default events in the default prediction model estimation.[4]

Table 1 reports the estimated α and β parameters of three default prediction models, an agency rating prediction model, and two bankruptcy prediction models. The three default prediction models are a short-term default prediction model with a time horizon of one year (SDP model), a long-term default prediction model with a time horizon of six years (LDP model), and a "marginal" default prediction model for a one-year period starting five years in the future (MDP model).[5] The estimation period is restricted to the 1981–95 period. (A 1981–2001 estimation period would underweight the observations of companies defaulting beyond one year in the estimation of the LDP and MDP models.)

We estimated two bankruptcy prediction models: the BP model for the 1970–95 period and the BPO model for the 1970–80 period. To ensure that the parameters of the default prediction models were not uniquely related either to the particular Standard & Poor's corporate bond dataset or Standard & Poor's definition of default, we constructed a new dataset to include all bankruptcies reported by Compustat.[6] The estimation methodology of the bankruptcy prediction models is identical to the estimation of the default prediction models, as described in the previous section, except for the omission of the *Age* variable in the BP and BPO models and the replacement of the default indicator, $p_{i,t}$, by a bankruptcy indicator. Time horizon T is one year for these bankruptcy models. In the remainder of the article, the BPO model is considered to be an *out-of-sample* model in testing default prediction performance in the 1981–2001 period.

The parameters of the credit-scoring models are robust over time in the 1981–99 period. No substantial differences are observed in parameter estimates between the 1981–90 and 1991–99 subperiods. The 2000–01 period is an exception. Most notable is the absence of the too-big-to-fail default protection in this period. When industry sector differences are controlled for, model parameters vary only slightly (with the exception of *WK/TA*).[7] A specific test of the AR model showed the robustness of the estimated parameters to a split of observations into non-investment-grade (BB+ and below) companies and investment-grade (BBB– and above) companies.[8] The AR parameters do not vary substantially with the agency rating level, which enabled us to model the entire agency rating scale with a single parameter set. These robustness tests demonstrate the universal character of the credit-scoring models that makes them a suitable benchmark for agency ratings.

Comparison of Credit-Scoring Models. All credit-scoring models use the same model variables, except the bankruptcy prediction models, which omit the *Age* variable. Therefore, a fair comparison can be made of the relative weights, RW_k, of model variables k:

$$RW_k = \frac{|\beta_k|\sigma_k}{\sum_{j=1}^{6}|\beta_j|\sigma_j}, \tag{3}$$

where β_k is the parameter estimate for model variable k and σ_k is the standard deviation of model variable k in the pooled sample distribution in the 1981–95 period.

Panel B of Table 1 shows the *RW* values for the estimated credit-scoring models. The *ME/BL* variable dominates in the SDP model, with a relative weight of 40.7 percent. This result is consistent with the Moody's KMV Corporation structural model, in which market equity and total liabilities also play a key role. Although the *ME/BL* variable is most important, accounting information—particularly the corporation characteristics of *Size* and *Age*—add substantially to the explanation of the default incidence. The *WK/TA* and *RE/TA* variables play a minor role. The weights of the model variables in the BP model, BPO model, and SDP model—all with a one-year time horizon—are comparable. The relative weights of the model variables in the default prediction models appear to be relatively robust to dataset choice, estimation period, and definition of the default event.

Table 1. Parameter Estimates of Default Prediction Models, Agency Rating Prediction Model, and Bankruptcy Prediction Models
(Z-statistics in parentheses; statistically significant parameters in boldface)

	Default Prediction Models			Agency Rating Prediction Model	Bankruptcy Prediction Models	
Prediction time:	One Year	Six Years	One-Year Period in Future	NA	One Year	
Horizon (0,*T*) or *T*:	(0,1)	(0,6)	(5,6)	NA	(0,1)	
Estimation period:	1981–95	1981–95	1981–95	1981–95	1970–95	1970–80
Results and Variables	SDP	LDP	MDP		BP	BPO
A. Regression results						
α Constant	**8.12**	**5.44**	**7.00**	[ordered-logit model[a]]	**7.61**	**10.15**
	(7.83)	(6.81)	(8.19)		(31.18)	(14.30)
$\beta_1(WK/TA)$	1.09	0.19	–0.30	**–2.25**	**0.60**	0.07
	(1.54)	(0.36)	(0.56)	(6.32)	(3.51)	(0.15)
$\beta_2(RE/TA)$	0.05	**1.02**	**1.07**	**3.59**	0.09	0.49
	(0.09)	(2.92)	(3.42)	(3.59)	(1.05)	(1.78)
$\beta_3(EBIT/TA)$	**5.39**	**2.81**	–1.31	**4.87**	**2.83**	3.55
	(3.51)	(2.41)	(1.12)	(7.88)	(10.75)	(4.29)
$\beta_4(ME/BL)$	**1.44**	**0.96**	**0.42**	**0.97**	**0.88**	**0.92**
	(9.86)	(9.51)	(3.85)	(14.38)	(23.28)	(8.95)
$\beta_5(Size)$	**0.53**	**0.50**	**0.35**	**0.91**	**0.29**	**0.50**
	(5.10)	(6.08)	(4.19)	(13.36)	(13.29)	(7.65)
$\beta_6(Age)$	**0.16**	**0.13**	0.05	**0.10**	—	—
	(4.14)	(4.69)	(1.80)	(6.92)		
Pseudo R^2	0.381	0.288	0.081	0.217	0.195	0.162
No. of observations	31,829	24,656	24,656	28,333	111,510	33,242
No. of default observations	278	1,677	343	—	720	119
B. Relative weight of model variables						
WK/TA	5.8%	1.2%	3.3%	8.7%	5.7%	0.5%
RE/TA	0.5	12.0	21.3	25.0	2.0	6.7
EBIT/TA	12.5	8.1	6.3	8.2	18.8	15.1
ME/BL	40.7	33.5	24.6	20.1	47.7	42.4
Size	24.0	28.2	33.4	30.0	25.7	35.3
Age	16.5	17.1	11.2	7.9	—	—

NA = not available.

Notes: *WK* = net working capital; *TA* = total assets; *RE* = retained earnings; *EBIT* = earnings before interest and taxes; *ME* = market value of equity; *BL* = book value of total liabilities. *Size* is the log transformation of total liabilities normalized by the total value of the U.S. equity market; *Age* is the number of years since the company was first rated by an agency. The three default prediction models are given for time horizons of one year, six years, and an annual period starting five years in the future. The standard errors in the logit regression estimation are a generalized version of the Huber and White standard errors. In a standard logit model setting, the error terms, ε_i, are assumed to be identically and independently distributed [$\text{var}_{(\varepsilon i)} = \sigma^2$, $\text{cov}_{(\varepsilon i,\, \varepsilon j)} = 0$ if $i \neq j$]. In reality, these conditions are violated. To obtain the correct statistics, Huber–White standard errors are used to relax the assumption of homoscedasticity. A generalization of Huber–White standard errors (Rogers 1993) also relaxes the assumption of independence among all observations. Instead, only independence between observations of different companies is assumed. "Pseudo R^2" is a measure of the goodness of the fit.

[a]Because of space considerations, the 15 boundary parameters in the ordered-logit model are not shown.

The time horizon has a significant impact on the relative weight of the model variables. Especially for the *RE/TA*, *ME/BL*, and *Size* variables, a clear shift in relative weight is observed in the SDP, LDP, MDP, and AR models, in that sequence. Not surprisingly, the short-term-oriented SDP model depends heavily on the variables that most closely follow the credit/business cycle, such as *ME/BL*, whereas the AR model and MDP model place relatively more weight on variables that are less sensitive to credit cycles, such as *RE/TA* and *Size*. (*RE/TA* is a measure of long-term historical performance and is relatively insensitive to short-term fluctuations in performance.)

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

Conversion of Credit Scores to Credit Model Ratings. We take SDP scores, LDP scores, AR scores, and BPO scores as representatives of various point-in-time perspectives. We converted these credit (model) scores to credit-model ratings (CM ratings) equivalent to agency ratings. We could then compare the dynamics of agency ratings unambiguously with the dynamics of credit scores. For the conversion, we ranked all companies at the end of each quarter by their credit scores. On the basis of this ranking, we assigned 16 credit model ratings (AAA/AA+, . . ., CCC/CC) that are equivalent to agency ratings to individual companies. So, at the end of each quarter, the number of companies in each agency rating class, N, equaled the number of companies in the equivalent CM rating class. The 16 rating classes are defined on a "notch" scale; rating classes are separated from their neighbors by one notch step.

Influence of Migration Policy on Rating Dynamics. SDP, LDP, AR, and BPO ratings are point-in-time ratings, in the sense that they reflect the most recently available credit-quality information without any delay from a migration policy. According to this definition, AR ratings are, in fact, point-in-time measures of the agencies' long-term view of default risk (after filtering out the cyclical component). AR ratings represent only one aspect of the through-the-cycle methodology—namely, the focus on the permanent credit risk component. The second aspect of the through-the-cycle methodology, the prudent migration policy, is not picked up when estimating the AR model by the static ordered-logit regression methodology.[9] To study the influence of the migration policy on rating dynamics, we adjusted the AR score following a particular migration policy model.

We modeled the migration policy of agencies by two parameters: a threshold parameter, TH, and an adjustment fraction, AF (for more details of this migration model, see Altman and Rijken 2004).[10] The threshold parameter specifies the size of a credit-quality interval ($-TH$, $+TH$) in which credit quality is allowed to fluctuate without triggering a rating migration.[11] This threshold prevents small fluctuations in credit quality from triggering a rating migration and thus reduces the probability of a rating migration. If a rating migration is triggered, ratings are adjusted by only a fraction, AF, to the actual credit-quality level. Adjustment fraction AF represents the partial adjustment of agency ratings. Partial adjustment (i.e., spreading of the target rating adjustment over time) is responsible for the observed drift in agency ratings.

After we adjusted the AR scores, we converted the adjusted AR scores to adjusted AR ratings. By varying threshold parameter TH and adjustment fraction AF, we could vary the migration probabilities and drift properties of adjusted AR ratings. We found a best match with the dynamics of agency ratings for a threshold of 1.8 notch steps and an adjustment fraction of 2/3. Allowing for different adjustment fractions on the upgrade side and the downgrade side reveals an adjustment fraction of 0.6 on the upgrade side and 0.7 on the downgrade side. Apparently, the agencies' migration policy is slightly more conservative on the upgrade side.

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

Computation of ARS ratings in a discrete-time setting with quarterly periods is one way to model agency rating dynamics. Lando and Skødeberg (2002) argued that modeling rating migrations in a continuous-time framework provides a better grip on rare migration events. The capture of rare events was not essential, however, in our study. Another alternative is to model the rating-migration process with probabilities of default, taken from a Merton model—for example, the expected default frequency (EDF) scores in the KMV model—instead of credit scores (see, for example, Das, Fan, and Geng 2002). We chose to use the credit scores defined previously in this section for two reasons. First, the volatility of these credit scores is significantly lower than probabilities of default based on price changes—for example, the EDF scores (see Kealhofer, Kwok, and Weng 1998). Second, ratings based on these credit scores predict default events better than agency ratings do for short prediction time horizons (which will be shown later). Ultimately, the best performing default prediction model is the best benchmark.

Benchmark Setup

We studied the influence of the through-the-cycle methodology on rating stability, rating timeliness, and default prediction by comparing the dynamic properties of agency ratings with various point-in-time CM ratings and ARS ratings. **Figure 1** shows how to position various CM ratings on the range between the agencies' through-the-cycle perspective and the one-year point-in-time perspective. SDP ratings and BPO ratings represent a one-year point-in-time perspective; ARS ratings and, of course, agency ratings represent the through-the-cycle perspective.

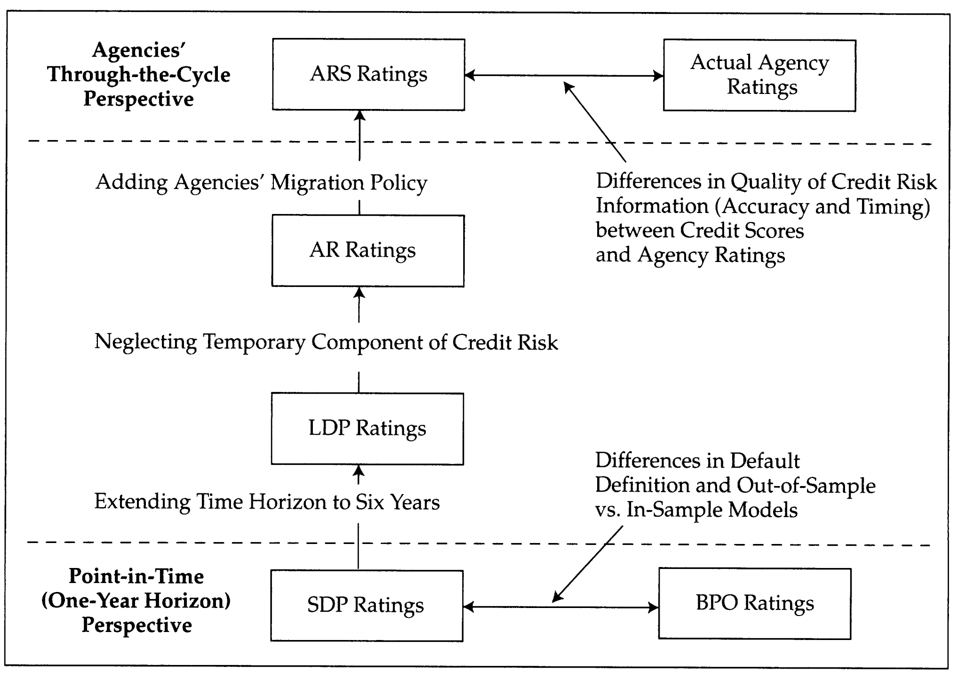
We studied the impact of the first aspect of the through-the-cycle methodology—focus on the permanent credit risk component—by noting the differences between SDP ratings (one-year horizon), LDP ratings (six-year horizon), and AR ratings. Differences between SDP ratings and LDP ratings illustrate the effect of extending the time horizon from one to six years, whereas differences between LDP ratings (sensitive to the temporary and permanent credit risk components) and AR ratings (sensitive only to the permanent credit risk component) illustrate the effect of neglecting the temporary credit risk component. A comparison of SDP ratings and BPO ratings provides a check on whether default definition and overlap in model estimation period and analysis period (in-sample versus out-of-sample analysis) affect the conclusions of the benchmark study.

Differences between AR ratings and ARS ratings quantify the influence of the second aspect of the through-the-cycle methodology—the prudent migration policy—on rating dynamics. ARS ratings reflect both aspects of the agencies' methodology, whereas AR ratings reflect only one aspect—namely, the point-in-time perspective that agencies have on the permanent credit risk component.

Because rating dynamics of ARS ratings and agency ratings are matched, differences between these ratings in timeliness and default prediction performance can be ascribed only to differences in the quality of *credit risk information* underlying these ratings. The quality of any credit risk information has two dimensions: accuracy and timing. Agency ratings are based on in-depth analysis of public and private information available to agency analysts, whereas credit scores are based on a limited set of six model variables that are available to the public. Therefore, agency ratings should have an accuracy advantage over credit-model ratings. The question is to what extent.

Because agency analysts have access to private information, agency ratings are expected also to have a timing advantage after the through-the-cycle effects are controlled for. This timing advantage

Figure 1. The Benchmark Study Setup



may be offset, however, by the time required by agencies to review their ratings after new information becomes available. Perhaps because of capacity restrictions, agency analysts do not update their credit risk analyses on a continuous basis. So, *a priori*, whether agency ratings have a timing advantage over ARS ratings is not clear. (Note that we are distinguishing this potential information timing advantage from the timeliness *disadvantages* introduced by the through-the-cycle methodology.)

Rating Stability and Rating Drift

In this section, we report the rating-migration probabilities and the mean migration figures—unconditional and conditional on a rating-migration event.

Rating-Migration Probabilities. We quantified rating stability by calculating rating-migration probabilities. Panel A of **Table 2** reports migration probabilities in a quarterly period for actual agency ratings and for the various credit-model ratings in the 1981–2001 period. For agency ratings and ARS ratings, the migration probability is 5.6 percent and 5.4 percent, respectively.[12] Elimination of the prudent migration policy (difference between ARS and AR ratings) increases the migration probability to 27.0 percent. Reducing the time horizon to one year and giving full weight to the short-term default-risk fluctuations (difference between AR and SDP ratings) increases the migration probability by a further 12.6 percentage points to 39.6 percent.[13] The prudent migration policy has more impact on migration probability than does the filtering out of the temporary credit risk component (the other aspect of through-the-cycle methodology).

The rating-migration probabilities of LDP ratings are between the rating-migration probabilities of AR ratings and SDP ratings. Because the LDP model weights the temporary component only moderately, this finding suggests that agencies put weight only on the permanent credit risk component. Like the results in Table 1, this empirical result supports the exclusive focus of agencies on the permanent credit risk component and their disregard of credit-quality cycles.

Unconditional Mean Rating Migration. Panel B of Table 2 reports the mean rating-migration data for upgrades and downgrades in the 1981–2001 period.[14] To calculate the mean migration values, we assigned a numerical scale to the ordinary notch scale of agency ratings and equivalent CM ratings. In this scale, D = 0, CCC/CC/C = 1, B– = 2, B = 3, . . ., AA– = 14, AA = 15, and AAA/AA+ = 16. This scale is an arbitrary but intuitive choice and is commonly found in the mapping of banks' internal ratings to agency ratings.

Without a prudent migration policy, as Panel B shows for the AR ratings, the average rating upgrade is 1.06 notch steps and the average rating downgrade is –1.14 notch steps. These values are slightly higher than a single notch step, which is to be expected when the necessary rating changes are made immediately and in full. The threshold of 1.8 notch steps, in combination with the moderating influence of the adjustment fraction, increases the average migration step for ARS ratings to 1.4 on the upgrade side and –1.5 on the downgrade side.

The unconditional mean rating migration, $\Delta R(u)$, in *each* quarter is about –0.02 for agency ratings and –0.01 for the CM ratings. Technically, this unconditional rating migration is equal to the difference in rating levels between companies entering the dataset and companies exiting the dataset divided by the number of quarters of

Table 2. Unconditional Rating Migration
(standard errors in parentheses)

Migration Event	Agency Ratings	ARS Ratings	AR Ratings	LDP Ratings	SDP Ratings	BPO Ratings
A. Probability of a rating-migration event in a quarterly period						
No migration	94.4%	94.6%	73.0%	67.0%	60.4%	62.7%
Upgrade	2.3	2.2	13.1	16.3	19.2	16.5
Downgrade	3.3	3.2	13.9	16.7	20.4	20.8
B. Mean rating migration for upgrades and downgrades						
Upgrade	1.36	1.44	1.06	1.09	1.17	1.14
	(0.03)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)
Downgrade	–1.56	–1.51	–1.14	–1.17	–1.26	–1.24
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)

Note: Mean rating-migration values are in notch steps.

unbroken stay in the dataset (on average, equal to 27 quarters). Defaulting companies are mainly responsible for this unconditional downward drift in ratings.

Rating-Drift Properties. Conditional on an upgrade, downgrade, or no migration in Q0, we computed the mean rating-migration values—$\Delta R(+)$, $\Delta R(-)$, and $\Delta R(0)$—for subsequent quarters: Q1 (the quarter immediately following Q0), Q2 (the quarter following Q1), and Q2 through Q8.[15] These conditional rating-migration values were corrected for unconditional values $\Delta R(u)$—that is, $\Delta R(+) - \Delta R(u)$ became $\Delta R(+)$ and $\Delta R(-) - \Delta R(u)$ became $\Delta R(-)$. Only the corrected $\Delta R(+)$ and $\Delta R(-)$ are of interest because $\Delta R(0) \approx \Delta R(u)$.

Table 3 reports the corrected $\Delta R(+)$ and $\Delta R(-)$ for Q1 (Panel A), Q2 (Panel B), and Q2–Q8 (Panel C). For point-in-time CM ratings, a short-term reversal (a change in sign compared with the migration direction of the agency and ARS rating changes) shows up. For these ratings, the average migration in Q1 is about +0.15 following a downgrade and about –0.2 following an upgrade. This rating reversal effect disappears in Q2. In Q2 through Q8, rating drift is absent, which suggests random behavior in point-in-time corporate credit quality beyond one quarterly cycle. The origin of the short-term reversal in Q1 needs further study.

Given the random behavior of the underlying credit risk fundamentals, rating drift is to be expected when ratings are partially adjusted to actual credit quality through the *AF* of 2/3 (see the section "Influence of Migration Policy on Rating Dynamics"). In the eight quarters after a downgrade or upgrade, agency ratings and ARS ratings drift with a steady rate to about –0.3 on the downgrade side and +0.3 on the upgrade side. Drift on both sides is expected because the underlying source of rating drift is effective in both directions. The asymmetry in rating drift as reported by Altman and Kao (1992) disappears when drift values $\Delta R(-)$ and $\Delta R(+)$ are corrected for the unconditional rating downward drift, $\Delta R(u)$.

Rating Timeliness

We investigated average changes in credit-model ratings surrounding an agency rating upgrade or downgrade. We considered the timing of conditional changes in CM ratings before the agency migration event to be an indication of the timeliness of agency ratings.

We recomputed the values for conditional migration, $\Delta R(+)$ and $\Delta R(-)$, by following exactly the same procedure as described in the previous section, except for one difference: All $\Delta R(+)$ and $\Delta R(-)$ values are conditional on an agency rating-migration event,

Table 3. Rating Drift: Mean Rating Migration Conditional on a Rating-Migration Event in Q0
(standard errors in parentheses)

Migration Event in Q0	Agency Ratings	ARS Ratings	AR Ratings	LDP Ratings	SDP Ratings	BPO Ratings
A. Mean rating migration in Q1 (first quarter following Q0)						
Upgrade	0.03	0.02	–0.22	–0.19	–0.18	–0.20
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Downgrade	–0.08	–0.10	0.18	0.17	0.12	0.12
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
B. Mean rating migration in Q2 (second quarter following Q0)						
Upgrade	0.04	0.08	0.00	0.01	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Downgrade	–0.04	–0.10	0.00	–0.02	–0.04	–0.05
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
C. Mean rating migration in periods Q2–Q8						
Upgrade	0.27	0.29	0.06	0.05	–0.04	–0.03
	(0.04)	(0.05)	(0.02)	(0.03)	(0.03)	(0.03)
Downgrade	–0.30	–0.27	0.01	–0.02	–0.06	–0.02
	(0.05)	(0.04)	(0.02)	(0.03)	(0.03)	(0.03)

Notes: Mean rating-migration values are in notch steps. Conditional mean rating-migration values $\Delta R(+)$ and $\Delta R(-)$ are corrected for the unconditional mean rating-migration values, $\Delta R(u)$.

ΔN, in Q0 instead of ΔR itself. The cumulative rating change, $\Delta R^C{}_t$, conditional on an agency rating-migration event in Q0 is given by

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

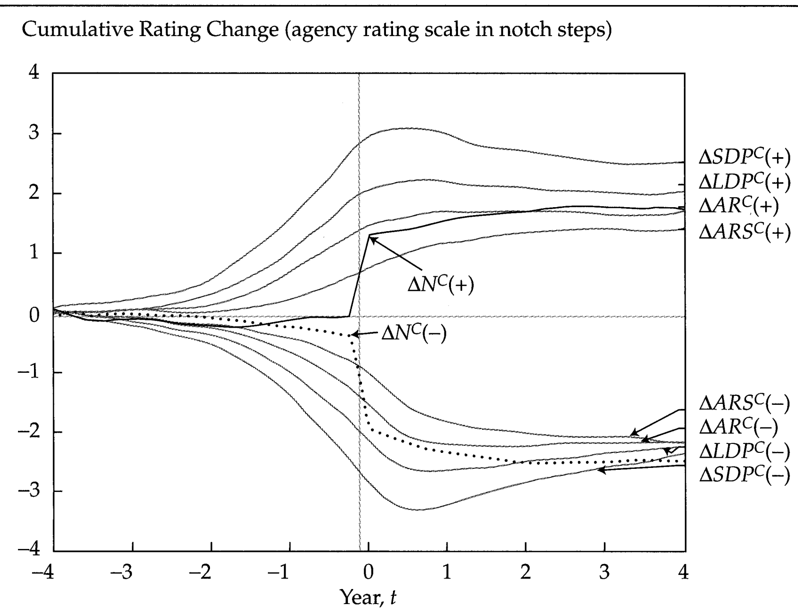
where v is positive (+) or negative (–). The starting time for the accumulation of rating migrations is $t = -4.25$.

Figure 2 shows cumulative rating changes ΔR^C as a function of time t conditional on an agency rating upgrade or downgrade in quarter (–0.25,0). The cumulative rating changes are plotted for agency ratings, ΔN^C; for ARS ratings, ΔARS^C; for AR ratings, ΔAR^C; for LDP ratings, ΔLDP^C; and for SDP ratings, ΔSDP^C. To allow a comparison of the data in terms of agency rating notch steps, the ΔR^C of all CM ratings were scaled by a factor of $1/\kappa_R$, where κ_R equals the slope in the regression equation: $CM = \kappa_R N$ + Constant. For the ARS, AR, LDP, SDP, and BPO ratings, κ_R is, respectively, 0.856, 0.841, 0.807, 0.744, and 0.756. Because of the strong variation of CM ratings within an agency rating class, with a "standard error" up to 3 notch steps, and the boundaries of the discrete rating scale, migrations in agency rating do not always show up in CM rating changes, even if all the agency rating migrations are picked up by changes in credit scores.[16]

On average, point-in-time CM rating changes clearly anticipate agency rating migrations (see Figure 2). Among the CM ratings, SDP ratings anticipate an agency rating-migration event most strongly. In the two-year period surrounding the agency rating-migration event, ΔSDP^C and (to a lesser extent) ΔLDP^C show "overshooting" behavior. Just after the agency migration date, ΔSDP^C clearly exceeds the change in the permanent credit risk component, as proxied by ΔAR^C and ΔN^C at t = 4. This overshooting behavior results from the sensitivity of SDP ratings to changes in the temporary credit risk component. As expected, the temporary changes in SDP ratings reverse; the overshooting behavior is less pronounced for LDP ratings because they are only moderately sensitive to the temporary credit risk component. The absence of overshooting in AR ratings is consistent with the disregard of short-term fluctuations in default risk by agencies.

The quantification of rating timeliness was based on the cumulative rating changes, ΔR^C, starting at t = –2.25 and ending at t = +2.00. The choice of this time interval is arbitrary, but most of the rating changes do take place in this period.[17] **Table 4** reports the total cumulative rating change, $\Delta AR^C{}_{TOT}$, in this time interval and the percentages of the cumulative rating change that happen in the two years before Q0 (–2.25, –0.25), in Q0 (–0.25, 0), and in the two years after Q0 (0, 2). ARS ratings do not show a timing advantage. In this case, a comparable fraction of total cumulative rating change happens before and after the

Figure 2. Cumulative Changes in Credit-Model Ratings Conditional on an Agency Rating-Migration Event in Quarter (–0.25, 0)



Note: Starting time is Year –4.25.

Table 4. Timeliness of Agency Ratings Relative to Credit-Model Ratings (standard errors in parentheses)

Measure	Agency Ratings	ARS Ratings	AR Ratings	LDP Ratings	SDP Ratings	BPO Ratings
A. Conditional on an agency rating upgrade						
ΔR^C_{TOT}	1.93	1.29	1.60	1.87	2.33	2.18
	(0.07)	(0.07)	(0.11)	(0.13)	(0.17)	(0.16)
$\%\Delta R^C_{TOT}(-2,-025)$[a]	10%	46%	77%	89%	98%	102%
$\%\Delta R^C_{TOT}(-0.25,0)$[a]	74%	15%	14%	11%	14%	16%
$\%\Delta R^C_{TOT}(0,2)$[a]	16%	40%	9%	0%	–12%	–18%
ΔR^C_{max}	1.93	1.29	1.61	2.02	2.68	2.44
	(0.07)	(0.07)	(0.10)	(0.11)	(0.14)	(0.13)
Timing: $1/2\ \Delta R^C_{max}$ reached	–0.11	–0.14	–0.72	–0.82	–0.90	–0.85
	(0.01)	(0.08)	(0.09)	(0.07)	(0.08)	(0.08)
Timeliness (credit-model rating vs. agency rating)	—	–0.03	–0.61	–0.71	–0.79	–0.74
B. Conditional on an agency rating downgrade						
ΔR^C_{TOT}	–2.49	–1.85	–2.01	–2.17	–2.47	–2.44
	(0.07)	(0.07)	(0.10)	(0.11)	(0.15)	(0.13)
$\%\Delta R^C_{TOT}(-2,-025)$[a]	16%	42%	60%	77%	92%	93%
$\%\Delta R^C_{TOT}(-0.25,0)$[a]	68%	15%	20%	20%	23%	22%
$\%\Delta R^C_{TOT}(0,2)$[a]	16%	43%	20%	3%	–15%	–15%
ΔR^C_{max}	–2.49	–1.85	–2.01	–2.33	–2.91	–2.80
	(0.07)	(0.07)	(0.09)	(0.09)	(0.11)	(0.11)
Timing: $1/2\ \Delta R^C_{max}$	–0.11	0.07	–0.25	–0.53	–0.67	–0.64
	(0.01)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)
Timeliness (credit-model rating vs. agency rating)	—	0.18	–0.14	–0.42	–0.56	–0.53

Note: ΔR^C values are in notch steps; timing and timeliness values are in years.

[a] The percentages of ΔR^C_{TOT} happening in the subperiods (–2,–0.25), (–0.25,0), and (0,2).

agency rating-migration event. However, for the point-in-time CM ratings, the majority of the conditional rating changes happen before Q0. After Q0, the cumulative rating changes are relatively low; for SDP^C and BPO^C, they are even negative—because of the overshooting behavior.

The timeliness of agency ratings relative to CM ratings is reported in the bottom three rows of Panel A and Panel B in Table 4. Timeliness in this case was defined as follows. The maximum cumulative rating change in interval (–2, 2), ΔR^C_{max}, is a proxy for the total change in both the permanent and the temporary components of default risk, conditional on an agency rating-migration event. The time when half of ΔR^C_{max} is reached is a proxy for the average timing of these permanent and temporary changes, conditional on the agency rating-migration event in interval (–0.25, 0). The difference between these two time moments is an indication of the timeliness of agency ratings relative to the timeliness of CM ratings. As Table 4 shows, the timeliness disadvantage of agency ratings compared with LDP, SDP, and BPO ratings is about 0.75 year on the upgrade side and 0.50 year on the downgrade side. This finding is consistent with the evidence that agencies are more conservative on the upgrade side.

After the through-the-cycle effects are controlled for, the difference in timing of agency rating migrations and corresponding changes in ARS ratings (based on public information) is negligible on the upgrade side (–0.03 year) whereas agency ratings are slightly more responsive than ARS ratings on the downgrade side (+0.18 year). On the upgrade side, agency ratings have no timing advantage over ARS ratings. Perhaps the expected timing advantage from the agencies' access to private information is offset by limitations in the processing of new information by agency analysts. The information timing advantage of 0.18 year on the downgrade side could be explained by the idea that companies with a potential downgrade are more closely watched by agency analysts.

On the downgrade side, the two aspects of the through-the-cycle methodology affect rating timeliness equally. Because of insensitivity to the temporary credit risk component, AR ratings are delayed by about 0.4 year in comparison with LDP, SDP, and BPO ratings; ARS ratings are delayed by another 0.3 year as a result of the prudent migration policy. On the upgrade side, the prudent migration policy has most impact on the timeliness of agency ratings. Perhaps changes in credit quality have a more permanent and less abrupt character on the upgrade side than on the downgrade side, which makes the time horizon in default risk less relevant on the upgrade side.

Default Prediction

In this section, we discuss the default probability of a company in a particular rating class and the ability of agency ratings and CM ratings to distinguish defaulters from nondefaulters.

Cumulative Default Rates. The probability of a company in a particular rating class defaulting *within* T years was measured from historical data by

$$\text{Cumulative default rate}(R,T) = \frac{\sum_{t=1981}^{2002-T}\sum_{i=1}^{N_{R,T,t}} D_{R,T,i,t}}{\sum_{t=1981}^{2002-T}\sum_{i=1}^{N_{R,T,t}}\left(S_{R,T,i,t}+D_{R,T,i,t}\right)}, \quad (5)$$

where $D_{R,T,i,t}$ and $S_{R,T,i,t}$ are binary variables identifying default observations ($p_{i,t} = 0$) and survival observations ($p_{i,t} = 1$) in rating class R with time horizon T. We characterized companies as surviving companies if they survived beyond T years. We characterized companies as defaulting companies if they defaulted within T years. Companies exiting the dataset by means other than default within T years (for example, through mergers or migration to a non-rated status) were excluded from the default-rate calculation. $N_{R,T,t}$ is the total number of defaulting and surviving observations in rating class R in year t with time horizon T.

Table 5 shows the *three-year* cumulative default rates for all 16 classes of agency ratings and credit-model ratings. In general, point-in-time CM ratings (the AR, LDP, SDP, and BPO ratings) perform slightly better than agency ratings in the non-investment-grade range (below BB+). In this range, the default rates of CM ratings are higher in the bottom rating classes (CCC/CC and B–) and lower in the B+, BB–, and BB rating classes. So, compared with agency ratings, Type I and Type II errors are lower for CM ratings if companies in the bottom rating classes are classified as defaulters and companies in the higher rating classes are classified as nondefaulters. In the investment-grade range (above BB+), the number of Type I errors is lower for agency ratings. For a three-year horizon, as Table 5 shows, the average investment-grade default rate is 0.31 percent for agency ratings and 0.41–0.58 percent for CM ratings.

Accuracy Ratios. Construction of a "cumulative accuracy profile" curve is a well-accepted methodology for measuring the overall default prediction performance of a rating scale, weighting Type I and Type II errors equally in distinguishing defaulters and nondefaulters. We obtained cumulative accuracy profile (CAP) curves by plotting for each rating class R the proportion of default observations in the *same and lower* rating classes, $F_D(R)$, against the proportion of all surviving and defaulting observations in the same and lower rating classes, $F_A(R)$:

$$F_A(R,T) = \frac{\sum_{C=1}^{R}\sum_{t=1981}^{2002-T}\sum_{i=1}^{N_{C,T,t}}\left(S_{C,T,i,t}+D_{C,T,i,t}\right)}{N_A(T)}, \quad (6)$$

where $F_A(0,T)$ is zero and $N_A(T)$ is the total number of default and survival observations with time horizon T in the dataset. The definition of $F_D(R,T)$ is similar but involves summing only the number of default observations up to rating R [equivalent to a replacement of $(S_{C,T,i,t} + D_{C,T,i,t})$ by $D_{C,T,i,}$ in Equation 6].

The higher the proportion of default events happening in the lower-grade classes—in other words, the higher the surface below the CAP curve—the better the performance of the rating scale. The accuracy ratio (ACR) measures the surface below the CAP curve relative to the surface below the CAP curve for a random rating scale with no prediction power (= 1/2). Based on cumulative default rates, ACR is given by

$$ACR(T) = \frac{\sum_{R=1}^{16}\left\langle\left[F_A(R,T)-F_A(R-1,T)\right]\left\{F_D(R-1,T)+(1/2)\left[F_D(R,T)-F_D(R-1,T)\right]\right\}\right\rangle-(1/2)}{1/2}. \quad (7)$$

Table 5. Cumulative Default Rates (Probabilities of Default) for a Three-Year Horizon (standard errors in parentheses)

		In-Sample Results					
Rating Class (or equivalent)	Agency Ratings	ARS Ratings	AR Ratings	LDP Ratings	SDP Ratings	BPO Ratings	No. of Obs.
CCC/CC	53.7%	56.1%	61.6%	71.9%	69.1%	73.5%	708
B–	39.3	37.0	41.6	41.0	41.6	40.4	1,098
B	32.1	26.7	31.1	30.9	31.3	32.3	2,192
B+	17.1	15.9	16.4	15.8	15.3	13.8	5,456
BB–	11.9	8.8	9.1	7.9	7.8	7.9	4,003
BB	5.3	6.0	4.8	5.0	4.7	4.7	3,163
BB+	2.7	2.9	3.6	3.4	2.8	2.9	2,103
BBB–	1.9	1.9	2.3	2.0	2.2	2.7	2,949
BBB	0.5	0.9	0.9	1.6	1.7	1.9	3,671
BBB+	1.2	2.0	1.5	1.0	1.4	1.3	3,126
A–	0.4	0.8	0.7	0.4	0.8	1.1	2,948
A	0.3	0.5	0.3	0.4	0.6	0.7	4,628
A+	0.0	0.0	0.0	0.0	0.0	0.0	2,746
AA–	0.0	0.0	0.0	0.0	0.0	0.0	1,550
AA	0.0	0.0	0.0	0.0	0.0	0.0	2,351
AAA/AA+	0.0	0.0	0.0	0.0	0.0	0.0	1,332
Default rate for investment-grade companies	0.31%	0.46%	0.42%	0.41%	0.51%	0.58%	
ACR for a pooled sample	73.0%	71.3%	73.4%	74.3%	73.0%	72.6%	
	(2.0%)	(2.0%)	(2.0%)	(2.0%)	(2.0%)	(2.0%)	
Mean rating of default observations	3.72	3.73	3.67	3.58	3.65	3.68	
	(1.90)	(2.11)	(2.07)	(2.09)	(2.22)	(2.31)	

Notes: In the bottom three rows, the default prediction performances of the various rating scales are compared (1) by the average default rate in the investment-grade range, which measures the Type I error in this range, (2) by the accuracy ratio, ACR, which weights Type I and Type II errors equally, and (3) by the mean rating of companies defaulting within three years. The analysis includes 2,214 default observations related to a total of 151 unique default events.

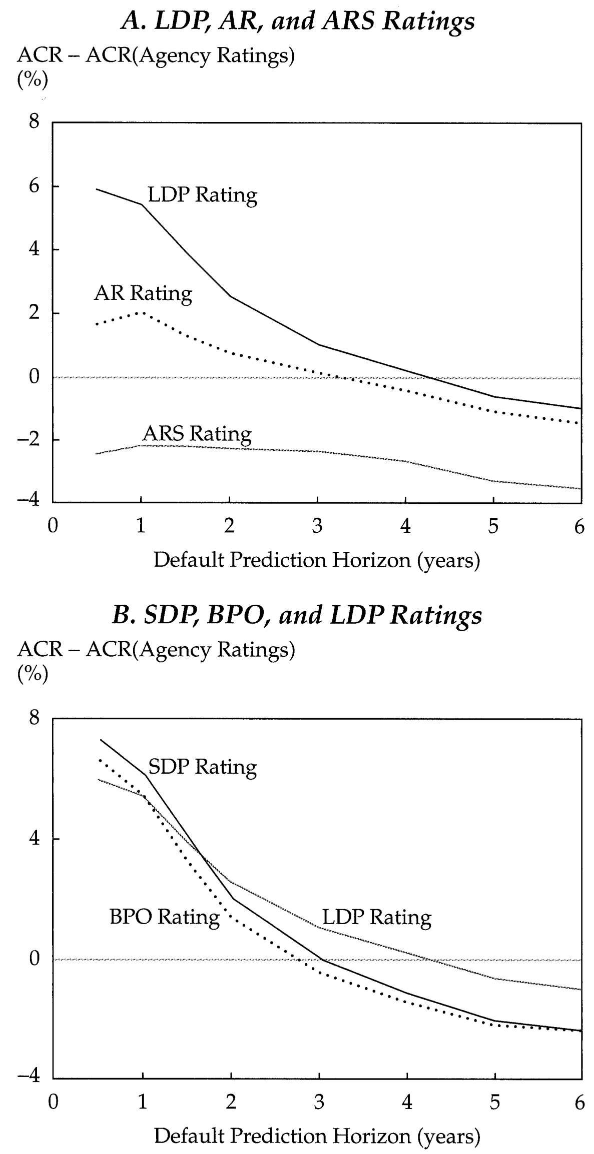
The ACR varies between 0 percent (random scale) and 100 percent (perfect prediction scale).[18] The standard error in ACR is 1.5 percent, 2 percent, and 2.5 percent for time horizons of, respectively, one year, three years, and six years.[19]

In general, we found that point-in-time CM ratings performed better than agency ratings up to a time horizon of three years. For a three-year horizon, as Table 5 shows, the accuracy ratios did not vary much between agency ratings and CM ratings. For *pooled* samples (Equations 6 and 7), the ACR is 73.0 percent for agency ratings and is 71–74 percent for CM ratings. Beyond a time horizon of three years, agency ratings show slightly better default prediction performance.[20]

Differences in accuracy ratios between agency ratings and various CM ratings, $\Delta ACR = ACR(\text{CM ratings}) - ACR(\text{Agency ratings})$, reveal the impact of the through-the-cycle methodology and the agencies' information advantage over credit scores on default prediction performance. **Figure 3** shows ΔACR as a function of time horizon T.[21] For a one-year horizon, the ACR of the agency ratings is 6 percentage points (pps) lower than the ACR of LDP ratings, which include both the permanent and temporary credit risk components. As shown by the difference in ACRs between LDP ratings and AR ratings, filtering out the temporary component reduces the accuracy ratio by 4 pps. The difference in ACRs between AR ratings and ARS ratings indicates that the prudent migration policy reduces the accuracy ratio by another 4 pps.

Differences between ARS ratings and agency ratings arise from quality differences in credit risk information underlying the ratings. An information advantage of 2 pps appears in the ACR of agency ratings compared with ARS ratings. The main conclusion is that the negative impact of the through-the-cycle methodology fully overshadows this 2 pp information advantage of agency

Figure 3. Default Prediction Performance of Credit-Model Ratings Relative to Agency Ratings



Note: Differences in accuracy ratio between agency ratings and the benchmark credit-model ratings.

ratings, resulting in a net ACR disadvantage of about 6 pps. In a comparable analysis, Fons and Viswanathan (2004) found a much smaller disadvantage in the ACR of actual Moody's ratings compared with point-in-time ratings based solely on accounting data.

As expected, for a six-year horizon, the negative impact of the through-the-cycle methodology is much less. At a six-year horizon, the temporary credit risk component evidently has little impact on the accuracy ratio and the prudent migration policy lowers the ACR by only 2 pps. For this time horizon, the information advantage of agency ratings over credit scores, producing +3.5 pps in the ACR, is only partly offset by the through-the-cycle effects, resulting in a net ACR advantage of about 1 pp for agency ratings over LDP ratings.

The superior performance of SDP ratings in the short term indicates that credit-model scores are complementary to agency ratings. This result supports the common practice of using credit scores or EDF scores in addition to agency ratings, not only as a second opinion but also because of the superior timeliness of credit scores.

The estimation period of the credit-scoring models largely overlaps the period used for analyzing default performance, except for the out-of-sample BPO model. The underlying credit-scoring models are robust, however, to dataset choice and default definition and are robust in time (see the section "Estimation Benchmark for Credit-Scoring Models"), so a distinction between in-sample and out-of-sample models is not very relevant. For example, in Panel B of Figure 3, the *ACR* of SDP ratings and the *ACR* of BPO ratings, both proxies for the one-year default probability, practically overlap.

Conclusion

The benchmark study presented here consists of two parts: first, the definition of the benchmark point-in-time ratings based on credit scores and, second, the benchmark study itself. From the definition of the benchmark point-in-time ratings, we conclude the following:

- Agency ratings focus exclusively on the permanent component of credit quality, which confirms the ability of agencies to look "through the cycle."
- An agency rating migration is triggered if the company's actual credit quality—permanent credit-quality component—exceeds a threshold of 1.8 notch steps relative to the average credit-quality level in the company's rating class. If a migration is triggered, agencies only partly adjust their ratings to the actual credit-quality level—60 percent on the upgrade side and 70 percent on the downgrade side.

In the benchmark study, we compared rating properties of agency ratings with those of point-in-time ratings based on a one-year default prediction model, serving as a proxy for the investor's perspective on credit quality. From this comparison, we draw the following conclusions (see **Table 6**):

- The agencies' through-the-cycle methodology lowers the rating-migration probability in a quarterly period by a factor of 7. The agencies' prudent migration policy, not the disregard of the temporary credit risk component, is the most important source of rating stability.

Table 6. Summary of Results

Subject	Rating Stability (reduction in rating-migration probability)	Rating Timeliness (timing of rating migrations): For Upgrades	Rating Timeliness (timing of rating migrations): For Downgrades	Default Prediction Performance (accuracy ratio): One-Year Horizon	Default Prediction Performance (accuracy ratio): Six-Year Horizon
Effects of agencies' through-the-cycle methodology from one-year point-in-time perspective					
Focus on permanent component of default risk	×1.5	−0.18	−0.32	−4%	−0.5%
Prudent migration policy	×5.0	−0.58	−0.42	−4	−2.0
Information advantage of agency ratings over credit scores after through-the-cycle effects controlled for					
Accuracy	—	—	—	} +2	} +3.5
Timing	—	−0.03	+0.18		
Total	×7	−0.79	−0.56	−6%	+1%

Notes: This summary provides the effects of the through-the-cycle methodology from a one-year point-in-time perspective as proxied by SDP ratings and the quantification of the information advantage of agency ratings over credit scores after the through-the-cycle effects are controlled for. The effect on rating stability is given by the reduction in rating-migration probability in a quarterly period for agency ratings compared to SDP ratings (see also Tables 2 and 3). The effect on rating timeliness is given by the delay in timing of agency rating migrations relative to related changes in SDP ratings (see also Figure 2 and Table 4). The effect on default prediction performance is given by the differences in accuracy ratio between SDP ratings and agency ratings for a one-year and a six-year prediction horizon (see also Figure 3). Where possible, differences in rating properties between agency ratings and SDP ratings are broken down into contributions of (1) the two aspects of the through-the-cycle methodology and (2) the accuracy and timing of the credit risk information underlying the ratings (in-depth analysis by agency analysts for agency ratings versus credit scores for SDP ratings).

- The through-the-cycle methodology delays the timing of rating migrations by 0.56 year on the downgrade side and by 0.79 year on the upgrade side. Rating agencies are slightly more responsive on the downgrade side than on the upgrade side, which suggests that rating analysts are more closely watching companies with potential downgrades. Another explanation could be that downturns in credit quality happen more rigorously than upturns.
- The through-the-cycle rating methodology affects the accuracy of default prediction. Both aspects of the through-the-cycle methodology—the disregard of the temporary credit risk component and the prudent migration policy—reduce the accuracy ratio for a one-year horizon by 4 pps.
- When the through-the-cycle effects are controlled for, the quality of the credit risk information underlying the ratings is, as expected, better for agency ratings than for ARS ratings based only on credit scores. This information advantage for agency ratings translates into a 2 pp accuracy ratio advantage. For a one-year horizon, however, this information advantage is fully overshadowed by the negative impact of the through-the-cycle methodology, resulting in a net disadvantage of 6 pps in the accuracy ratio of agency ratings. The two aspects of the through-the-cycle methodology fully offset the agencies' information advantage for time horizons up to three years.

The main purpose of this study was to quantify the effects of the through-the-cycle methodology on rating stability, rating timeliness, and default prediction. It is up to investors themselves to judge whether this balance between rating stability and timeliness best matches their interests.

This article qualifies for 1 PD credit.

Notes

1. This critique focused on the timeliness properties of agency ratings, not on the accuracy of ratings. The survey by the Association for Financial Professionals (www.afponline.org) reported that 83 percent of investors believe that agency ratings accurately reflect the issuer's creditworthiness most of the time.
2. In this article, the rating agencies are Standard & Poor's, Moody's, and Fitch Ratings.
3. The *Age* variable was set to 10 for observations with *Age* values above 10 and for all observations of companies already rated at the start of the dataset in 1981.
4. Companies with a non-rated status are continuously monitored for default events by Standard & Poor's. If they default, their status changes to the "default" status.
5. The time horizon of 6 years in the LDP and MDP models is a result of a compromise. A 6-year horizon is just beyond the length of credit-quality cycles in the temporary credit risk component, typically about 3–4 years. At the same time, the length of the estimation period, 14 years, is kept at an acceptable level. Robustness checks for time horizons between 5 and 8 years produced no substantial changes in the parameters of the MDP model.
6. The bankruptcy dataset covers the 1970–95 period and contains 111,510 survival observations and 720 bankruptcy observations, which are defined in a similar manner as survival and default observations in our Standard & Poor's corporate bond dataset. Only a small fraction of these bankruptcy observations overlap the default observations in the Standard & Poor's corporate bond dataset.
7. These results are not reported here and are available on request.
8. The only significant difference is the absence of a significant parameter for the *Age* variable for non-investment-grade companies.
9. Ratings predicted by the AR model are slightly overstated as a result of the prudent migration policy in the following manner: Temporarily, ratings may be either understated or overstated because of the migration policy. If the numbers of overstated and understated ratings are equal over the sample period—neutralizing the variation in overstated and understated ratings arising from the migration policy and business cycles—the migration policy will not affect the parameter estimates. In that case, it will only widen the distribution of the error term in the logit regression. Historically, however, the number of downgrades is 30 percent higher than the number of upgrades and the average agency rating migration shows a downward trend, so the number of overstated ratings is expected to be slightly higher than the number of understated ratings. When AR scores are converted to AR ratings, however, the shift in AR scores caused by overstatement is not relevant; only the ranking of AR scores matters. Consequently, AR rating dynamics are insensitive to the migration policy.
10. A detailed description of this simulation experiment is beyond the scope of this article.
11. The minimum threshold level imposed by a discrete agency rating scale is a half notch step.
12. These figures are obtained by adding the upgrade and downgrade probabilities.
13. Similar outcomes resulted when the migration policy was eliminated and when the time horizon was reduced in reverse order: first, a change from the agencies long-term view to a one-year horizon by shifting from ARS ratings to "adjusted" SDP ratings (SDP ratings were adjusted in a way similar to the adjustment of AR ratings to ARS ratings following the migration policy with a threshold parameter of 1.8 notch steps and an adjustment fraction of 0.7 on the downgrade side and 0.6 on the upgrade side) and, subsequently, elimination of the prudent migration policy by shifting from "adjusted" SDP ratings to SDP ratings.
14. With quarterly data, the sign of a migration event in Q0 is strictly defined by the net rating migration of all actual rating-migration events in Q0. More than one migration event happening in one quarter is rare, however, so designating the net migration in quarters as single events is appropriate.
15. For each company-quarter observation (company i and Q0), we computed the net rating change, $\Delta R_{-0.25,0}$, in Q0 and in the 32 quarters surrounding Q0 [Q–16, . . ., Q–1, Q0, Q+1, . . . Q+16, $t \in (-4, -3.75, \ldots, 3.75, 4)$]. Because of dataset boundaries—defaulting companies, new companies entering the dataset, and so forth—the time series is not complete for 50 percent of the 40,440 company-quarter observations. The mean $\Delta R_{t-0.25,t}$ for all available observations is the *unconditional* average rating migration, $\Delta R(u)_{t-0.25,t}$. In addition, we calculated the *conditional* average rating migration, $\Delta R_{t-0.25,t}$, for observations with an upgrade in Q0 $[\Delta R(+)_{t-0.25,t}]$, for observations with a downgrade in Q0 $[\Delta R(-)_{t-0.25,t}]$, and for observations with a zero migration in Q0 $[\Delta R(0)_{t-0.25,t}]$.
16. A large fraction of agency rating migrations match changes in ARS ratings. After scaling ΔARS^C by $1/\kappa_R$, ΔN_C and ΔARS^C converge in the years after the agency migration event. At $t = 4$, $\Delta ARS^C(+)/\kappa_R$, $\Delta N^C(+)$, $\Delta ARS^C(-)/\kappa_R$, and $\Delta N^C(-)$ become, respectively, 1.40, 1.72, –2.16, and –2.48 notch steps. So, on the upgrade side, 19 percent of the agency rating migrations are not picked up by changes in ARS ratings, and on the downgrade side, only 11 percent are missing.
17. Longer time intervals do not change the results substantially and are at the cost of statistical significance.
18. An alternative to the ACR methodology is to measure the average rating of companies defaulting within T years. This average rating methodology weights Type I and II errors proportionally to numerical rating scale R, whereas the ACR methodology weights these errors proportionally to $F_A(R)$. The lower the average rating figure is, the better ratings anticipate a possible default event. The average agency rating is 3.72 for companies defaulting within three years, and the average varies between 3.58 and 3.73 for CM ratings (see Table 5). As with the ACR methodology, differences in default prediction performance between agency ratings and CM ratings are small.
19. The stochastic defaulting process can be modeled by the exponential distribution function, $\alpha[\exp(-\alpha F_A)]$. With this distribution function, the CAP curve can be modeled by $1 - \exp(-\alpha F_A)$ with $F_A < 1$. The surface below the CAP curve is $1 - 1/\alpha$ when approximating $\exp(-\alpha) \approx 0$. In that case, ACR equals $1 - 2/\alpha$. In a sampling experiment with n defaulting events, the average F_A for default events is $1/\alpha$ and $\text{var}(F_A)$ is $1/(n\alpha^2)$. In that case, the standard error in ACR is $2/\alpha\sqrt{n}$. For a time horizon of three years, a best fit with the actual CAP curve is obtained for $\alpha = 8.5$, so the standard error in ACR(3) is 0.020 ($n = 151$). The standard error in ACR(6) is slightly higher: 0.025 ($n = 130$ and $\alpha = 7$); the standard error in ACR(1) is 0.015.
20. "CreditWatch" and "Rating Outlook" information has not been included in this analysis. Hamilton and Cantor (2004) conducted some initial tests that show that the accuracy ratio of agency ratings improves significantly when "Outlook" and "CreditWatch" information is added.

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

References

Altman, E.I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance*, vol. 23, no. 4 (September):589–609.

Altman, E.I., and D.L. Kao. 1992. "The Implications of Corporate Bond Ratings Drift." *Financial Analysts Journal*, vol. 48, no. 3 (May/June):64–75.

Altman, E.I., and H.A. Rijken. 2004. "How Rating Agencies Achieve Rating Stability." *Journal of Banking & Finance*, vol. 28, no. 11 (November):2679–2714.

Baker, H.K., and S.A. Mansi. 2002. "Assessing Credit Agencies by Bond Issuers and Institutional Investors." *Journal of Business Finance & Accounting*, vol. 29, nos. 9–10 (November/December):1367–98.

Basel Committee on Banking Supervision. 2000. "Range of Practice in Banks' Internal Rating Systems." Discussion article.

Cantor, R. 2001. "Moody's Investors Service Response to the Consultative Article Issued by the Basel Committee on Bank Supervision: 'A New Capital Adequacy Framework'." *Journal of Banking & Finance*, vol. 25, no. 1 (January):171–185.

Cantor, R., and C. Mann. 2003. "Are Corporate Bond Ratings Procyclical?" Special Comment, Moody's Investors Service (October).

Das, S.R., R. Fan, and G. Geng. 2002. "Bayesian Migration in Credit Ratings Based on Probabilities of Default." *Journal of Fixed Income*, vol. 12, no. 3 (December):17–23.

Ellis, D. 1998. "Different Sides of the Same Story: Investors' and Issuers' Views of Rating Agencies." *Journal of Fixed Income*, vol. 7, no. 4 (March):35–45.

Fons, J.S., and J. Viswanathan. 2004. "A User's Guide to Moody's Default Predictor Model: An Accounting Ration Approach." Special Comment, Moody's Investors Service (December).

Fons, J.S., R. Cantor, and C. Mahoney. 2002. "Understanding Moody's Corporate Bond Ratings and Rating Process." Special Comment, Moody's Investors Service (May).

Hamilton, D., and R. Cantor. 2004. "Rating Transition and Default Rates Conditioned on Outlooks." *Journal of Fixed Income*, vol. 14, no. 2 (September):54–70.

Kealhofer, S., S. Kwok, and S. Weng. 1998. "Uses and Abuses of Default Rates." *CreditMetrics Monitor* (1st Quarter):37–55.

Lando, D., and T.M. Skødeberg. 2002. "Analyzing Rating Transitions and Rating Drift with Continuous Observations." *Journal of Banking & Finance*, vol. 26, nos. 2–3 (March):423–444.

Löffler, G. 2004. "An Anatomy of Rating Through-the-Cycle." *Journal of Banking & Finance*, vol. 28, no. 3 (March):695–720.

"Moody's Mulls Changes to Its Ratings Process." 2002. *Financial Times*, Companies & Finance Section (19 January):15.

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

Standard & Poor's Corporation. 2003. "Corporate Ratings Criteria." Standard & Poor's Corporation: www.standardandpoors.com.

Treacy, W.F., and M. Carey. 2000. "Credit Rating Systems at Large US Banks." *Journal of Banking & Finance*, vol. 24, nos. 1–2 (January):167–201.

