The Complementary Nature of Ratings and Market-Based Measures of Default Risk

GUNTER LÖFFLER

GUNTER LÖFFLER
is professor of finance at Institute of Finance, University of Ulm in Ulm, Germany.
gunter.loeffler@uni-ulm.de

What is the optimal combination of agency credit ratings and quantitative estimates of default risk that are based on stock market prices and balance-sheet data? When it comes to default prediction, the answer I derive from an empirical study is very simple: Put equal weight on both measures and you can hope to get the best results.

For many readers, this answer may be surprising because the literature suggests a different one. Kealhofer [2003] reports that market-based forecasts are more accurate predictors of default than are ratings. Other studies show that ratings take months to react to stock market information (Delianedis and Geske [1999]), or contain as much information for bond prices as a simple measure such as stock market volatility (Campbell and Taksler [2002]).

The answers I obtain differ for several reasons. One is that I pose different questions. The literature is dominated by studies that consider the merits or shortcomings of individual measures of default risk; they usually stop if they find that a system is efficient or deficient with respect to some criterion that optimal forecasts are expected to fulfill. This way of looking at the data distracts from the important question of how different measures could be optimally combined. Investors and financial institutions rarely opt for only one system, but rather combine different sources of information to arrive at their own assessments. If one forecast is superior to another, it does not follow that one should neglect the other one altogether. It may be possible to combine the two forecasts to form an even better one.

In addition, I extend the prior literature by exploring differences in default prediction power across forecast horizons and borrowers, and by using a comprehensive data set. It contains both agency ratings and market-based measures of default risk for more than 4,000 issuers over up to 25 years. To estimate optimal weights for default prediction, I employ straightforward logit regressions.

Most closely related is research by Miller [1998]; Kealhofer and Kurbat [2001], and Kealhofer [2003]. Miller shows that market-based measures add default prediction power to ratings; Kealhofer and Kurbat conclude that ratings have no incremental value for default prediction. The data samples used in these articles are relatively small, however. Kealhofer's [2003] analysis of the incremental power of S&P ratings, for example, rests on 67 defaults. The data used for the present articles contains 534 defaults.

DATA AND MODELING

I use monthly data on Moody's long-term ratings and a market-based measure of default risk—Expected Default Frequencies (EDFs)—computed by Moody's KMV (MKMV). The

EDFs estimate a firm's one-year probability of default using an extension of Merton's [1974] model. EDFs, which are available through a commercial product, have been shown to provide powerful default risk forecasts (cf. Kealhofer [2003]).

Issuers contained in the database are made up of the intersection of U.S. and non-U.S. corporate bond issuers that have a Moody's rating and traded equity. The data extend from January 1980 to April 2005. Since Moody's refined its rating system in 1982, I start the analysis at the end of 1982. The resulting data set comprises 336,729 firm-months.

I build my analysis on logit regressions. Through maximum likelihood, they determine the coefficients β_j in

$$\text{Prob}(\text{Default}_{t,T}) = \Lambda\left[\sum_j \beta_j x_{jt}\right] \qquad (1)$$

where Λ denotes the logistic function exp(.)/(1+exp(.)) and j indexes explanatory variables x_j. The indicator variable $Default_{t,T}$ takes the value one if a borrower defaults in the $(T\text{-}t)$ years after observing the variables x_j at dates t.

In this article, the forecast horizons $(T\text{-}t)$ will be one, three, or five years; explanatory variables x will be based on EDFs and ratings. The first issue one must address is how to enter these measures into the analysis. Ratings are on an ordinal scale from Aaa to Caa3; EDFs are on a cardinal scale from 0.02% to 20%. In order to combine ratings and EDFs, one or both have to be transformed. I transform both, choosing the transformations such that two requirements are fulfilled. First, the transformations should not favor one measure over the other; second, the transformations should be such that combinations of the two measures are easy to interpret in terms of relative weights.

It turns out that simple transformations work very well. I take logarithmic EDFs and use the common conversion of ratings into cardinal numbers from 1(=Aaa) to 21 (=Caa3). To allow easy interpretation of results, I linearly transform ratings again so that the rating variable has the same sample mean and variance as logarithmic EDFs. Note that this linear transformation changes only the size of the estimated coefficients, not the fit that is achieved through this variable. In exhibits and formulae, I will refer to the transformed variables as *T-EDFs* and *T-Ratings*. Exhibit 1 compares the (pseudo) R^2 that these transformed variables produce in logit regressions to the one that obtains if the transformation is estimated through a third-order polynomial. For all three forecast horizons, log EDFs lead to a higher R^2 than a polynomial in EDFs. Log EDFs also produce a slightly better fit than using the inverse of the logistic distribution—i.e., the transformation of choice if EDFs were identical to the correct underlying probabilities of default. In the case of ratings, the polynomial leads to a somewhat better fit but its default probability prediction is non-monotonic in the underlying ratings, which is an implausible feature likely due to overfitting. Thus, the chosen transformations do not seem to lead to a loss of fit compared to alternative transformations.

Exhibit 2 presents descriptive statistics for the original variables and their transformations. The exhibit shows that the empirical distribution of transformed

EXHIBIT 1

Benchmarking Variable Transformations: Pseudo-R^2s from Logit Regressions with Default Indicator as Dependent Variable

T-EDF denotes transformed EDFs (=ln(EDF)), T-Rating denotes transformed ratings (ratings are first converted to integers from 1 to 21 and then linearly transformed to match the mean and variance of T-EDF).

Horizon for default prediction

x_j. If, for example, the regression results in a
t of 0.8 and 0.4 for *T-Ratings* and *T-EDFs*,
ly, my interpretation will be that ratings and
uld be weighted 2 to 1. As shown in the bottom
xhibit 2, *T-Ratings* and *T-EDFs* have very sim-
outions for subsets of the data as well (here for
ns with rating better than Ba1 (=11) and EDF
an 0.5%). Therefore, statements on optimal
within subsets can also be based on logit
ts.

stical inference within the logit regression (1) has
h two problems. First, one cannot rule out that
errors are contemporaneously correlated—e.g.,

accuracy profiles, which obtain by plotting the proportion of defaults that occurred among issuers ranked x or worse, against the proportion of issuers ranked x or worse (see Sobehart et al. [2000]); the more northwestern the curve, the better the rating system. Accuracy ratios relate the area above the diagonal to the maximum area a curve can enclose above the diagonal.

DEFAULT PREDICTION PERFORMANCE AND OPTIMAL WEIGHTS FOR COMBINATIONS

I start by estimating the following logit regression:

score $\Sigma_j \beta$
coefficie
respectiv
EDFs sho
panel of
ilar distri
observati
smaller t
weights
coefficie

Stat
to deal w
regressio

EXHIBIT 3

The Relative Weight of EDFs and Ratings in Default Prediction: Logit Regressions with Non-overlapping Observations

T-EDF denotes transformed EDFs (=log(EDF)), T-Rating denotes transformed ratings (ratings are converted to integers from 1 to 21 and then linearly transformed to match the mean and variance of T-EDF). Accuracy ratios are for the fit of the logit models. T-values (in parentheses) are corrected for heteroscedasticity and clustering within observations belonging to the same year. Non-overlapping forecast horizons start at the end of December and are spaced by default horizon.

	1-year horizon		3-year horizon		5-year horizon	
T-EDF (β_1)	1.162	0.714	0.615	0.185	0.460	0.162
	(19.82)	(3.09)	(3.89)	(1.76)	(2.38)	(2.32)
T-Rating (β_2)	0.495	0.865	0.655	0.921	0.712	0.880
	(7.54)	(3.29)	(4.83)	(5.00)	(4.59)	(5.04)
T-EDF×$I_{EDF>0.5}$ (β_3)		0.487		0.577		0.468
		(1.92)		(4.81)		(2.62)
T-Rating×$I_{Spec.grade}$ (β_4)		-0.430		-0.364		-0.265
		(-1.33)		(-1.34)		(-1.86)
Observations	28506	28506	7535	7535	4521	4521
Pseudo R^2	0.400	0.401	0.315	0.320	0.283	0.288
Accuracy ratio	0.895	0.897	0.797	0.803	0.745	0.750
$p(\beta_1 = \beta_2)$	0.000	0.730	0.889	0.001	0.468	0.003
$p(\beta_1 + \beta_3 = 0)$		0.000		0.000		0.003
$p(\beta_2 + \beta_4 = 0)$		0.000		0.004		0.002

The relative weights of ratings versus EDFs are 0.4 to 1, 1.1 to 1, and 1.5 to 1 for forecast horizons of one, three, and five years respectively. However, the null hypothesis that the coefficients of *T-EDF* and *T-Rating* are identical can only be rejected for the one-year horizon. That is, for longer horizons one cannot reject the hypothesis that an equal weighting of the two measures provides the best fit.

The regression framework can easily accommodate more variables. In order to check whether the relative default prediction power changes with the magnitude of default risk, I estimate

$$\begin{aligned}\text{Prob}(\text{Default}_{t,T}) = \Lambda[&\beta_0 + \beta_1 \textit{T-EDF}_t + \beta_2 \textit{T-Rating}_t \\ &+ \beta_3 \textit{T-EDF}_t \times I_{EDF>0.5\%} \\ &+ \beta_4 \textit{T-Rating}_t \times I_{Specgrade}] \qquad (3)\end{aligned}$$

where $I_{EDF>0.5\%}$ and $I_{Specgrade}$ are indicator variables that allow coefficients to vary if the EDF is larger than 0.5 and the rating is speculative grade (i.e., below Baa3), respectively. This roughly cuts the sample into two halves differing in average default risk. The coefficients β_1 and β_2 give the optimal weights on EDFs and ratings in this low-risk group; β_3 and β_4, the coefficients on the interacted variables, show how the weights change when moving to the high-risk group. Results are also presented in Exhibit 3. For low-risk issuers, ratings appear to carry more information for default prediction than EDFs ($\beta_2 > \beta_1$, statistically significant for three- and five-year horizons). The optimal weight of ratings versus EDFs is 1.2 to 1, 5 to 1, and 5.4 to 1 for one-, three-, and five-year horizons, respectively. For below-investment-grade issuers, the informativeness of EDFs is higher ($\beta_1 + \beta_2 > \beta_2 + \beta_4$), but ratings continue to add significant prediction power. This can be checked by testing whether the sum of the coefficients β_2 and β_4 is zero; this hypothesis can be rejected at levels better than 0.4%. The improvement brought about by the interaction variables, however, is small. Individual t-statistics of the interaction variables are relatively low, and the maximum increase in accuracy ratios is 60 basis points.

While the analysis of the logit coefficients allows straightforward statements on statistical significance, conclusions about their economic significance may not be obvious. This is why I now focus on the analysis of accuracy ratios. Accuracy ratios are also statistical measures, of course, but there are studies that relate them to economic significance (Stein [2005]). Roughly speaking, a difference of 1 percentage point can have moderate but visible consequences, while a difference of three or more percentage points can be expected to have clear economic value. Exhibit 4 sets the accuracy ratios associated with the use of just one measure (i.e., either EDFs or ratings) against the ones attainable if the fit from the above logit regressions is used as a measure of default risk. In addition, I report accuracy ratios for a naïve combination of EDFs and ratings, which is $0.5 \times (T\text{-}EDF + T\text{-}Rating)$. The exhibit is split into three panels for the three forecast horizons.

The first column of the exhibit reproduces one of the results pointed out by Kealhofer [2003]: On a one-year horizon, EDFs perform considerably better than ratings. Notwithstanding this fact, combining EDFs with ratings increases the accuracy ratio by up to 1.5 percentage points. The benefit of combining the two measures is more pronounced for three- and five-year horizons. In those, accuracy ratios of combined measures are up to 4.1 (three years) and 4.0 (five years) percentage points better than the maximum attainable with a single measure. Compared to these gains in accuracy, the differences between the three combined measures (naïve, logit without differentiation across risk classes, and logit with differentiation) are small.

The other columns of Exhibit 4 show whether the results are stable across time and also valid out of sample. In columns two and three, I compute accuracy ratios for two subsamples that split the sample period in halves, but continue to take the logit weights from the regression over the entire sample. This provides a first indication of the stability of the logit results. In general, the picture does not change very much from the first to the second half of the sample period. Ratings seem to perform better

EXHIBIT 4

Accuracy Ratios for EDFs, Ratings, an Equal Weight Combination, and Fitted Logit Scores

Optimized logit weights are either determined through one regression covering the entire sample (*in sample*) or through regressions that use only information available at the start of the forecast horizon (*out of sample*).

	1982-2005 (in sample)	1982-1994 (in sample)	1994-2005 (in sample)	1994-2005 (out of sample)
1-year horizon				
EDF	0.881	0.875	0.885	0.885
Rating	0.826	0.805	0.841	0.841
EDF+Rating	0.893	0.887	0.899	0.899
Logit(Edf, Rating)	0.896	0.889	0.900	0.899
Logit(Edf, Rating, EDF>0.5%, Rating>Baa3)	0.896	0.892	0.900	0.896
3-year horizon				
EDF	0.767	0.786	0.759	0.759
Rating	0.758	0.721	0.784	0.784
EDF+Rating	0.807	0.800	0.812	0.812
Logit(Edf, Rating)	0.807	0.802	0.811	0.796
Logit(Edf, Rating, EDF>0.5%,Rating>Baa3)	0.808	0.801	0.813	0.794
5-year horizon				
EDF	0.691	0.719	0.678	0.678
Rating	0.713	0.680	0.745	0.745
EDF+Rating	0.750	0.746	0.754	0.754
Logit(Edf, Rating)	0.751	0.744	0.757	0.733
Logit(Edf, Rating, EDF>0.5%, Rating>Baa3)	0.752	0.744	0.760	0.730

in the second half of the sample period, though. This is worth noting because it was only in the second half of the sample period that EDFs were commercially sold. To understand why the naïve combination rule can surpass the optimized forecasts in one or the other subsample, note that the weights are optimized over the entire sample, not just the first or second half.

The last column presents the out-of-sample performance in period 1994–2005. Weights are determined with regressions that use only data from 1982 until the time of the forecast. For example, the optimized rating at the end of 1998 uses logit regressions whose forecast horizon ends at the end of 1998. Out of the sample, naïve forecasts perform best. The difference to optimized forecasts is small on a one-year horizon but increases to more than 2 percentage points on a five-year horizon.

However, it would be inappropriate to conclude from these results that logit regressions are ineffective for determining optimal weights. First, the results of the logit regressions already indicate that a naïve combination may not be inferior. (Recall that for three- and five-year horizons, the weights on EDFs and ratings were not significantly different). The fact that an equal weighting works should thus be regarded as a property of the data, not as a shortcoming of the regression approach. Second, if one considers the combination of default risk measures today, the optimization can be based on more observations than were used in the out-of-sample experiment reported in Exhibit 4.

DO THE TWO MEASURES OF DEFAULT RISK REALLY CONTAIN DIFFERENT INFORMATION?

EDFs and ratings are conceptually different. EDFs are model-driven, quantitative estimates that use current stock price information; the estimates used in this article are calibrated to yield the best default prediction on a one-year horizon. Ratings, by contrast, are judgmental assessments of long-term credit quality. For rating agencies, accurate default prediction is not the only objective; they trade off timeliness and accuracy against stability (cf. Fons [2002]). These observations are consistent with the results of the previous section. The value of ratings increases with the forecast horizon as agencies employ a long-term horizon, but it decreases with decreasing credit quality, because the higher a company's default risk, the more important short-term developments are for its survival.

While EDFs and ratings capture different aspects of rating quality, it is not clear that one has to follow two distinct rating philosophies and then combine them in order to achieve an optimal default prediction. Perhaps the information that a particular philosophy does *not* focus on can be extracted from the rating history. Ratings, for example, exhibit drift (Altman and Kao [1992]), which is likely to be partly explained by the rating philosophy. Controlling rating drift might improve default prediction performance without having to bring in a new rating philosophy. Likewise, using one-year EDFs for long-term default prediction may cause someone to miss trends or reversals in default risk, but this can be mended through time-series information on EDFs. To explore these ideas, I augment regression (2) from above with variables that capture autocorrelation in ratings and EDFs. For ratings, I examine

$$\begin{aligned}\text{Prob}(\text{Default}_{t,T}) = \Lambda[&\beta_0 + \beta_1 T\text{-}EDF_t + \beta_2 T\text{-}Rating_t \\ &+\beta_3(T\text{-}Rating_t - T\text{-}Rating_{t\text{-}12}) \\ &+ \beta_4(T\text{-}Rating_t - T\text{-}Rating_{t\text{-}24})] \quad (4)\end{aligned}$$

I examine two different lag lengths (12 and 24 months) for past rating changes because it is not clear a priori which one is relevant. For EDFs, I use the same approach: I augment (2) by $\beta_5(T\text{-}EDF_t - T\text{-}EDF_{t\text{-}12}) + \beta_6(T\text{-}EDF_t - T\text{-}EDF_{t\text{-}24})$. The results are reported in Exhibit 5. Note that the number of observations is lower than that of Exhibit 3 because the new explanatory variables are missing in many instances.

For ratings, the regressions show that the rating drift affects default prediction performance, in particular for the one-year horizon. A downgrade in the last 12 months results in a higher default probability, because the current rating is still too good. The lagged 24-month change commands negative coefficients, which would point to mean reversion in ratings, but the statistical significance of this is weak. The result that is most relevant for this section, however, is that controlling for lagged rating changes does not greatly affect the influence of EDFs; the size of the coefficients and their statistical significance is similar to the ones in Exhibit 3. Intriguingly, the picture is very similar if one introduces lagged EDF changes into the regression. Again there is drift towards default. Past increases in EDF are associated with higher default risk, even on the one-year horizon to which EDFs are calibrated. There is no indication that the use of one-year EDFs for long-term prediction neglects mean reversion,

EXHIBIT 5

Do Trends in EDFs and Ratings Improve Default Prediction? Logit Regressions with Non-Overlapping Observations

	1-year horizon		3-year horizon		5-year horizon	
T-EDF	1.291	1.110	0.800	0.654	0.534	0.397
	(16.94)	(10.55)	(5.73)	(4.51)	(2.39)	(2.46)
T-Rating	0.339	0.544	0.493	0.611	0.582	0.690
	(3.62)	(5.54)	(3.34)	(5.34)	(3.84)	(5.90)
$T\text{-}Rating_t$ - $T\text{-}Rating_{t-12}$	0.374		0.494		0.134	
(=1-year rating trend)	(2.68)		(1.77)		(1.44)	
$T\text{-}Rating_t$ - $T\text{-}Rating_{t-24}$	-0.044		-0.251		-0.108	
(=2-year rating trend)	(-0.29)		(-1.60)		(-1.24)	
$T\text{-}EDF_t$ - $T\text{-}EDF_{t-12}$		0.444		0.199		0.258
(=1-year EDF trend)		(5.09)		(2.56)		(2.59)
$T\text{-}EDF_t$ - $T\text{-}EDF_{t-24}$		0.003		0.119		0.068
(=2-year EDF trend)		(0.04)		(2.41)		(0.66)
Observations	21678	21733	5473	5475	3301	3301
Pseudo R^2	0.434	0.442	0.361	0.364	0.284	0.291
Accuracy ratio	0.912	0.917	0.832	0.834	0.745	0.751

Notes: see Exhibit 3.

and the influence of ratings is similar to the previous regressions, which did not include lagged EDF changes.

What do we learn from these estimates? Rating drift reduces the default prediction power of ratings and thus offers scope for improvement through alternative measures of default risk. However, EDFs are more than ratings without rating drift. Also, ratings are more than just EDFs corrected for trends or reversals. It appears that the long-term qualitative assessment inherent in ratings is something that cannot easily be replaced by a statistical analysis of the process that governs short-term default risk. Likewise, market-based measures seem to provide information not conveyed by the traditional rating.

SEASONALITIES AND AGING EFFECTS

Over time and across issuers, the informativeness of EDFs and ratings can fluctuate. EDFs are aimed to provide up-to-date measures of credit quality, but one input to EDFs—balance sheet information on corporate liabilities—is only published in quarterly intervals. Ratings, by contrast, are not meant to be "high-frequency sources of information" (Fons [2002], p. 13). The design of the rating process and the rating philosophy tend to reduce the speed with which ratings adjust to new information. In consequence, a newly issued rating may have a larger predictive content than an aged rating that was assigned several years ago.

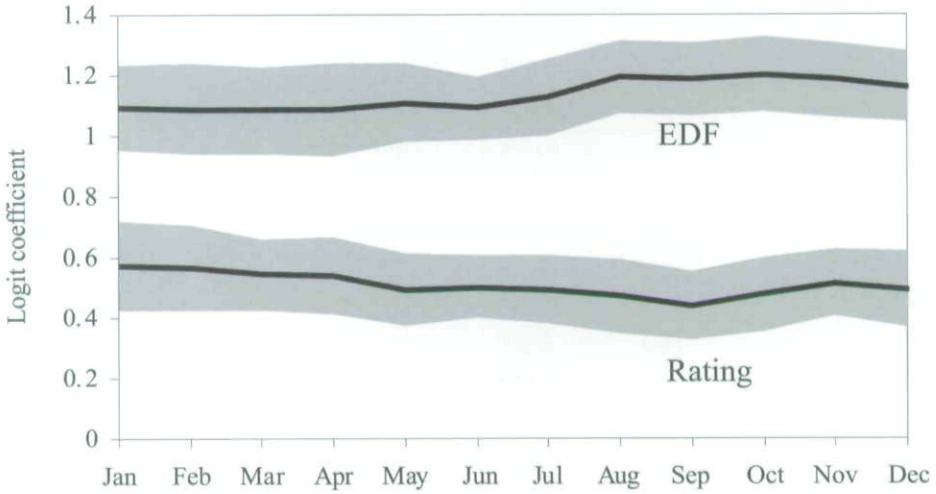
In this section, I explore the relevance of such effects, starting with an investigation of seasonalities. In periods in which many firms have just published financial reports, the average quality of EDF might be relatively high. Recall that the starting point of the analysis was a logit regression that only used observations from the end of December. As a quick check for the presence of seasonalities, I now run 12 separate logit regressions,

$$\text{Prob}(\text{Default}_{t,T}) = \Lambda[\beta_0 + \beta_1\ T\text{-}EDF_t + \beta_2\ T\text{-}Rating_t]$$

where the dates t are restricted to January, February, ... or December. Exhibit 6 shows the coefficients β_1 and β_2 together with their associated 95% confidence intervals. Coefficients of ratings and EDFs change over calendar months, but the changes are smooth and within the confidence bounds. Therefore, there is little evidence of aggregated seasonal effects in the informativeness of EDFs and ratings.

EXHIBIT 6

Seasonalities in the Importance of EDFs and Ratings for 1-Year Default Prediction: Coefficients (with 95% Confidence Intervals) from Logit Regressions Using Only Observations from One Calendar Month



To model the aging effect of ratings, I include an additional explanatory variable in the logit regressions. It is an interaction variable defined as

$$\ln(1 + \text{rating age}) \times \text{T-Rating}$$

where rating age is the number of months that have passed since the rating was assigned. Thus, ln(1+rating age) is zero for a rating that was changed in the current month.

In addition, I augment the regressions by information on rating outlooks and watchlists. These signals are an integral part of the rating system used for timely information about changes in the rating agency's assessment. Therefore, it is important to control for these signals when examining aging effects. Through an outlook, the rating agency expresses an opinion on the likely direction of medium-term rating actions. The watchlist contains issuers whose rating is actually on review for a possible rating change. It therefore tends to provide stronger signals than outlooks. I code outlook and watchlist information into one variable. It takes the value 2 (–2) for issuers that are on watch for upgrade (downgrade), 1 (–1) for issuers with a positive (negative) outlook, and zero for all others. (The coding follows the one in research by Cantor and Hamilton [2004].) Outlook and watchlist information is only available from 1991 onward. In regressions run over the entire data set, its information content will therefore be underestimated.

The regressions including these two additional variables have the following form:

$$\begin{aligned}\text{Prob}(\text{Default}_{t,T}) = \Lambda[&\beta_0 + \beta_1 T\text{-}EDF_t + \beta_2 T\text{-}Rating_t \\ &+ \beta_3 Outlook/Watchlist_t \\ &+ \beta_4(1 + rating\ age) \times T\text{-}Rating_t] \quad (5)\end{aligned}$$

Results are shown in Exhibit 7. Regressions are run for the entire sample and for the second half of the data only. The coefficient on outlook/watchlist is significant and negative, which is the correct sign, as positive values of this variable (e.g., 2 for watch-for-upgrade) then imply a lower default probability. In absolute terms, the coefficient increases when the analysis is restricted to the second half of the sample, in which outlook and watchlist information is available for more issuers. Bond investors wishing to assess default risk should definitely take outlook and watchlist information into account.

The coefficient on the variable that interacts rating age with ratings is significant and negative, so the optimal weight that should be put on ratings decreases with their age. Overall, the weight attached to a rating is $\beta_2 + \beta_4$ (1 + rating age); Exhibit 8 plots this overall weight against rating age. At first, the value of rating information declines quickly. When predicting one-year defaults, a newly assigned rating should command a weight twice as large as the one for a rating that was assigned 16 months ago.

EXHIBIT 7

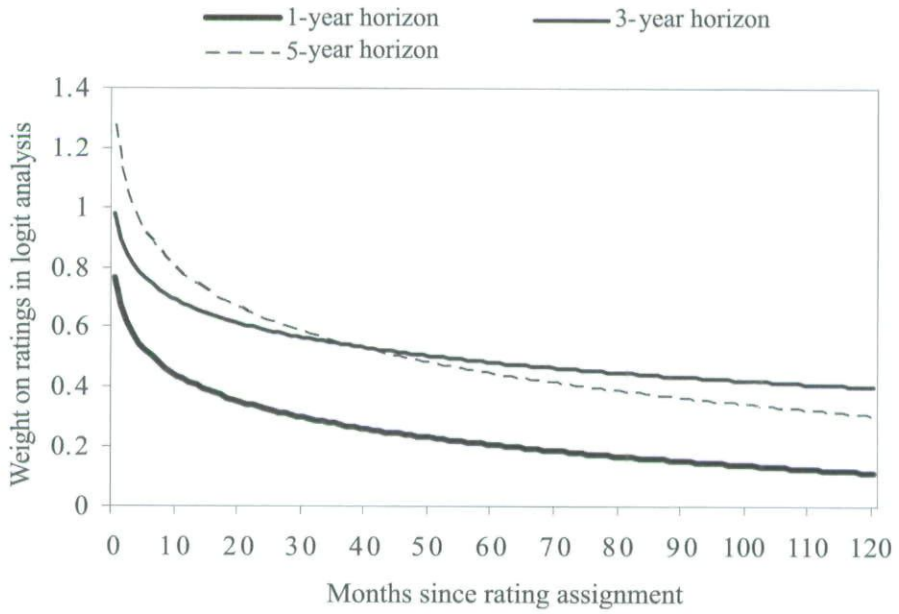
The Impact of Outlook/Watchlist Information and of Rating Age

Positive (negative) outlooks are coded as 1 (–1); positive (negative) watches are coded as 2 (–2). Rating age is the number of months since the rating assignment date. For further information see notes to Exhibit 3.

	1-year horizon		3-year horizon		5-year horizon	
	1982-2005	1994-2005	1982-2005	1994-2005	1982-2005	1994-2005
T-EDF (β_1)	1.128	1.053	0.593	0.504	0.463	0.313
	(17.41)	(16.72)	(3.76)	(7.29)	(2.28)	(2.62)
T-Rating (β_2)	0.762	0.844	0.976	0.987	1.278	0.957
	(10.04)	(14.14)	(14.04)	(7.51)	(11.49)	(12.68)
Outlook/Watchlist	-0.379	-0.527	-0.202	-0.483	-0.282	-0.327
	(-4.44)	(-7.80)	(-2.02)	(-5.99)	(-2.04)	(-12.94)
ln(1+rating age)×T-Rating	-0.135	-0.141	-0.120	-0.092	-0.203	-0.006
	(-9.17)	(-7.68)	(-2.02)	(-1.49)	(-3.85)	(-0.11)
Observations	28506	17640	7535	5140	4521	3352
Pseudo R^2	0.413	0.436	0.321	0.358	0.295	0.324
Accuracy ratio	0.901	0.902	0.801	0.814	0.754	0.770
$p(\beta_1 = \beta_2)$	0.005	0.065	0.010	0.001	0.004	0.000

EXHIBIT 8

Impact of Rating Age on Importance of Ratings for Default Prediction: Weight on Ratings in Logit Regressions Using the Entire Sample



The decline in informativeness slows down, however. Ratings that were set 5 or 10 years ago are still useful for default prediction.

CONCLUSION

Combining ratings and market-based measures of default risk improves the prediction of defaults. The shorter the horizon, the greater the influence of the market-based measure should be, but a simple equal-weight combination of ratings and market-based measures is hard to beat out of sample. Why, then, don't providers of rating information combine the merits of the two approaches into a single rating product? At least until now, rating agencies have chosen instead to offer separate quantitative products in addition to their traditional qualitative ratings, "just like Daimler offers gas and diesel engines" (Cass [2002], p. 7).

Here's one possible answer: Uses of rating information include risk measurement, investment management, or pricing—all three of which could be further branched off according to the time horizon or the liquidity of the credit-risky instrument. It seems likely that the optimal weighting of ratings and market-based measures is not constant across these usages. Clients of rating agencies are perhaps better off if they themselves combine different rating information according to their own needs, or if they opt for one or the other depending on their situation. In addition, keeping measures separate fosters diversity in interpretations—providing a fruitful ground for active bond managers.

ENDNOTE

I am grateful to Richard Cantor, Roger M. Stein, and seminar participants at Moody's Investors Service for their valuable comments and to Moody's Investors Service and Moody's KMV for providing the data.

REFERENCES

Altman, E., and D.L. Kao. "Rating Drift in High-Yield Bonds." *Journal of Fixed Income* (March 1992), pp. 15–20.

Campbell, J.Y., and G.B. Taksler. "Equity Volatility and Corporate Bond Yields." *Journal of Finance*, 58 (2002), pp. 2321–2349.

Cantor, R., and D.T. Hamilton. "Rating Transitions and Defaults Conditioned on Outlooks." *Journal of Fixed Income* (September 2004), pp. 54–70.

Cass, D. "KMV Acquisition Throws up Rating Industry Questions." *Risk*, (March 2002), p. 7.

Delianedis, G., and R. Geske. "Credit Risk and Risk Neutral Default Probabilities: Information about Rating Migrations and Defaults." Working paper, UCLA, 1999.

Fons, J. "Understanding Moody's Corporate Bond Ratings and Rating Process." Special comment, Moody's Investors Service, 2002.

Kealhofer, S. "Quantifying Credit Risk I: Default Prediction." *Financial Analysts Journal,* 59 (2003), pp. 30–44.

Kealhofer, S., and M. Kurbat. "The Default Prediction Power of the Merton Approach, Relative to Debt Ratings and Accounting Variables." Moody's KMV, 2001.

Merton, R.C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance*, 29 (1974), pp. 449–470.

Miller, R. "Refining Ratings." *Risk* (August 1998), pp. 97–99.

Sobehart, J.R., S.C. Keenan, and R.M. Stein, "Benchmarking Quantitative Default Risk Models: A Validation Methodology." Moody's Investors Service, 2000.

Stein, R.M. "The Relationship Between Default Prediction and Lending Profits. Integrating ROC Analysis and Loan Pricing." *Journal of Banking & Finance*, 29 (2005), pp. 1213–1236.

To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.

©Euromoney Institutional Investor PLC. This material must be used for the customer's internal business use only and a maximum of ten (10) hard copy print-outs may be made. No further copying or transmission of this material is allowed without the express permission of Euromoney Instituitonal Investor PLC.
The most recent two editions of this title are only ever available at http://www.euromoneyplc.com