Löffler, Gunter; Posch, Peter N.

Working Paper

How do rating agencies score in predicting firm performance

SFB 649 Discussion Paper, No. 2007,043

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Löffler, Gunter; Posch, Peter N. (2007) : How do rating agencies score in predicting firm performance, SFB 649 Discussion Paper, No. 2007,043, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/25215

SFB 649 Discussion Paper 2007-043

How do Rating Agencies Score in Predicting Firm Performance

Gunter Löffler*
Peter N. Posch*

* University of Ulm, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

How do rating agencies score in predicting firm performance?

Gunter Löffler*, Peter N. Posch

Draft Version: 30th of July 2007

Abstract We use dynamic panel analysis to examine whether credit rating agencies achieve what they claim to achieve, namely, look into the future when assigning their ratings. We find that Moody's ratings help predict individual financial ratios over a horizon of up to five years. Ratings also predict a multivariate credit score, again over five years. The contribution of ratings appears to be economically significant and robust for different specifications.

Keywords: Credit Ratings, Predictive ability, Dynamic Panel Model

JEL-Classification: C33, G20, G33

*Institute of Finance, University of Ulm, Helmholtzstrasse 18, 89081 Ulm, Germany. email: {gunter.loeffler,peter.posch}@uni-ulm.de. Phone: +49-731-50-23597.

Acknowledgement We thank Werner Smolny for helpful comments. This research was supported by the Deutsche Forschungsgemeinschaft (DFG) through the SFB 649 "Economic Risk". Peter N. Posch is associated member of the DFG Research Training Group 1100 "Modelling, analysis and simulation in economy mathematics".

1 Introduction

Rating agencies claim their credit ratings to be forward-looking over a horizon of several years. So far, however, there has been no attempt to examine the agencies' forward-looking ability in detail. There are many studies on the quality of ratings but most of them focus on short-term performance. Kealhofer (2003), for example, shows that there are better predictors of one-year default behavior than ratings. Other papers document that rating changes can be predicted using firm fundamentals or market data (Delianedis and Geske (1999), Jianming Kou and Varotto (2007), Hull et al. (2004).) The latter evidence is in line with the widespread notion that rating agencies are too slow to change their ratings (Baker and Mansi (2002)). A slow response to recent information, however, does not rule out that rating agencies are useful for long-term prediction.

Our research strategy is as follows: If ratings are indeed forward-looking they should contain information about the future values of financial ratios that are relevant for assessing the default risk of a firm. Therefore, we test whether today's rating of a company helps explain future financial ratios of this company. Using several financial ratios we show which determinants of credit risk can be explained by past ratings and which cannot. To benchmark the predictive power of ratings we include the history of the financial ratio that is going to be predicted.

Tests are performed using panel regressions in which the future realization of the financial ratio is the dependent variable; explanatory variables are the lagged rating as well as lagged values of the dependent variable. The latter brings a bouquet of problems. Primarily, it yields inconsistent estimates when using linear regression. Thus we employ a generalized method of moments

estimator which is the state-of-the-art solution in such dynamic panel models. We seek for robust standard errors using heteroskedasticity and autocorrelation consistent variance estimators (HAC). Although this technique promises stable results, we conduct a simulation study to derive critical t-values in our setting.

Another challenge is the assessment of a possible selection bias. There are cases in which we do not observe future financial ratios. Many of these cases are likely to be performance-related, e.g. occur because of a default or delisting. Ignoring such non-random selection could lead to biased coefficients. In the paper, we test for selection bias and correct it using a generalization of the Heckman procedure.

Our contribution to the literature is twofold. First, we contribute to the understanding of the information content of ratings. Our results suggest that rating agencies indeed foresee credit quality indicators over a horizon of several years. Compared to naïve predictors ratings perform remarkably well. With respect to their forward-looking ability, rating agencies therefore seem to deliver what they promise to.

Second, we introduce a new methodology to this branch of literature. The class of dynamic models, although widely used in the economic literature, has not yet been applied to the analysis of ratings and financial ratios. They allow straightforward statements on the contribution of a predictor conditional on the history of the variable that is going to be predicted.

Related literature includes papers which show that the relative default prediction performance increases with the prediction horizon (Löffler (2007) and Altman and Rijken (2006)). Our selection of financial ratios draws on the credit scoring literature (Altman (1968), Shumway (2001)) as well as on papers that use contemporaneous financial ratios to predict ratings, e.g. Blume et al. (1998) or Kisgen (2006).

The remainder of this paper is organized as follows. In the next section we briefly introduce the dataset and econometric specification. Afterwards we present the result from a multivariate scoring model and comment on univariate results and the overall robustness. The final section concludes.

2 Data and model

Financial ratios are derived from Compustat North America. The database used covers the years 1985 to 2005 and contains information on 24,161 companies of which 3,848 companies can be matched to an issuer-level rating by Moody's.[1]

Overall there are 40,207 firm-years with ratings as well as information from Compustat; the average number of years is 12.48 per company. Ratings enter the analysis as categorical numbers from 1 (Aaa) to 21 (C). This may not be the optimal way of coding rating information. Research by Kisgen (2006) or Löffler (2007), however, shows that the simple numerical conversion performs well in rating or default prediction. More importantly, since we later find that ratings are useful predictors a better coding could only strengthen our results.

To assess the relative impact of a lagged rating compared to a variable's history we use a *dynamic panel model*, where a lagged dependent variable is included. The model to be estimated is of the following form:

$$y_{it} = \beta y_{i,t-l_1} + \gamma RAT_{i,t-l_2} + \eta_{it} \text{ , with } \eta_{it} = u_i + \tau_t + \varepsilon_{it}. \tag{1}$$

Here y is a specific financial ratio or a score built through a linear combination of financial ratios, u_i is a firm-specific error term, τ_t a time specific error term and ε_{it} the usual residual.

[1] Rating data have been kindly provided to us by Moody's Investors Service.

Bond (2002) provides a comprehensive overview on the use of dynamic models for panel data and Davidson and MacKinnon (2004), ch. 13 discusses their use for cross-sectional time series. The classical studies of dynamic panels are Arellano and Bond (1991) and Arellano and Bover (1995) who introduce a *general method of moments (GMM)* estimator, which we employ here.

A dynamic panel model has important implications on the interpretation of the coefficients as they are now not only conditioned on all information at a point in time, but also include the history. The coefficient γ captures the influence of a lagged rating on the current value of y_{it}. Since we include $y_{i,t-l_1}$ the impact of the coefficient γ measures the impact of rating information beyond that contained in the predicted variable's own history. If we set $l_1 = 1 < l_2$ we would not mimic the economic situation. We are interested in whether the lagged rating provides new information for the contemporary variable given all available information at the time the lagged rating was assigned. So we set $l = l_1 = l_2$, varying the two lag lengths in lock step. For a given lag length l, γ then indicates whether rating agencies foresee the variable's realization over a horizon of l, given all history up to the rating assignment date and conditional on the instruments of $y_{i,t-l}$. The instrument equation itself is of the following form

$$y_{i,t-l} = \sum_{j=\kappa_1}^{\mathbf{K}_1} y_{i,t-l-j}. \tag{2}$$

In all of the following regressions one further lag ($\mathbf{K}_1 = \kappa_1 = 1$) was sufficient to achieve valid instruments with respect to the tests described below.

Note that this specification benchmarks the predictive ability of the rating agencies against a time-series modeling of the variable of interest. In other words the question answered by this model is whether ratings contribute pre-

dictive power over a variable's own history. If we find, e.g., that the three-year lagged rating is significant in this setting we can conclude that current ratings help predict the realization of the variable three years from now.

It could be, however, that the lagged rating influences today's value of the variable rather than just predicting it. For example, if the rating is good despite a currently high leverage a firm may find it easier to reduce the leverage in order to maintain the rating (see Kisgen (2006) for rating-targeting behavior of firms). Including the rating and the variable itself with the same lag accounts for this aspect of simultaneity since this conditions the results on the information set at that time.

Estimating dynamic panel models with GMM brings in a bouquet of specification test.[2] All results presented in the next section include relevant instruments (Anderson (1984)) which are not weakly identified (Cragg and Donald (1993) and Yogo and Stock (2005)) and the model is identified (Hansen (1982) or Sargan (1958) for the two-stage settings). The reported t-statistics are based on *HAC Covariances* proposed by Newey and West (1987).

Another matter, which is not specific to the dynamic model used here, is the stationarity of variables. In time-series analysis the standard test for the presence of an unit root is the *augmented Dickey-Fuller (ADF)* test, proposed by Dickey and Fuller (1979). Based on the ADF Maddala and Wu (1999) propose a Fisher test (Fisher (1932)) for stationarity of unbalanced panels. Testing the variables used below the null hypothesis of unit root can be rejected.

Selection bias One important issue to deal with in this analysis is the *selection bias*. The term "selection bias" refers to the case where variables are only observed if some criteria in terms of a selection process H is met. Consider

[2]For a good overview and more details on the following tests see Baum et al. (2003).

the case where the rating of a company is no longer available. This could be due to a withdrawal or a default. In both cases the company might still have non missing observations for some of the explanatory variables included in the regression. However, it could as well be the case that these observations are missing and thus excluded from the regression. This means we have to deal with the case of *non randomly missing data*.

Wooldridge (1995) proposes a two step technique based on a fixed effects model. He allows the errors η_{it} to be serially correlated and unconditionally heteroskedastic.[3] The selection is made by an indicator vector $s_i = (s_{i1}, ..., s_{iT})$ for each panel member i. If $s_{it} = 1$ we observe the dependent variable y_{it}. Now he distinguishes between two cases: First, if the latent variable triggering selection is (partially) observed and secondly if it is not observed at all. Vella and Verbeek (1999) generalize the approach proposing a two-step estimation which allows for many panel models. Their approach can be considered as a *two-stage conditional maximum likelihood* estimation. The well known cross-sectional estimator of Heckman (1979) obtains as a special case. In contrast to other panel estimators, e.g. Hausman and Taylor (1981) and Honore (1993), where the bias is contributed to time-constant individual effects, the Vella and Verbeek (1999) approach allows for selectivity due to an individual time specific component (Vella and Verbeek (1999), p. 240). The efficiency loss using two-stage estimation[4] in contrast to maximum likelihood is relatively small as shown by Lee (2001) for the Wooldridge (1995) approach which we employ here.

In our study we have the case that we do observe the selection s. We start by setting s to one if there is a rating by Moody's for firm i at time t ($s_{it} = 1$).

[3] He assumes only that $\eta_{it}|\mathbf{x} \underset{iid}{\sim} N(0, \sigma_t^2)\ , \forall t$

[4] For a discussion see Newey and West (1987).

s is set to zero if the firm is in default at time t ($s_{it} = 0$) as recorded by Moody's. This also controls for the case where a rating is withdrawn before a default. Furthermore the selection is set to zero depending on the availability of the examined variable. If the variable of firm i at time t is missing the selection is set to zero ($s_{it} = 0$). Hence each variable can have a different definition of the selection process s and the correction is thus calculated for each variable.[5]

We do not directly observe the selection mechanism s^*, however, we can model it using a set of variables $\mathbf{z}$.

The two stages are as follows:[6]

1. For each time t=1,2,...,T estimate a standard probit of the latent selection model s^* and calculate the inverse Mill's ratio[7] $\widetilde{\lambda_{it}} = \lambda(\xi' \mathbf{z}_{it})$ for $s_{it} = 1$.

2. Estimate the panel model including $\zeta \widetilde{\lambda_{it}}$ for $s_{it} = 1$

To test for selectivity one simply tests $H_0 : \zeta = 0$ using t-statistics. Of course in the second stage one should use standard errors which are robust

[5]Note that in the literature on sample selection the selection process is defined upon the dependent variable only. The case of selection bias in a dynamic panel model is hardly addressed. Our special case where the selection depends on a further independent variable (the rating) has to our best knowledge not been assessed yet. Preliminary robustness tests on the specification of s indicate that the results do not change qualitatively.

[6]See Baltagi (2001), p. 222-224 for methods of testing and controlling structural breaks and Vella and Verbeek (1999) for a more detailed treatment on the used procedure.

[7]For any probability density function $f(x)$ and cumulative density function $F(x)$ the inverse Mill ratio is given as $\lambda(x) = f(x)/(1 - F(x))$, see Greene (2003) p. 759. Note that there exist a lot of definitions in the literature, most of them are interchangeable in most cases, noting that $F(-x) = 1 - F(x)$ and $f(x) = f(-x)$ if $f()$ is standard normal.

for serial correlation and heteroskedasticity. If the null cannot be rejected, the outcome equation is likely to be biased due to sample selection. In this case the estimator from the first-step is used as additional regressor as this controls for the unobserved heterogeneity which is responsible for the selection bias (Baltagi (2001), p. 224).

The choice of the s^* model is not obvious. One has to find variables to explain missing values. The variables in the selection equation do not need to be uncorrelated with the explanatory variables in the outcome equation. Almost all available candiates are from the annual reports of a company and therefore subject to selection as well. We therefore use the logarithm of the market value deflated by the CPI as univariate predictor of the s^* model as this variable gives the highest coverage of data and yields a McFadden-R^2 of 32.78%.[8]

The Wooldridge (1995) test does not reject the null hypothesis of sample selection for the vast majority of the following specifications. Thus the prediction from the first stage is included in the outcome Equation 1 to correct for the selection bias. We omit reporting this variable in full detail in the results below.

[8]Interpolating over missing values in the selection equation or imputing them based on auxilary regressions, e.g. explaining missing market values by book values etc., is generally possible. Since the reasons for a missing value, however, vary and might include the onset of default, e.g. if the company stops being traded actively, the methods have to be choosen carfully and their effect on the outcome model should be examined. A closer examination of the selection process, also with respect to possible feedback effects, should be part of future research on this topic.

3 Multivariate results using a scoring model

Ratings are assessments of current and future credit quality. We argued in the introduction that ratings should therefore predict future credit quality indicators. Since credit quality is determined by a set of indicators rather than just one, the link between ratings and specific indicators would not be perfect even if agencies had perfect foresight. For example, the agency may expect an improvement in profitability and still keep its rating at a low grade - because it may also expect leverage to further increase, neutralizing the effect of improved profitability.

We therefore start by assessing whether ratings predict a credit score. A credit score is a linear combination of financial ratios whose weights are chosen such that the score optimally predicts default. We determine a score by estimating a logit model (cf. Shumway (2001)) with four commonly used variables, i.e. book leverage as measured by total liabilities over total assets (TL2TA), working capital over total assets (WC2TA), retained earnings over total assets (RE2TA) and earnings before interest and taxes over total assets (EBIT2TA).[9]

The concept used in the literature to measure the quality of a rating system or a scoring model with respect to default prediction is the accuracy ratio (AR) based on the cumulative accuracy profile (see Sobehart and Keenan (2001) or Engelmann et al. (2003)). To construct the CAP all debtors are sorted according to their rating, starting with the debtor with the worst rating (highest probability of default) down to the debtor with the best rating. A CAP is then obtained by plotting the proportion of defaulted debtors against the proportion of all debtors. The accuracy ratio is defined as the area between the

[9]Sales over total assets is insignificant and thus excluded. Robustness checks show no difference when S2TA is included.

CAP of the analyzed scoring system and the non-informative system divided by the area between the CAP of a 'perfect' scoring model and the CAP of the non-informative rating model. A rating model with high discriminative power has an accuracy ratio close to 100%, while the minimum value of the AR is 0% for the random rating model. The accuracy ratio over a one-year horizon of the scoring model used here is at 79.5%.

Figure 1: *Coefficients on lagged ratings in a dynamic panel model explaining today's score..* These figures show the coefficients γ for Moody's ratings along with 95% confidence bands.

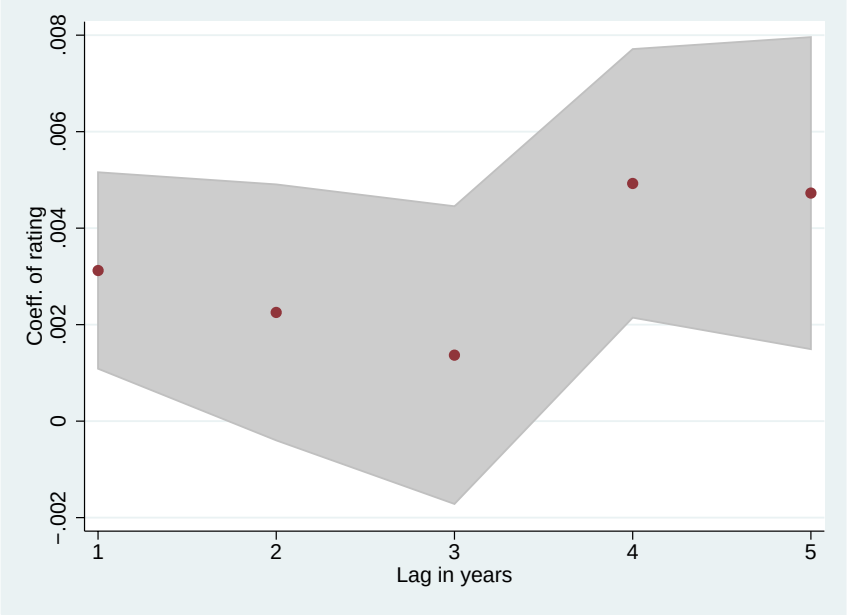


Figure 1 shows the results. All coefficients have the expected sign because a higher rating number is associated with a higher probability of default, as is a higher score. The coefficient is significant for lag lengths of one, four and five years, and marginally significant for two and three years. The results indicate that the rating agency has a significant predictive ability in forecasting future credit quality.

Before refining the specification we shall assess the robustness of this result. Although the method of GMM and the use of HAC-covariances should provide robust t-values, we conduct a Monte Carlo study to simulate critical t-values. In a bootstrap study, we replace each company's rating history randomly by another company's rating history. The set of companies from which the new history is drawn contains all companies whose rating history includes the time span of the ratings which are to be replaced. Conducting 1,000 repetitions and estimating the dynamic panel model in each of the repetitions gives an average critical value t-value at 95% confidence of 2.38. This does not necessitate a qualitative change in the interpretation of the results.

The accuracy ratio of the model's prediction is at 75.7% on average over all lags (max: 80.1%, min: 72.6%). Excluding the rating from the model, i.e. explaining the score solely by its history, significantly reduces the accuracy ratio for all lags to an average of 73.5% (max: 75.9%, min: 72.1%). Table 1 shows the accuracy ratio for different model specifications and lag lengths. It also shows the accuracy ratio for the lagged score, which is a readily available but simple forecast.

Table 1: Accuracy Ratio for different models and lag lengths. AR refers to accuracy ratio, OLS means ordinary least squares regression. The accuracy ratio is calculated for a default over the next period using the prediction of the model $Score_t = \beta Score_{t-lag} + \gamma Rating_{t-lag} + \eta$ either including the rating $Rating_{t-lag}$ for the lefthand columns or excluding the rating for the righthand ones. The three different specifications of η are fixed effect, time series without the panel error term u_i and in a standard pooled setting. The former two models are estimated with GMM, while the latter is a linear regression.

		Prediction using information from t-lag					
		Incl. Rating			Excl. Rating		
Lag	Score (t-lag)	Fixed Effects	Time Series	OLS	Fixed Effects	Time Series	OLS
0	0.803						
1	0.779	0.803	0.760	0.764	0.793	0.771	0.774
2	0.734	0.782	0.722	0.724	0.742	0.674	0.683
3	0.704	0.756	0.691	0.693	0.725	0.619	0.631
4	0.771	0.759	0.712	0.715	0.738	0.584	0.597

Apart from the robustness of the model itself, one could argue that the choice of a fixed effects dynamic panel model is not appropriate when dealing with ratings. Since rating agencies claim that they *rank* issuers according to their creditworthiness, (e.g. Cantor and Mann (2003), p. 6) a rating is a *relative* risk measure. Employing fixed effects controls for differences between issuer and thus might not fully mimic the agency's goal. In fact switching to a time series dynamic panel, i.e. dropping the u_i from the error term's specification, increases the significance of the rating's coefficient. However, the accuracy ratio of the time series model's prediction is lower, see Table 1.

As another of predictive performance, Table 2 reports the rank correlation between the actual score $Score_t$ and its predictions. Again, rank correlations of the time series model are significantly lower than with the panel model.

Table 2: Rank Correlation for different models. The rank correlation with the actual score $Score_t$ is given for predictions of three different specifications of η in $Score_t = \beta Score_{t-lag} + \gamma Rating_{t-lag} + \eta$ as fixed effect (A), time series (B) without the panel error term u_i and in a standard pooled setting (C). The former two models are estimated with GMM, while the latter is an OLS regression. The rightmost column shows the rank correlation using the true lagged score.

	(A)	(B)	(C)	(D)
Lag	Fixed Effects	Time Series	OLS	$Score_{t-lag}$
1	0.946	0.930	0.930	0.926
2	0.930	0.882	0.881	0.876
3	0.928	0.853	0.853	0.849
4	0.931	0.834	0.833	0.828

Finally, comparing the root mean squared error (RMSE) in-sample favors the panel model on a high level for all lags. Table 3 shows the t-values for the three specifications.

A second alternative model type would be the use of standard (pooled) ordinary least squares (OLS) regression. Although the coefficients would be biased due to endogeneity and heteroskedasticty, the model's prediction might still outperform the econometric correct specification. Although the OLS is slightly better than the time series setting in terms of the accuracy ratio and the RMSE, it performs worse than the panel model in terms of all these measures.

Closely related with the choice of model type is the question of out-of-sample performance. All robustness checks so far relied on the in-sample performance and one could argue that this good in-sample performance is due to the fact that trends of the score are captured by the fixed effect. Thus we

Table 3: Test of equality of the root mean squared error (RMSE) for different specifications. This table compares the RMSE for different specifications and lag lengths. The prefered model is given in paranthesis. The RMSE is given for predictions of three different specifications of η in $Score_t = \beta Score_{t-lag} + \gamma Rating_{t-lag} + \eta$ as fixed effect (A), time series (B) without the panel error term u_i and in a standard pooled setting (C). The former two models are estimated with GMM, while the latter is an OLS regression.

Lag	(A) = (B) Fixed Effect = Times Series	(A) = (C) Fixed Effect = OLS	(B) = (C) Time Series = OLS
1	-10.05	-10.21	1.00
	(A)	(A)	(C)
2	-15.44	-15.88	2.28
	(A)	(A)	(C)
3	-17.75	-18.22	2.73
	(A)	(A)	(C)
4	-17.01	-17.38	2.39
	(A)	(A)	(C)

split the dataset into the estimation period 1985-2000 and the testing period 2001-2005. The fixed effects model performs better than a naive prediction of extrapolating the scoring model's prediction using the highest lag of each estimation step. On average the rank correlation of the 'true' score and the fixed effects prediction is 3.2 percentage points higher than the correlation with the extrapolated score.[10]

Concluding we can say that the panel specification used here is stable with respect to the coefficients significance, the models in-sample performance is good both in terms of statistical and economically relevance and the out-of-sample performance is still appropriate.

[10]Note that the fixed effects model is estimated on de-meaned variables and the prediction error includes changes in each panel unit's mean over time.

Refining the specification It is interesting to examine whether the results obtained so far differ across the rating universe. The most obvious split of the rating range is into investment grade and sub-investment grade ratings. Including an interaction term of both categories together gives the following model, where $\mathbb{I}$ denotes an indicator function:

$$y_{it} = \beta y_{i,t-l} + \gamma_1 \mathbb{I}\{RAT_{i,t-l} \leq 10\} RAT_{i,t-l} + \gamma_2 \mathbb{I}\{RAT_{i,t-l} > 10\} RAT_{i,t-l} + \eta_{it}. \tag{3}$$

Table 4 reports the results. For none of the five lags, the hypothesis of equal coefficients can be rejected at a significance of 5%. There is thus no clear evidence that the predictive ability of the rating agency differs across the rating spectrum.

Another refinement is the examination of the agency's predictive ability over the size of the company. The idea is that larger companies tend to issue more public information. Furthermore they often issue relatively more debt and thus are under closer control of both the rating agencies and the financial market. Here we measure size as the logarithm of total assets, but the results are stable for other definitions such as total sales etc. The universe of companies is split into ten deciles according to a company's average size across the sample period.

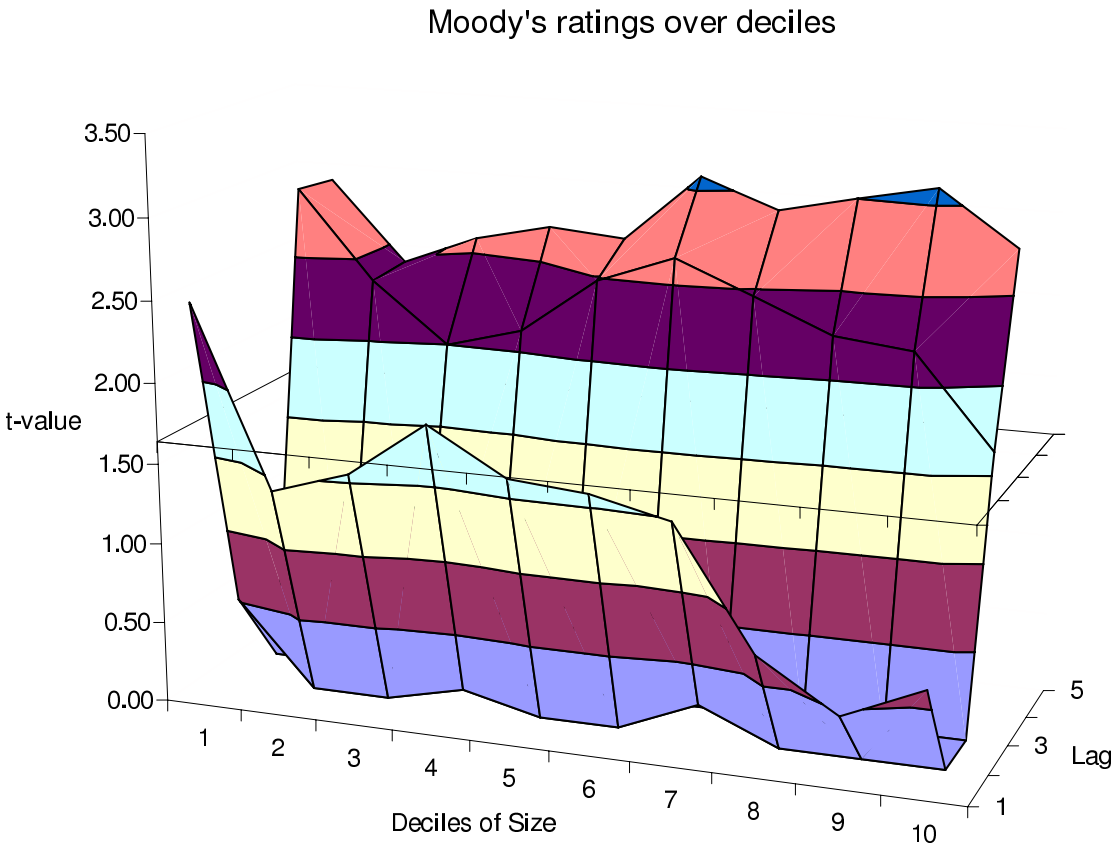
Figure 2 shows the t-values of the rating's coefficient over different lags and those deciles of size.

While the lagged rating becomes more significant with longer lags for all deciles, there is a tendency of a better predictive ability in the smaller companies for smaller lag-lengths and for larger companies and higher lag-lengths. However the results are not too pronounced, so one is to conclude that size does not matter much in our setting.

Table 4: *Multivariate predictive ability across the rating universe.* This table shows the coefficients γ_1 for investment grade ratings (inv.) and γ_2 for speculative grade ratings (spec.) Robust t-values are given in parenthesis. The p-values for the null hypothesis of equal coefficients $\gamma_1 = \gamma_2$ are given in the column "p".

Lag	Inv. (γ_1)	Spec. (γ_2)	p($\gamma_1 = \gamma_2$)
1	0.004	0.003	0.09
	(2.98)	(2.96)	
2	0.002	0.002	0.68
	(0.89)	(0.89)	
3	0.001	0.001	0.72
	(0.41)	(0.47)	
4	0.006	0.005	0.05
	(3.24)	(3.21)	
5	0.007	0.007	0.11
	(3.05)	(3.06)	

Figure 2: t-values of rating for $Score_t = \beta Score_{t-lag} + \gamma Rating_{t-lag}$ over deciles of size.



4 Univariate results of the dynamic panel model

We now turn to the question whether ratings also predict individual financial ratios which are relevant for the credit risk of a company. Using several financial ratios proposed in the literature, Table 5 presents the coefficients of the lagged rating together with the robust t-statistics in parenthesis. The maximal lag significant on a two-sided 5% error level is marked bold.

The logarithm of assets (LN(A)) is significant for the whole five years.[11]

[11]The term "significance" refers to a confidence level of 5%. While the term "marginally significant" denotes a 10% confidence level.

Table 5: *Univariate results of the dynamic panel model.* This table shows the coefficient of the rating when estimating Equation 1. Robust t-values are given in parenthesis. Bold coefficients indicate the highest lag with significance on a 5% level, while italic coefficients mark the highest lag significant on a 10% level.

	LN(A)	ME2D	D2C	LTD2TA	TD2TA	TL2TA	WC2TA	EBIT2TA	NI2TA	OI2S	RE2TA	S2TA	PTIC
1	-0.018	0.018	0.005	0.005	0.000	0.002	0.002	0.001	-0.016	0.000	-0.007	0.005	-0.313
	(-13.43)	(0.40)	(1.94)	(6.43)	(-0.11)	(1.01)	(3.52)	(2.44)	(-0.34)	(-0.22)	(-4.42)	(4.65)	(-2.32)
2	-0.026	-0.014	0.008	0.006	0.006	0.001	0.003	0.002	0.014	0.001	-0.008	0.007	*-0.238*
	(-14.49)	(-0.24)	(2.69)	(6.67)	(0.83)	(0.37)	(3.98)	(4.78)	(1.87)	(1.29)	(-3.33)	(4.93)	(-1.85)
3	-0.030	-0.057	0.012	0.006	0.018	0.003	0.003	**0.002**	0.002	0.002	-0.011	0.009	*-0.219*
	(-13.94)	(-0.87)	(4.03)	(5.88)	(1.65)	(1.29)	(3.04)	(4.37)	(0.47)	(1.35)	(-3.86)	(5.30)	(-1.70)
4	-0.032	**-0.171**	0.010	0.004	**0.023**	0.007	**0.002**	0.001	-0.001	-0.001	-0.013	0.008	-0.206
	(-11.48)	(-2.53)	(3.08)	(3.96)	(2.31)	(2.55)	(2.06)	(1.54)	(-0.31)	(-0.60)	(-4.84)	(4.56)	(-1.72)
5	**-0.025**	*-0.138*	**0.007**	**0.003**	0.012	**0.007**	*0.002*	*0.001*	**0.005**	0.001	**-0.012**	**0.007**	-0.135
	(-7.83)	(-1.90)	(2.45)	(3.06)	(1.04)	(2.51)	(1.93)	(1.70)	(2.33)	(1.37)	(-4.55)	(3.30)	(-1.03)

This variable proxies the size of a company, which changes relatively slowly over the years. Although the size's history explains the contemporary firm size well, the rating adds explanatory power across the five different prediction horizons.

The next five ratios of Table 5 proxy the leverage of a company, the extent to which it is financed with debt. The market leverage (ME2D) defined as market equity divided by debt (sum of short-term and long-term debt) is predicted with statistical significance for prediction horizons of four and five years. The rating is insignificant for shorter horizons.

Debt to total capitalization (D2C) is based on book values, it is debt divided by the sum of debt and book value of equity. Ratings contribute significantly on all five lags. Long-term debt divided by total assets (LTD2A) measures the proportion of loans and obligations with a maturity of more than one year on the total book value of assets. Again, all five lags are significant.

The next leverage ratio, total liabilities over total assets (TL2TA) sums all liabilities including e.g. deferred payments and divides them by total assets. The rating again shows a more long-term predictability, with significance in the fourth and fifth lag.

The next ratio, working capital over total assets (WC2TA), measures the short-term liquidity of the company. The rating's significance ends at the fourth lag, the fifth however is very close to the 5% confidence level.

EBIT over total assets (EBIT2TA) measures the current profitability, as does net income over total assets (NI2TA) and operating income over sales (OI2S). In many cases, coefficients are not significant, but when they are they do not have the expected sign. We would expect a negative relationship between profitability and ratings because higher profitability should be associated with a lower rating value (i.e. a better rating), and vice versa.

Given that profitability is only one component of credit quality, and not the most important one, the results should not cause too much concern. They could reflect a missing-variable bias as other credit quality indicators that influence the rating are missing in the regression equation.

The next ratio's coefficients show the theoretically predicted sign, with significance at a high level. The ratio retained earnings over total assets (RE2TA) generally proxies for the historic profitability. In the regression here, we control for lagged retained earnings so what is left to explain for the rating is the change in RE2TA. More precisely, $RE2TA_t$ as seen at time $t-l$ proxies for the average profitability from $t-l$ to t because retained earnings are previous year's retained earnings plus profits minus pay-outs. Since it should be easier to predict the average profitability over five years rather than, say, the change in profitability from year 0 to year 5, it is not surprising that the predictive ability of ratings is better than for the other profitability ratios discussed above.

Sales over total assets (S2TA) measures the cash-generating ability of a company. Coefficients of ratings are significant but have the wrong sign. Similar to the profitability measures, this could be due to missing-variable biases. S2TA does not significantly contribute to default prediction in our sample (See section 3).

Pretax interest coverage (PTIC), finally, is defined as pretax earnings over interest expenses and measures to what extent interest payments can be covered out of current earnings. The first four lags are marginally significant.

One way of bringing the ratio-specific results together is to examine whether ratings predict ratios relatively well if they are relatively important for default prediction. We measure the default prediction relevance of ratio XYZ through the Pseudo - R^2 from a logit regression of defaults on ratio XYZ. We correlate

these R^2s with the maximum significant lag length from Table 5 and get a correlation of 48.7% with a p-value of 6.56%. The rank correlation is at 54.57% (p-value of 3.54%).

The differences in significance that we see in Table 5 therefore meet our expectations as Moody's ratings tend to exhibit longer foresight for variables that are more relevant for credit quality.

5 Conclusion

We have used a dynamic panel model to examine whether credit rating agencies achieve what they claim to achieve, namely, provide forward-looking assessments of credit quality. We find that Moody's ratings are useful for predicting individual financial ratios over a horizon of up to five years. Ratings also predict a multivariate credit score, again over five years. The contribution of ratings appears to be economically significant and robust for different specifications.

The results are consistent with the agencies' claims and with prior empirical studies. When it comes to default prediction, the common finding is that ratings underperform alternative predictors in the short run. With a default prediction horizon of three to five years, the difference becomes smaller, and ratings can even outperform alternative models.

The results of our analysis suggest that rating agencies are relatively good at long-term default prediction because they are able to predict the future evolution of credit risk drivers (cf. Altman and Rijken (2006), Löffler (2007)). Through their qualitative analysis, rating analysts therefore achieve what the literature on quantitative default prediction is just beginning to work on, i.e. model the dynamics of credit risk drivers (cf. Duffie et al. (2007)).

References

Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankcrupty. *Journal of Finance*, 23:589–609.

Altman, E. I. and Rijken, H. (2006). A point-in-time perspective on through-the-cycle ratings. *Financial Analyst Journal*, 62:54–70.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 2nd edition.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics*, 68:29–51.

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

Baltagi, B. H. (2001). *Econometric Analysis of Panel Data*. Wiley, 2nd edition.

Baum, C. f., Schaffer, M. E., and Stillman, S. (2003). Instrumental variables and gmm: Estimation and testing. Working Paper No. 545, Boston College.

Blume, M. E., Lim, F., and MacKinlay, A. C. (1998). The declining credit quality of US corporate debt: Myth or reality? *Journal of Finance*, 53(4):1389–1413.

Bond, S. (2002). Dynamic panel data models: A guide to micro data methods and practice. Working Paper CWP09/02, Cemmap. Centre for microdata methods and practice.

Cantor, R. and Mann, C. (2003). Measuring the performance of corporate bond ratings. Special comment, Moody's Investor Service.

Cragg, J. and Donald, S. (1993). Testing identifability and specification in instrumental variables models. *Econometric Theory*, 9:222–240.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford University Press, Oxford.

Delianedis, G. and Geske, R. (1999). Credit risk and risk neutral default probabilities: Information about migrations and defaults. Working Paper, Anderson School, UCLA, Los Angeles.

Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate failure prediction with stochastic covariates. *Journal of Financial Economics*, 83:635–665.

Engelmann, B., Hayden, E., and Tasche, D. (2003). Testing rating accuracy. *Risk*, 16:82–86.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 4th edition.

Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall, New Jersey, 5th edition.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054.

Hausman, J. and Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica*, 49:1377–1398.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Honore, B. (1993). Orthogonality conditions for tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics*, 59:35–61.

Hull, J., Predescu, M., and White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking and Finance*, 28(11):2789–2811.

Jianming Kou, J. and Varotto, S. (2007). Timeliness of spread implied ratings. *European Financial Management*, forthcoming.

Kealhofer, S. (2003). Quantifying credit risk I: Default prediction. *Financial Analysts Journal*, 59(1):30–44.

Kisgen, D. J. (2006). Credit ratings and capital structure. *Journal of Finance*, 61:1035–1072.

Lee, M. J. (2001). First-difference estimator for panel censored-selection model. *Economic Letters*, 70:43–49.

Löffler, G. (2007). The complementary nature of ratings and market-based measures of default risk. *Journal of Fixed Income*, 17:38–47.

Maddala, G. and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, 61:631–652.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Sargan, J. D. (1958). The estimation of economic realtionships using instrumental variables. *Econometrica*, 26:393–415.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74:101–124.

Sobehart, J. R. and Keenan, S. C. (2001). Measuring default accurately. *Risk (Credit Risk Special Report)*, 14:31–33.

Vella, F. and Verbeek, M. (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics*, 90:239–263.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68:115–132.

Yogo, M. and Stock, J. H. (2005). Testing for weak instruments in linear IV regression. In Andrews, D. and Stock, J., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, Cambridge.

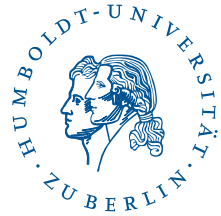
SFB 649 Discussion Paper Series 2007

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Trade Liberalisation, Process and Product Innovation, and Relative Skill Demand" by Sebastian Braun, January 2007.
002 "Robust Risk Management. Accounting for Nonstationarity and Heavy Tails" by Ying Chen and Vladimir Spokoiny, January 2007.
003 "Explaining Asset Prices with External Habits and Wage Rigidities in a DSGE Model." by Harald Uhlig, January 2007.
004 "Volatility and Causality in Asia Pacific Financial Markets" by Enzo Weber, January 2007.
005 "Quantile Sieve Estimates For Time Series" by Jürgen Franke, Jean-Pierre Stockis and Joseph Tadjuidje, February 2007.
006 "Real Origins of the Great Depression: Monopolistic Competition, Union Power, and the American Business Cycle in the 1920s" by Monique Ebell and Albrecht Ritschl, February 2007.
007 "Rules, Discretion or Reputation? Monetary Policies and the Efficiency of Financial Markets in Germany, 14th to 16th Centuries" by Oliver Volckart, February 2007.
008 "Sectoral Transformation, Turbulence, and Labour Market Dynamics in Germany" by Ronald Bachmann and Michael C. Burda, February 2007.
009 "Union Wage Compression in a Right-to-Manage Model" by Thorsten Vogel, February 2007.
010 "On σ−additive robust representation of convex risk measures for unbounded financial positions in the presence of uncertainty about the market model" by Volker Krätschmer, March 2007.
011 "Media Coverage and Macroeconomic Information Processing" by Alexandra Niessen, March 2007.
012 "Are Correlations Constant Over Time? Application of the CC-TRIG$_t$-test to Return Series from Different Asset Classes." by Matthias Fischer, March 2007.
013 "Uncertain Paternity, Mating Market Failure, and the Institution of Marriage" by Dirk Bethmann and Michael Kvasnicka, March 2007.
014 "What Happened to the Transatlantic Capital Market Relations?" by Enzo Weber, March 2007.
015 "Who Leads Financial Markets?" by Enzo Weber, April 2007.
016 "Fiscal Policy Rules in Practice" by Andreas Thams, April 2007.
017 "Empirical Pricing Kernels and Investor Preferences" by Kai Detlefsen, Wolfgang Härdle and Rouslan Moro, April 2007.
018 "Simultaneous Causality in International Trade" by Enzo Weber, April 2007.
019 "Regional and Outward Economic Integration in South-East Asia" by Enzo Weber, April 2007.
020 "Computational Statistics and Data Visualization" by Antony Unwin, Chun-houh Chen and Wolfgang Härdle, April 2007.
021 "Ideology Without Ideologists" by Lydia Mechtenberg, April 2007.
022 "A Generalized ARFIMA Process with Markov-Switching Fractional Differencing Parameter" by Wen-Jen Tsay and Wolfgang Härdle, April 2007.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".



023 "Time Series Modelling with Semiparametric Factor Dynamics" by Szymon Borak, Wolfgang Härdle, Enno Mammen and Byeong U. Park, April 2007.
024 "From Animal Baits to Investors' Preference: Estimating and Demixing of the Weight Function in Semiparametric Models for Biased Samples" by Ya'acov Ritov and Wolfgang Härdle, May 2007.
025 "Statistics of Risk Aversion" by Enzo Giacomini and Wolfgang Härdle, May 2007.
026 "Robust Optimal Control for a Consumption-Investment Problem" by Alexander Schied, May 2007.
027 "Long Memory Persistence in the Factor of Implied Volatility Dynamics" by Wolfgang Härdle and Julius Mungo, May 2007.
028 "Macroeconomic Policy in a Heterogeneous Monetary Union" by Oliver Grimm and Stefan Ried, May 2007.
029 "Comparison of Panel Cointegration Tests" by Deniz Dilan Karaman Örsal, May 2007.
030 "Robust Maximization of Consumption with Logarithmic Utility" by Daniel Hernández-Hernández and Alexander Schied, May 2007.
031 "Using Wiki to Build an E-learning System in Statistics in Arabic Language" by Taleb Ahmad, Wolfgang Härdle and Sigbert Klinke, May 2007.
032 "Visualization of Competitive Market Structure by Means of Choice Data" by Werner Kunz, May 2007.
033 "Does International Outsourcing Depress Union Wages? by Sebastian Braun and Juliane Scheffel, May 2007.
034 "A Note on the Effect of Outsourcing on Union Wages" by Sebastian Braun and Juliane Scheffel, May 2007.
035 "Estimating Probabilities of Default With Support Vector Machines" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, June 2007.
036 "Yxilon – A Client/Server Based Statistical Environment" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, June 2007.
037 "Calibrating CAT Bonds for Mexican Earthquakes" by Wolfgang Härdle and Brenda López Cabrera, June 2007.
038 "Economic Integration and the Foreign Exchange" by Enzo Weber, June 2007.
039 "Tracking Down the Business Cycle: A Dynamic Factor Model For Germany 1820-1913" by Samad Sarferaz and Martin Uebele, June 2007.
040 "Optimal Policy Under Model Uncertainty: A Structural-Bayesian Estimation Approach" by Alexander Kriwoluzky and Christian Stoltenberg, July 2007.
041 "QuantNet – A Database-Driven Online Repository of Scientific Information" by Anton Andriyashin and Wolfgang Härdle, July 2007.
042 "Exchange Rate Uncertainty and Trade Growth - A Comparison of Linear and Nonlinear (Forecasting) Models" by Helmut Herwartz and Henning Weber, July 2007.
043 "How do Rating Agencies Score in Predicting Firm Performance" by Gunter Löffler and Peter N. Posch, August 2007.

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".