DISCUSSION PAPER SERIES

No. 3415

THE ROLE OF NON-FINANCIAL FACTORS IN INTERNAL CREDIT RATINGS

Jens Grunert, Lars Norden and Martin Weber

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP3415.asp

ISSN 0265-8003

THE ROLE OF NON-FINANCIAL FACTORS IN INTERNAL CREDIT RATINGS

Jens Grunert, Universität Mannheim
Lars Norden, Universität Mannheim
Martin Weber, Universität Mannheim and CEPR

Discussion Paper No. 3415
June 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3415

June 2002

ABSTRACT

The Role of Non-financial Factors in Internal Credit Ratings*

Internal credit ratings are expected to gain in importance because of their potential use for determining regulatory capital adequacy and banks' increasing focus on the risk-return profile in commercial lending. Therefore, the components of internal credit ratings merit not only a qualitative but also a quantitative analysis. Whereas the eligibility of financial factors as inputs for credit ratings is widely accepted, the role of non-financial factors remains ambiguous. Analysing credit file data from four major German banks, we find evidence that the combined use of financial and non-financial factors leads to a more accurate prediction of current and future default events than the single use of each of these factors respectively.

JEL Classification: G21
Keywords: credit ratings, credit risk, debt default and PROBIT analysis

Jens Grunert
Lehrstuhl für ABWL
Finanzwirtschaft
Insbesondere Bankbetriebslehre
Universität Mannheim, L5, 2
68131 Mannheim
GERMANY
Email: grunert@bank.bwl.uni-mannheim.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157748

Lars Norden
Lehrstuhl für ABWL
Finanzwirtschaft
Insbesondere Bankbetriebslehre
Universität Mannheim, L5, 2
68131 Mannheim
GERMANY
Email: norden@bank.bwl.uni-mannheim.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157749

Martin Weber
Lehrstuhl für ABWL
Finanzwirtschaft
Insbesondere Bankbetriebslehre
Universität Mannheim, L5, 2
68131 Mannheim
GERMANY
Tel: (49 621) 181 1532
Fax: (49 621) 181 1534
Email: weber@bank.bwl.uni-mannheim.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=119269

* We thank Gunter Löffler, Achim Machauer, Steven Ongena, and seminar participants at the University of Mannheim for valuable comments and insights.

Submitted 17 May 2002

1 Introduction

Similar to capital market investors that use credit ratings provided by rating agencies banks assign internal credit ratings to appraise the creditworthiness of their borrowers. In both cases ratings can be interpreted as a screening technology that is used to alleviate asymmetric information problems between borrowers and lenders. Whereas external ratings are well established since the beginning of the twentieth century, internal ratings have been increasingly used by banks during the nineties (see English/Nelson (1998), Treacy/Carey (2000)). Internal credit ratings for corporate borrowers can be described as an aggregation of a valuation procedure of various financial and non-financial factors. Ratings are generally used for loan approval, pricing, monitoring, and loan loss provisioning. While considerable research has proved the suitability of financial factors to predict borrower insolvency (see, for example, Altman (1968)), the role of non-financial factors remains ambiguous. Although the use of non-financial factors such as management quality and industry perspectives is beyond controversy (see Basel Committee on Banking Supervision (2000a, 2001) and Günther/Grüning (2000)) there is a lack of quantitative research on this issue. With respect to these "soft" factors bankers often refer to their experience and to a distrust in the sole use of financial criteria. A first investigation of the importance of soft information in borrower-bank relationships is conducted by Berger et al. (2002) and Stein (2002). Depending on bank size, Berger et al. (2002) explore a bank's ability to act in projects that require the evaluation of soft information. They find that small banks are more capable of collecting and acting on soft information than large banks. Stein (2002) points out that decentralized banking hierarchies are likely to be more attractive when projects' soft factors are to be evaluated.

This paper explores the role of non-financial factors in credit ratings. Fur this purpose we examine empirically if the combined use of financial and non-financial factors leads to a more accurate prediction of default events than their single use respectively. The indicator variables for default events defined hereinafter are consistent with the Basel II definition and can be regarded as a benchmark to test the prediction accuracy of different rating categories. Our study has implications for both banks and bank supervisors: Banks will be able to better understand the role of quantitative and qualitative factors in credit ratings and supervisors will be supported in claiming a “mixed” credit rating to determine regulatory capital requirements (see Basel Committee on Banking Supervision (2001)).

The paper proceeds as follows. Section 2 provides an overview of related literature, in particular on the structure of internal rating systems and the properties of non-financial sub-ratings. Section 3 describes the data, the variables and deduces a testable hypothesis. Section 4 analyzes whether a combination of financial and non-financial factors leads to a more accurate prediction of default events than the single use of each of these factors respectively. The paper concludes in section 5.

2 Overview of related literature

In modern theory of financial intermediation the existence of intermediaries is explained with an improvement of welfare that results from a reduction in costs of asymmetric information (see, for example, Leland/Pyle (1977), Diamond (1984), and Bhattacharya/Thakor (1993) for a detailed survey). Many of these models presume that banks screen and monitor borrowers at a given cost but they rarely specify the technology that is or should be applied. Since the latter issue is closely connected to our study, we outline three lines of related literature in the following. First, research on the prediction of corporate bankruptcy on the basis of financial factors is presented. Secondly, empirical and normative research on banks’ internal credit rating systems is reviewed. Finally, literature concerning the components of credit ratings, both quantitative and qualitative, is described.

Much work has been dedicated to develop models that are able to predict corporate bankruptcy on the basis of financial factors (see Altman (1968), Altman/Haldeman/Narayan (1977), Platt/Platt (1990), Baetge (1998)). These factors typically concern the capital structure, profitability and liquidity of a firm. Models are based on linear discriminant analysis, on logit/probit regression analysis or, more recent ones, on neural networks. Because of their relative high discriminary power these models are widely accepted but they nevertheless show some disadvantages (see Basel Committee on Banking Supervision (2000b, pp. 107-110)). Few of them are based on a theory that explains why and how certain financial factors are linked to corporate bankruptcy. As financial factors are mostly backward-looking point-in-time measures these models are inherently constrained and it is not clear how well these models perform out-of-sample (time, firm, industry etc.). This research area is relatively well developed but still has to overcome the above mentioned problems.

Research on banks' internal credit rating systems is still scarce but growing considerably. It can be divided into an empirical and a normative part. On the one hand, empirical analyses of banks' internal rating systems examine the structure and the use of ratings (see English/Nelson (1998), Machauer/Weber (1998), Treacy/Carey (2000), Crouhy/Galai/Mark (2001), Ewert/Szczesny (2001), Norden (2001)). These studies and an overview of international best practice rating standards in the banking industry (see Basel Committee on Banking Supervision (2000a)) show that internal rating systems are based on either statistical methods, constrained expert judgment-based techniques or exclusively expert judgments. These systems tend to include similar types of risk factors, typically a mix of quantitative and qualitative factors (e.g. leverage, profitability and liquidity ratios, management experience, industry perspectives). The weighting schemes of these risk factors differ considerably across banks. Ratings are used for loan approval, management reporting, pricing, limit setting, and loan loss provisioning. Additionally, the frequency and the extent of banks' rating disagreement for a given borrower are analyzed (see Risk Management Association (2000) and Carey (2001)). In addition to the reasons given in these studies we argue that differences in opinion about borrower quality are more likely to stem from a different valuation of non-financial factors than from financial factors.

On the other hand, Krahnen/Weber (2001) present a normative set of „Generally accepted rating principles" that points out the necessity of a link between credit rating and probability of default. Requirements concerning completeness, definition of default, monotonicity, back testing etc. of a rating system are developed. They describe credit ratings as being a „mixture of mathematical models and management intuition", but they say nothing about the risk factors, the factor weights and the value function to be included in a „good" rating. Based on the first consultation period and several own studies the Basel Committee on Banking Supervision (2001) released a second Consultative Document in January 2001 which contains the proposal of an internal ratings-based approach for regulatory capital adequacy. This document includes an extensive list of normative requirements banks have to meet if they want to calculate regulatory risk weights based on their internal credit ratings.

Finally, Günther/Grüning's (2000) survey reports that 70 of 145 German banks use not only quantitative but also qualitative factors in credit risk assessment with management quality being the most important "soft" factor. 77.6 % of these banks answer that qualitative factors clearly improve default prediction. However, nothing is said about the degree of improvement. Hesselmann (1995) and Blochwitz/Eigermann (2000) incorporate qualitative variables (for example accounting behavior or discrete cover ratio classes) in discriminant analysis to differentiate between subsequently defaulting and non-defaulting German companies. They find that the use of qualitative variables improve the percentage of companies correctly classified. These results support the requirement of the Basel Committee on Banking Supervision (2001) that banks have to consider not only quantitative but also qualitative factors, for example, the availability of audited financial statements, depth and skill of management, the position within the industry and future prospects (see n° 265 in the second Consultative Document). Furthermore, analyses of sub-ratings using different sets of credit file data from German banks (see Weber et al. (1999), Brunner et al. (2000)) show that qualitative sub-ratings exhibit significantly better grades with less dispersion around their mean, that they change less often than quantitative sub-ratings and that the origins of rating changes stem mainly from changes in the quantitative sub-ratings. They leave open the question whether the important role of "soft" information in internal credit ratings is a desirable or problematic feature.

Against the background of the portrayed literature it becomes clear that more needs to be known about the specific role of and interaction between different risk factors in internal credit rating systems. Whereas the importance of financial factors is widely accepted because its impact is measurable, the relevance of non-financial factors is mainly considered in a holistic manner. These factors are usually chosen on the basis of experts' judgments and common industry knowledge but how much do they contribute to an accurate forecast of borrower quality? We intend to answer this question in the remainder.

3 Data, variables and hypothesis

Our data on bank-borrower relationships consists of credit files from four major German banks including 160 corporate borrowers from the period January 1992 to December 1996.[1] The data set was restricted to medium-sized firms with an annual turnover between EUR 25 and 250 million and a minimum loan size of EUR 1.5 million. To avoid the influence of the restructuring process in the eastern part of Germany only customers of the western part are included. The meta rating scale with grades from 1 to 6 was created to make internal ratings comparable between banks (for details see Elsas et al. (1998)). Grade 1 means very good, 2 good or above average, 3 average, 4 below average, 5 problematic and 6 very much in danger of default. Some variables were not documented in the credit files because not all relationships lasted for five years and the creditworthiness of high quality borrowers was not checked annually but every second year at one bank.

In our analysis an observation consists of a borrower's financial, non-financial, and overall rating in a particular year as well as default information for the year of the rating assignment and the subsequent year. All variables used in the further analyses are summarized in table 1.

[1] See Elsas et al. (1998) for a detailed description of the original sample which consists of two randomly taken sub-samples (A and P) of credit files from six major German banks including 240 borrowers. Sample A was randomly drawn from a predefined population and sample P was randomly drawn from a sub-population which consists of borrowers in financial distress during 1992-1996. We merged these sub-samples A and P, controlling for a potential oversampling bias, to obtain a higher number of observations, especially in order to increase the number of default events. In our study bank 5 was eliminated due to a lack of non-financial factors and bank 6 because of the small number of observations. Since all firms in the sample borrow exclusively from one of the four banks (or other banks that are not in the sample) we could not compare multiple rating assignments of borrowers from different lenders as done by Risk Management Association (2000) and Carey (2001).

Table 1: Description of variables

Variable	Description
Default dummy variables	
FDEF	= 1 if default occurred in the year following the one of the considered rating
SDEF	= 1 if default occurred in the same year as the considered rating
Rating categories	
FR	Financial rating with grades 1 to 6
NFR	Non-financial rating with grades 1 to 6
OR	Overall rating with grades 1 to 6
Non-financial factors	
MGT	Non-financial factor "Management quality"
MKT	Non-financial factor "Market position"
Financial factors	
LTA	Logarithm of total assets
ER	Equity-to-assets ratio
CR	Current ratio
CFNL	Cash flow-to-net liabilities
CFTA	Cash flow-to-total assets
ICR	Interest coverage ratio
CIR	Capital intensity ratio
ROA	Return on assets
ROS	Return on sales
ROE	Return on equity
Bank dummy variables	
B1, B2, B3, B4	= 1 if bank 1, 2, 3, 4 is the lender
Year dummy variables	
Y1992, Y1993, Y1994, Y1995	= 1 if observation is from 1992, 1993, 1994, 1995

The variables FDEF and SDEF are indicators for default events. Consistent with the definition given by the Basel Committee of Banking Supervision (2001) (see n° 272 in the second Consultative Document), the variable FDEF equals 1 if one or more of the following sub-events occur in the year following the one of the considered rating assignment and otherwise zero: Moratorium, allowance of loan loss provisions, withdrawal of a credit, disposition of collaterals, liquidation, formation of a bank pool, recapitalization. The variable SDEF is defined identically except that it refers to default events in the same year as the considered rating assignment. The financial, non-financial and overall ratings are directly adopted from the original credit files of each bank and transformed accordingly to the overall rating on the meta rating scale. We obtain the non-financial factors (management quality, market position) directly from the credit files, whereas we have to compute the financial factors, some of them are integral parts of the financial rating, because they are not in the dataset.[2] These factors cover all categories of the C's of credit (excepting collateral), a familiar credit analysis concept in commercial lending (see Collins (1966)). Dummy variables are created to control for bank and year specific effects. Table 2 shows the distribution of the default variable FDEF by banks, years and overall rating classes. Panel A shows that default events are agglomerated at bank 2 but quite evenly distributed acrosss banks 1, 3 and 4. Note that this agglomeration of default events at bank 2 is not a problem because our results are not sensitive to the omission of bank 2 from the sample (see robustness checks in section 4.1 and 4.2).

Table 2: Distribution of default events

Panel A, B, and C present the distribution of the default variable FDEF by banks, years and overall rating classes. FDEF takes the value 1 if default occurred in the year following the one of the considered rating assignment and 0 otherwise.

Panel A: Default events by banks

Bank	FDEF=0	FDEF=1	Total	% of all obs.
1	66	3	69	16.87
2	110	52	162	39.61
3	84	11	95	23.23
4	80	3	83	20.29
total	340	69	409	100

Panel B: Default events by years

Year	FDEF=0	FDEF=1	Total	% of all obs.
1992	68	11	79	19.32
1993	94	11	105	25.67
1994	98	18	116	28.36
1995	80	29	109	26.65
total	340	69	409	100

Panel C: Default by overall rating classes

Overall rating	FDEF=0	FDEF=1	Relative default frequency (%)	Total	% of all obs.
1	18	0	0.00	18	4.40
2	61	1	1.61	62	15.16
3	120	11	8.40	131	32.03
4	99	25	20.16	124	30.32
5	36	20	35.71	56	13.69
6	6	12	66.67	18	4.40
total	340	69		409	100

[2] See appendix for detailed definitions.

Whereas Panel B indicates a relatively even distribution of the default events across years a monotonous increase of the relative default frequency from rating class 1 to 6 can be observed in Panel C.

Table 3 displays descriptive statistics of different rating categories. The means of all three credit rating categories are higher for defaulters than for non-defaulters. This is a first hint for a robust relation between credit ratings and default status. The standard deviations of the different rating categories indicate that the dispersions of defaulters' ratings are lower, which may be caused by the fact that default events occur mainly in the grades 5 and 6. Similar to the study of Weber et al. (1999) the standard deviation of non-financial ratings is lower than the one of financial ratings. Furthermore, non-financial ratings are significantly better at the 0.01-level than financial ratings using a Wilcoxon ranksum test.

Table 3: Descriptive statistics for credit rating categories

Panel A: Credit ratings and default status in the following year (variable FDEF)

	Mean	Std.Dev.	Mean FDEF=0	Std.Dev.	Mean FDEF=1	Std.Dev.
Financial rating	3.72	1.58	3.45	1.50	5.07	1.26
Non-financial rating	3.51	1.15	3.30	1.07	4.54	0.96
Overall rating	3.47	1.17	3.27	1.09	4.45	1.01
No. of observations	409		340		69	

Panel B: Credit ratings and default status in the same year (variable SDEF)

	Mean	Std.Dev.	Mean SDEF=0	Std.Dev.	Mean SDEF=1	Std.Dev.
Financial rating	3.72	1.58	3.54	1.51	5.45	1.15
Non-financial rating	3.51	1.15	3.36	1.06	4.93	0.86
Overall rating	3.47	1.17	3.30	1.07	5.00	0.82
No. of observations	409		369		40	

We now turn to the formulation of our hypothesis. Since the objective of appraising a borrower's creditworthiness is to specify his probability of default over a given time horizon (usually one year), banks should not only use backward-looking "hard" financial data but also some forward-looking "soft" information. Internal ratings of banks are usually based on borrowers' current condition (point-in-time) whereas rating agencies follow a "through the cycle" approach projecting borrowers' condition on an entire economic cycle (see Treacy/Carey (2000) and Löffler (2001)). Accordingly, we infer the following hypothesis claiming that the combined use of financial and non-financial factors improves default prediction relative to a single use of either financial or non-financial factors respectively:

> *A combination of financial and non-financial factors leads to a more accurate prediction of default than the single use of either financial or non-financial factors.*

4 Measuring the relation between credit ratings and default events

Our main issue is whether an additional inclusion of non-financial factors in a bank's internal credit rating is beneficial or not. It can be deemed beneficial if it leads to a more accurate prediction of default events. Therefore, we test the above established hypothesis by comparing the explanatory power of the overall rating with that of the financial rating for default events that occur in either the year following the one of the rating assignment (section 4.1) or in the same year (section 4.2).

4.1 The relation between credit ratings and default events in the following year

The purpose of a credit rating is to classify prospects and borrowers according to their probability of default over a given time horizon. As banks typically assign credit ratings for a one-year horizon (see Treacy/Carey (2000)) we analyze how different rating categories are related to the default status in the year following a rating assignment. For this purpose, we compare credit ratings assigned in the year t with the variable FDEF (default in t+1).

The relation between ratings and default events can be measured in several ways. Starting with the most simple approach, we use appropriate statistical association measures, then we turn to probit analysis to estimate default probability models. Since default is a dichotomous variable and credit ratings are ordinal variables it is not reasonable to apply Pearson's correlation coefficient because it is exclusively designed to compare two metrical variables. Instead, rank correlations and concordance coefficients are suitable. Especially, Kendall's τ_b is a convenient measure because it takes into account the existence of ties[3] in grouped data (see Kendall/Gibbons (1990)).[4] Since we intend to explore the additional benefit of non-financial factors we contrast the financial rating FR and the overall rating OR (consisting of financial and non-financial factors). Table 4 compares the strength of association between default and the financial rating on the one hand with that of default and the overall rating on the other hand using Spearman's ρ_s and Kendall's τ_b:

[3] Ties between comparison pairs can either be present within the rating variable or within the default variable.

[4] Kendall's τ_b calculates the difference between concordant and discordant pairs relative to the total number of comparison pairs without ties and can take values from –1 (maximal discordance) to 1 (maximal concordance). To classify a pair of observations as concordant or discordant two different criteria are needed (in our context FDEF and rating). For example (FDEF=0, rating=2) vs. (FDEF=1, rating=3) is considered as concordant because the value of both criteria shifts into the same direction (deteriorates) whereas (FDEF=1, rating=4) vs. (FDEF=0, rating=6) is considered as discordant.

Table 4: Rank correlation and concordance between ratings and future default events

ρ_s is Spearman's rank correlation coefficient and τ_b is Kendall's coefficient of concordance (correcting for ties). The dummy variable FDEF indicates a default event one year ahead, FR is a borrower's financial rating, and OR is the overall credit rating from the indicated year respectively.

(1) Year	(2) Obs.	(3) $\rho_{s(FDEF,FR)}$	(4) $\rho_{s(FDEF,OR)}$	(5) $\tau_{b(FDEF,FR)}$	(6) $\tau_{b(FDEF,OR)}$	(7) Relation
1992	79	0.3110	0.2717	0.2811	0.2496	–
1993	105	0.3299	0.3668	0.2942	0.3321	+
1994	116	0.4180*	0.3253	0.3704	0.2919	–
1995	109	0.4393	0.4706	0.3908	0.4270	+
Pooled data	409	0.3840	0.3678	0.3414	0.3327	–

***, **, * Significantly different from $\rho_{s(FDEF,OR)}$ in column (4) at the 0.01, 0.05, and 0.10 level.

A positive sign in column (7) indicates that the overall rating exhibits both a higher rank correlation and concordance with the default variable than the financial rating. In three of four years and for the pooled data the rank correlation between the financial rating FR and FDEF is not significantly different from the rank correlation between the overall rating OR and FDEF. Given this equivocal picture we proceed in a more detailed manner using regression analysis.

In the following, probit regression models with FDEF as dependent variable and the financial rating FR (model 1), the non-financial rating NFR (model 2), and the overall rating OR (model 3) respectively as independent variables are estimated.[5] In each model we control for bank and year specific influences with dummy variables using bank 1 and year 1992 as reference categories. The models can be evaluated by using different criteria.[6] We decided to take McFadden's R^2,[7] the Brier-Score[8], the percentage of correctly classified observations, and type I and type II error

[5] In a preparatory analysis dummy variables for the financial, the non-financial and the overall rating (merging rating classes 1 and 2 and taking these ratings as reference category) were used. As this specification basically yields the same results we use the credit rating variables (coded on a scale from 1 to 6).

[6] See Hosmer/Lemeshow (2000), pp. 143-200.

[7] Since the conventional R^2 cannot be calculated for probit and logit models McFadden's R^2 (Pseudo R^2) is employed. It is defined as 1 – (unrestricted log-likelihood function/restricted log-likelihood function).

[8] See Brier (1950). The Brier Score (BS) is a measure of prediction accuracy that is well-known in meteorology and medical science. It is calculated as $BS = \frac{1}{n}\sum_{i}^{n}(\theta_i - p_i)^2$ where θ_i is a binary indicator for the actual realization of

rates[9] as evaluation criteria because they represent a good mix of goodness of fit and classification accuracy measures. We compare the accuracy measures with those of a naive forecast and between models.[10]

Regression results and evaluation criteria for models 1-3 are presented in table 5. All three rating variables have positive coefficients and are significant at the 0.01-level indicating the strong relation between default and credit ratings. Concerning the dummy variables, bank 2 and bank 3 have significant influence on the prediction of default which is consistent with the fact that these two banks show higher average default frequencies than the two other banks. None of the year dummies is significant at the 0.10-level which is consistent with the relatively even distribution of default events over time. The null hypothesis that all three models are compatible with the observed outcomes cannot be rejected using Spiegelhalter's z-statistic (see Spiegelhalter (1986)). All models are more accurate than the naive forecast which leads to a Brier Score of 0.1402. Finally, model evaluation results shown in Panel B confirm that model 3 is superior to model 1 and 2 with respect to all criteria. We find that the Brier Score of model 3 is significantly smaller than the ones of both other models.[11] Moreover, the type I error rate of model 3 (40.58%) is lower than the one of model 1 (43.48%) which roughly means that model 3 detects actual defaulters more accurately than model 1 in three of 100 cases.

---

the default variable (1 if default, 0 if no default) and p_i is the estimated probability of default. The difference between the Brier Score and the percentage of correctly classified observations is that the former is more sensitive to the level of the estimated probabilities. These measures do not process equally the predicted probabilities because the Brier Score takes them directly into account whereas the percentage of correctly classified observations transforms probabilities equal to or higher than 0.50 to 1 and others to 0. Hence, in extreme cases both measures can produce contradictory evaluation results.

[9] Type I error is the percentage classified as "not default" of all observations that actually did default. Type II error is the percentage classified as "default" of all observations that actually did not default.

[10] The Brier Score of a naive forecast is obtained by taking the average relative default frequency (ADF) of the entire sample as default probability for each individual observation: $BS = \frac{1}{n}\left[n_{DEF=1}(ADF-1)^2 + n_{DEF=0}(ADF-0)^2\right]$

[11] The significance test is based on the Williams-Kloot-statistic z_{wk} which is described in detail by Redelmeier/Bloch/Hickam (1991) and Vinterbo/Ohno-Machado (1999).

Table 5: Regressions results and evaluation criteria for models 1-3

The sample used in all three probit regressions is the same and consists of 409 observations from the period 1992-1995. The dependent variable, FDEF, takes the value one if default occurs in the year following the one of the rating assignment and zero otherwise. Model 1 uses in addition to bank and year dummy variables the financial rating FR, model 2 the non-financial rating NFR, and model 3 the overall rating OR as independent variable (instead of "Rating" as indicated in the first column) to estimate the probability of a default event. Coefficients are estimated using the maximum likelihood method.

Panel A: Regression results

Variable	Model 1 (financial rating)		Model 2 (non-financial rating)		Model 3 (overall rating)	
FDEF	Coefficient	Std.Err.	Coefficient	Std.Err.	Coefficient	Std.Err.
Rating	0.5250***	0.0709	0.7884***	0.1008	0.8941***	0.1109
B2	1.8064***	0.3505	1.2130***	0.3622	2.2382***	0.4038
B3	1.2537***	0.3862	0.4345	0.3910	1.0678***	0.4116
B4	0.3411	0.4249	-0.6469	0.4463	0.2862	0.4397
Y1993	-0.4631	0.2934	-0.6495**	0.3131	-0.5865*	0.3002
Y1994	-0.3056	0.2768	-0.2061	0.2769	-0.3354	0.2820
Y1995	0.3666	0.2572	0.4305	0.2646	0.3272	0.2679
Intercept	-4.3714***	0.5296	-4.6006***	0.5570	-5.6879***	0.6705

***, **, * Significantly different from zero at the 0.01, 0.05, and 0.10 level.

Panel B: Evaluation criteria

The null hypotheses BS(model 1)=BS(model 3) can be rejected with a p-value 0.0030 and BS(model 2)=BS(model 3) can be rejected with a p-value 0.0232 using the Williams-Kloot statistic z_{wk} (two-tailed test).

Evaluation criterion	Model 1 (financial rating)	Model 2 (non-financial rating)	Model 3 (overall rating)
McFadden's R^2	0.3562	0.3859	0.4165
Brier Score	0.0861	0.0823	0.0707
% of obs. correctly classified	88.02	89.00	91.69
Type I error %	43.48	47.83	40.58
Type II error %	5.59	3.53	1.76

To avoid an overfit of the models to our sample we subsequently divide the data into an estimation and validation sample and then perform out-of-the-sample validation. As each observation can be characterized by borrower number, bank number, and year, a split up-procedure has to ensure that observations in both samples are independent. This independence criterion is respected by drawing randomly 50% of all borrowers and considering them as the estimation sample, leaving the remainder for the validation sample. A drawback of this method is that one random draw can lead to an extremely favorable or unfavorable distribution of default events between both samples. To overcome this problem the random 50%/50%-split up procedure is repeated 100 times to average out favorable and unfavorable sample distributions. In this context, we compare the extreme cases (models 1 and 3) which is sufficient to investigate the impact of an additional consideration of non-financial factors. All steps of this procedure are summarized below:

1. Random draw of 80 borrowers (estimation sample)
2. The remaining 80 borrowers constitute the validation sample
3. Estimation of models 1 and 3 with observations from the estimation sample
4. Comparison of models using McFadden's R^2, Brier Score, percentage of observations correctly classified[12], and type I and II error rates for the estimation and validation sample separately
5. External Validation of each model by comparing its Brier Score and its percentage of observations correctly classified in- vs. out-of-the-sample
6. Repetition of steps 1 to 5 for 100 times.

Table 6 summarizes aggregated results for models 1 and 3.

[12] For the calculation of predicted values we always use a cut-off point of 0.50 since our interest is model comparison and not the optimization of model sensitivity and specifity.

Table 6: Results of the split-up procedure

The whole sample of 409 observations (160 borrowers) is subdivided in an estimation and validation sample. The split up is done by drawing randomly 80 borrowers and considering them as estimation sample, leaving the remaining 80 borrowers for the validation sample. To avoid extreme favorable or unfavorable partitions the split-up procedure is repeated 100 times and probit models with FDEF as dependent variable are estimated. Model 1 uses the financial rating FR and model 3 the overall rating OR as independent variable.

	Estimation sample		Validation sample	
Mean number of obs.	203.3		205.7	
Mean Brier Score (naive forecast)	0.1408		0.1389	
Mean % of obs. correctly classified (naive forecast)	82.96		83.26	
	Model 1	Model 3	Model 1	Model 3
Mean McFadden's R^2	0.3770	0.4400	-	-
Mean Brier Score	0.0866	0.0716	0.0971	0.0796
Mean % of obs. correctly classified	88.24	91.42	86.83	90.50
Mean type I error %	43.09	39.77	46.60	41.84
Mean type II error %	5.24	2.05	6.07	2.68

According to all three mean evaluation criteria model 3 outperforms model 1. This is also true for medians that are not shown in table 6. In the estimation sample model 3 always has a lower Brier Score and in 97% of the cases a higher percentage of correctly classified observations than model 1. In the validation sample model 3 shows in 99% of the cases a lower Brier Score and in 98% of the cases a higher percentage of correctly classified observations than model 1. Again, mean Brier Scores of both models are lower than the mean Brier Scores of naive forecasts.[13] Additionally, model 3 exhibits in both samples a lower mean type I error (the model classifies an observation as "not default" that actually did default) and mean type II error (the model classifies an observation as "default" that actually did not default) than model 1. For example, in the validation sample model 3 exhibits a type I error rate of 41.84% compared to model 1 with 46.60% [14] Note that in commercial banking the type I error is more important than

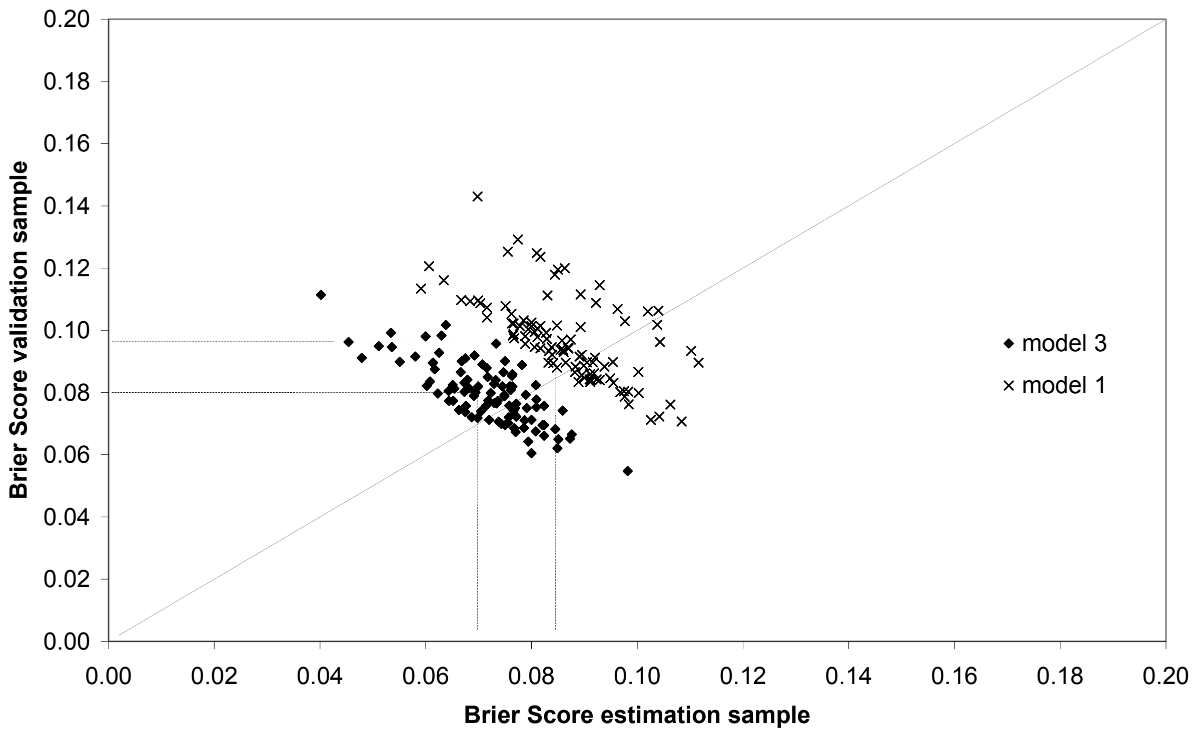
[13] In the estimation sample the Brier Scores of model 1 and 3 are always lower than the Brier Score of the naive forecast. In the validation sample the Brier Scores of model 1 and 3 are in 99% of the cases lower than the Brier Score of the naive forecast.

[14] See Carey/Hrycay (2001). Their logit default prediction model (based on four financial factors) produces a type I error of 68% in the sample and 65% out of the sample.

the type II error because of its higher costs. Eventually, the repeated split-up procedure leads to the outcome that the unequal number of observations in both samples almost average out.

The repetition of the split-up procedure leads to 100 Brier Score estimation-validation pairs for model 1 and model 3 respectively. Figure 1 shows these estimation-validation pairs:

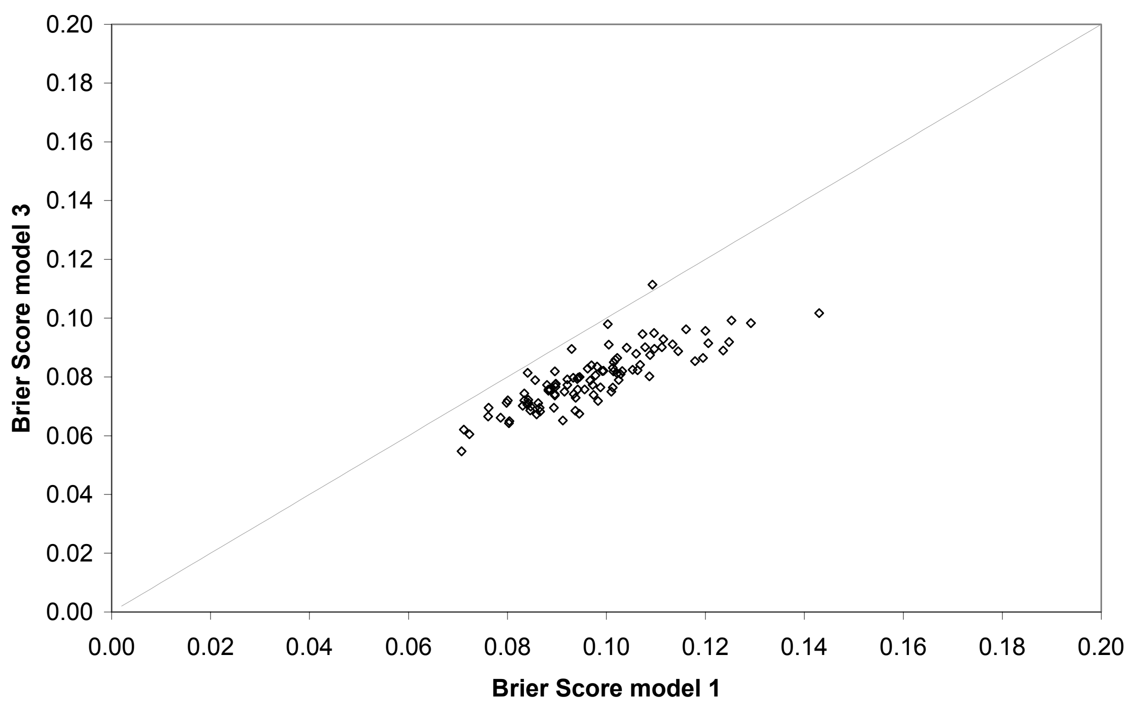
Figure 1: Brier Score estimation-validation pairs for model 1 and 3



For both models prediction accuracy in the validation sample is slightly worse than in the estimation sample because each model exhibits more estimation-validation pairs that lay above the 45°-line than below. As reported in table 6, it is visible that the mean Brier Score estimation-validation pair of model 3 is closer to the origin than that of model 1 reflecting a more accurate prediction of model 3.

Similarly, for each sample it is possible to verify separately in how many cases model 3 generates a better forecast than model 1. Using a diagram in which the Brier Score of model 1 is indicated on the horizontal axis and the Brier Score of model 3 on the vertical axis we obtain again 100 dots of Brier Score comparison pairs. It is important to mention that the Brier Scores of model 1 and 3 are calculated on the basis of the same observations. The 45°-line indicates model pairs of equal Brier Scores. Note that this analysis constitutes an inter-model comparison for a given sample and not an estimation-validation comparison for a given model. Figure 2 illustrates these comparison pairs in the validation sample :

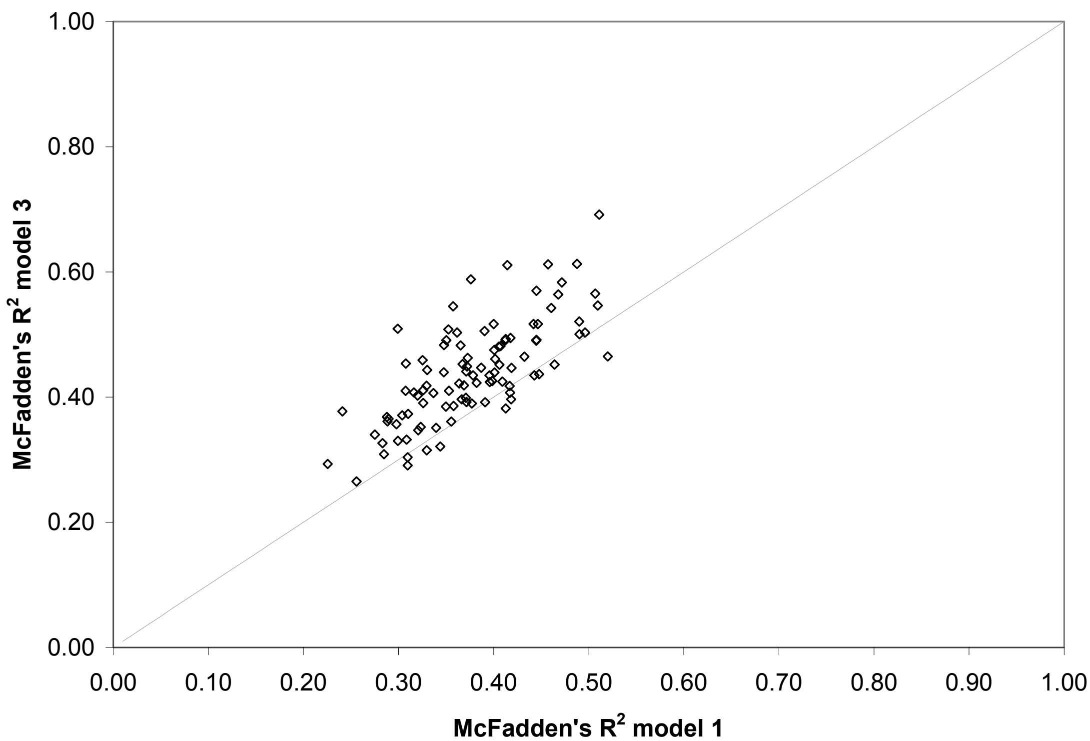
Figure 2: Brier Score pairs for model 1 and 3 in the validation sample



This figure gives a clearer impression than a comparison of the dot clouds in figure 1. It can be seen that only 1 of 100 pairs in the validation sample does not support the hypothesis that model 3 is superior to model 1 in the sense that it leads to a lower Brier Score.

Finally, in figure 3 McFadden's R^2 of model 1 is plotted against that of model 3. Note that this could only be done for the estimation sample.

Figure 3: McFadden's R^2 of model 1 and 3 in the estimation sample



McFadden's R^2 of model 3 is in 89 out of 100 cases higher than that of model 1 indicating that the use of the overall rating –instead of the pure financial rating– leads to a better fit.

However, given these results it is not clear why model 3 performs better than model 1. One reason might be the additional inclusion of non-financial factors. Another reason might be that the independent variables in both models are based on different (one optimal and one sub-optimal) weighting schemes. In particular, it might be problematic that we use the financial rating which is based on a weighting scheme optimized for the overall rating.

To investigate the influence of weighting schemes we compare a probit model to explain default events on the basis of financial factors (regression 1) with a probit model to explain default events on the basis of financial *and* non-financial factors (regression 2).

Essentially, in regression 1 we find that cash flow-to-total assets (CFTA) and the equity ratio (ER) are significant at the 0.01-level. In regression 2 CFTA and the non-financial factor management quality (MGT) are significant at the 0.01-level in regression 2. Evaluation critieria are reported in Table 7.[15]

Table 7: Prediction of default events in year t+1 with different factor types

The dependent variable FDEF indicates if default occurs in the year following the one of the rating assignment. Regression 1 uses all financial factors described in table 1 as independent variables (LTA, ER, CR, CFNL, CFTA, ICR, CIR, ROA, ROS, ROE and dummies for banks and years) whereas regression 2 uses all financial (and dummies for banks and years) *and* non-financial factors (MGT, MKT). Due to lacking data the sample is reduced to 278 observations. McFadden's R^2 is adjusted because of the different number of regressors. Coefficients are estimated using the maximum likelihood method. Testing the null hypothesis BS(regression 1)=BS(regression 2) leads to a p-value 0.0348 (two-tailed test) using the Williams-Kloot statistic z_{wk}.

Evaluation criterion	Regression 1	Regression 2
McFadden's R^2 (Adjusted)	0.189	0.283
Brier Score	0.0845	0.0704
% of obs. correctly classified	87.77	90.65
Type I error %	69.05	50.00
Type II error %	2.12	2.12

Since here the weighting of financial and non-financial factors is not predetermined (as it was the case in the financial ratings used in model 4 and 6) but rather estimated in the regressions, weighting schemes do not seem to be critical to our previous results. In our model, the combined use of financial and non-financial factors leads to a significantly better prediction of default events than the single use of financial factors even if the latter are not based on a predetermined weighting.

[15] Note that due to the smaller number of observations the absolute values of the evaluation criteria are not comparable to the previous analyses.

Eventually, we adress the marginal impact of each banks' rating-default-set by omitting successively one bank from the whole sample. The results confirm our previous findings: Model 3 always has a higher McFadden's R^2, a higher percentage of correctly classified observations and a lower Brier score than model 1. We also carried out regression analyses on the individual bank level. Due to the number of observations from bank 1 and 4 being too small individual models can only be estimated for bank 2 and bank 3. The obtained results are consistent with previous ones.

Summarizing, model 3 leads to a more accurate prediction of default events than the two other models or naive forecasts do. The result is significant, not sensitive to the omission of any bank from the sample, and robust on the individual level for two banks. Hence, the inclusion of non-financial factors in a credit rating can be called beneficial.

4.2 The relation between credit ratings and default events in the same year

We now investigate whether the additional inclusion of non-financial factors leads to a more accurate prediction of default events that occur in the same year as the rating assignment. Hence, the following analysis looks at the intra-year predictive power of rating components. Note that the distribution of default events over the years slightly changes because default events of 1992 are included and those of 1996 are discarded from the dataset. Also, note that in this context there is one problem. Since we compare default events and ratings in the same year we cannot draw conclusions about the direction of cause and effect. Due to missing default dates during a year, we do not know which variable changed at first and which one followed. Nonetheless, this problem is not critical to the earlier analysis performed in section 4.1 because there a rating assignment in year t is always compared with the default status in year t+1.

Using the same statistical association measures as above, we find evidence for a stronger relation between the overall rating and default compared to the financial rating and default (see table 8).

Table 8: Rank correlation and concordance between credit ratings and default events

ρ_s is Spearman's rank correlation coefficient and τ_b is Kendall's coefficient of concordance (correcting for ties). The dummy variable SDEF indicates a default event, FR is a borrower's financial rating, and OR is the overall credit rating from the indicated year respectively.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year	Obs.	$\rho_{s(SDEF,FR)}$	$\rho_{s(SDEF,OR)}$	$\tau_{b(SDEF,FR)}$	$\tau_{b(SDEF,OR)}$	Relation
1992	79	0.2796	0.3304	0.2528	0.3035	+
1993	105	0.3820	0.3988	0.3406	0.3611	+
1994	116	0.3030^{**}	0.3711	0.2685	0.3331	+
1995	109	0.4342^{*}	0.4954	0.3863	0.4495	+
Pooled data	409	0.3582^{***}	0.4090	0.3184	0.3700	+

***, **, * Significantly different from $\rho_{s(SDEF,OR)}$ in column (4) at the 0.01, 0.05, and 0.10 level.

In every year and for the pooled data the overall rating is more closely linked to the default variable SDEF than the financial rating is. Here, the result is significant in two of four years and for the pooled data by comparing Spearman's ρ. This could be considered as a first indice that non-financial factors are beneficial in the sense explained above.

To clarify this finding we proceed again with probit regression analysis in order to estimate models that correspond to those of section 4.1 (labeled 4, 5, and 6 hereafter). Note that the dependent variable here is SDEF. Regression results and evaluation criteria for models 4-6 are reported in table 9.

Table 9: Regression results and evaluation criteria for models 4-6

The sample used in all three probit regressions is the same and consists of 409 observations from the period 1992-1995. The dependent variable, SDEF, takes the value one if default occurs in the same year as the rating assignment and zero otherwise. Model 4 uses in addition to bank and year dummy variables the financial rating FR, model 5 the non-financial rating NFR, and model 6 the overall rating OR as independent variable (instead of "rating" as indicated in the first column) to estimate the probability of a default event. Coefficients are estimated using the maximum likelihood method.

Panel A: Regression results

Variable	Model 4 (financial rating)		Model 5 (non-financial rating)		Model 6 (overall rating)	
DEF	Coefficient	Std.Err.	Coefficient	Std.Err.	Coefficient	Std.Err.
Rating	0.5076***	0.0820	0.8914***	0.1300	1.4686***	0.1990
B2	1.3273***	0.3759	0.8822**	0.4426	3.2837***	0.6658
B3	0.9714**	0.4283	0.2935	0.4900	2.0429***	0.6871
B4	0.2731	0.4773	-0.6238	0.5454	1.0842	0.6992
Y1993	0.2073	0.3576	0.02533	0.3935	-0.1180	0.4499
Y1994	0.0986	0.3490	0.1248	0.3762	-0.2849	0.4420
Y1995	0.4832	0.3393	0.5759	0.3666	0.2227	0.4388
Intercept	-4.7179***	0.6051	-5.6340***	0.7324	-9.8370***	1.2887

***, **, * Significantly different from zero at the 0.01, 0.05, and 0.10 level.

Panel B: Evaluation criteria

The null hypotheses BS(model 4)=BS(model 6) can be rejected with a p-value 0.0022 and BS(model 5)=BS(model 6) can be rejected with a p-value 0.0102 using the Williams-Kloot statistic z_{wk} (two-tailed test).

Evaluation criterion	Model 4 (financial rating)	Model 5 (non-financial rating)	Model 6 (overall rating)
McFadden's R^2	0.3325	0.4111	0.5923
Brier Score	0.0576	0.0542	0.0386
% of obs. correctly classified	92.18	92.91	94.62
Type I error %	75.00	60.00	37.50
Type II error %	0.54	1.36	1.90

As shown in panel A, all rating variables are statistically significant at the 0.01-level indicating the strong relation between default and credit ratings. All models are more accurate than the naive forecast which leads to a Brier Score of 0.0882. Moreover, the null hypothesis that all three models are compatible with the observed outcomes can not be rejected using Spiegelhalter's z-statistic. Finally, the model evaluation results presented in panel B confirm that model 6 is superior to model 4 and 5 with respect to most of the criteria. The Brier Score of model 6 is significantly smaller than the one of the two other models at the 0.01-level. Note that the type I error of model 6 (37.50%) is half of the type I error of model 4 (75.00%).

In correspondence with section 4.1, we continue by dividing the entire sample into an estimation and a validation sample and estimate model 4 and 6. Aggregated results of this split-up procedure are summarized in table 10.

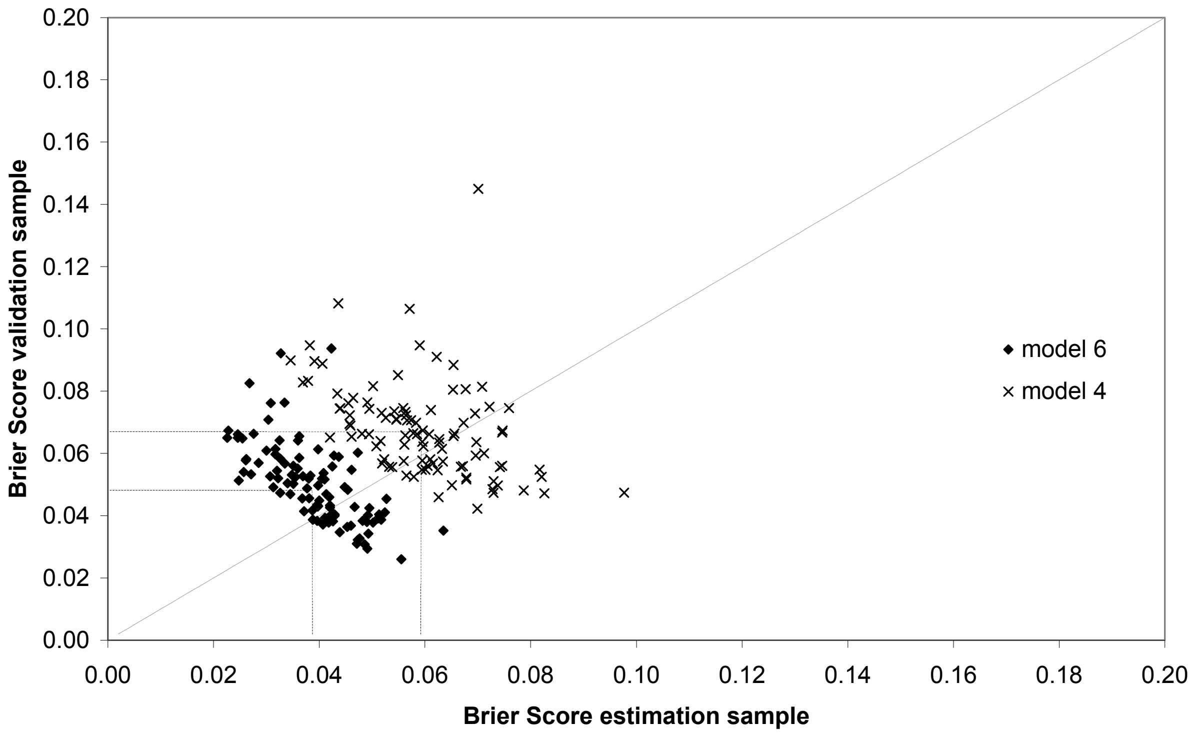
Table 10: Results of the split-up procedure

The whole sample of 409 observations (160 borrowers) is subdivided in an estimation and validation sample. The split up is done by drawing randomly 80 borrowers and considering them as estimation sample, leaving the remaining 80 borrowers for the validation sample. To avoid extreme favorable or unfavorable partitions the split-up procedure is repeated 100 times and probit models with SDEF as dependent variable are estimated. Model 4 uses the financial rating FR and model 6 the overall rating OR as independent variable.

	Estimation sample		Validation sample	
Mean number of obs.	203.3		205.7	
Mean Brier Score (naive forecast)	0.0899		0.0876	
Mean % of obs. correctly classified (naive forecast)	89.89		90.25	
	Model 4	Model 6	Model 4	Model 6
Mean McFadden's R^2	0.3776	0.6211	-	-
Mean Brier Score	0.0594	0.0390	0.0677	0.0501
Mean % of obs. correctly classified	91.56	94.50	90.29	93.54
Mean type I error %	66.22	34.14	69.46	34.89
Mean type II error %	1.78	2.18	2.67	3.22

With respect to most of the mean evaluation criteria (except mean type II error) model 6 is not only superior to model 4 in the estimation sample but also in the validation sample. Comparing medians (not shown in table 10) instead of means does not change any of the previous findings. In the estimation sample model 6 has in 99% of the cases a lower Brier Score and a higher percentage of correctly classified observations than model 4. In the validation sample model 6 has in 95% of the cases a lower Brier Score and in 94% of the cases a higher percentage of correctly classified observations than model 4. Moreover, model 6 exhibits a considerably lower mean type I error than model 1. Mean Brier Scores of both models are lower than the mean Brier Scores for naive forecasts in both samples.[16] Figure 4 depicts the Brier Score estimation-validation pairs for model 4 and 6:

Figure 4: Brier Score estimation-validation pairs for model 4 and 6



It turns out that the dot clouds of model 4 and 6 lay slightly above the 45°-line indicating that external validity is somewhat worse than internal validity. Moreover, since the mean brier score comparison pair of model 6 is closer to the origin it leads to a more accurate default prediction than model 4.

Similar to section 4.1 we study whether the weighting scheme of the financial factors included in the financial rating influences our previous results. It results that the equity ratio (ER)

---

[16] In the estimation sample the Brier Scores of model 4 and 6 are always lower than the Brier Scores of the naive forecasts. In the validation sample the Brier Scores of model 4 and 6 are in 97% of the cases lower than the Brier Scores of the naive forecasts.

and cash flow-to-total assets are significant in regressions 3 and 4, whereas both non-financial factors are additionally significant in regression 4 (MGT at the 0.01-level and MKT at the 0.10-level). Results of regressions 3 and 4 are reported in Table 11:

Table 11: Prediction of default events in year t with different factor types

The dependent variable SDEF indicates if default occurs in the year of the rating assignment. Regression 3 uses all financial factors described in table 1 as independent variables (LTA, ER, CR, CFNL, CFTA, ICR, CIR, ROA, ROS, ROE and dummies for banks and years) whereas regression 4 uses all financial (and dummies for banks and years) *and* non-financial factors (MGT, MKT). Due to lacking data the sample is reduced to 278 observations. McFadden's R^2 is adjusted because of the different number of regressors. Coefficients are estimated using the maximum likelihood method. Testing the null hypothesis BS(regression 3)=BS(regression 4) leads to a p-value 0.1118 (two-tailed test) using the Williams-Kloot statistic z_{wk}.

Evaluation criterion	Regression 3	Regression 4
McFadden's R^2 (Adjusted)	0.100	0.203
Brier Score	0.0462	0.0387
% of obs. correctly classified	93.53	94.24
Type I error %	78.95	63.16
Type II error %	1.16	1.54

We obtain that the additional inclusion of non-financial factors leads to a more accurate prediction of default events than the single use of (unweighted) financial factors with respect to most of the evaluation criteria. However, in comparison to section 4.1 (see table 7) the improvement in prediction accuracy measured by the Brier Score is only marginally significant in a two-tailed test.

To check the robustness of the results we again analyze the marginal impact of each banks' rating-default-set by discarding successively one bank from the whole sample. Proceeding in this manner, model 6 has always a higher McFadden's R^2, a higher percentage of correctly classified observations and a lower Brier score than model 4. Even if bank 2 is withdrawn from the sample the evaluation criteria are better for model 6 than for model 4, but at a lower level compared to the whole sample. Finally we run regressions on the individual bank level using the whole sample. Due to a too small number of observations from bank 1 and 4 only the remaining two banks can be analyzed individually. Individual regressions yield the same results as those for the pooled data which indicates the robustness of our findings. Just to restate, we find evidence that non-financial factors improve the accuracy of an intra-year default prediction, and thus can be deemed as beneficial.

5 Conclusion

Over the past ten years banks' uses of internal credit ratings have multiplied. In the near future ratings will be recognized by banking supervision authorities to determine banks' capital adequacy, converging considerably the internal and the external perspective of credit risk management. Given this rising importance of credit ratings, the design of sound rating systems is in the interest of banks, borrowers, and supervisors. Whereas the relevance of financial factors for rating purposes is widely accepted, the use of non-financial factors is equally beyond controversy but it has often been justified only holistically.

This paper constitutes a first attempt to explore the role of non-financial factors in credit ratings. The main result is that the combined use of financial and non-financial factors leads to a significantly more accurate default prediction than the single use of financial or non-financial factors. This is true for default in the year of the rating assignment and default in the subsequent year. Default is defined consistently with the definition of the Basel Committee on Banking supervision, accuracy of default prediction is measured using the Brier Score, the percentage of correctly classified observations and type I and II error rates.

Although our results are limited in some ways due to the used data, they essentially confirm banking practice (see Günther/Grüning (2000)) and show that the holistic justifications for the use of non-financial factors can be approved with quantitative arguments. However, since only the benefits of non-financial factors have been analyzed, it is not possible to conclude that their additional use represents a net advantage because we have not examined the costs of acquiring and processing non-financial information. The latter may be left to future research that should proceed with an integrated cost benefit analysis of internal credit rating systems on the individual bank level. Using more extensive data, it would be interesting to differentiate our analysis with respect to age and size of borrowing firms since both characteristics might be linked to the degree to which non-financial factors improve default prediction. Additionally, in particular for pricing issues, it might be instructive to study whether non-financial factors in credit ratings can improve the differentiation between borrowers disposing of an acceptable degree of creditworthiness. Collecting data from different types of banks, our results could also be tested depending on bank size and organizational structure following Berger et al. (2002) and Stein (2002). Finally, a promising extension of Carey's (2001) and our research could be to investigate if and how multiple lenders' rating disagreement for common borrowers is related to non-financial factors in credit ratings.

Appendix: Definitions of financial factors

This table shows the formulae to calculate the financial factors used in section 4.1 and 4.2.

Variable and formula
Logarithm of total assets (LTA) = $\log(\text{total assets})$
Equity-to-assets ratio (ER) = $\frac{\text{equity}}{\text{total assets}} \cdot 100$
Current ratio (CR) = $\frac{\text{current assets}}{\text{current liabilities}} \cdot 100$
Cash flow-to-net liabilities (CFNL) = $\frac{\text{cash flow}}{\text{total liabilities} - \text{current assets}} \cdot 100$
Cash flow-to-total assets (CFTA) = $\frac{\text{cash flow}}{\text{total assets}} \cdot 100$
Interest coverage ratio (ICR) = $\frac{\text{interest expenses}}{\text{net earnings}} \cdot 100$
Capital intensity ratio (CIR) = $\frac{\text{fixed assets}}{\text{equity} + \text{long - term liabilities}} \cdot 100$
Return on assets (ROA) = $\frac{\text{net earnings}}{\text{total assets}} \cdot 100$
Return on sales (ROS) = $\frac{\text{net earnings}}{\text{sales}} \cdot 100$
Return on equity (ROE) = $\frac{\text{net earnings}}{\text{equity}} \cdot 100$

The factors ER, CR, CFNL, CFTA, ICR, CIR, and ROA are parts of the internal credit ratings systems of banks 1-4.

References

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

Altman, E.I./Haldemann, R.G./Narayan, P. (1977): Zeta^{TM} Analysis – A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1, 29-54.

Baetge, J. (1998): Empirische Methoden zur Früherkennung von Unternehmenskrisen. Opladen/Wiesbaden, 1998.

Basel Committee on Banking Supervision (2000a): Range of Practice in Banks' Internal Rating Systems. Discussion Paper, January 2000.

Basel Committee on Banking Supervision (2000b): Credit ratings and complementary sources of credit quality information. Working Paper No. 3, August 2000.

Basel Committee on Banking Supervision (2001): The New Basel Capital Accord. Consultative Document, January 2001.

Berger, A.N./Miller, N.H./Petersen, M.A./Rajan, R.G./Stein, J.C. (2002): Does function follow organizational form? Evidence from the lending practices of large and small banks. Working Paper 8752, National Bureau of Economic Research.

Bhattacharya, S./Thakor, A. (1993): Contemporary Banking Theory. *Journal of Financial Intermediation*, 3, 2-50.

Blochwitz, S./Eigermann, J. (2000): Unternehmensbeurteilung durch Diskriminanzanalyse mit qualitativen Merkmalen. *Zeitschrift für betriebswirtschaftliche Forschung*, 52, 58-73.

Brier, G.W. (1950): Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1-3.

Brunner, A./Krahnen, J.P./Weber, M. (2000): Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking. Working Paper No. 2000/10, Center for Financial Studies, Frankfurt/Germany.

Carey, M. (2001): Some Evidence on the Consistency of Banks' Internal Credit Ratings. Working Paper, Federal Reserve Board.

Carey, M./Hrycay, M. (2001): Parameterizing credit risk models with rating data. *Journal of Banking & Finance*, 25, 197-270.

Collins, N.J. (1966): Credit Analysis – Concepts and Objectives. In: Baughn, W.H./Walker, C.E. (eds.): The Banker's Handbook, 1966, 279-289.

Crouhy, M./Galai, D./Mark, R. (2001): Prototype risk rating system. *Journal of Banking & Finance*, 25, 47-95.

Diamond, D.W. (1984): Financial Intermediation and Delegated Monitoring. *Review of Economic Studies*, 51 , 393-414.

Elsas, R./Henke, S./Machauer, A./Rott, R./Schenk, G. (1998): Empirical Analysis of Credit Relationships in Small Firms Financing: Sampling Design und Descriptive Statistics. Working Paper No. 1998/14, Center for Financial Studies, Frankfurt/Germany.

English, W.B./Nelson, W.R. (1998): Bank Risk Rating of Business Loans. Working Paper, Federal Reserve Board.

Ewert, R./Szczesny, A. (2001): Countdown for the new Basle Capital Accord. Working Paper No. 2001/05, Center for Financial Studies, Frankfurt/Germany.

Günther, T./Grüning, M. (2000): Einsatz von Insolvenzprognoseverfahren bei der Kreditwürdigkeitsprüfung im Firmenkundenbereich. *Die Betriebswirtschaft*, 60, 39-59.

Hesselmann, S. (1995): Insolvenzprognose mit Hilfe qualitativer Faktoren. Aachen, 1995.

Hosmer, D.W./Lemeshow, S. (2000): Applied logistic regression. New York, 2000.

Kendall, M./Gibbons, J.D. (1990): Rank Correlation Methods. London, 1990.

Krahnen, J.P./Weber, M. (2001): Generally accepted rating principles: A Primer. *Journal of Banking & Finance*, 25, 3-23.

Leland, H.E./Pyle, D.H. (1977): Information asymmetries, financial structure, and financial intermediation. *Journal of Finance*, 32, 371-387.

Löffler, G. (2001): An Anatomy of Rating through the Cycle. Working Paper, University of Frankfurt/Germany.

Machauer, A./Weber, M. (1998): Bank behavior based on internal credit ratings of borrowers. *Journal of Banking & Finance*, 22, 1355-1383.

Norden, L. (2001): Spezialbanken und Basel II: Eine empirische Untersuchung interner Ratingsysteme. *Die Betriebswirtschaft*, 62, 271-286, forthcoming.

Platt, H.D./Platt, M.B. (1990): Development of a Class of Stable Predictive Variables: The Case of Bankruptcy Prediction. *Journal of Business, Finance and Accounting*, 17, 31-51.

Redelmeier, D.A./Bloch, D.A./Hickam, D.H. (1991): Assessing predictive accuracy: How to compare brier scores. *Journal of Clinical Epidemiology*, 44 (11), 1141-1146.

Risk Management Association (2000): EDF Estimation: A „Test-Deck" Exercise. *RMA Journal*, November 2000, 54-61.

Spiegelhalter, D.J. (1986): Probabilistic prediction in patient management and clinical trials. *Statistics in Medicine*, 5, 421-433.

Stein, J.C. (2002): Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms. *Journal of Finance*, forthcoming.

Treacy, W.F./Carey, M. (2000): Credit risk rating systems at large US banks. *Journal of Banking & Finance*, 24, 167-201.

Vinterbo, S./Ohno-Machado, L. (1999): A recalibration method for predictive models with dichotomous outcomes. In: Vinterbo, S.: Predictive Models in Medicine: Some Methods for Construction and Adaptation. PhD thesis, Norwegian University of Science and Technology, 1999.

Weber, M./Krahnen, J.P./Voßmann, F. (1999): Risikomessung im Kreditgeschäft: Eine empirische Analyse bankinterner Ratingverfahren. *Zeitschrift für betriebswirtschaftliche Forschung*, Sonderheft 41, 117-142.