No. 2000/10

Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking

Antje Brunner/Jan Pieter Krahnen/Martin Weber

No. 2000/10

Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking

Antje Brunner/Jan Pieter Krahnen/Martin Weber

Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking[1]

Antje Brunner[*], Jan Pieter Krahnen[#] and Martin Weber[§]

August 2000

Abstract: This paper discusses the role of internal corporate ratings as a means by which commercial banks condense their informational advantage and preserve it vis-à-vis a competitive lending market. In drawing on a unique data set collected from leading universal banks in Germany, we are able to evaluate the extent to which non-public information determines corporate ratings. As a point of departure, the paper describes a sample of rating systems currently in use, and points at methodological differences between them. Relying on a probit analysis, we are able to show that the set of qualitative, or soft, factors is not simply redundant with respect to publicly available accounting data. Rather, qualitative information tends to be decisive in at least one third of cases. It tends to improve the firms' overall corporate rating. In the case of conflicting rating changes, i.e. when qualitative and quantitative rating changes have opposing signs, quantitative criteria dominate the overall rating change. Furthermore, the more restrictive the weighting scheme as part of the rating methodology is, the stronger is the impact of qualitative information on the firms' overall rating. The implications of our results underline the need to define stringent rating standards, from both a risk management and a regulatory point of view.

Keywords: Corporate rating, long-term lending, financial intermediation, information aggregation, credit risk management.

JEL classification: G21

---

[*] Humboldt Universitaet Berlin and CFS Center for Financial Studies, Frankfurt (Main). e-mail: abrunner@wiwi.hu-berlin.de

[#] University of Frankfurt and CFS Center for Financial Studies and CEPR Centre for Economic Policy Research. Tel.: +49-69-798 22568, e-mail: krahnen@wiwi.uni-frankfurt.de (corresponding author)

[§] University of Mannheim. Tel.: +49-621-181-1532, e-mail: weber@bank.bwl.uni-mannheim.de

1   Introduction

The technology of risk assessment through corporate ratings has received considerable attention recently. Risk managers and regulators around the world now share the view that banks ought to meet minimum capital requirements that are sensitive to all the risks borne, including interest rate risk and credit risk [Basel Committee on Banking Supervision 1999, 2000]. Not surprisingly, the past decade has seen a surge of academic and non-academic studies on banks' risk management. This work has been concerned primarily with the measurement and management of interest rate risk. Recently, the emphasis has shifted towards work, both theoretical and policy oriented, on the management of credit portfolios, and on the integration of market and credit risk [Saunders 1999]. For all these issues, the accessability of individual rating data and their histories, either from rating agencies or from bank-internal rating activities, is a basic requirement.

In this paper, we will take a closer look at the methodologies and experiences used in the German banking industry in order to assess expected default risk of their corporate clients. In particular, we want to know what types of information are being used to determine corporate ratings, and how important they are in defining the final rating attribution. The empirical part of the paper builds on a special data set compiled at the Frankfurt-based Center for Financial Studies [see Elsas et.al. 1998 for details]. It contains detailed information about corporate ratings generated by commercial banks. The present study along with a companion paper [Weber/Krahnen/ Vossmann 1999] explore the economic function of internal corporate ratings. The paper also discusses the economic function of these ratings in the context of the theory of financial intermediation.

The remainder of this paper is organized as follows. Section 2 will derive a working hypothesis for the empirical sections. In doing so, we will survey the relevant literature and relate it to the theory of banking. Section 3 gives a brief overview of the rating methodologies currently used by German commercial banks, explaining their common features. Emphasis is on the use of qualitative and quantitative data within the framework of the more general scoring model. Section 4 presents basic descriptive statistics of our data set. Section 5 concentrates on the main issue of this paper, i.e. the

significance of hard (quantitative) and soft (qualitative) factors as determinants of the overall rating[2]. Finally, Section 6 summarizes the results and draws conclusions.

2 Rating as a core bank business: Towards a working hypothesis

A helpful first theoretical consideration is the comparison between a corporate bond and a bank loan. On this issue, there is an extensive literature that concentrates on the informational role played by banks and markets [see Thakor 1995 for a survey]. Diamond (1991) stresses the importance of credibility of information. It serves to explain why firms with ample public information, notably large firms, will have lower capital costs by tapping the bond market. Firms with less public information, in contrast, will prefer financial intermediaries as financiers. As argued by Rajan (1992), their comparative advantage with respect to the production of information is concerned with pre-contractual screening and post-contractual monitoring. We will discuss bond financing and loan financing in turn.

Bond financing relies, according to this literature, on a high degree of public information. It is typically provided by external agencies, like Standard & Poor's and Moody's. Their corporate ratings are regularly updated, and they are publicly known. The specific rating notches, ranging from AAA down to D in the case of Moody's, for example, can be regarded as the attribution of an estimated default probability (POD) over a well-defined horizon (the next 12 months, say). Note that the value of the expected loss given default is not part of the rating information itself.

Of course, publication of a particular corporate rating conveys information to ordinary market participants, and helps to reduce asymmetries of information on the market. Based on the contractual details of the loan agreement, information on the default probability (POD) forms an important factor in determining the expected repayment and becomes thus a basis for determining the nominal interest rate in a lending agreement. Public rating will contribute directly to the price formation process on the market, provided that they are reliable [Hand/Holthausen/Leftwich 1992]. The question of a rating's reliability is very important here. It depends ultimately on the competitive structure of the market for

---

[2] The terms corporate rating, creditworthiness and overall rating are used synonymously in the text.

rating information. If it is competitive and rating companies are free to set their prices, theory predicts an inverse relationship between the rating class (ranging from high to low quality) and the effective interest rate of the bond in question (ranging from high to low rates). The inverse relationship between rating and bond return reflects the economic value of an agency's reputation. Empirical studies have, to a large degree, confirmed the hypothesized inverse relationship between agency ratings and bond spreads [Blume/Lim/MacKinlay 1998, Datta et al. 1999].

Now let us turn to loan financing. With no public rating information available, the development of an active secondary market is severely hampered. Since investors do not have specific estimates of default probabilities, an adverse selection process will lead to market failure. Even if there is a financial intermediary involved, the argument remains unchanged. Assuming that an intermediary who, based on his private information, is able to estimate the risk without bias, enters the market wishing to sell part of his loan portfolio, he is likely to have to accept a major discount. These mark-downs compensate the buyers of such loans for the expected risk taken, including the behavioral risk resulting from the intermediary's incentive to provide false information. In the absence of public ratings, loans can therefore only be sold on the market if the price is marked down from its "true" value. For this reason, loans are often referred to as *illiquid assets* in the literature on financial intermediation [see Diamond 1991][3].

According to this argument, the illiquidity of the loan portfolio is a typical feature of commercial banking. Efficiently selecting and monitoring borrowers explains a positive spread between the secondary market price of loans and its primary market price. While the former price is observable, resulting from the purchase and sale of the loan on a credit market, the latter is unobservable, depending on the bank's portfolio. Any company-specific information obtained in the course of the lending relationship is aggregated into the bank's corporate credit rating. Internal ratings should therefore be seen as private information. Typically, banks do not inform their customers of the internal ratings or the implied PODs, nor do they publicize the criteria and methods used in deriving them.

---

[3] An example supporting our claim is the emerging market of *asset-backed securities*, with credit portfolios as underlyings, because, in ABS transactions, only part of the entire portfolio (typically with the exception of the equity tranche) is actually put on the market.

At this point, we conclude that for both types of debt, market-traded bonds and non-traded loans, the information available can be processed into a rating, either public or private. Both types of ratings give an estimate of the probability of default (POD). For the remainder of our analysis we will concentrate on private ratings, generated by banks. These ratings result from the aggregation of all borrower-related information privately available in the bank. Private corporate ratings thus may contain the information-related substance of the lending relationship. We can summarize our informal theoretical discussion as a working hypothesis:

Private corporate ratings (internal ratings) reflect the core business of commercial banks, whose superior information as compared to an external assessment by the market allows a more precise estimate of the POD.

Our working hypothesis motivates the direction of the empirical analysis. What is eventually needed is a measure for the private information content in internal ratings. In deriving an appropriate measure, the different rating methodologies used by the banks have to be described and calibrated (see also Weber/Krahnen/Vossmann 1999). On a general level, internal bank ratings are based on the concept of a scoring procedure, as outlined in further detail in Section 3. Typically, the overall rating for a specific borrower results from the aggregation of a number of sub-ratings (criteria), using a linear weighting function. In its most general form the weighting may be achieved in a non-linear fashion, e.g. by a neural network .

A distinction is usually made between two types of ratings: the borrower rating and the actual loan rating. The former indicates borrower's risk of default, while the latter also takes into account the individually agreed loan collateral. Therefore, a loan rating generally cannot be worse than the borrower rating. In this study, we concentrate on borrower ratings because they contain information on the borrower's risk of going bankrupt only, while the loan ratings also incorporate details of the loan agreement. The latter may therefore vary enormously for different lenders to the same customer, when for instance collateral differs among lenders.

Assuming that ratings imply an estimate for the borrower's PODs, the question to be answered is which criteria determine the POD forecast. Although there are many different sub-ratings, we will concentrate on two classes of sub-ratings and their contribution to the total score: objective and subjective criteria. The former class is based on objective information, including "hard" accounting

data (balance sheet and income statement), which is supposed to indicate a company's economic situation, and which contains mainly past information. The latter class is more subjective in nature, leaving judgemental discretion to the person in charge of the rating process. It comprises assessments of the present and future trends of the industry and the company itself. Most importantly, this class also includes evaluations of the management of the company being assessed. Typically, the information used here is "soft" and forward-looking.

Since private information not yet available in the public domain will, by its very nature, be primarily used with subjective criteria, these are of particular importance in identifying the informational value of bank-internal ratings. To the extent that lenders can obtain information on the basis of their long-term and intensive relationship with the customer earlier than the (bond) market, the "soft" criteria may contribute to an informational lead with respect to those assessments based solely on publicly available information. "Soft" criteria thus play a decisive role in determining the *value added* of bank-internal ratings.

3   How German commercial banks rate corporate clients: A first look

The 1990s saw a substantial change in the methodology of internal bank rating procedures, thus calling for a comparison of different rating systems and development trends. During the course of this decade, all five of the German commercial banks in our sample either revised their rating systems or introduced standardized rating procedures for the first time. Proceeding from estimates of creditworthiness, some of which are only verbal, or ratings based on holistic assessments, or even the existing scoring-based rating systems, current internal rating procedures used by the banks listed here are now all based on what is known as the scoring procedure [see Weber/Krahnen/Vossmann 1999 for further details].

$$v(a) = \sum_i k_i v_i (a_i) \tag{1}$$

Scoring models can be characterized by the criteria i to be assessed, and a pre-specified value functions $\{v_i\}$. In addition, the scoring model is characterized by an aggregation rule, which assigns weights k_i to the individual criteria and aggregates these by means of an algebraic rule, usually additive, to form the overall score $v(a)$ [Weber/Krahnen/Weber 1995]. If the overall score and the

rating are not the same by definition, the rating is determined by assigning the overall score to a given risk class.

The bank-internal rating procedures based on scoring methods differ in terms of their criteria catalogues, rating scales, aggregation rules, and in particular, the choice of weighting factors.

The criteria catalogue

The criteria catalogues of the individual procedures differ greatly. For example, the banks mentioned here incorporate between 4 and 18 criteria in their scoring function.[4] Among these evaluation criteria there are, for each bank, both quantitative and qualitative criteria. We define a criterion to be quantitative, if its evaluation is based on published accounting data, as for instance cash flow, earnings, debt-equity ratio, short-term and long-term debt, and so forth. These criteria are typically backward-looking. On the other hand, we define a criterion to be qualitative if its evaluation is based partly or entirely on non-published or subjective attributions; it is typically forward-looking. Examples comprise the general prospects of an industry, sales forecasts, liquidity planning, marketing strength and, most importantly, management quality and continuity, the general evaluation of the customer relationship.

A comparison in Table 1 shows that banks typically differ very little from the international rating agencies S&P and Moody's with regard to the catalogue of criteria they use in assessing borrowers' creditworthiness. The criteria catalogue of the "typical bank" was created based on the five bank-internal rating procedures we analyzed. Account maintenance is the only bank-specific criterion and it is particularly relevant in cases of long-term customer relationships.

---

[4] If the rating procedure contains a small number of criteria, each criterion usually has a list of determining factors that are to be considered in the evaluation, without each factor having to be evaluated by being assigned to a risk class.

S&P	Moody's	"Representative bank"
Financial risk: • Balance sheet and accounting • Financial policy • Profitability • Capital structure • Cash flow • Financial flexibility	**Finance management risks**: • Cash flow • Liquidity • Debt structure • Equity capital and reserves	**Economic circumstances**: • Operating position (cash flow, profitability, etc.) • Financial position (equity ratio, liquidity, etc.)
Business risk: • Industry features • Competitive position • Management	**Competitive and operating risks**: • Relative market share / competitive situation • Diversification • Sales, costs, operating result • Sales volume and purchasing **Corporate structure and legal risks**: • inclusion of associated companies	**Corporate situation**: • Sector evaluation • Market position/competition • Product/range • Special risks • Forecasts/sales revenue and liquidity planning • Corporate structure
	Management quality: • Planning and control • Management experience • Organizational structure • Successor organization	**Management**: • Experience • Succession • Quality of accounting/ controlling
		Customer relationship / Account management

Table 1:   Criteria catalogues of major rating agencies and a representative bank[5]

The rating scale

The rating scales of the criteria and the overall scores of the individual models in our sample comprise between 5 and 10 risk classes, less than the 16 risk classes found at Moody's and S&P system. The rating scales for the sub-criteria and the total score do not necessarily have the same number of notches.

---

[5]   The evaluation criteria given in Table 1 aim to highlight the areas of agreement of the criteria lists of the various models. These show the main points of the individual procedures, the number of which cannot be assessed. Cf. Berblinger 1996, Meyer-Parpart 1996.

The aggregation mechanism

Some of the methods used are single-stage and therefore correspond to the scoring procedure in its simplest form, i.e. the criteria are assigned scoring points, which are then aggregated by means of the weighting function to form the overall score (equation 1), which either corresponds to the borrower rating or can be used to derive it (by assigning a risk class). However, four of the five banks studied use two-stage models, containing two aggregation rules. The scores assigned to the criteria are, as shown in equation 2, aggregated to form a number of intermediate scores $v_j(a_j)$, which are then collated by means of a second rule to form the overall score.

Single-stage rating:
$$v(a) = \sum_i k_i v_i (a_i) \tag{1}$$

Two-stage rating:
$$v(a) = \sum_j k_j v_j (a_j) = \sum_j k_j \sum_i k_{ji} v_{ji} (a_{ji}) \tag{2}$$

An important structural feature of bank-internal rating procedures is the weighting scheme used. Apart from the identification of criteria of relevance to the borrower rating, the choice of weighting is decisive for a reliable estimation of the expected probability of default (POD).

In practice, the aspect of weighting is usually given very little consideration. Although, based purely on intuition, the impact of individual criteria on the expected probability of default should differ in extent, an undifferentiated weighting can be seen in our data set in 2 out of the 5 cases (i.e. uniform weighting). Another bank leaves the choice of the weighting factors to the decision-maker, on a case by case basis. This is virtually an holistic estimation of the company's creditworthiness, unless the decision maker's freedom in setting the weighting function is restricted in another way.

The success of the POD forecasts also depends on the structural features of the rating model, and on its procedural efficiency and unbiasedness. Here, the evaluation of the individual criteria through assignment to a particular risk class plays a special role. In the case of qualitative criteria, the decision-maker has comparatively more freedom in making an assessment, due to the subjectivity of such an evaluation and the inability to verify the result. By utilizing this leeway, a weighting fixed ex-ante can effectively be transformed, yielding a biased POD forecast. For this reason as well, the following section will focus particularly on the differences in the evaluation of objective and subjective criteria.

4 Database and Descriptive Statistics

4.1 Database

The analyses performed in this study are based on the rating data from three banks[6] (A, B and C) which were collected in accordance with the following requirements.

The parent population of borrowers comprised only corporate customers, subject to commercial accounting requirements, from the trading, mechanical engineering and software industries who had taken out loans with our banks totalling at least DM 100,000. The sampling period was to cover at least three successive years. The scope of survey was such that each bank was to take a random sample of 200 borrowers from each of the three sectors. This was to be done either by random selection on the basis of the last digit of the account number or by counting every k^{th} borrower. If the parent population in a sector was less than 200 cases, the survey was to be performed on them all.[7] In addition to the overall ratings, the data to be recorded also included all sub-ratings in keeping with the relevant bank-internal rating procedure.

The following Table 2 shows an overview of the number of borrowers with respect to the industry, bank, and year for which a borrower rating was observed. The column "1993-97" contains the overall number of borrowers per sector and bank for which there is at least one rating observation between 1993 and 1997. The subsequent Table 3 contains the rating changes observed per bank and industry.

---

[6] The remaining three of the six banks involved in the survey are not included in this empirical study because of the lack or incompleteness of their data. They are, however, included in the illustration of the various internal bank rating procedures.

[7] The exceptionally high number of Bank C borrowers from the trade sector occurred due to matters of data collection in Bank C. Implications cannot be drawn from this observation.

Bank	Industry	1993-97	1993	1994	1995	1996	1997
A	Trade	226	72	163	170	185	65
	Mechanical engineering	190	54	134	154	160	56
	Software	112	44	81	86	94	38
	Total	528	170	378	410	439	159
B	Trade	4,267	-	2,847	3,262	307	-
	Mechanical engineering	559	-	348	439	45	-
	Software	147	-	83	116	9	-
	Total	4,973	-	3,278	3,817	361	-
C	Trade	200	-	-	200	196	158
	Mechanical engineering	200	-	-	200	199	175
	Software	137	-	-	99	126	107
	Total	537	-	-	499	521	440

Table 2: Number of ratings observed for individual borrowers per bank, industry, year[8]

Bank	Industry	1993/94	1994/95	1995/96	1996/97
A	Trade	49	129	136	44
	Mechanical engineering	33	116	129	38
	Software	34	61	69	23
	Total	116	306	334	105
B	Trade	-	1,891	229	-
	Mechanical engineering	-	234	33	-
	Software	-	54	7	-
	Total	-	2,179	269	-
C	Trade	-	-	196	154
	Mechanical engineering	-	-	199	174
	Software	-	-	97	103
	Total	-	-	492	431

Table 3: Observed rating changes over a one year horizon

For all three banks, the sub-ratings and overall ratings were observed, although for Bank B this related only to the two sub-ratings at the second aggregation stage after the prior aggregation of the original sub-ratings. As Tables 2 and 3 also show, the data set of Bank B –with a total of 4,973

---

[8] If several ratings were calculated for a borrower in one year, this study uses only the last rating drawn up in that year.

borrowers– is a lot larger than the that from the other banks. Nevertheless, the rating changes observed for 1995/96 decrease to 269.

4.2   Descriptive Statistics

Systematic differences can be identified in the evaluation of individual criteria using the distribution of the borrowers' sub-ratings. As an example, Figures 1 to 3 show the frequency distributions of a quantitative sub-rating and a qualitative sub-rating for each of the three banks (A, B, C), as well as the median, the 25% and 75% quantile as measures of central tendency and the variance as the parameter of dispersion. The two sub-ratings are "operating profitability" (an objective assessment) and "management quality" (a subjective assessment) for Bank A, "economic situation" (objective) and "corporate situation" (subjective)[9] for Bank B, and "debt-equity ratio" (objective) and "management assessment" (subjective) in the case of Bank C.

The rating system of Bank B consists of five risk classes, with risk class 1 indicating the lowest risk of default and risk class 5 indicating the highest risk. Bank C's rating procedure has two separate rating scales for the sub-ratings (4 risk classes) and the overall rating (10 risk classes). We have standardized the rating scale of the overall rating from ten to four risk classes. Bank A has eight risk classes for both sub-ratings and the overall rating. Here, rating scales were transformed to four risk classes in order to achieve more comparability with other rating systems. [10]. Figures 1 to 3 illustrate the assignment of the individual risk classes for each sub-rating viewed and also for the overall ratings in the form of relative frequencies.

---

[9]   The overall rating for Bank B is composed only of these two sub-ratings.

[10]   To make the procedures clearer and easier to compare, the number of risk classes was reduced from eight to four by combining every two classes. The sub-ratings of Bank C each comprise four risk classes, while the level of creditworthiness has ten. Here, too, classes were combined to reduce them to four in number. Classes A+, A, A- were combined to form risk class 1, B+, B, B- were combined to form risk class 2, etc.
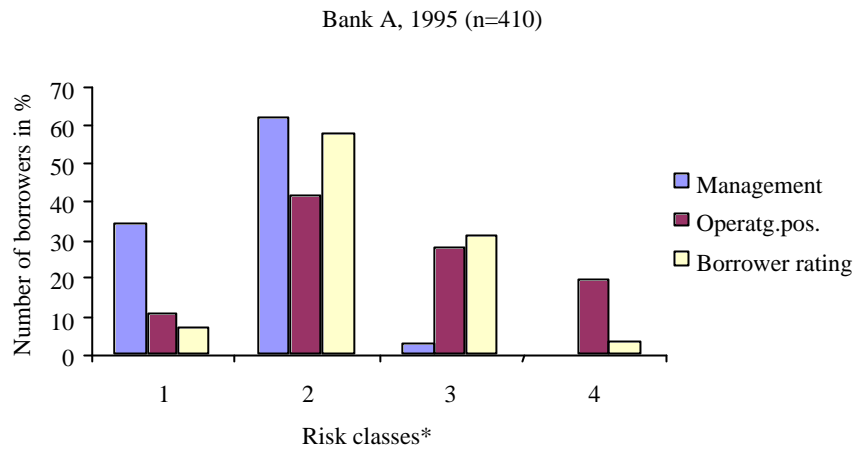
Bank A, 1995 (n=410)

n=410	Average	Variance	25% quantile	Median	75% quantile
Management	1.68	0.28	1	2	2
Operatg. pos.	2.56	0.86	2	2	3
Borrower rating	2.30	0.43	2	2	3

Figure 1: Relative frequencies of Bank A's rating assignments, 1995 (n=410)
* The number of rating classes has been transformed to from eight to four.
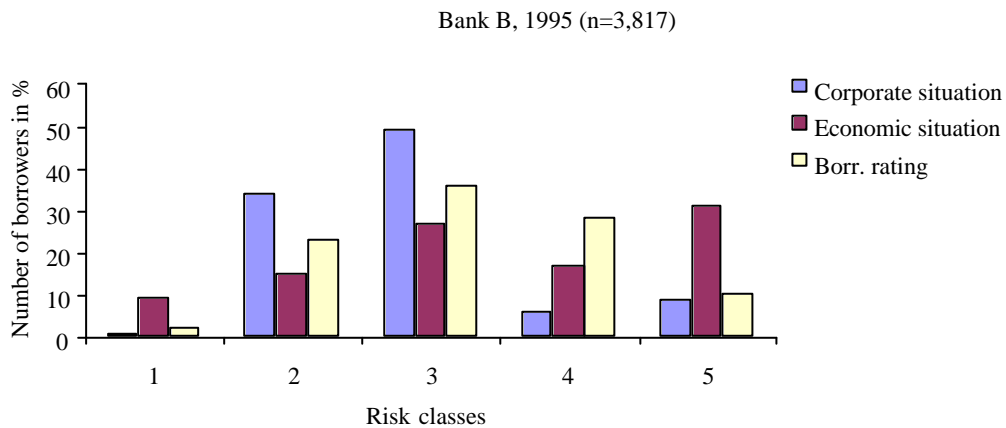
Bank B, 1995 (n=3,817)

n=3,817	Average	Variance	25% quantile	Median	75% quantile
Corporate situation	2.84	0.73	2	3	3
Economic situation	3.47	1.75	3	3	5
Borr. rating	3.22	0.97	2	3	4

Figure 2: Relative frequencies of Bank B's rating assignments, 1995 (n=3.817)
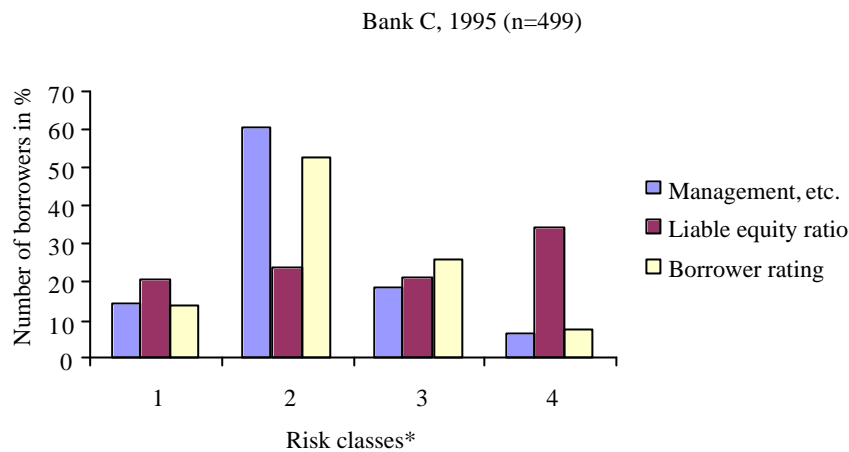
Bank C, 1995 (n=499)

n=499	Average	Variance	25% quantile	Median	75% quantile
Management, etc.	2.17	0.55	2	2	2,5
Liable equity ratio	2.69	1.32	2	3	4
Borrower rating	2.27	0.62	2	2	3

Figure 3:   Relative frequencies of Bank C's rating assignments, 1995 (n=499)
            * The number of overall rating's risk classes has been transformed from ten to four.

Evaluation differences between hard and soft criteria are evident in the mean value of the sub-ratings, and through differences in dispersion. All figures support the claim that soft criteria are rated more positively on average than hard criteria. Moreover, the dispersion of hard sub-ratings, measured by the variance, is more than twice the dispersion of the soft sub-ratings in all three cases. The decision-makers are not making full use of the evaluation range available in the subjective area and tend to allocate better risk classes, with lower implied probabilities of default. In addition, for Bank B it can be noted that the extreme categories 1 and 5 are avoided in the evaluation of soft criteria.

The hypothesis of a different evaluation of hard and soft criteria can be checked using the Kolmogoroff-Smirnoff test statistic which is a distribution-free test of differences between distribution functions. The two sub-ratings are taken as two independent random samples of frequency data. We test the null that both samples originate from the same parent population and are thereby subject to the same distribution, i.e. H_0: $F(x_{QN}) = F(x_{QL})$, where $F(x_{QN})$ designates the distribution function of

the quantitative sub-rating and $F(x_{QL})$ that of the qualitative sub-rating.[11] For all three banks, the null hypothesis can be rejected at the 1% level. Thus, in all three cases, the qualitative sub-rating is stochastically smaller –indicating less default risk– than the quantitative sub-rating. This result confirms the hypothesis of a more positive evaluation of qualitative criteria [Weber/Krahnen/Vossmann 1999].

5   Private information as rating determinants

This section will examine the main hypothesis on the role of qualitative information in the derivation of ratings. To this end, we propose a technique that permits us to record the contribution of individual sub-ratings to the aggregated overall rating. This contribution will then serve as a measure of the decision-oriented importance of the criteria in question.

The relevance of a criterion to the overall rating results not only from its absolute or relative weight in the scoring function but also from its "marginal" contribution[12] to the overall rating. The marginal contribution of any criterion X may be zero even though it has a positive weight in the aggregation function. As an example, let us suppose criterion X has the value 4 (on a scale from 1 to 10) and the weight 0.8. It is thus considered to be of major importance. Criteria Y and Z, considered to be of minor importance, have values 3 (0.1) and 5 (0.1) respectively. The overall score S for all three criteria is the weighted sum of all criteria, S(X,Y,Z) = 4. If one criterion is omitted and the weights of all others are adjusted proportionally, we have S(X,Y)=3.89, S(X,Z)=4.11 and S(Y,Z)=4. The exclusion of criterion X from the set of all criteria does not alter the overall score. Criterion X is said to be irrelevant for the overall rating, despite its 80% contribution to the weighting function.

The analysis below is based on the differentiation into qualitative and quantitative sub-ratings as the first layer below the overall ratings. We have chosen this level as the subject of the study because we are primarily interested in the explanatory contribution of the totality of the qualitative criteria. At present, we have no reasonable indication for a further differentiation of the qualitative criteria in

---

[11]   The term "quantitative (qualitative) sub-rating" describes the sub-ratings shown in Figures 1 to 3. The rating procedure of Bank A contains several other quantitative and qualitative sub-ratings in addition to these, while the rating procedure of Bank C contains another qualitative sub-rating.

[12]   "Marginal" is in quotation marks because the characteristic values are discrete.

terms of their potential informational contribution within the framework of long-term credit relationships.

5.1 The impact of qualitative criteria on borrower ratings

First, we will investigate the hypothesis that the soft sub-rating contributes no additional information to determining the overall rating and is thereby redundant. The test is performed first by means of descriptive statistics and then by a more formal econometric test. Figure 4 shows for Bank B the frequency distribution of the rating deviations λ between the hard sub-rating and the overall rating (borrower rating), where as before the hard sub-rating collates all quantitative, financial statement-related information on the borrower[13]. The rating deviation λ is defined as sub-rating QN ./. borrower rating BR. Therefore, a numerical value of $\lambda=1$ implies that the borrower rating is one risk class lower, i.e. better, than the quantitative sub-rating. Figures 5 and 6 show corresponding overviews for the other two banks (A and C).
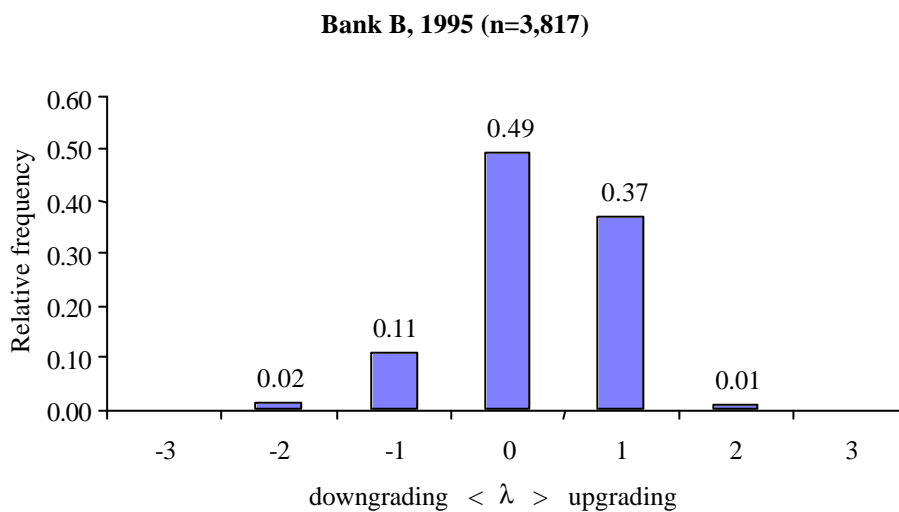
Bank B, 1995 (n=3,817)

Figure 4:   Rating deviations λ for Bank B, 1995 (n=3.817)

---

[13]   The hard sub-rating differs from the borrower rating (overall rating) in that it does not take any soft information into account. For one bank, with no fixed weights in the scoring function, weights were approximated by an OLS estimate.
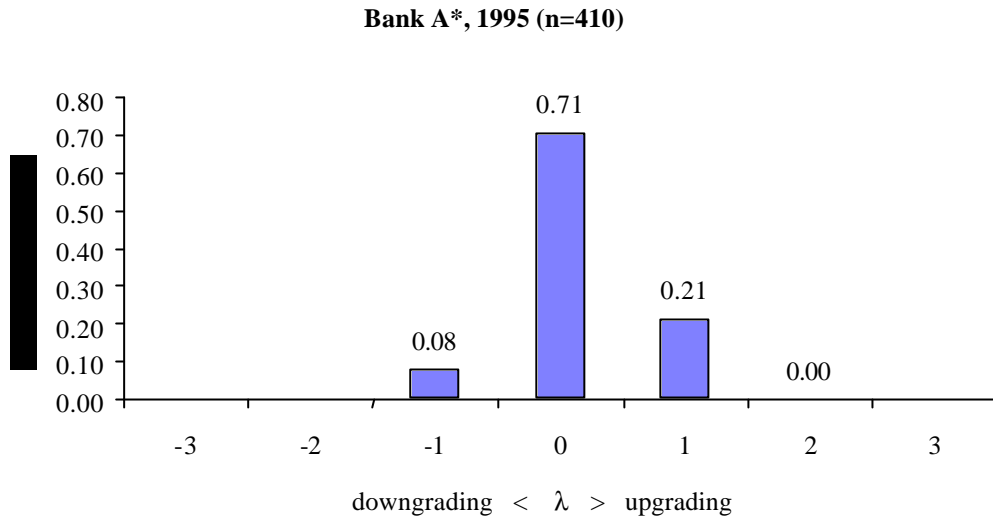
Bank A*, 1995 (n=410)

Figure 5: Rating deviations λ for Bank A*, 1995 (n=410)
* Number of risk classes has been transformed to four classes.
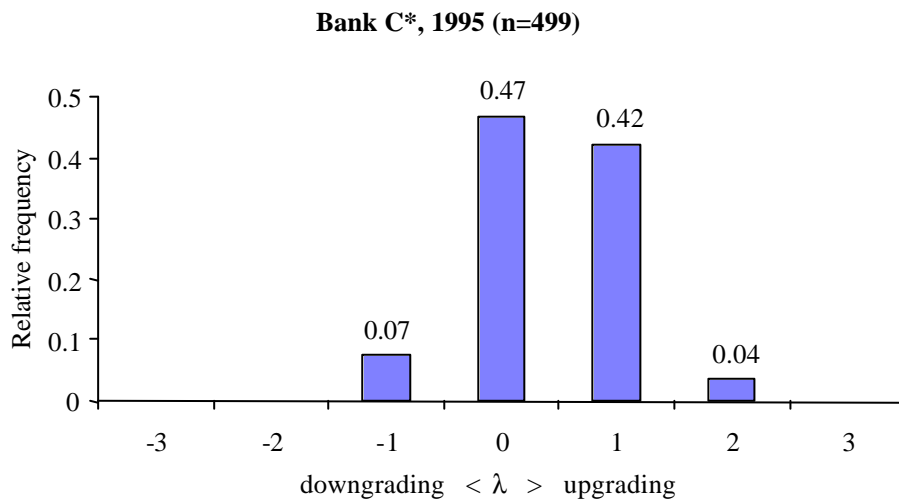
Bank C*, 1995 (n=499)

Figure 6: Rating deviations λ for Bank C*, 1995 (n=499)
* Number of risk classes has been transformed to four classes.

Given the distributions of the rating deviations in the diagrams, we can make two statements. First, qualitative criteria are particularly significant in causing deviations of the overall rating from the quantitative sub-rating. In the unweighted average of the three banks, they have an impact on the overall borrower rating for 45 of 100 borrowers, measured by the deviation of the overall rating

from the quantitative sub-rating,. The individual values are 29% for Bank A*, 51% for Bank B and 53% for Bank C*.[14]

The second point is the distribution of the deviations of the overall rating from the quantitative sub-rating induced by the qualitative sub-rating, i.e. the marginal effect of the qualitative sub-ratings, distributed asymmetrically around zero. On average, this results in an upgrade of the overall rating through the assignment of a lower rating class ($\overline{I} > 0$). It can be seen that both the frequency of such rating deviations as well as the positive bias are hogher in rating systems with fixed weights (Bank B and C). For these systems, this is $\overline{I}_B = 0.24$ and $\overline{I}_C = 0.43$ respectively, as opposed to a value of $\overline{I}_A = 0.13$ for the system used at Bank A, which does not use preset weights.

The Wilcoxon signed-rank test[15] can be used to test whether the observed rating deviations are distributed symmetrically around a median of zero. Under the null hypothesis, the differences λ stem from a parent population with the distribution function F(λ) and the density function f(λ), where F(+λ)+F(-λ)=1 or f(+λ)=f(-λ). The null hypothesis of a symmetrical distribution around a zero median can be rejected in all three cases at the 0.1% level.[16]

5.2   The impact of qualitative criteria on rating changes

In the second stage of the analysis, the contribution of qualitative criteria to changes in the borrower rating is observed. The dependent variable ΔBR is defined as the difference between the overall ratings for two successive observation times. Three attributes are encoded: rating improvement (u), rating constancy (c), and rating deterioration (d). The factors to be studied that influence the borrower rating are the indicators for changes in both the qualitative and the quantitative sub-rating

---

[14]   Refer to the results for the original 8 and 10 risk classes of Bank A and Bank C respectively. In this case, the relative frequency of a deviation is 48% for Bank A and 82% for Bank.

[15]   This test is also known as the Wilcoxon test for pair differences or the *Wilcoxon matched pairs signed rank test*. Cf. Sachs 1997.

[16] In this test the absolute amounts of the m ($m \leq n$) of the rating deviations different from zero ($|I| \neq 0$) are arranged in ascending order of rank, with the smallest observation ranked 1, the largest ranked *m* (in case of the same amounts, the middle rank numbers are assigned). Then the sum of positive ranks (*I*>0) and the sum of negative ranks (*I*<0) is formed. The test statistic used is the smaller of the two rank totals. If this is smaller than or equal to the critical value, the null hypothesis is rejected. For *m*>25, the critical value is approximately:

$$R(m; a) = \frac{m(m+1)}{4} - z\sqrt{\frac{1}{24} m(m+1)(2m+1)}$$

[up (u) / constant (c) / down (d)]. The reference variable in each case is the constant qualitative sub-rating. Using this procedure, we can determine the effect of a qualitative improvement or deterioration depending on the attribute of the quantitative variable.[17] As already explained in Section 5.1, the quantitative sub-rating QN contains all quantitative rating information of the relevant bank-internal procedure. Correspondingly, the qualitative sub-rating QL comprises all qualitative rating information of the procedure. Of particular interest are the estimated coefficients of the indicator variables, which describe changes in the qualitative and quantitative sub-ratings in the opposite direction (uQNdQL; dQNuQL). Table 10 shows that, in Bank A, 4% of all cases and, in Banks B and C, 16-18% of all cases fall under one of the two "contradictory" categories. It can be noted for Bank A that the largest numbers occur in cases of a constant qualitative sub-rating (uQNcQL, dQNcQL, cQNcQL together make up 237 of 334 cases).

| | ΔBR | | | | | | | | | | | |
| | Bank A | | | | Bank B | | | | Bank C | | | |
	u	c	d	Total	u	c	d	Total	u	c	d	Total
uQNuQL	17	8	0	25	21	5	0	26	45	8	0	53
uQNdQL	0	8	1	9	12	8	1	21	16	27	3	46
uQNcQL	17	27	2	46	18	12	0	30	26	11	0	37
dQNuQL	1	2	1	4	3	11	7	21	3	27	12	42
dQNdQL	0	2	24	26	0	6	20	26	0	11	77	88
dQNcQL	0	26	33	59	0	8	8	16	1	20	15	36
cQNuQL	10	8	1	19	8	14	0	22	6	10	0	16
cQNdQL	2	10	2	14	0	21	4	25	0	15	6	21
cQNcQL	10	111	11	132	0	82	0	82	0	153	0	153
Total	57	202	75	334	62	167	40	269	97	282	113	492

Table 10: Frequencies of the changes in sub-rating and overall rating, Banks A, B and C, 1995/96.

The "contradictory" cases are particularly interesting because they allow us to compare the dominance of quantitative or qualitative criteria. For example, Table 10 shows for Bank B that the resulting creditworthiness deteriorates in only one of 21 cases in which the quantitative sub-rating is upgraded while the qualitative sub-rating is downgraded. In 12 of the 21 cases the quantitative

---

[17] The exogenous indicator variables are cQNuQL, cQNdQL, uQNuQL, uQNdQL, dQNuQL and dQNdQL, whereby cQNuQL indicates cases with a constant hard sub-rating and a soft upgrading, uQNdQL comprises cases with a quantitative upgrading and qualitative downgrading, and accordingly for the other variables.

upgrade dominates the change in borrower rating. Similar observations are made for Bank C. From this, we can derive the result that, when the sub-ratings change in opposite directions, the impact of the quantitative change tends to dominate the effect on the overall rating.

The hypothesis of the dominating impact of hard rating changes on the overall rating will be studied below by means of a probit analysis. As the implementation of the dependent variable ΔBR can take more than two values (u, c, d), an ordered probit analysis will be used. In keeping with the interpretation in the introduction to this article, rating classes are interpreted as an ordinal proxy value for the default probability of the loan. The latter is a continuous, latent variable. Through their algebraic sign, the estimated coefficients show the direction of the change in probability for a jump into one of the two boundary categories (up, down). With a minus sign, the probability of upgrading (downgrading) increases (decreases). The results of the analysis for the three Banks A, B and C are given in Table 11.

The dependent variable "change in the borrower rating" ΔBR can take 3 indicator values. The parentheses contain z-values, where *** indicates the 1% significance level and ** the 5% significance level.

| Variables | Ordered Probit (dependent variable ΔBR) | | |
	Bank A	Bank B	Bank C
cQNuQL	-1.18 (-4.08***)	-0.72 (-2.51**)	-0.83 (-2.60***)
cQNdQL	-0.17 (-0.53)	0.65 (2.30**)	0.92 (3.17***)
uQNuQL	-1.69 (-6.12***)	-1.89 (-6.11***)	-2.14 (-9.22***)
uQNdQL	0.07 (0.18)	-1.07 (-3.63***)	-0.58 (-2.89***)
dQNuQL	-0.17 (-0.29)	0.68 (2.35**)	0.71 (3.35***)
dQNdQL	2.29 (6.14***)	2.18 (7.04***)	2.56 (12.47***)
n	334	269	492
LR statistics	120.21	139.20	375.53
Pseudo R^2	0.19	0.28	0.39

Table 11: ΔBR as a function of changes in the hard and soft sub-ratings, Banks A, B and C, 1995/96.

The results can be interpreted as follows. The isolated changes in the qualitative sub-rating in both a positive and a negative direction (with the quantitative sub-rating remaining constant) have, in the rating procedures of Banks B and C in the sample studied, a significant impact (1% and 5% significance level) on the probability of a change being observed in the overall rating. In case of Bank A, this only applies for an upgrading of the qualitative sub-rating; when viewed in isolation, a downgrading of the qualitative sub-rating has no significant effect on the probability of a change in the

overall rating. The negative sign of the coefficient of cQNuQL for all three banks means that an improvement in the qualitative sub-rating while the quantitative sub-rating remains constant increases the probability of an upgrading of the overall rating and lowers that of a downgrading. The significantly positive sign of the coefficient of cQNdQL for Banks B and C describes an increase in the probability of a downgrading of the overall rating at the same time as a qualitative deterioration and a constant quantitative sub-rating.

Changes in the same direction in both the quantitative and the qualitative sub-rating (uQNuQL, dQNdQL) have a significant impact at the 1% level on the probability of a change in the overall rating in the expected direction. An improvement in the two sub-ratings increases (decreases) the probability of an upgrading (downgrading) of the overall rating, with a negative sign for the coefficient. A deterioration in both sub-ratings increases the probability of a downgrading accordingly.

The indicator variables for opposite changes in the quantitative and the qualitative sub-rating (uQNdQL, dQNuQL) are significant for Banks B and C at the 1% and 5% level respectively. The improvement in the qualitative sub-rating with a deterioration in the quantitative sub-rating at the same time increases the probability of a downgrading of the overall rating (significantly positive sign of the coefficient of dQNuQL for B and C), and a qualitative downgrading at the same time as a quantitative improvement lowers the probability of a downgrading of the overall rating (significantly negative sign of the coefficient of uQNdQL for B and C). In the rating procedures of Banks B and C, therefore, the direction of the quantitative change dominates the overall effect when the sub-ratings change in opposite directions. For Bank A, however, neither of the two indicators of opposite changes in the sub-rating is significantly different from zero.

One might also ask how important soft and hard factors are in explaining rating changes in periods preceding a the financial distress of the borrower. The underlying hypothesis maintains that soft criteria in the borrower rating are particularly significant in predicting future rating deterioration. Therefore, in the period before distress and given informational efficiency, soft factors should cause a rating downgrade since their assessment is based on economic and financial forecasts for the borrower in question. Informational efficiency thus implies a quicker perception of an ongoing corporate crisis through soft information. On the other hand, the relationship manager who evaluates

the client may have an incentive to hide the true state of the borrower from the senior bank manager if, for instance, his bonus payment depends on the quality of the loan portfolio he is in charge of, measured by the rating. In this case, he could try to compensate a decline in the objective quantitative sub-rating by a overoptimistic evaluation of qualitative criteria which can hardly be verified. Thus, there are two hypotheses on qualitative rating changes in the period before financial distress to be investigated. The qualitative sub-rating may either go down, due to informational efficiency, or go up due to incentive problems. Using a data set on 130 distressed credit relationships[18] from a second survey [see Brunner/Krahnen 2000], we did not find any evidence that qualitative sub-rating behave in a way suggested by the incentive problem hypothesis. Instead, changes of qualitative sub-ratings are correlated with the downward trend of overall borrower quality. This result supports the hypothesis of information efficiency.

6   Results and Conclusions

The aim of this paper is to contribute to the ongoing discussion on the significance of internal ratings for the credit business in commercial banking. Our point of departure is theoretical considerations on the functioning of financial markets facing asymmetric information, and the roles of intermediaries therein [see Thakor 1995, or Rajan 1992 for an introduction]. From this perspective, banks are specialists in selecting and monitoring borrowers, as well as in designing efficient contracts, and –if necessary– renegotiating with them. The ongoing collection, processing and utilization of information on the borrower and his business prospects constitute a value added service for which intermediaries on the credit should earn a return, in equilibrium. The fee for this value added banking service can be collected either directly through a higher lending rate [Sharpe 1990] or indirectly through cross-selling, the sale of associated services [Greenbaum/Kanatas/Venezia 1989, Elsas/Krahnen 1998]. Obviously, the competitive situation on any particular credit market will also play an important role in determining the most effective pricing policy [Broecker 1990, Boot/Thakor 1999].

---

[18]   The selection mode of this survey required borrowers to show a negative rating (the last 2 out of 6 risk classes on a transformed rating scale for all banks included) at least once during the observation period from 1992-1997.

Our working hypothesis claims that internal ratings can be seen as the core competency of commercial banking. Without ratings, banks lose their lead in terms of information and assessment which, in an efficient lending market, define their competitiveness. A more precise view of the standard rating procedures directs our attention to the qualitative criteria. These, unlike quantitative criteria, are said to reflect the important private information, which is the quintessence of the economic advantage implied by close, long-term bank/customer relations (Ongena/Smith 1998). Using a new data set of internal bank ratings, we have tried to assess the role of qualitative sub-ratings in determining the overall rating of borrowers' creditworthiness.

Four results have been established :

- First, qualitative factors have a significant impact on the overall rating in the rating systems we analyzed.

- Second, qualitative and quantitative criteria are evaluated differently. The frequency distributions of qualitative and quantitative sub-ratings differ significantly. Qualitative criteria are systematically assessed more positively than quantitative criteria and show less dispersion across risk classes.

- Third, the impact of qualitative sub-ratings on the overall rating, which is measured as the difference between the quantitative sub-rating and overall rating, is generally positive. The qualitative sub-ratings upgrade the overall rating compared to the quantitative rating in about one third of cases . In an average of 55% of cases the qualitative sub-rating does not cause a a difference between quantitative sub-rating and overall rating. The distribution of the difference between the overall rating and the quantitative sub-rating shows asymmetry around its median zero. The frequency of rating adjustments caused by qualitative factors is higher in systems with fixed, preset weightings (Banks B and C) than in Bank A's system, where the weighting is intentionally left up to the individual expert. Moreover, systems of Banks B and C show higher rating activity, i.e. the probability that a rating is changed over a given period of time (one year), is significantly higher [see also Weber/Krahnen/Vossmann 1999]. Both observations together indicate the superior information processing of rating systems with preset weightings.

- Fourth, in cases of rating changes, the partial effects of quantitative and qualitative sub-rating changes are significant. In comparison, the cross effect of a qualitative rating change on an

(opposing) quantitative rating change within systems featuring fixed aggregation weights shows an asymmetry in which the quantitative rating change is dominant. At first sight, this result contradicts the assumption sometimes expressed that the scope of discretion in assessing qualitative factors causes an incentive problem and results in an underestimation of the default risk by compensating for the deterioration of quantitative criteria with a more positive evaluation of qualitative criteria. Whether there remains a compensation that weakens the quantitative effect, but does not fully compensate it, has yet to be studied. This would have to reflect the extent of the rating change (number of risk classes) instead of its effect in the outline used here, featuring the three attributes: up/constant/down.

We can therefore conclude that qualitative sub-ratings do play a significant role in our data, as is also the case in theory. However, it remains to be seen whether the heavy weighting of "soft" information is a desirable or problematic variable. In the present article we are satisfied to have developed and laid the groundwork for this question. Answering it would require a valid calibration of bank-internal ratings. Such a comparison, which assigns actual defaults and the corresponding transitional probabilities to rating dasses and their changes, requires a long-term rating practice as well as a "good" rating model. In general, neither condition is met to a satisfactory degree at the present time. There are no generally accepted rating standards available, nor do there exist records of credit ratings covering lengthy timespans and including the actual default rates in such a way that they could be evaluated statistically. In our opinion, this represents an economically fascinating and a commercially relevant area to be researched over the coming years.

References

Altman, Edward I., and Robert Haldeman (1995): "Corporate credit scoring models: Approaches and tests for successful implementation." *The Journal of Commercial Lending*, 10-22.

Altman, Edward I., and Anthony Saunders (1996): "Credit risk measurement: developments over the last 20 years." *Journal of Banking and Finance* 21, 1721-1742.

Basel Committee on Banking Supervision (1999): "A New Capital Adequacy Framework." Consultative Paper, Bank for International Settlements, Basel.

Basel Committee on Banking Supervision (2000): "Range of Practice in Banks' Internal Rating Systems." Discussion Paper, Bank for International Settlements, Basel.

Blume, Marshall E., Felix Lim and A. Craig MacKinlay (1998): "The declining credit quality of U.S. corporate debt: Myth or reality." *Journal of Finance* 53, 1389-1413.

Boot, Arnoud, and Anjan V. Thakor (1999): "Can relationship banking survive competition?." *Journal of Finance*, forthcoming.

Berblinger, Jürgen (1996): "Marktakzeptanz des Rating durch Qualität." in: *Handbuch Rating*, Büschgen, Hans E. / Everling, Oliver (editors), Gabler Verlag, Wiesbaden, 21-110.

Broecker, Thorsten (1990): "Credit-worthiness tests and interbank competition." *Econometrica* 58, 429-452.

Brunner, Antje, and Jan Pieter Krahnen (2000): "Private debt restructuring: On the role of coordination risk in financial distress." Working Paper, Center for Financial Studies, Frankfurt.

Datta, Sudip, Mai Iskandar-Datta and Ajay Patel (1999): "Bank monitoring and the pricing of corporate public debt." *Journal of Financial Economics* 51, 435-449.

Diamond, Douglas (1991): "Monitoring and reputation: The choice between bank loans and directly placed debt." *Journal of Political Economy* 99, 689-721.

Elsas, Ralf, Ralf Ewert, Jan Pieter Krahnen, Bernd Rudolph and Martin Weber (1999): "Risikoorientiertes Kreditmanagement deutscher Banken." *Die Bank* 3/99, 190-199.

Elsas, Ralf, Sabine Henke, Achim Machauer, Roland Rott and Gerald Schenk (1998): "Empirical analysis of credit relationships in small firms financing: Sampling design and descriptive statistics." Working Paper No. 98/14, Center for Financial Studies, Frankfurt.

Elsas, Ralf, and Jan Pieter Krahnen (1998): "Is relationship lending special? Evidence from credit-file data in Germany." *Journal of Banking and Finance* 22, 1283-1316.

Elsas, Ralf, and Jan Pieter Krahnen (1999): "Collateral, default risk, and relationship lending: An empirical study on financial contracting." Working Paper, Center for Financial Studies,

Frankfurt.

Everling, Oliver (1991): *Credit Rating durch internationale Agenturen*. Gabler Verlag, Wiesbaden.

Gorton, Gary (1994): "Bank regulation when "banks" and "banking" are not the same." *Oxford Review of Economic Policy* 10, 106-119.

Greenbaum, Stewart I., George Kanatas and Itzhak Venezia (1989): "Equilibrium loan pricing under the bank-client relationship." *Journal of Banking and Finance* 13, 221-235.

Greene, William H. (1997): *Econometric Analysis*. 4th edition, Prentice-Hall, New Jersey.

Gite, Gailen, and Arthur Warga (1997): "The effect of bond-rating changes on bond price performance." *Financial Analysts Journal* 53, 35-51.

Hackethal, Andreas, Reinhard Schmidt and Marcel Tyrell (1999): "Disintermediation and the role of banks in Europe: An international comparison." *Journal of Financial Intermediation* 8, 36-67.

Hand, John, Robert Holthausen and Richard Leftwich (1992): "The effect of bond rating agency announcements on bond and stock prices." *Journal of Finance* 47, 733-752.

Krahnen, Jan Pieter, and Martin Weber (2000): "Generally accepted rating principles: A primer." *Journal of Banking and Finance*, forthcoming.

Machauer, Achim, and Martin Weber (1998): "Bank behavior based on internal credit ratings of borrowers." *Journal of Banking and Finance* 22, 1355-1383.

Machauer, Achim, and Martin Weber (2000): "Number of bank relationships and competition." Working Paper, Center for Financial Studies, Frankfurt.

Mester, Loretta J., Leonard I. Nakamura and Micheline Renault (1999): "Checking accounts and bank monitoring." Working Paper No. 99-02, Wharton Financial Institutions Center, University of Pennsylvania.

Meyer-Parpart, Wolfgang (1996): „Ratingkriterien für Unternehmen." in: *Handbuch Rating*, Büschgen, Hans E. and Oliver Everling (editors), Gabler Verlag, Wiesbaden, 111-173.

Ongena, Steven, and David C. Smith (1998): "Bank relationships: A review." unpublished manuscript, Norwegian School of Management (BI), Sandvika.

Paul-Choudhury, Sumit (1997): "Choosing the right box of credit tricks." *Risk* 10, No. 11, 28-35.

Rajan, Raghuram G. (1992): "Insiders and outsiders: The choice between relationship and armslength debt." *Journal of Finance* 47, 1367-1400.

Sachs, Lothar (1997): *Angewandte Statistik*. Springer Verlag, Berlin Heidelberg.

Saunders, Anthony (1998): "Credit risk: The new models from CreditMetrics to KMV to Credit

Risk^{+}. An introduction and overview." unpublished manuscript.

Sharpe, Stephen A. (1990): "Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships." *Journal of Finance* 45, 1069-1087.

Thakor, Anjan V. (1995): "Financial intermediation and the market for credit." in: *Handbooks in Operations Research and Management Science* 9, Jarrow et al. (editors), 1073-1103.

Urban, Dieter (1993): *Logit-Analyse*. Gustav Fischer Verlag, Stuttgart.

Wahrenburg, Mark (1999): „Vergleichende Analyse alternativer Kreditrisikomodelle", Finance Discussion Papers, Universitaet Frankfurt.

Weber, Martin, Jan Pieter Krahnen and Frank Vossmann (1999): „Risikomessung im Kreditgeschäft: Eine empirische Analyse bankinterner Ratingverfahren." *Zeitschrift für betriebswirtschaftliche Forschung,* Sonderheft 41, 117-142.

Weber, Martin, Jan Pieter Krahnen and Adelheid Weber (1995): „Scoring-Verfahren – häufige Anwendungsfehler und ihre Vermeidung." *Der Betrieb* 48, 1621-1626.

Wilson, Thomas C. (1997): „Measuring and managing credit portfolio risk: A two part series." manuscript, McKinsey & Co., Zurich.