Generally Accepted Rating Principles: A Primer[*]

Jan Pieter Krahnen

*Goethe Universität Frankfurt; CFS Center for Financial Studies, and CEPR ,
Mertonstr. 17-21, D-60054 Frankfurt am Main, Germany. Tel.:+49-69-79822568, Fax: +49-69-79828951,
eMail: <krahnen@wiwi.uni-frankfurt.de>.*

Martin Weber

*Universität Mannheim,
L 5.2, 68131 Mannheim, Germany, Tel.: +49-621-181 1532, Fax: +49-621-181 1534,
eMail: <weber@bank.bwl.uni-mannheim.de>.*

First version: November 2, 1999
This version: February 14, 2000

Abstract

Bank internal ratings of corporate clients are intended to quantify the expected likelihood of future borrower defaults. This paper develops a comprehensive framework for evaluating the quality of standard rating systems. We suggest a number of principles that ought to be met by "good rating practice". These "generally accepted rating principles" are potentially relevant for the improvement of existing rating systems. They are also relevant for the development of certification standards for internal rating systems, as currently discussed in a consultative paper issued by the Bank for International Settlements in Basle, entitled "A new capital adequacy framework". We would very much appreciate any comments by readers that help to develop these rating standards further.

---

1 Objectives of the paper

The rating of borrowers is a widespread practice in capital markets. It is meant to summarize the quality of a debtor and, in particular, to inform the market about repayment prospects. Apart from so-called external ratings by agencies, there are also internal ratings by banks and other financial intermediaries providing debt finance to corporates. While external ratings by agencies are available since many years, in fact since 1910 for Moody's, the oldest agency, internal ratings by commercial banks are a more recent development. Their history in most cases does not exceed 5-10 years.

This paper is a first attempt to answer a simple question: What are criteria for good rating practice? We will propose a consistent set of rules that an appropriate rating system should meet. These rating standards, presented below, are not only a collection of "best practice"-rules. Instead, the standards will also be motivated from a decision-theoretic and a statistical perspective, and by examining internal ratings systems currently used in Germany. We have derived insights on properties of actual rating systems from the investigation of two special data sets that contain detailed information on corporate ratings, see Elsas et.al. (1998), Elsas/Krahnen (1998) and Machauer/Weber (1998) for the first data set, Weber/Krahnen/Vossmann (1998) and Brunner/Krahnen/Weber (2000) for the second. By itself, these standards will provide a guideline for the development of new rating systems, and they will help to improve existing systems. Furthermore, they will help to evaluate established systems, as it is already practiced by auditing firms, rating agencies and, occasionally, by supervisory authorities.

It is common among practitioners to distinguish between borrower rating and facility rating. The former relates to the borrower as a legal entity, while the latter relates to a specific loan-cum-collateral. In this paper we concentrate on borrower ratings alone. The empirical basis for our analyses and suggestions is derived from internal rating systems common among major German universal banks. Internal rating systems are therefore the primary fields of application for our principles; at this stage we leave open the question of their applicability to external ratings.

On a broader level, the paper also wants to contribute to the economics of ratings. In particular, we will discuss the ability of ratings to establish credibility vis-à-vis external observers as, for instance, supervisory authorities, and market participants. Of course, credibility of rating information is closely related to acceptable rating requirements. A consultative paper by the Bank for International Settlements in Basle (1999) has put the discussion about the proper role of ratings, notably internal ratings, in the forefront of financial policy debate. Under the title "A New Capital Adequacy Framework", the Basle commission issued a report on how to modify the current international standards on capital adequacy of financial institutions. The current standards, dating back to 1988, require banks to put a 8% equity position against its risky assets, in particular its corporate loans. No consistent distinction is made between high risk and low risk assets.

In the proposed new equity standards, the capital to be held against assets should match implied default risk. A variety of ways how to account for differences of default risk in the loan book can be thought of. Information provided by rating agencies is one way how to deal with different risk categories in a bank's loan book, ratings by lending institution's own internal models is another one. This paper attempts to propose a set of rules sound rating practice should respect. In doing so, we mainly rely on our experiences with internal rating systems in Germany.

Its remaining parts are organized as follows. Section 2 outlines the economic background for an understanding of rating methodologies. Section 3 contains our main contribution. It presents and discusses a list of 14 rating principles that, in our opinion, every rating system should fulfill. Section 4 discusses further implications for the credibility of ratings, and points at a number of open research questions.

2 Economic background

2.1 Why ratings matter

Rating categories, typically letter labeled (AAA or Aaa for prime quality), or simply numbered (1 to 10, say), are a shorthand to quantify credit risk. On the basis of historical data, ratings can be related to the relative frequency of defaults (default-mode paradigm), or they become the basis for the valuation of an asset (mark-to-model paradigm). The most prominent application relates to corporate asset-liability management, where RAROC- (risk-adjusted return on capital) numbers are used to benchmark divisional performance. Ratings allow to measure credit risk, and to manage consistently a bank's credit portfolio, i.e. to alter the banks' exposure with respect to type of risk. In particular, ratings are useful for the pricing of a bond or a loan, reflecting an intended positive relation between expected credit risk and nominal return.

For all these reasons, the quality of a financial institution's rating system has attracted attention from many parties. Auditing firms discuss the risk reporting systems of a corporation in the annual report, rating agencies evaluate the risk assessment system of a borrower who wants to issue asset backed securities, and supervisory authorities are expected to start soon to certify institutional rating systems and credit risk models.

A final remark is in order about the differences between two types of ratings: internal and external. External ratings are generated by rating agencies. These agencies specialize in the production of rating information about corporate or sovereign borrowers, they do not engage in the underwriting of these risks. The rating information is made public, while the rating process itself remains non-disclosed. Internal ratings, in contrast, are produced by financial intermediaries (notably banks) to evaluate the risks they take into their own books. The rating information is seen as a source of competitive advantage, because it is believed to contain proprietary information, and is therefore not made public. Even the firm being rated is typically not informed about its current internal rating.

While there is a growing empirical literature on the validity and the reliability of external ratings (see notably Ederington/Yawitz/Roberts (1987), and Blume/Lim/MacKinlay 1998), and on the informational content of external rating changes (see Hand/Holthausen/Leftwich 1992

and Liu/Seyyed/Smith 1999), there is still very little published work on the methodology and the empirics of internal ratings. A notable exception that relies on data from the US is Treacy/Carey (1998) and Carey (1998). We will base our subsequent discussion on our experiences and insights derived primarily from internal ratings of major German banks.

2.2 Ratings and default risk

We define a rating of a corporate as the mapping of the POD, the expected probability of default, into a discrete number of quality classes, or rating categories[1]. The POD is a continuous variable, bounded by zero from below and by one from above.

$$\text{POD:} \quad \text{Companies} \rightarrow [0, 1] \tag{1}$$

An POD is the expected relative frequency of a credit event, where the latter is defined as a non-payment of principal or interest due (over a period of at least 30 days, say). The POD is one component of a lenders' expected loss, as in (2).

$$E(L) = POD \cdot E(LGD) \tag{2}$$

Here, *E(L)* is expected loss, and *E(LGD)* is the expected loss given default. The expectations are taken over a common time interval, usually one year in the future. Expected loss is thus the average amount a lender is expecting to loose over the next twelve months.

---
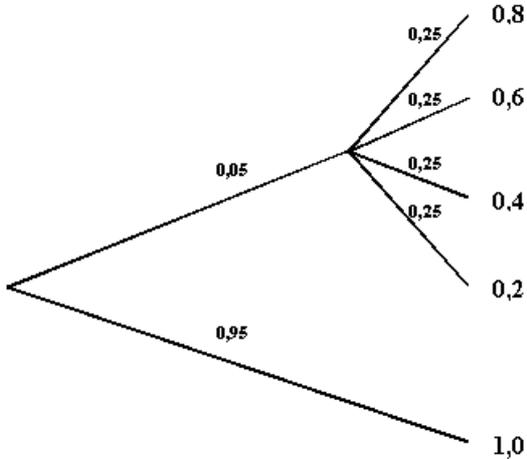
[1] This definition is less innocent as it may first appear. In particular, if rating captures POD (expected probability of default), but not LGD (loss given default), then in general there will be no direct relation between rating and credit spread. To see this, consider the following simple example of two firms A and B with an identical POD: Assume firm A to have a low LGD, while B's LGD is high. In equilibrium, the observable spreads for A loans and B loans have to be set such that the creditor breaks even in expected values. Therefore, the B-spreads have to be larger than the A-spreads. Note the tradeoff: Either we define ratings to measure expected default probability, or we let ratings proxy for expected loss. While the former definition is in line with the interpretation given by credit officers, and by the agencies as implicit in their historical default rates tables (see Moody's 1999b, Standard & Poor's 1998), it does not allow to relate statistically ratings to spreads.

Figure 1 exemplifies the calculation of expected loss on the assumption that default probability is .05, and recoveries vary discretely between 80% and 20%, with equal probability. All values are expressed as percentages of the loan outstanding at the time of default.

Expression (2) and Figure 1 essentially assume that the incidence of default and the severity of a given default are independently distributed random variables. Thus, losses will typically vary between zero and one hundred percent (in our example: between 20% and 100%). A more general expression allows for a non-zero covariance between *POD* and *LGD*:

$$E(L) = E(POD \cdot LGD) \tag{3}$$

The distribution of losses around their expected value is an important measure of overall (institutional) value at risk. The unexpected risk is the number of standard deviations a given quantile (99%) lies away from its expected value. Unexpected losses are addressed by recent value-at-risk tools[2].

Though in theory, PODs are mapped in rating classes, in practice it is the other way round. Rating classes are mapped into PODs on the basis of historical data. The established agencies, notably S&P and Moody's, use historical default rates to calibrate their model. The default rate is the percentage of all bond issues outstanding at *t* that will have a credit event between *t* and *t+1*, e.g., a 12 months period. Conceptually, there is no simple direct relation (linear or log-linear, say) between ordinal rating notches and cardinal PODs. Empirical studies using studies of S&P and Moody's have found a exponential relation between POD and rating notch.

In this paper we focus exclusively on PODs as the objective of rating systems. The typical client we have in mind is a corporate entity, a firm, not an individual borrower, nor a financial instrument, like an asset backed transaction[3]. A distinct set of principles may has to be developed in order to deal with LGD, which is beyond the scope of this paper. In differentiating between POD and LGD, and in clarifying the objective of rating systems, we are in line with the ideas expressed in the BIS consultative paper (Bank for International Settlements 1999, Annex 2)

2.3 Rating models

There are a variety of procedures to arrive at a rating, i.e. a discretized POD-measure. The typical procedure used today is the scoring method. It relies on a well-defined set of criteria, each of which is scored separately. The individual scores relating to the set of criteria are weighted and then added up, yielding the overall score. This score is translated in one of the rating classes, defined as an interval on the real line that extends from minimum overall score to its maximum.

A well known example is the z-score proposed by Edward Altman in 1977 (see Altman/Saunders 1997 for a survey and further references). This author has suggested to

---

[2]   See Saunders 1999, Wahrenburg 1999.

[3]   Though our principles might apply to individuals and financial instruments as well, we will confine our discussion in this paper exclusively to corporate loans.

regress historical default experience on a set of accounting variables (mostly balance sheet and P&L) in order to determine an optimal separating function between issuers that defaulted later on and those that survived. The weights of the estimated function are then used to predict default probability for an individual firm, called the z-score. This z-score may again be translated into a rating class (see Caouette/Altman/Narayanan 1998, chapter 10).

A different approach to rating is exemplified by KMV's public firm model. Building on option pricing theory KMV, a data vendor, derives default estimates from expected movements of stock prices over a specified period of time, typically one year. In contrast to the scoring approach, there is no need here to collect a variety of firm-related, fundamental information, nor is there any weighting function needed. It only requires a time series of observable stock market prices and an estimate of firm indebtedness.

We will next turn to internal rating systems, as exemplified by the systems in place at major German banks. A recent study contains a more detailed description of their individual rating models, and presents an empirical analysis of the determinants of ratings. (see Brunner/Krahnen/Weber 2000). All institutions included in the study apply the scoring methodology, as defined in (4). It specifies a number of distinct criteria a_i, an equal number of value functions v_i, and an aggregation rule, typically linear, with weights k_i.

$$v(a) = \sum_i k_i v_i (a_i) \tag{4}$$

Differences across institutions refer to the list of criteria, particularly the importance given to so-called soft factors, or qualitative criteria. This includes the assessment of management quality, or a general forecast of the prospects of the firm in its market. Table 1 gives a summary assessment of these criteria. It can be seen that the banks typically draw on a list of criteria comparable to the one utilized by S&P, or Moody's.

S&P	Moody's	Typical bank in our sample
Financial risk: ➢ Balance sheet and P&L ➢ Financial policy ➢ Return ➢ Capital structure ➢ Cash-flow ➢ Financial flexibility	**Financial risk**: ➢ Cash-flow ➢ Liquidity ➢ Debt structure ➢ Equity and reserves	**Economic situation**: ➢ Earnings (Cash-flow, return,...) ➢ Financial situation (Capital structure, liquidity,...)
Business risk: ➢ Industry code ➢ Competitive situation ➢ Management	**Competition and business risk**: ➢ Relative market share, competitive position ➢ Diversification ➢ Turnover, costs, returns ➢ Sales and purchases **Legal structure and legal risk**: ➢ Consolidation of related firms **Quality of management**: ➢ Planning and controlling ➢ Managerial track record ➢ Organizational structure ➢ Entrepreneurial succession	**Business situation**: ➢ Industry assessment ➢ USP and competition ➢ Product mix ➢ Special risks ➢ Forecasts: earnings and liquidity ➢ Legal structure **(Quality of) management**: ➢ Experience ➢ Succession ➢ Quality of accounting and controlling
		Customer relationship, account management

Table 1: *Rating criteria of agencies and banks.*
> Sources: Moody's 1999a, Standard & Poor's 1999, Brunner/Krahnen/Weber 2000, see also IMF 1999, annex V.

Table 1 suggests that internal and external rating systems are relying on a similar set of explanatory variables. With respect to the number of rating classes, Standard & Poor's and Moody's each have 22 rating notches (excluding the watchlist), whereas internal rating systems of commercial banks typically have less, e.g. 6-10 rating classes.

Though we have no information about the aggregation process by which agencies derive their final ratings from the underlying criteria, we proceed under the assumption that general accepted rating requirements may apply to external agencies and internal models alike.

3 Rating requirements: What should a good rating system be like?

In the following, we will derive properties good rating systems should obey. These properties can be a foundation for what we propose to be "generally accepted rating principles". We will call these principles occasionally "requirements". Altogether, we come up with 14 requirements, some of which are formally derived, some of which are empirically founded,

some of which are inspired by the recent publication of the Basle Committee on Banking Supervision, and some of which we learnt from talking to high level practitioners.

3.1 A rating system is a mapping

Rating systems are what is mathematically called a function:

R: {companies}   →   {Rating-values},

meaning that the rating system R is a function which assigns each element of the set of companies to a rating value. These rating values, or short ratings, can be categories, i.e. {A, B+, B, B-, ...}, or values of an interval $[r_{min} , r_{max}.]$. R(company X) = .67 means that the rating system R assigns the rating value of .67 to company X. We will assume that rating categories and values can be ranked, i.e. A ⊃ B means, that rating category A is better, in the sense of a lower default probability, than rating category B. The symbol „~" means that both ratings are identical.

This simple mathematical definition of rating systems as functions allows us to define the first requirements without specifying at this point, what rating really means.

> Requirement 1 (Comprehensiveness): A bank's rating system should be able to rate all past, current and future clients.

This requirement defines the potential set of companies to be rated. A bank's rating system should be able to cope with all clients possible. Of course, this requirement is quite general, and hard to meet. There may be future clients, and risk criteria, a given bank may not even imagine. There may be past clients who do not exist any more. However, a bank should make any effort possible to ensure that its rating system is flexible enough to cope with all foreseeable types of risk. It should not happen, e.g., that foreign companies can not be rated or that the rating system is not able to handle certain industries.

> Requirement 2 (Completeness): A bank should rate all current clients and keep on rating its past clients.

The requirement states that a bank should rate all its current clients. This is rather trivial and will in most cases be current management practice. In addition we require that a bank should keep on rating its past clients. This might not be easy and in certain cases it might not even be possible. Accounting data as well as qualitative data from talking to the companies' management might not be available for past clients. Nevertheless, we think that a bank should put effort in maintaining its rating data base. It is of central importance for any type of back-testing and further development of the bank's rating that the bank has an ongoing set of rating data. If the bank stops rating clients which, e.g., defaulted, the set of companies which are in the rating database can be biased. Such a bank would know nothing about the probabilities of events that happen after a default: how likely is the success of restructuring, etc. The survivorship bias (to consider "surviving companies" only) is well known from empirical work in capital markets.

> Requirement 3 (Complexity): A bank should have as many different rating systems as necessary and as few as possible. The reasons for choosing the number of rating systems should be made transparent.

We have to ask, if there should be one function R or if we allow for different functions. From a mathematical perspective it does not matter. We can make one function R so complex that it can be applied to all companies or we can "split" this function up into different functions. In practice, however, there are different aspects to be considered. One function would be a rating system which could be applied to foreign real estate companies as well as to medium sized companies in Southern Germany. The complexity of such a system, however, would make it difficult to use in an organization. Quite a number of aspects are important in evaluating real estate companies which are of no interest if a manufacturer is considered. On the other hand, one should not divide the set of companies into too many subsets, i.e. construct too many different rating systems. Certain companies might fall into more than one system, too many rating systems might ask too much from the credit officers, and the rating systems might be difficult to backtest due to relatively small data-pools. It is for this reason that we recommend to balance both aspects. In addition, we suggest the reason for choosing a certain number of rating systems to be made transparent.

3.2 Rating systems map probabilities of default

In section 2 we have argued that the probability of default is the central variable to be considered when a bank wants to judge the risk of a single loan. In this section, we will define requirements that link rating systems to probabilities of default (POD).

Requirement 4 (POD-definition): Probabilities of default have to be well defined.

This requirement states that a bank has to have a proper definition what its PODs mean. The bank has to define what it considers to be a default event. We found that financial institutions rely on a variety of definitions of a default event, e.g. loan loss provision, or failure to pay interest, or principal, over a specified time span. Note that without a harmonization of default definitions, it will prove difficult to pool POD-data across banks. We therefore suggest that the industry works towards a common definition of POD, which is both transparent and reasonable. In addition, financial institutions have to state the time horizon within which a default is considered. Some banks just consider one time horizon (mostly one year), some other consider multiple time horizons which lead to different sets of PODs. Still other institutions, notably ratings agencies, estimate PODs by averaging over a complete business cycle[4]. The ultimate goal should be a term structure of ratings or, for that matters, PODs that capture default risk beyond the one year horizon. For example, a company might have a small POD over the next two years, and a large POD for year three (when a patent will have expired).

Requirement 5 (Monotonicity):

i) $POD(\text{company } X) = POD(\text{company } Y) \Rightarrow R(\text{company } X) \sim R(\text{company } Y)$,

ii) $POD(\text{company } X) < POD(\text{company } Y) \Rightarrow R(\text{company } X) \subseteq R(\text{company } Y)$,

iii) $R(\text{company } X) \supset R(\text{company } Y) \Rightarrow POD(\text{company } X) < POD(\text{company } Y)$.

---

[4]   The averaging of default estimates of a cycle is, in our view, problematic if the objective of a POD-assessment lies in specifying minimum equity requirements.

This requirement defines the relation between ratings and expected default frequencies. As discussed before, we take POD as the primitive and derive rating from there on. If two PODs are identical, the ratings also have to be identical (case i). If the POD of company X is smaller than that of company Y (case ii), the rating of company X has to be at least as good as that of company Y. To illustrate the weak inequality for ratings, let us consider a bank which only has two rating categories {good, bad} with good $\supset$ bad. This might be a bank which only wants to know if a credit should be given or not. Case (ii) allows two different PODs to yield the same rating. If the rating of a company is better than that of another company (case iii), the POD of the first company should be smaller than the POD of the second one. Note that (iii) is implied by (i) and (ii).

> Requirement 6 (Fineness): The rating system can vary in the degree of fineness. It should always be as fine as necessary.

Looking back at requirement 5, the central question for the definition of a rating system now remains, how fine a rating system should be, i.e. how many categories it should have. It could be as fine as the POD itself, being basically identical to POD, or it could map PODs into a finite number of categories. Of course, a rating system which models POD would be the most exact one. However, for quite a number of situations a less fine rating system would be sufficient and more appropriate in an organizational context. The fineness of a rating system can not be considered independently from back-testing (see Requirement 8). There is no use in defining a large number of rating categories, if a bank is not able to back-test consistently, due to lack of data.

Thus the fineness of the rating system is a function of its intended use. It is therefore that one should allow rating systems to communicate PODs in different degrees of fineness. For pricing the rating system should be finer than for defining credit limits. Some banks, e.g., use traffic lights (three categories: red, yellow, green) to attract the attention of the credit officer to more or less risky credits. Knowing the conversion of rating into POD will always allow us to transform one way of communication into the other.

> Requirement 7 (Reliability): The rating system should be reliable.

Suppose, that a company has some true POD. Then the rating should be identical regardless of the person who rates, or the point in time when the rating is done. Note, that this requirement does not assume that the rating does not change. The rating might change with the creditworthiness of the client, or along the economic cycle. However, it should stay constant, if the creditworthiness does not change. An example to test for the stationarity property of the data set is explained in Blume/Lim/MacKinlay 1998.

3.3 Do rating systems *really* map probabilities of default?

Now that we have defined some first key requirements for rating systems, the question remains how a bank or even a supervisory agency makes sure that the rating system is correct. Thus it is required that ratings (or PODs) are rational forecasts on the basis of all available information, being the best ex-ante predictor of credit risk.

Credit ratings can be technically incorrect, i.e. even if applied properly their values do not correspond to the (ex-post) number of realized defaults. In addition, rating systems which are

technically correct can be used in a way that the resulting ratings do not mimic PODs anymore. We will discuss the first class of problems first.

A POD is based on an ex-ante point of view. It states that a company with an POD of 0.7% has a 7 in a 1000 chance to default within a given time period. We know from research on capital markets that testing expectations is always tricky. In order to relate (ex-ante) expectations to (ex-post) observed data, we have to assume that the structure of the problem under consideration remains constant from the date where expectations are formed to the date where observations are taken. This assumption is called the stationarity assumption. We will assume stationarity for the next requirement. Nevertheless, we are aware that in the future, statistical methods will have to be introduced that account for possible non-stationarities.

> Requirement 8 (Back-testing): The (ex-ante) probability of default should not be significantly different from the (ex-post) realized default frequency.

Requirement 8 basically states that what you expect is what you should get. It also stresses the need of a data-base to fulfill this requirement. Back-testing in credit-management is especially difficult because first, there are no market prices for most types of credits and second, there are so few historical data of credit defaults. As we will argue in more detail in section 4, it might be useful to pool resources across different banks to create a better data-base which allows for an improved back-testing.

Since back-testing is central for validating a rating, the need for it yields some important implications for the design and use of ratings. As already mentioned, a bank should not have too many rating systems (i.e. define to many subsets of companies) and it should not change the rating system too often.

There are numerous ways of testing rating systems, and apparently a number of them are already used in the industry. Test procedures are related to back-testing and they may be seen as defining necessary conditions for the appropriateness of the rating system:

- Ex-post default rates within any given rating category should be larger than that of a higher (i.e. better) rating category.

Even if we do not know whether a cardinal relation between rating and POD can be assumed, the above condition will test ordinality.

- Ex-post default rates should increase with the time horizon.

It is obvious that the default rates of companies based on a time horizon of five years have to be equal or greater than those based on a time horizon of one year.

- For companies with corporate bonds outstanding, credit spreads may be compared to internal credit ratings.

Across companies, the bank will be able to compare the risk-ordering implied by the market with the risk-ordering implied by credit-ratings.

Besides back-testing, credit ratings have to obey certain structural and technical necessities (see Weber/Krahnen/Voßmann 1998 for further details).

> Requirement 9 (Informational efficiency): Ratings should be informationally efficient, i.e. it should not be possible to predict rating changes based on rating history. All the available information should be modeled correctly in the rating. The rating system

should cope with biases known from the general literature on rating (splitting bias, range bias, etc.).

As mentioned before, a rating should correctly incorporate all information available to the bank, both public and private, i.e. it should be efficient. This requirement is identical to the use of the term "information efficiency" in financial markets. Today's rating should be the best predictor for tomorrow's rating, i.e. it should not be possible to get information about tomorrow's rating by knowing which rating the company had yesterday (or in earlier periods). In addition, quite a number of biases known from the psychological literature on judgment have to be taken care of when designing a rating system. Credit officers may, e.g., have the tendency to rate qualitative criteria of a rating system better than quantitative ones and they tend to change qualitative variables less than quantitative (Brunner/Krahnen/Weber 2000)[5].

Requirement 10 (System development): A rating system has to be improved over time.

It might sound trivial but after a bank has seen deficiencies in its rating, it should be willing to change it. Such a change can result from back-testing and from ex-ante management insight. Management might know that the structure and the aggregation of variables to estimate creditworthiness have changed, i.e. stationarity is violated. One should not wait until (ex-post) back-testing forces system modifications, provided that ex-ante insights had suggested these changes already. A modification of the system has to be carefully considered. There are large costs (back-testing is more difficult, education of credit staff, etc.) and in some cases uncertain benefits.

Requirement 11 (Data management): Past and current rating data should be easily available.

A modern data management is a prerequisite for successful back-testing as well as successful system development. Any type of statistical analysis requires data to be (easily) available. Even if the fulfillment of this requirement seems to be easy on a first glance, we are well aware of problems which can arise in practice. The change of a bank's computer system, the further development of an existing rating system, the introduction of a finer rating system, a change in the organizational structure of the rating process, a merger of two banks are just examples to demonstrate that the requirement can pose a serious challenge. However, without a well maintained data management, no testing of a rating system will be possible.

3.4 Good rating systems account for incentive problems

Ratings compile objective and subjective information. The higher the share of subjective, or soft information, the more difficult is the detection of untruthful reporting. This may be a considerable problem, because credit officers in charge of rating a particular client may have an incentive to underestimate the risk of a loan, e.g. to overestimate the quality of a particular management. For instance, in some institutions, loan responsibility migrates from the credit officer to a special work-out group, once the rating falls below a critical value. This organizational rule may induce the credit officer to adjust his or her risk assessment to the

---

[5] Even on an efficient market, due to the categorial nature of ratings, first differences (i.e. rating changes) will not necessarily be distributed like independent random variables.

point where control over the customer is not migrating. Another example of how an organizational rule may affect reporting incentives relates to bonus systems, where performance measures depend on ratings.

> Requirement 12 (Incentive compatibility): The rating process has to be embedded in the organization of credit business such that the risk of misrepresentation by credit officers is minimized.

We know of no simple test of organizational incentive compatibility, but several rules of thumb are available. First, and inspired by the above example, possible "critical values" of rating assessments that trigger action have to be recorded and followed up. In particular, measures of statistical similarity and significance may help to identify unusual frequencies of specific rating decisions, or rating migrations. Second, the internal reward system of the institution may or may not be related to past rating performance of loan officers. As a rule, an officer's rating history should "stick to him". For example, a significant, above average frequency of rating revisions after the officer in question has moved from his post, or authority for certain loans been moved away from him, could have a predictable (and negative) impact on his overall evaluation. The fulfillment of Requirement 12 can be checked by asking to what extent management has thought about possible incentive conflicts caused by the organizational design of the lending process, and what it has done to control for its behavioral consequences.

However, we do not advocate the minimization of discretionary decisions in the rating process, because the specific value added (in terms of incremental information) by internal ratings mainly consist of aggregating "soft", or subjective information produced by the loan officer. A certain degree of consistency check may help to improve incentives, and to establish credibility of the overall rating process. This is summarized in the following requirement:

> Requirement 13 (Internal compliance): The distribution of rating outcomes is constantly monitored by controllers, assisted by random inspections.

In order to identify systematic biases in the evaluations of loan officers, all ratings and their histories are to be kept in a back-testing file (see Requirement 11). Rating quality maintenance has to develop (and, of course, to apply) statistical test routines that are capable of identifying significant variations in rating decisions over time, or across firms. The task resembles a statistical quality control as it is common in, e.g. production management.[6] The follow-up to these statistical tests could be a partial or complete replication of past ratings.

Fulfillment of Requirement 13 would not only allow the detection of specific behavioral patterns, but also would strengthen the Incentive Compatibility (Requirement 12). In order to have some deterrent effect, the algorithms of the sampling plan must not be completely transparent to loan officers. Again, outside rating quality assessment would try to clarify to what extent sampling plans have been developed, and are applied consistently.

---

[6] Building on well established methodologies of random quality inspections, a continuous sampling plan may prove helpful (see Shirland 1993, or Krishnaiah/Rao 1988). Such a plan specifies a set of algorithms that would analyze the similarity of specific rating subsets pertaining to, e.g., a cross section of ratings given within an industry, or a time series of ratings given by a particular officer.

The adherence of a bank's management to its agreed rating standards is monitored by neutral (uninterested) outside controllers, either on a continuous, or on a random basis.

Requirement 14, though similar in nature to the preceding Requirement 13, is the keystone for establishing credibility to rating data produced by an interested party. Here, interested party refers to, e.g. banks as providers of internal ratings. A bank's "interest" derives from the underwriting of credit risk vis-à-vis the customer that has been rated. Requirement 14 involves an evaluation by an outside party, e.g. a supervisory authority. Past ratings have to be shown to be without biases, or deliberate misrepresentation. Therefore, external compliance is not about the informational value of any particular rating, but rather it is about the consistency of its use. The methodology applied to control external compliance is likely to be similar to the one used in Requirement 13[7].

4 Policy considerations and agenda

In the concluding section we want to address two questions. First, is regulation of the rating process really needed? We will argue that indeed some type of outside regulation is required to safeguard credibility of internal ratings. Second, we will point out additional needs in two areas which are of great importance to the future acceptance of rating as a risk measurement instrument, namely a need for research, and a need for better, and larger data-bases.

4.1 Is there a need for external supervision of internal ratings?

The BIS consultative paper as of June 1999 gives some consideration to ratings as a basis for the assessment of bank capital requirements. The regulatory importance of ratings do apply not only for external ratings, but also for internal ratings. To answer the question of whether or not internal ratings should be certified and constantly supervised by a regulator, or an auditor, we will first compare the processes by which internal and external ratings earn credibility. There are basically two models: In the first, professional (external) rating agencies produce public rating information without doing any underwriting; their credibility derives from reputation in the market place. In the second model, bank loan departments produce private (or internal) rating information on the basis of an underwriting business. Here, credibility derives from the shareholder value interest of bank management, and hence its credit department, in a proper loan repayment.

Let us start with external ratings. Default probability estimates by specialized agencies (S&P, Moody's, Fitch IBCA, a.o.) draw on the agency's reputation as a provider of accurate default predictions. Reputational value stems from the impact ratings exert on credit spreads. Thus, reputational value is high (low), if a change of published ratings has a significant (an insignificant) influence on corporate cost of capital. This means that ratings have to be accepted as a proxy for true fundamental information in order to be valuable in the market.

---

[7] The methodological question of reliability of rating decisions is not trivial - even if it comes to large data sets, as those assembled by the agencies.

The market value of a rating agency, its franchise value, is therefore directly related to the discounted stream of cost of capital effects that are due to its corporate ratings. Firms are willing to pay a fee to an agency up front in order to receive public rating. By the same line of argument faulty ratings will, if detected, eventually destroy the franchise value of a rating agency.

Thus, the reputational argument developed in the preceding paragraph claims that agencies have a proper incentive to produce true and unbiased corporate ratings. Of course, the reputational model of incentive compatibility is subject to an important caveat. It relies on the market being able to detect faulty ratings ex-post. What is needed, therefore, is a statistical methodology to spot changes of the distribution function (or, for that matter, of the rating behavior of an Agency) relatively soon after its onset.

Note that the reputational argument for external ratings will have to take into account the fact that rated companies usually pay the Agency for providing the rating label. There is a natural incentive problem here, and ratings probably derive much of their value from the Agency's reputation for being unwavering in their high standards.

Let us now turn to internal ratings and possible determinants of their credibility. The basic credibility explanation is simple: Internal ratings are private information, typically not even communicated to the rated firm itself. Of great importance for internal ratings is their ability to incorporate all types of information accessible that may contribute to a good default forecast. In particular, a relationship-based financial system may be in a position to exploit not-easy-to-measure qualitative information, and thereby improve estimates. This includes insider information due to, e.g., account surveillance, and advisory business. The comparative advantage of internal ratings in our view refers precisely to this fact, the incorporation of soft information. If one assumes, for simplicity, no incentive conflicts within the financial institution itself, then there is good reason to believe in the unbiasedness of internal ratings [8]. Since a bank underwrites credit risk, she essentially takes a bet on the creditworthiness of any particular borrower. Any bias in POD-estimations would harm the banks competitive position and would eventually impair equity value. True and fair private ratings are therefore in the proper interest of the bank.

However, there is a caveat here as well. The proposed new equity standards outlined in the BIS consultative paper attach, in fact, a sort of shadow price to the internal ratings of bank borrowers. Since the amount of equity required to be held against a given structure of bank assets will then be affected by their internal ratings, there may well be pressure to accommodate rating decisions in the future. Once ratings fulfill a regulatory task, they have a dual function, measuring risk *and* triggering equity charges. These two functions are likely to have opposite incentive effects.

To sum up: In the light of the emerging new equity standards both external and internal ratings constantly have to prove their unbiasedness, and their neutrality. While there is a market test for external ratings (which, in fact, has been effective for many years already), there is no external check for internal ratings so far. One way to test for neutrality of internal ratings is a serious test of rating methodology and rating performance, see Requirement 13. Both may be

---

[8] A within-firm conflict of interest may arise when, e.g., a loan officer tries to avoid a shift of responsibility from himself to a work-out group. They may then accept a better-than-justified rating.

elements in a certification process carried out by a supervisory authority (which, for that matter, may also be delegated to a specialized entity or auditing firm, say).

4.2 What else is needed: Data and research

As pointed out a couple of times, data is key to successfully maintain and develop rating systems. Due to the number of rating classes, the long time periods and the small probabilities of failure, statistical analysis at the level of a single bank might be limited. We advocate to think about the need of a shared data base. Such a data base could aggregate the ratings for a company across different banks (of course, with full confidentiality of each bank's private rating). Based on the joint data, each bank might be able to analyze and validate its internal rating system against some average rating. A combined data base would allow for a more elaborate back-testing thereby preparing the ground for an official recognition of internal rating systems.

In addition, for companies which are rated by quite a number of banks, the aggregate risk rating would reflect something like a "market opinion" of the default risk of a particular company, a specific industrial sector or even the whole economy of a country (thus creating a default risk index of certain entities). A joint data base would also allow to derive correlations between the credit risks of companies, industry sectors and countries as well. Such information is of great importance for the development of credit portfolio models.

Finally, we want to point out some research needs. We have tried to state some requirement, a good rating system should obey. Nevertheless, we have said very little (on purpose) on how to construct a good rating system. Which factors should be elements of the scoring rule? How should the weights for each factor within the scoring model be derived? With respect to the value function, what number should be attached to an average vs. an excellent management? Today, most banks use a mixture of mathematical models and management intuition to construct their systems. We do think this is a good approach, but we would like to know more. Along these lines, it would be interesting to analyze in greater detail how LGD (loss given default) depends upon the state of default (see section 2). Furthermore, methods for back-testing rating systems are not yet well developed. Sophisticated statistical sampling plans are needed to check for internal and external compliance. Equally, statistical tools can be used to correct for any trends (like industry cycles) and biases (like survivorship bias) that are to be found in the raw data. Finally, one would like to see more research being undertaken on the validation of ratings. Given the low data frequency, and thus the long duration for a detection of faulty ratings, the reputational justification of true and fair ratings would also benefit from this exercise. Finally, it may be of interest to use credit rating for optimizing portfolio risk. For this end, estimates of correlations across borrowers, and over time are needed. The correlation structure of corporate loans will help to model risk migration, and the dependence of risk ratings on the economic cycle.

We conclude with a general disclaimer concerning Generally Accepted Rating Principles: Their development is seen as work in progress. And we expect it to remain work in progress for quite some time, since such principles have to be developed jointly by regulators, researchers and experts in the field.

References and recommended readings

Altman, Edward I., and Anthony Saunders (1997): Credit risk measurement: developments over the last 20 years, *Journal of Banking and Finance* 21, 1721-1742.

Bank for International Settlements (1999): A new capital adequacy framework. Consultative paper issued by the Basel Committee on Banking Supervision, Basel, June 1999.

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

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 (1997): Can relationship banking survive competition?, Discussion Paper No. 1592, Center for Economic Policy Research, London.

Boyes, William.J., Dennis L. Hoffman, and Stuart A. Low (1989): An econometric analysis of the bank credit scoring problem, *Journal of Econometrics* 40, 3-14.

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

Brunner, Antje, Jan P. Krahnen and Martin Weber (2000): Information production in lending relationships: On the role of corporate ratings in commercial banking, CFS-Working Paper, in progress.

Caouette, John B., Edward I. Altman and Paul Narayanan (1998): *Managing credit risk: The next great financial challenge*, Wiley Frontiers in Finance.

Carey, Mark S. (1998): Credit risk in private debt portfolios, *Journal of Finance* 53, 1363-1387.

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.

Ederington, Louis H., Yawitz, Jess B., and Brian E. Roberts (1987): The informational content of bond ratings, *Journal of Financial Research* 19, 211-261.

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 98/14, Center for Financial Studies, Frankfurt/Main.

Elsas, Ralf, and Jan Pieter Krahnen (1998): Is relationship lending special? Evidence from credit-file data in Germany, *Journal of Banking and Finance* 22, Nos. 10-11, 1283-1316.

English, William B., and William R. Nelson (1998): Bank risk rating of business loans, Working Paper, Washington: Federal Reserve Board, November.

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

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.

Hite, Gailen, and Arthur Warga (1997): The effect of bond-rating changes on bond price performance,

Financial Analysts Journal 53, May-June, 35-51.

International Monetary Fund (1999): International capital markets: Developments, prospects, and key policy issues, September (www: imf.org/external/pubs/ft/icm/1999/index.htm).

Jackson, Patricia, and William Perraudin (1999): Regulatory implications of credit risk modelling, Working Paper.

Krishnaiah, P.R., and C.R. Rao, Eds. (1988): Handbook of statistics, vol. 7: Quality Control and Reliability, Amsterdam: North-Holland.

Liu. Pu, Seyyed, Fazal J., Stanley D. Smith (1999): The independent impact of credit rating changes – The case of Moody's rating refinement on yield premiums, *Journal of Business Finance and Accounting* 26, 337-363.

Machauer, Achim, and Martin Weber (1998): Bank behavior based on internal credit ratings of borrowers, *Journal of Banking and Finance* 22, Nos. 10-11, 1355-1383.

McAllister, Patrick H., and John J. Mingo (1994): Commercial loan risk management, credit scoring, and pricing: The need for a new shared database, *Journal of Commercial Lending*, May, 6-22.

Moody's Investor Service (1999a): Measuring private firm default risk, Special Comment, June.

Moody's Investor Service (1999b): Historical default rates of corporate bond issuers, 1920-1998, Special comment, January.

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.

Saunders, Anthony (1999): Credit risk measurement. New approaches to value at risk and other paradigms, New York: Wiley.

Shirland, L.E. (1993): Statistical quality control with microcomputer applications. New York: Wiley.

Standard & Poor's (1999): 1999 Corporate ratings criteria, (www: standardandpoors.com).

Standard & Poor's (1998): CreditPro, May.

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

Treacy, William F., and Mark S. Carey (1998): Credit risk ratings at large U.S. banks, Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin*, November, 897-921.

Wahrenburg, Mark (2000): Vergleichende Analyse alternativer Kreditrisikomodelle, *Kredit und Kapital,* Heft 2, S. 235-257.

Weber, Martin, Jan Pieter Krahnen and Frank Voßmann (1998): Risikomessung im Kreditgeschäft: Eine empirische Analyse bankinterner Ratingverfahren, *Zeitschrift für betriebswirtschaftliche Forschung,* Sonderheft 41, S. 117-142.