ELSEVIER Journal of Banking & Finance 24 (2000) 167–201

Journal of BANKING & FINANCE

www.elsevier.com/locate/econbase

Credit risk rating systems at large US banks ☆

William F. Treacy, Mark Carey *

Federal Reserve Board, Washington, DC 20551, USA

Abstract

Internal credit risk rating systems are becoming an increasingly important element of large commercial banks' measurement and management of the credit risk of both individual exposures and portfolios. This article describes the internal rating systems presently in use at the 50 largest US banking organizations. We use the diversity of current practice to illuminate the relationships between uses of ratings, different options for rating system design, and the effectiveness of internal rating systems. Growing stresses on rating systems make an understanding of such relationships important for both banks and regulators. © 2000 Published by Elsevier Science B.V. All rights reserved.

JEL classification: G20; G21

Keywords: Ratings; Credit risk; Risk management; Bank risk

1. Introduction

Internal credit ratings are an increasingly important element of credit risk management at large US banks. Their credit-related businesses have become progressively more diverse and complex and the number of their counterparties has grown rapidly, straining the limits of traditional methods of controlling

☆ The views expressed herein are the authors' and do not necessarily reflect those of the Board of Governors or the Federal Reserve System.

* Corresponding author. Tel.: +1-202-452-2784; fax: +1-202-452-5295.
E-mail address: mcarey@frb.gov (M. Carey).

PII: S0378-4266(99)00056-4

and managing credit risk. In response, many large banks have introduced more structured or formal systems for approving loans, portfolio monitoring and management reporting, analysis of the adequacy of loan loss reserves or capital, and profitability and loan pricing analysis. Internal ratings are crucial inputs to all such systems as well as to quantitative portfolio credit risk models. Like a public credit rating produced by agencies such as Moody's or Standard & Poor's, a bank's internal rating summarizes the risk of loss due to failure by a given borrower to pay as promised. However, banks' rating systems differ significantly from those of the agencies, partly because internal ratings are assigned by bank personnel and are usually not revealed to outsiders.

This article describes the internal rating systems presently in use at the 50 largest US banking organizations. We use the diversity of current practice to illuminate the relationships between uses of ratings, different options for rating system design, and the effectiveness of internal rating systems.

An understanding of such relationships is useful to banks, regulators, and researchers. Such understanding can help banks manage transitions to more complex and demanding uses of ratings in risk management. US regulatory agencies already use internal ratings in supervision. Moreover, the Basle Committee is beginning to consider proposals to make international bank capital standards more sensitive to differences in portfolio credit risk, and internal ratings play a key role in several such proposals, two of which are sketched by Mingo (2000). Regulatory reliance on internal ratings would introduce new and powerful stresses on banks' internal rating systems which, if not addressed, could disrupt credit risk management at many banks.

The specifics of internal rating systems currently differ across banks. The number of grades and the risk associated with each grade vary, as do decisions about who assigns ratings and about the manner in which rating assignments are reviewed. To a considerable extent, such variations are an example of form following function. Banks in different lines of business or using internal ratings for different purposes design and operate different systems that meet their needs. For example, a bank that uses ratings mainly to identify deteriorating or problem loans to ensure proper monitoring may find that a rating scale with relatively few grades is adequate, whereas a bank using ratings in computing the relative profitability of different loans may require a scale with many grades in order to achieve fine distinctions of credit risk.

As described by Altman and Saunders (1997), much research on statistical models of debt default and loss has been published over the past few decades. Many banks use statistical models as an element of the rating process, but rating assignment and review almost always involve the exercise of human judgment. Because the factors considered in assigning a rating and the weight given each factor can differ significantly across borrowers, banks (like the rating agencies) generally believe that the current limitations of statistical models are such that properly managed judgmental rating systems deliver more

accurate estimates of risk. Especially for large exposures, the benefits of such accuracy may outweigh the higher costs of judgmental systems, and banks typically produce internal ratings only for business and institutional loans and counterparties. [1] In contrast, statistical credit scores are often the primary basis for credit decisions for small exposures, such as consumer credit. [2]

Given the substantial role of judgment, potentially conflicting staff incentives are an important consideration in rating system design and operation. In the absence of effective internal rating review and control systems, rating assignments may be biased. The direction of such bias tends to be related to a bank's uses of ratings in managing risk. For example, at banks that use ratings in computing risk-adjusted profitability measures or pricing guidelines, the staff may be tempted to assign ratings that are more favorable than warranted. Most banks rely heavily on loan review departments and informal disciplines associated with corporate culture to control incentive conflicts.

Although form generally follows function, rating system design and operation is a complex task, involving considerations of cost, efficiency of information gathering, consistency of ratings produced, and staff incentives, as well as the uses to which ratings are put. Changes in a bank's business and its uses of ratings can cause form and function to diverge, placing stresses on its rating systems that are neither anticipated nor immediately recognized. Failure to relieve severe stresses can compromise the effectiveness of a bank's credit risk management. Outlined below are a number of recommended practices for both banks and regulators. Such practices can help limit stresses and can improve the operation and flexibility of internal rating systems.

This article is based on information from internal reports and credit policy documents for the fifty largest US bank holding companies, from interviews with senior bankers and others at more than 15 major holding companies and other relevant institutions, and from conversations with Federal Reserve bank examiners. The institutions we interviewed cover the spectrum of size and practice among the fifty largest banks, but a disproportionate share

[1] Credit risk can arise from a loan already extended, loan commitments that have not yet been drawn, letters of credit, or obligations under other contracts such as financial derivatives. We follow industry usage by referring to individual loans or commitments as "facilities" and overall credit risk arising from such transactions as "exposure". Throughout this article, we ignore issues of "loan equivalency", that is, the fact that some portion of the unfunded portion of a commitment is exposed to loss because the borrower may draw on the commitment prior to default.

[2] At most large banks, internally rated assets include commercial and industrial loans and facilities, commercial leases, commercial real estate loans, loans to foreign commercial and sovereign entities, loans and other facilities to financial institutions, and sometimes large loans to individuals made by "private banking" units. In general, ratings are produced for exposures for which underwriting requires large elements of subjective analysis.

of the banks we interviewed have relatively advanced internal rating systems. [3]

Although a large literature has examined public rating agency procedures and the properties of their ratings (see Cantor and Packer, 1994; Ederington and Yawitz, 1987; Altman and Saunders, 1997; and references therein), this article is the first to provide a detailed analysis of internal credit risk rating systems. [4] Udell (1987,1989) examined the internal rating systems of a sample of Midwestern US banks as part of a broader study of such banks' loan review systems. Brady et al. (1998) and English and Nelson (1998) offer some information about the internal rating scales of a sample of US banks of all sizes and also report both distributions of loans across grades and relationships between grades and loan pricing for a stratified sample of banks. Robert Morris Associates (1997) and Santomero (1997) surveyed internal rating systems as part of larger studies of banks' credit risk management practices. Machauer and Weber (1998) employ German banks' internal ratings in studying loan pricing patterns.

Sections 2 and 3 describe the architecture and operating design of large banks' internal rating systems, while Section 4 briefly compares such systems to those of Moody's and Standard and Poor's. Section 5 describes the current difficulty of measuring the riskiness of exposures in any given grade and the difficulty of tuning rating systems so that grades have specified loss characteristics. Section 6 presents an estimate of the aggregate credit quality distribution of large US banks' commercial loans. Section 7 describes the uses of internal ratings, Section 8 offers recommendations to both banks and regulators, and Section 9 offers concluding remarks.

2. Architecture

In choosing the architecture of its rating system, a bank must decide which loss concepts to employ, the number and meaning of grades on the rating scale corresponding to each loss concept, and whether to include "Watch" and "regulatory" grades on such scales. The choices made and the reasons for them vary widely, but the primary determinants of bank rating system architecture appear to be the bank's mix of large and smaller borrowers and the extent to which the bank uses quantitative systems for credit risk management and profitability analysis.

[3] Internal rating systems are typically used throughout US banking organizations. For brevity, we use the term "bank" to refer to consolidated banking organizations, not just the chartered bank.

[4] A related article, Treacy and Carey (1998), includes some topics touched on only briefly in this article while omitting other topics.

In principle, banks must also decide whether to grade borrowers according to their current condition or their expected condition under stress. The rating agencies employ the latter, "through the cycle", philosophy, which involves projecting the borrower's condition and probability of default at the trough of an economic or industry cycle and setting the rating accordingly. In contrast, all banks we interviewed set grades to reflect the probability of default over a period of one or a few years based on the borrower's current condition. This difference in philosophy, which is not widely understood, is important to take into account in a variety of circumstances, as discussed further below and in Treacy and Carey (1998). [5]

2.1. Loss concepts and their implementation

The credit risk on a loan or other exposure over a given period involves both the probability of default (PD) and the fraction of the loan's value that is likely to be lost in the event of default (LIED). LIED is always specific to a given exposure. PD, however, is often associated with the borrower, the presumption being that a borrower will default on all obligations if it defaults on any. [6] The product of PD and LIED is the expected loss rate (EL) on the exposure.

The banks at which we conducted interviews generally fall into two categories with regard to loss concept. About 60% have one-dimensional rating systems, in which ratings are assigned only to facilities. In such systems, ratings approximate EL. The remaining 40% have two-dimensional systems, in which the borrower's general creditworthiness (approximately PD) is appraised on one scale while the risk posed by individual exposures (approximately EL) is appraised on another; invariably the two scales have the same number of rating categories. The policy documents of banks we did not interview indicate that they also have one- or two-dimensional rating systems, and it is our impression that the systems use the same loss concepts as the banks we interviewed.

A number of banks would no doubt dispute our characterization of their single-scale systems as measuring EL; in interviews, several maintained that their ratings primarily reflect the borrower's PD. However, collateral and loan structure play a role in grading at such banks both in practical terms and in the definitions of grades. Moreover, certain specialty loans such as cash-collater-

[5] The agencies' through-the-cycle philosophy at least partly accounts for the fact that default rates for any given agency grade vary with the business cycle. The agencies' projections of creditworthiness are most stringently tested at the trough of cycles, and thus it is natural that any errors of optimism in their ratings are most likely to be revealed then.

[6] PD might differ across transactions with the same borrower. For example, a borrower may attempt to force a favorable restructuring of its term loan by halting payment on the loan while continuing to honor the terms of a foreign exchange swap with the same bank. However, for practical purposes, estimating a single probability of any default by a borrower is usually sufficient.

alized loans, those with guarantees, and asset-based loans, can receive relatively low risk grades, reflecting the fact that the EL of such loans is far less than for an "ordinary" loan to the same borrower. Such single-grade systems might be most accurately characterized as having an ambiguous or mixed conceptual basis rather than as clearly measuring either PD or EL. Although an ambiguous basis may pose no problems when ratings are used mainly for administrative and reporting purposes and when the nature of the bank's business is fairly stable over time, a clear conceptual foundation becomes more important as models of portfolio risk and profitability are used more heavily and during periods of rapid change.

In two-dimensional systems, the usual procedure is to first determine the borrower's grade (its PD) and then to set the facility grade equal to the borrower grade unless the structure of the facility makes likely a LIED that is substantially better or worse than normal. Implicitly, grades on the facility scale measure EL as the PD associated with the borrower grade multiplied by a standard or average LIED (an example appears in Table 1). Thus, most bank systems include ratings that embody the EL concept. Two-dimensional systems are advantageous in that they promote precision and consistency in grading by separately recording a rater's judgments about PD and EL rather than mixing them together.

Since our interviews were conducted, a few banks have introduced systems in which the borrower grade reflects PD but the facility grade explicitly measures LIED. In such systems, the rater assigns a facility to one of several LIED categories on the basis of the likely recovery rates associated with various types of collateral, guarantees, or other considerations associated with the facility's structure. EL for a facility can be calculated by multiplying the borrower's PD by the facility's LIED. [7]

2.2. Loss concepts at Moody's and S&P

At the agencies, as at many banks, the loss concepts (PD, LIED, and EL) embedded in ratings are somewhat ambiguous. Moody's Investors Service (1991, p. 73) states that "ratings are intended to serve as indicators or forecasts of the potential for *credit loss* because of failure to pay, a delay in payment, or partial payment." Standard and Poor's (1998, p. 3) states that its ratings are an

[7] Two-dimensional systems recording LIED rather than EL as the second grade appear especially desirable. PD–EL systems typically impose limits on the degree to which differences in loan structure permit an EL grade to be moved up or down relative to the PD grade. Such limits can be helpful in restraining raters' optimism but, in the case of loans with a genuinely very low expected LIED, such limits can materially limit the accuracy of risk measurement. Another benefit of LIED ratings is the fact that raters' LIED judgments can be evaluated over time by comparing them to loss experience.

Table 1
Example of a two-dimensional rating system using average LIED values

Grade	Borrower scale: borrower's probability of default (PD) (%) (1)	Assumed average loss on loans in the event of default (LIED) (%) (2)	Facility scale: expected loss (EL) on loans (%) (1×2)
1 – Virtually no risk	0.0	↑	0.00
2 – Low risk	0.1	|	0.03
3 – Moderate risk	0.3	|	0.09
4 – Average risk	1.0	|	0.30
5 – Acceptable risk	3.0	30	0.90
6 – Borderline risk	6.0	|	1.80
7 – OAEM[a]	20.0	|	6.00
8 – Substandard	60.0	|	18.0
9 – Doubtful	100	↓	30.0

[a] Other assets especially mentioned.

"opinion of the general creditworthiness of an obligor, or ... of an obligor with respect to a particular ... obligation ... based on relevant risk factors." A close reading of the agencies' detailed descriptions of rating criteria and procedures gives the impression that both agencies' ratings incorporate elements of PD and LIED but are not precisely EL measures.

2.3. Administrative grades

All the banks we interviewed maintain some sort of internal "Watch" list as well as a means of identifying assets that fall into the "regulatory problem asset" grades other assets especially mentioned (OAEM), substandard, doubtful, and loss (all other assets are collectively labeled "Pass"). [8] Although Watch and regulatory problem-asset designations typically identify high-risk credits, they have administrative meanings that are conceptually separate from risk per se. Special monitoring activity is usually undertaken for such assets, such as formal quarterly reviews of status and special reports that help senior bank management monitor and react to important developments in the portfolio. However, banks may wish to trigger special monitoring for credits that are not high-risk and thus may wish to separate administrative indicators from risk measures (an example would be a low-risk loan for which an event that might influence risk is expected, such as a change in ownership of the borrower).

[8] Bank examiners, among other responsibilities, identify high risk and troubled loans and ensure they are properly classified into the regulatory problem asset categories. The volume of assets in such categories has important implications for loan loss reserve requirements and for examiners' appraisal of the general quality of a bank's assets. Definitions of these categories are specified by regulators (see Treacy and Carey, 1998), although banks and regulators sometimes disagree about the proper classification of individual assets into the regulatory grades.

Among the 50 largest banks, all but two include in their rating systems grades corresponding to the regulatory problem-asset categories. US bank supervisory agencies do not specifically require that banks maintain regulatory categories on an internal scale but do require that recordkeeping be sufficient to ensure that loans in the regulatory categories can be quickly and clearly identified. The two banks that use procedures not involving internal grades appear to do so because the regulatory asset categories are not consistent with the conceptual basis of their own grades. [9]

Watch credits are those that need special monitoring but that do not fall in the regulatory problem-asset grades. Only about half the banks we interviewed administer the Watch list by including a Watch grade on the internal rating scale. Others add a Watch flag to individual grades, such as 3W versus 3, or simply maintain a separate list or identifying field in their computer systems.

2.4. Number of grades on the scale

Although the vast majority of the fifty largest US banking organizations include three or four regulatory problem asset grades on their internal scales, the number of Pass grades varies from two to the low 20s, as shown in Fig. 1. The median is five Pass grades, including a Watch grade if any. Among the 10 largest banks, the median number of Pass grades is six and the minimum is four. Even where the number of Pass grades is identical on two different banks' scales, the risk associated with the same grades (for example, two loans graded 3) is almost always different. The median bank in Udell's (1987) sample had three Pass grades, implying that the average number of grades on internal scales has increased during the past decade.

Although internal rating systems with larger numbers of grades are more costly to operate because of the extra work required to distinguish finer degrees of risk, banks with relatively formal approaches to credit risk management are likely to choose to bear such costs. Finer distinctions of risk are especially valuable to formal profitability, capital allocation, and pricing models, and many banks are beginning to use ratings in such analytical applications, accounting for the trend toward more grades.

The proportion of grades used to distinguish among relatively low risk credits versus the proportion used to distinguish among the riskier Pass credits tends to differ with the business mix of the bank. Among banks we interviewed,

[9] Although the definitions are standardized across banks, we learned that banks vary in their internal use of OAEM. Most loans identified as OAEM pose a higher-than-usual degree of risk, but banks' opinions about the degree of such risk vary. Moreover, some loans may be placed in this category for lack of adequate documentation in the loan file, which may occur even for loans not posing higher-than-usual risk. In such cases, once the administrative problem is resolved, the loan can be upgraded.

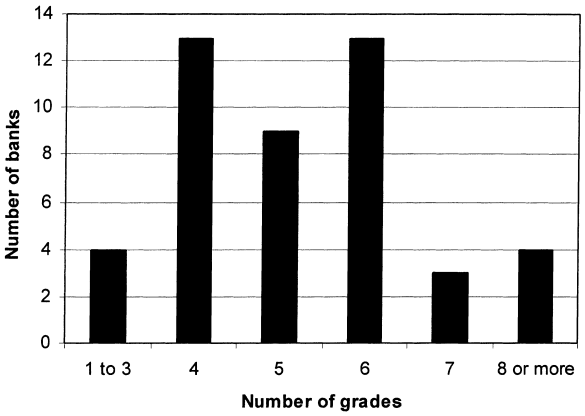

Fig. 1. Large US banks, distributed by number of Pass grades (shown are the 46 banks for which this measure was available).

those that do a significant share of their commercial business in the large corporate loan market tend to have more grades reflecting investment-grade risks. The allocation of grades to investment-grade and below-investment-grade tends to be more even at banks doing mostly middle-market business. [10] The differences are not large: The median middle-market bank has three internal grades corresponding to agency grades of BBB−/Baa3 or better and three riskier grades, whereas the median bank with a substantial large-corporate business has four investment grades and two junk grades. An ability to make fine distinctions among low-risk borrowers is quite important in the highly competitive large-corporate lending market, but such distinctions are less crucial in the middle market, where fewer borrowers are perceived as posing AAA, AA, or even A levels of risk.

A glance at Table 2 reveals that an ability to distinguish risk in the below-investment-grade range is important for all banks. Risk tends to increase nonlinearly on both bank and agency scales. Using bond experience as a guide, default rates are low for the least risky grades but rise rapidly as the grade worsens. The range of default rates spanned by the agency grades BB+/Ba1 through B−/B3 is orders of magnitude larger than the range for A+/A1 through BBB−/Baa3. However, the median large bank we interviewed uses only two or three grades to span the below-investment-grade range, one of

[10] The term "large corporate" includes nonfinancial firms with large annual sales volumes as well as large financial institutions, national governments, and large nonprofit institutions. Certainly the Fortune 500 firms fall into this category. Middle-market borrowers are smaller, but the precise boundary between large and middle-market and between middle-market and small business borrowers varies by bank.

Table 2
Moody's and Standard & Poor's bond rating scales[a]

Category	Moody's			Standard & Poor's		
	Full letter grade	Modified grades	Average default rate (PD) (%, 1970–1995)[b]	Full letter grade	Modified grades	Average default rate (PD) (%, 1981–1994)[b]
Investment grade	Aaa		0.00	AAA		0.00
	Aa	Aa1, Aa2, Aa3	0.03	AA	AA+, AA, AA−	0.00
	A	A1, A2, A3	0.01	A	A+, A, A−	0.07
	Baa	Baa1, Baa2, Baa3	0.13	BBB	BBB+, BBB, BBB−	0.25
Below investment grade, or "Junk"	Ba	Ba1, Ba2, Ba3	1.42	BB	BB+, BB, BB−	1.17
	B	B1,B2,B3	7.62	B	B+,B,B−	5.39
	Caa, Ca, C		n.a.	CCC, CC, C		19.96
Default	n.a.[c]			D		

[a] *Sources*: Moody's Investors Service Special Report, "Corporate Bond Defaults and Default Rates 1938–1995", January 1996. Standard & Poor's Creditweek Special Report, "Corporate Defaults Level Off in 1994," May 1, 1995.
[b] Average default rates are over a one-year horizon. The periods covered by the two studies are somewhat different.
[c] Defaulted issues are typically rated Caa, Ca, or C.

them perhaps being a Watch grade. As with the number of grades on scales, an ability to make finer distinctions among relatively risky assets becomes more important as a bank makes more use of its internal ratings in applications like profitability models.

Systems with many Pass categories are less useful when loans or other exposures tend to be concentrated in one or two grades. Among large banks, 16 institutions, or 36%, assign half or more of their rated loans to a single risk grade, as shown in Fig. 2. Such systems appear to offer relatively modest gains in terms of understanding and tracking risk posture relative to systems in which all exposure is in a single Pass grade.

The majority of the banks that we interviewed (and, based on discussions with supervisory staff, other banks as well) expressed at least some desire to increase the number of grades on their scales and to reduce the extent to which credits are concentrated in one or two grades. Two kinds of plans were voiced (but few were

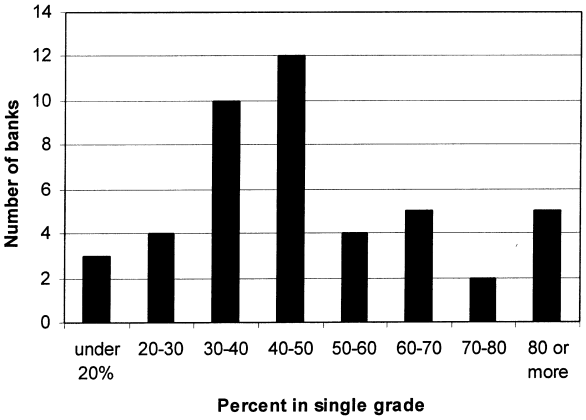

Fig. 2. Large US banks, distributed by percentage of outstandings placed in the grade with the most outstandings (shown are the 45 banks for which this measure was relevant).

actively pursuing such plans): Addition of a ± modifier to existing grades, and a split of existing riskier grades into a larger number, leaving the low-risk grades unchanged. The ± modifier approach is favored by many because grade definitions are subdivided rather than completely reorganized. For example, the basic meaning of a 5 stays the same, but it becomes possible to distinguish between a strong and a weak 5 with grades of 5+ and 5−. This limits the extent of disruption of staff understanding of the meaning of each grade (as noted below, such understanding is largely cultural rather than being formally written).

3. Operating design

At essentially all large banks, the human judgment exercised by experienced bank staff is central to the assignment of a rating. Banks design the operational flow of the rating process in ways that are aimed at promoting accurate and consistent ratings while not unduly restricting the exercise of judgment. Key aspects of operating design include the organizational division of responsibility for grading (line staff or credit staff), the nature of reviews of ratings to detect errors, the organizational location of ultimate authority over grade assignments, the role of external ratings and statistical models in the rating process, and the formality of the process and specificity of formal rating definitions. Design decisions depend on the relative costs of the alternatives, the nature of the bank's commercial business lines, the bank's uses of ratings, and the role of the rating system in maintaining the bank's credit culture.

Ratings are typically assigned (or reaffirmed) at the time of each underwriting or credit approval action. Analysis supporting a rating is inseparable

from that supporting the underwriting or credit approval decision. Moreover, the rating and underwriting processes are formally intertwined. The rating assignment influences the approval process in that underwriting limits and approval requirements depend on the grade, while approvers of a credit are expected to review and confirm the grade. For example, an individual staff member typically proposes a grade as part of the pre-approval process for a new credit. The proposed grade is then approved or modified at the same time that the transaction itself receives approval. In nearly all cases, approval requires assent by individuals with requisite "signature authority" rather than by a committee. The number and level of signatures needed for approval typically depend on the size and (proposed) risk rating of the transaction: In general, less risky loans require fewer and perhaps lower-level signatures. In addition, signature requirements may vary according to the line of business involved and the type of credit being approved. [11]

After approval, the individual that assigned the initial grade is generally responsible for monitoring the loan and for changing the grade promptly as the condition of the borrower changes. Exposures falling into the regulatory problem asset grades are an exception at some institutions, where monitoring and grading of such loans becomes the responsibility of a separate unit, such as a workout or loan review unit.

3.1. Who assigns and monitors ratings, and why?

Ratings are assigned and monitored either by relationship managers (RMs) or the credit staff. RMs are lending officers (line staff) responsible for the marketing of banking services. Depending on the bank's organization, they may be attached to units defined by the size of the business customer, by the customer's primary industry, or by the type of product they sell (for example, commercial real estate loans). All banks evaluate the performance of RMs – and thus set their compensation – on the basis of the profitability of the relationships in question, although the sophistication of methods of assessing profitability and determining compensation varies. Even where profitability measures are not risk-sensitive, ratings assigned by an RM can affect his or her compensation. [12] Thus, in the absence of sufficient controls, RMs may have incentives to assign ratings in a manner inconsistent with the bank's interests.

[11] If those asked to provide signatures believe that a loan should be assigned a riskier internal rating, additional signatures may be required for loan approval. Thus, disagreement over the correct proposed rating can alter the approval requirements for the loan in question.

[12] For example, because loan policies often include size limits that depend on ratings, approval of a large loan proposed by an RM may be much more likely if it is assigned a relatively low risk rating.

The credit staff is generally responsible for approving loans and rating assignments, especially in the case of larger loans; for monitoring portfolio credit quality and sometimes for regular review of individual exposures; and sometimes for directly assigning ratings of individual exposures. The credit staff is genuinely independent of sales and marketing functions when the two have separate reporting structures (that is, "chains of command") and when the performance assessment of the credit staff is linked to the quality of the bank's credit exposure rather than to loan volume or business line or customer profitability. [13]

The primary responsibility for rating assignments varies widely among the banks we interviewed. RMs have the primary responsibility at about 40% of the banks, although in such cases the credit staff may review proposed ratings as part of the loan approval process, especially for larger exposures. [14] At 15% of interviewed banks the credit staff assigns all initial ratings, whereas the credit staff and RMs rate in partnership at another 20% or so. About 30% of interviewed banks divide the responsibility between the credit staff, which has sole responsibility for rating large exposures, and RMs alone or in partnership with the credit staff, which rate middle-market loans. In principle, both groups use the same rating definitions and criteria, but the different nature of loans to large and medium-size borrowers may lead to some divergence of practice.

A bank's business mix appears to be a primary determinant of whether RMs or the credit staff are primarily responsible for ratings. Those banks we interviewed that lend mainly in the middle market usually give RMs primary responsibility for ratings. Such banks emphasized informational efficiency, cost, and accountability as key reasons for their choice of organizational structure. Especially in the case of loans to medium-size and smaller firms, the RM was said to be in the best position to appraise the condition of the borrower on an ongoing basis and thus to ensure that ratings are updated on a timely basis. Requiring that the credit staff be equally well informed adds costs and may introduce lags into the process by which ratings of such smaller credits are updated.

Banks at which an independent credit staff assigns ratings tend to have a substantial presence in the large corporate and institutional loan markets.

[13] Some banks apportion the credit staff to specific line-of-business groups. Such arrangements allow for closer working relationships but in some cases could lead to an implicit linkage of the credit staff's compensation or performance assessment with profitability of business lines; in such cases, incentive conflicts like those experienced by RMs can arise. At other banks, RMs and independent credit staff produce ratings as partners and are held jointly accountable. Whether such partnerships work in restraining incentive conflicts is not clear.

[14] At most banks, RMs have signature authority for relatively small loans, and the credit staff might review the ratings of only a fraction of small loans at origination.

Incremental costs of having the credit staff perform all analysis are smaller relative to the revenues for large loans than for middle-market loans, and independent credit staff typically achieve greater accuracy in their rating assignments, which is especially valuable for large exposures. Their ratings are less likely to be colored by considerations of customer or business line profitability and, because the credit staff is small relative to the number of RMs and is focused entirely on risk assessment, it is better able to achieve consistency (to assign similar grades to similarly risky loans, regardless of their other characteristics). [15]

Almost all the banks we interviewed are at least experimenting with consumer-loan-style credit scoring models for small commercial loans. For exposures smaller than some cutoff value, such models are either a tool in the rating process or are the sole basis for the rating. In the latter case, performing loans are usually assigned to a single grade on the internal rating scale rather than making grade assignments sensitive to the score value.

3.2. How do they arrive at ratings?

Both assigners and reviewers of ratings follow the same basic thought process. The rater considers both the risk posed by the borrower and aspects of the facility's structure. In appraising the borrower, the rater gathers information about its quantitative and qualitative characteristics, compares them with the standards for each grade, and then weights them in choosing a borrower grade. The comparative process often is as much one of looking across borrowers as one of looking across characteristics of different grades: that is, the rater may look for already-rated loans with characteristics very similar to the loan being rated and then set the rating to that already assigned to such loans.

Raters nominally base their decisions on criteria specified in written definitions of each internal grade, but usually the definitions are very brief and broadly worded and give virtually no guidance regarding the weight to place on different factors. Moreover, although most banks require some sort of written justification of a grade as part of the loan approval documents, such writeups have no formally specified structure. According to interviewees, such brevity and informality arises partly because some risk factors are qualitative but also because the specifics of quantitative factors and the weights on factors can differ a great deal across borrowers and exposures. Some noted that the number of permutations is so great that any attempt to produce complete

[15] Middle-market lending probably represents a much larger share of the business of banks we did not interview, and thus the proportion of the all large banks using RM-centered rating processes is probably higher than among our interviewees.

written definitions would be counterproductive. Instead, raters learn to exercise judgment in selecting and weighting factors through training, mentoring, and especially by experience. The specifics of rating assignment procedures at such banks are common, unwritten knowledge embedded in the bank's credit culture. In contrast, a few banks believe that greater formalism is both possible and warranted. Such banks' rating definitions are brief, but their rating process involves forms or grids on which the rater identifies relevant risk factors and their weights. Such forms serve to structure the analysis, remind the rater to consider a broad set of risk factors and to rate them appropriately, and provide those approving the transaction with clear and concise information about the basis for the rating assignment.

The rating criteria that define each grade are articulated as standards for a number of specific risk factors. For example, a criterion for assignment of a grade "3" might be that the borrower's leverage ratio must be smaller than some value. The risk factors are generally the same as those considered in deciding whether to extend a loan and are similar to the factors considered by the rating agencies. Financial statement analysis to determine the borrower's debt service capacity is central, but the details of such analysis vary with the borrower's other characteristics. For example, cash flow, interest coverage, leverage and other characteristics are typically compared to norms for the borrower's industry. Industry also influences ratings in that market leaders are often considered less risky because they are thought less vulnerable to competitive pressure, and firms in declining industries are considered more risky other things equal. Even if industry and financial factors are favorable, medium-size and smaller firms often are assigned relatively risky grades because they have limited access to external finance and frequently have few assets that can be sold in an emergency without disrupting operations. Similarly, at many banks the borrower's grade may be no less risky than the grade assigned to the borrower's country of domicile or operations (such country grades are typically assigned by a special unit in the bank, and may be influenced by country risk grades assigned by regulators). Other risk factors include the reliability of the borrower's financial statements and the quality of its management; elements of transaction structure (for example, collateral or guarantees); and miscellaneous other factors such as exposure to litigation or environmental liability. See Treacy and Carey (1998) for a more detailed description of the complexities of internal rating criteria.

Although in principle the analysis of risk factors may be done by a mechanical model, in practice banks appear hesitant to make models the centerpiece of their rating systems for four reasons: (1) some important risk factors are subjective, such as quality of borrower management; (2) the complex interaction of risk factors implies that different models would be required for each asset class and perhaps for borrowers in different industries or geographic regions; (3) data to support estimation of such models are currently very difficult to obtain; (4) the reliability of such models would become apparent

only over time, exposing the bank to possibly substantial risks in the interim. Those few banks moving toward heavy reliance on models appear to feel that models produce more consistent ratings and that, in the long run, operating costs will be reduced in that less labor will be required to produce ratings.

As part of their judgmental evaluation, most of the banks we interviewed either use statistical models of borrower default probability as one consideration (about three-fourths do so) or take into consideration any available agency rating of the borrower (at least half, and probably more, do so). Such use of external points of comparison is common for large corporate borrowers because they are most likely to be externally rated and because statistical default probability models are more readily available for such borrowers. As described further below, many banks also use external ratings or models in quantifying the loss characteristics of their grades and in identifying likely mistakes in grade assignments.

3.3. Rating reviews and reviewers

Reviews of ratings are threefold: monitoring by those who assign the initial rating of a transaction, regularly scheduled reviews of ratings for groups of exposures, and occasional reviews of a business unit's rating assignments by a loan review unit. Monitoring may not be continuous, but is intended to keep the rater well enough informed to recommend changes to the internal risk grade in a timely fashion as needed. All the banks we interviewed emphasized that failure to recommend changes to risk grades when warranted is viewed as a significant performance failure by the rater and can be grounds for internally imposed penalties. Updates to the risk grade usually require approvals similar to those required to initiate or renew a transaction.

The form of regularly scheduled quarterly or annual reviews ranges from a periodic signoff by the relationship manager working alone to a committee review involving both line and credit staff. Banks with substantial large-corporate portfolios tend to review all exposures in a given industry at the same time, with reviews either by the credit specialist for that industry or by a committee. Such industry reviews were said to be especially helpful in revealing inconsistent ratings of similar credits.

Ratings are also checked by banks' independent loan review units, which usually have the final authority to set grades. Such departments conduct periodic examinations of each business unit's underwriting practices and adherence to administrative and credit policies on a one- to three-year cycle (see Udell (1987,1989)). Not unlike bank examiners, the loan review staff inspects a sample of loans in each line of business. Although the sampling procedures used by different institutions vary somewhat, most institutions weight samples toward loans perceived to be riskier (such as those in high-risk loan grades), with a primary focus on regulatory problem asset categories. In general,

however, an attempt is made to review some loans made by each lender in the unit being inspected.

At a few banks, the loan review unit inspects Pass loan rating assignments only to confirm that such loans need not be placed in the Watch or regulatory grades. Thus, as a practical matter, the loan review unit at these banks has little role in maintaining the accuracy of assignments within the Pass grades. Such institutions tend to make relatively little use of Pass grade information in managing the bank.

In part because operational rating definitions and procedures are embedded in bank culture rather than written down in detail, the loan review unit at most institutions is critical to maintaining the discipline and consistency of the overall rating process. As the principal entity looking at ratings across business lines and asset types, loan review often bears much of the burden of detecting discrepancies in the operational meaning of ratings across lines. Moreover, the loan review unit at most institutions has the final say about ratings and thus can exert a major influence on the culturally understood definition of grades. Typically, when the loan review staff finds grading errors, it not only makes corrections but works with the relevant staff to find the reasons for the errors. Misunderstandings are thus corrected as they become evident. Similarly, when a relationship manager and the credit staff are unable to agree on a rating for a new loan, they turn to the loan review unit for help in resolving the dispute. Thus, the loan review staff guides the interpretations of rating definitions and standards and, in novel situations, establishes and refines the definitions.

Loan review units generally do not require that all ratings produced by the line or credit staff be identical to the ratings they judge to be correct. At almost all banks we interviewed, only two-grade discrepancies for individual loans warrant discussion. With a typical large bank having four to six Pass categories, such a policy permits large discrepancies for individual exposures, potentially spanning ranges of risk corresponding to two or more whole letter grades on the Standard & Poor's or Moody's scales. However, most banks we interviewed indicated that a pattern of one-grade disagreements within a given business unit – for example, a regional office of a given line of business – does result in discipline of the unit and changes in its behavior.

Interviewees indicated that differences of opinion tend to become more common when the number of ratings on the scale is greater, creating more situations in which "reasonable people can disagree". More direct linkage between the risk grade assigned and the incentive compensation of relationship managers also tends to produce more disagreements. In both cases, resolution of disagreements may consume more resources.

All interviewees emphasized that the number of cases in which the loan review staff changes ratings is usually relatively small, ranging from essentially none to roughly 10% of the loans reviewed, except in the wake of large cultural disruptions such as mergers or major changes in the rating system. This fact, as

well as competitive pressures to reduce expenses, has led to suggestions at a few banks that loan review activity be curtailed. Although reviews can be curtailed or eliminated in the short run without apparent damage to rating system integrity, inadequate review activity may result in biased and inconsistent rating assignments over the longer term. Naturally, a low percentage of discrepancies does not imply that the loan review function is unimportant but rather that, in well-functioning systems, the cultural meaning of ratings tends to remain stable and widely understood. One element of a well-functioning system is the rater's expectations that the loan review staff will be conducting inspections. [16]

Because of its central role in maintaining the integrity of the rating system, the loan review unit must have substantial independence and staff members who are well versed in the bank's credit culture and the meaning of ratings. All loan review units at banks we interviewed report to the chief auditor or chief credit officer of the bank, and many periodically brief the board (or a committee thereof) on the results of their reviews.

Loan review units may be less critical to the integrity of rating systems at banks that are primarily in the business of making large corporate loans and at which all exposures are rated by a relatively small, very independent credit staff. Although few banks currently fit this description, they provide an interesting contrast. Such banks' credit units tend to conduct the annual industry-focused reviews mentioned previously and thus are likely to detect rating discrepancies. Having such reviews conducted by broadly based committees rather than only by industry specialists tends to restrain any drift in the meaning of ratings as applied to different industries. In such circumstances, the small credit staff is in a good position to function as the "keeper of the flame" with regard to the credit culture because it essentially carries out the key rating oversight functions of traditional loan review units.

3.4. Rating systems and credit culture

"Credit culture" refers to an implicit understanding among bank personnel that certain standards of underwriting and loan management must be maintained. Such maintenance can be difficult, especially at very large banks serving many customers over a wide area. Of necessity, substantial authority must be delegated to mid-level and junior personnel, and a relaxation of standards may not appear in the form of loan losses for some time.

[16] Another possible expense-reduction strategy is to rely more heavily on statistical models in assigning ratings, reducing the degree of judgment, and thus the amount of labor required to produce each rating. The long-run success of such a strategy depends on the adequacy of the models, including their ability to incorporate subjective factors and their robustness over the business cycle. Our impression is that, at present, such adequacy is uncertain.

At some of the banks we interviewed, senior managers indicated that the internal rating system is at least partly designed to promote and maintain the overall credit culture. At such banks, relationship managers are held accountable for credit quality partly by having them rate all credits, including large exposures that might be more efficiently rated by the credit staff. Review processes aim to identify and discipline relationship managers that produce inaccurate ratings. Such a setup provides incentives for the individual most responsible for negotiating with the borrower to assess risk properly and to think hard about credit issues at each stage of a credit relationship rather than relying entirely on the credit staff. An emphasis on culture as a motivation for rating system design choices was most common among institutions that had suffered serious problems with asset quality in the past 10 or 15 years.

Tensions can arise when rating systems both maintain culture and support sophisticated modeling and analysis. As noted, the latter applications introduce pressures for architectures involving fine distinctions of risk, and the frequency of legitimate disagreements about ratings is likely to be higher when systems have a large number of Pass grades. If not properly handled by senior management and the loan review unit, a rating system redesign that increases the number of grades may make cultural norms fuzzier and the rating system less useful in maintaining the credit culture.

4. Bank systems relative to rating agency systems

Agency and bank rating systems differ substantially, mainly because rating agencies themselves make no investments and thus are neutral parties to transactions between borrowers and lenders. Their revenue comes from the sale of publications and from fees paid by issuers of debt. Such fees can be substantial: S&P's fee for rating a public corporate debt issue ranges from US $25 000 to more than US $ 125 000, with the usual fee being 0.0325% of the face amount of the issue. Fees are a reflection of the substantial resources the agencies typically devote to producing each rating, especially the initial rating.

At banks, the costs of producing ratings must be covered by revenues on credit products. Thus, although a bank might expend resources at a rate similar to that of the rating agencies when underwriting and rating very large loans, the expenditure of so much labor for middle-market loans would make the business unprofitable.

Agency ratings are used by a large number and variety of parties for many different purposes. To ensure wide usage (and thus their ability to collect fees), the agencies attempt to be deliberate, accurate, and evenhanded. They also produce relatively fine distinctions of risk on rating scales having forms and meanings that are stable over time. Accuracy and evenhandedness are crucial to the rating agency business – for example, an agency suspected of producing

the most favorable ratings for those that pay the highest fees would soon be out of business: investors would cease paying attention to its ratings, and issuers would thus have no incentive to pay.

Similarly, changing the rating scale can confuse the public and at least temporarily degrade the value of an agency's product. The agencies also have incentives to be relatively open about their process and to produce written explanations of each rating assignment or change. Clarity helps investors use the ratings and helps assure issuers that the process is as objective as possible.

At banks, ratings are kept private and the costs and benefits of rating systems are internal; hence, pressures for accuracy, consistency, and fine distinctions of risk are mainly a function of the ways in which ratings are used in managing the portfolio. Moreover, the rating system can be tailored to fit the requirements of the bank's primary lines of business and can be restructured whenever the internal benefits of doing so exceed the costs.

Agencies and banks both consider similar risk factors, and both rely heavily on judgment and cultural elements rather than on detailed and mechanical guidance and procedures. However, the agencies publish supplementary descriptions of rating criteria that are much more detailed than banks' internal guidance, partly because agency ratings must be understood by outsiders. In addition, the agencies track the financial characteristics of borrowers receiving their ratings and publish both default histories for each grade and financial profiles of the ''typical'' borrower in each grade, thus providing additional referents to outsiders seeking to understand the meaning of their ratings.

Agencies have nothing comparable to a bank's loan review unit. The rating culture at agencies is maintained instead by a combination of market discipline and a committee system. Market discipline arises because the agencies stand between investors and issuers, with the former typically preferring conservative ratings and the latter preferring optimism. Thus, the agencies quickly hear from investors or issuers about any perceived tendency toward excessive optimism or pessimism. Although a single agency analyst is primarily responsible for proposing a rating, committees make the final determinations. The membership of a committee changes from one rating action to the next so that agency staff members participate in many rating decisions and a cultural understanding of the meaning of each grade is maintained.

5. Tuning rating criteria, quantifying loss characteristics, and the lack of data

In order to use internal ratings in quantitative applications, such as reserving, profitability analysis, or capital allocation, banks must estimate appropriate quantitative loss characteristics for each internal grade. For example, in Table 1, the bank must somehow obtain the probability of default estimates shown in the second column. As described previously, banks assign ratings

using criteria that are thought to be predictive of loss (PD, LIED, or EL), but the process of setting up the criteria is usually judgmental and does not automatically yield quantitative values of PD, LIED, or EL for each grade. Moreover, if internal ratings are to be accurate and consistent, different assets posing a similar level of risk should receive the same grade, and thus rating criteria must be "tuned" both over time and across asset classes to promote accuracy and consistency in terms of PD, LIED, or EL.

The most obvious methods of quantifying and tuning involve use of historical loss experience for the bank's own portfolio. For example, the probability of default for each grade might be estimated as the average of annual default rates for assets in each grade observed over many years. Similarly, if the default rate for commercial real estate loans assigned a given grade were observed to differ systematically from the rate for industrial loans assigned the same grade, the criteria used to rate one or both classes of asset might be adjusted to achieve better consistency.

Unfortunately, to the best of our knowledge, few if any banks have available the necessary data, especially for a variety of asset classes. At a minimum, information on the performance of individual loans and their rating histories is required. Because rating criteria have changed over time at most banks and because tuning requires that criteria be related to loan outcomes, information about borrower and loan characteristics is also required. However, banks have historically retained performance data by loan type (for example, data provided on Call Reports) or by line of business in the aggregate, but not by risk grade. Even at banks that have tracked performance by grade, frequent mergers and acquisitions result in the detailed data covering only one predecessor institution rather than the experience of the whole. Mergers also cause upheaval in both rating processes and data systems and often lead to loss or obsolescence of historical data.

Although data collection is costly, many large banks have recognized its importance and have begun projects to build databases of loan characteristics and loss experience. However, the costs of extracting from archival files historical data on the performance of individual loans appear to be prohibitively high. Thus, those banks that are collecting data indicated that they are several years away from having data sufficient to support empirical analyses on their own portfolios that are comparable to available studies of publicly issued bond experience. [17]

[17] The situation is somewhat better with respect to loss in the event of default (LIED) in that historical studies require information only on the bad assets. Often their number is small enough that gathering data from paper files is feasible, and thus many banks are beginning to accumulate reasonable LIED information from their own portfolio experience. A few publicly available studies have also appeared. Estimating PD and EL requires much more data in that information on both performing and nonperforming assets are required. Studies with LIED statistics include Carty and Lieberman (1996), Asarnow and Edwards (1995), and Society of Actuaries (1998).

5.1. Tuning criteria

The task of tuning rating criteria may be split in two: ensuring that criteria are calibrated so that different assets of the same general type in the same grade have the same loss characteristics, and addressing diversity among asset types. Within a narrowly defined asset class, such as loans to large commercial firms in the same industry, comparisons across firms are relatively manageable, so the main problem is defining the boundaries of rating classes and inferring the default or loss rates for each class. That by itself is not easy, but the problem becomes much more difficult when very different types of assets must be compared. For example, how would a loan to a well-established commercial real estate developer, featuring a 70% loan-to-value ratio, compare with a term loan to a firm in a relatively stable manufacturing industry with a current debt to equity ratio of 1:1 and an interest coverage ratio of 3?

In the absence of data, it is our impression that the traditional means of tuning both rating criteria and underwriting standards relies heavily on the judgment and experience of senior credit staff with long experience at their institution. Over a period encompassing multiple credit cycles, such staff accumulate an individual and collective memory of the credit problems experienced by the institution and of the implications for risk of various borrower and loan characteristics. Such experience is very likely sufficient to support meaningful tuning of rating systems that have small numbers of Pass grades (each covering a broad band of risk) and that are used to rate traditional banking assets. The precision with which systems involving a large number of Pass grades can be tuned by experience alone is not clear.

5.2. Mapping to agency grades as a partial solution

Many banks have estimated the quantitative loss characteristics of their ratings by using the extensive data available on the loss performance of publicly issued bonds. As noted, rating agencies and others frequently publish studies of historical bond default and loss experience by grade covering many years, and publicly available databases of bond issuer characteristics make it possible to relate loss experience to potential rating criteria. Indeed, S&P occasionally publishes tables of indicative or average financial ratio values by grade (while noting that many other factors enter into its rating decisions).

To use data on bond loss experience, a bank must develop or assume some correspondence between agency ratings and its own internal grades. Interviews suggest that the basis of such mappings is threefold: (1) the internal grades assigned to borrowers who have also issued publicly rated bonds; (2) analysis of the "typical" financial characteristics of bank borrowers in each internal grade vis-a-vis the characteristics of the firms with bonds in each agency grade; (3) subjective analysis.

When the mapping is done by comparing the internally assigned grades of publicly rated borrowers with ratings assigned by agencies, there exists a possibility of circularity. In most cases, agency grades are a rating criterion, and even when agency grades do not appear in written rating scale definitions, assigners of ratings always know the agency grade for a given borrower and have an idea of the borrower's likely position on the internal scale. Obviously, if the agency rating is the sole criterion used in assigning internal grades to rated borrowers, publicly rated and unrated borrowers within a given internal grade might differ substantially in risk. In such circumstances the mapping is circular because borrowers are assigned to internal grades based on the agency rating and the agency rating corresponding to each internal grades is inferred only from such rating assignments. Even when circularity is avoided, heavy use of bond experience in defining criteria for each grade might lead to exclusion of criteria needed to capture the risk of unrated borrowers, such as middle-market firms. The banks we interviewed maintain that agency ratings are used only as a starting point in their rating processes, not as the sole criterion.

Another potential pitfall of using bond experience to quantify loss characteristics of internal ratings is that the default and loss experience of loans and bonds may differ. Altman and Suggitt (2000) and Society of Actuaries (1998) present evidence that both default rates and loss in event of default differ significantly across the two asset classes, especially for the riskier grades.

Taking another approach, several large banks use statistical models that estimate the probability of default on the basis of the financial characteristics of the firm or the behavior of the borrower's stock price. Such models provide an "external" estimate of the probability of default. The primary use of such estimates, however, appears to be determining whether the default probability of a given borrower is significantly out of line with that of the agency grade associated with the internal rating.

5.3. Mapping problems caused by inconsistent architectures

Because the major rating agencies rate borrowers with the expectation that the rating will be stable through normal economic and industry cycles, only those borrowers that perform much worse than expected during a cyclical downturn will be downgraded (will "migrate" to riskier grades). In contrast, rating systems that focus on the borrower's current condition (virtually all bank systems) are likely to feature much more migration as cycles progress but, in principle, should exhibit somewhat less cyclical variation in default rates for each individual grade.

Though apparently subtle, the difference in architectures has important implications for mapping exercises and the inference of default probability values for internal grades. Both the point in the economic cycle at which the mapping exercise is done and the exact nature of the PD statistics drawn from the agencies'

studies of long-term default history can have a dramatic effect on the mapping. Values of PD attributed to internal grades can differ by several percentage points depending on how the mapping is done. PDs are most likely to be badly estimated for the higher-risk Pass grades, but reasonable precision is also especially important for such grades in that the aggregate dollar amounts of allocated reserves and capital are most sensitive to assumptions about riskier assets.

As shown in a detailed example in Treacy and Carey (1998), obtaining reasonably accurate mappings appears to be mainly a matter of paying attention to the stage of the cycle at which the mapping is being done and of using historical average PD values from either good-experience or bad-experience years as appropriate. However, interviews left us with the impression that few banks carefully consider cyclical issues when mapping their internal grades to agency grades.

6. An aggregate bank risk profile

As part of the analysis leading to this article, we reviewed internal reports showing distributions of rated assets across internal grades for the 50 largest consolidated domestic bank holding companies. In addition, we obtain mappings of internal grades to agency equivalents from 26 of them. The mappings allow us to allocate internally rated balances to grades on a rating agency scale. To our knowledge, this is the first time that such a characterization of the overall risk profile of a large portion of the banking industry's commercial loan portfolio has been possible.

The 26 banks accounted for more than 75% of aggregate banking industry assets at year-end 1997. Rated loans outstanding at such banks usually represent 50% to 60% of total loans (total loans include consumer loans, which are rarely rated).

In general, we cannot judge whether the mappings provided by banks are correct. Inaccuracy can arise from errors or inconsistency in assigning the internal ratings themselves, problems of cyclicality or circularity in the mapping process, inconsistencies between large corporate and middle market lines of business, or other difficulties. In addition, mappings at some institutions are more precise in form in that they distinguish among modified agency grades, such as BB and BB+. Still, such mappings are an element of banks' day-to-day operating procedures and analysis, which suggests that the 26 banks have endeavored to make them reasonably accurate given the properties of their ratings systems. We believe that aggregation and comparison of mapped loan balances represents a reasonable-albeit crude and broad-first approximation of the actual risks in banks' portfolios.

Fig. 3 displays the distribution of internally rated outstanding loans at year-end 1997 for the 26 consolidated bank holding companies (the proportions are

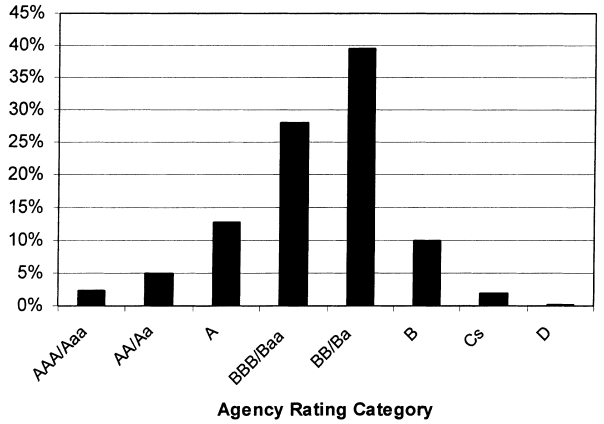

Fig. 3. Percentage of aggregate internally rated outstandings placed in each agency rating category at banks mapped to agency scale, year-end 1997. (*Note*. Twenty-six of the 50 largest banks are included.)

weighted averages, being the sum of dollar outstandings in each grade at all 26 banks divided by the sum of all rated outstandings). About half of aggregate rated loans pose below-investment-grade risks (were rated the equivalent of BB+/Ba1 or riskier), and about 65% of outstandings were concentrated around the boundary between investment and below-investment grades (rated BBB or BB).

Banks' loan loss experience during 1997 is consistent with the credit quality distribution shown in Fig. 3. Using the 1997 default frequencies for each grade drawn from S&P's latest annual study, and an assumption that the average LIED for loans is about 30%, an aggregate portfolio with the quality distribution for the 26 banks would be expected to have an annual credit loss rate of roughly 0.20%. Although this is roughly equal to the actual loan loss experience of the banking industry's aggregate commercial loan portfolio during 1997 (0.21%), this simple exercise should not be taken as proof that the distribution in Fig. 3 is representative; nonetheless, the results are supportive. [18]

Fig. 4 displays the percentages of internally rated assets that are below investment grade for three peer groups as of year-end 1997. For purposes of this analysis, the 26 banks with mappings were divided into major loan syndication agents; smaller banks (less than US $25 billion in total assets at year-end 1997); and the rest, labeled "regionals" (many other peer groupings are possible, of

[18] Actual loss experience is measured as the average annualized net charge-off rate for bank loans in the commercial and industrial, commercial mortgage, and agricultural loan categories as reported on the quarterly Report of Condition ("Call Report") filed by all banks.

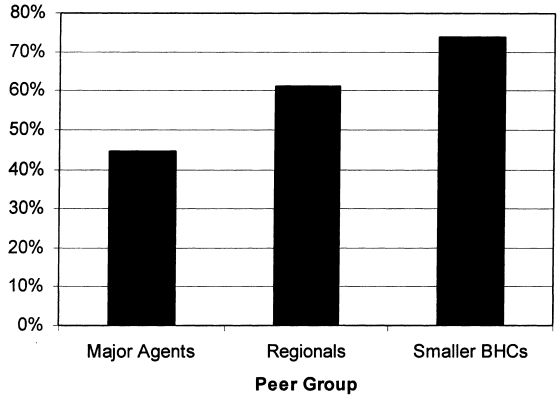

Fig. 4. Percentage of aggregate internally rated outstandings below investment grade at banks mapped to agency scale, by peer group, year-end 1997. (*Note*. Twenty-six of the 50 largest banks are included. Smaller BHCs are those with less than US $25 billion in total assets. Regionals are those that are not major syndication agents or smaller banks.)

course). The three peer groups show systematic differences in risk posture. On average, the major agents have 45 percent of rated assets in categories corresponding to BB and riskier, compared to about 60 percent for regionals and 75 percent for smaller banks. [19]

7. Uses of internal ratings

Banks use internal ratings in two broad categories of activity: analysis and reporting, and administration. Analytic uses include reporting of risk postures to senior management and boards of directors; loan loss reserving; and economic capital allocation, profitability measurement, product pricing, and (indirectly) employee compensation. Administrative uses include guiding loan origination and loan monitoring processes and regulatory compliance. [20] In addition, over time external entities may become more significant users of internal ratings information in the future. Different uses place different stresses on the rating system, and may have different implications for the internal controls that are needed to maintain the system's integrity.

[19] That some institutions have a much larger-than-average fraction of loans posing below-investment-grade risks than others does not imply that such institutions are taking excessive risks, but risk management demands, reserving, and economic capital requirements might differ.

[20] That rating systems and the discipline associated with them are a vehicle for maintaining lending standards and discipline might also be considered an administrative use.

7.1. Portfolio reporting

Virtually all large banks generate internal reports of total asset balances falling in the regulatory problem-asset grades, which are seen by senior management and boards of directors. About 80% also internally report breakdowns of balances in each of their Pass grades. Pass grade breakdowns appear to be used either by management or credit staff as a means of detecting changes in portfolio mix, and are only infrequently shown to boards of directors. [21] Balances in the regulatory grades give a sense of the share of bank assets that are troubled, whereas a profile of balances in Pass grades can provide a forward-looking sense of trends in the bank's risk posture (as long as Pass grade assignments meaningfully distinguish risks). Internal reports are less likely to be informative when a large share of rated assets falls into only one or two Pass grades.

7.2. Reserving

Although many accounting and regulatory policies influence setting of loan loss reserves and provisions, balances in the regulatory grades are integral to reserve analysis at all banks. Supervisors require a specific reserve of at least 50% of doubtful loans plus 20% of substandard loans; banks set the amount of additional reserves for OAEM and Pass loans according to their judgment, subject to evaluation by examiners. Many banks develop reserve factors specific to each Pass category. According to accounting and regulatory standards, loan loss reserves are to cover losses already "embedded in the portfolio," and the generally accepted interpretation is that reserves for Pass loans should cover expected losses over a horizon of one year. Thus, if an institution can identify a reasonable estimate of expected loss for each Pass grade, a reserve analysis sensitive to balances in the different Pass grades provides a good estimate of embedded losses.

A significant number of the banks we interviewed do not differentiate among the Pass grades in performing reserve analysis. In such cases, a single expected-loss (EL) weight is applied to balances in all Pass grades. Such a simplification is least costly in terms of accuracy of the reserve analysis when loan balances are concentrated in a single category or when the composition of the Pass portfolio by risk grade is very stable.

[21] At some banks, portfolio composition is reported as a weighted-average risk grade. Such averages weight the balances by the grade's numeric designator. For example, assets in grade 4 are treated as being twice as risky as assets in grade 2. This can produce misleading averages, because risk – whether PD or EL – tends to increase more than linearly with grade, as shown in Table 2. At those banks we interviewed that used this measure, there appeared to be a clear understanding that this measure does not reflect portfolio risk. It is used as an indicator that the mix has changed.

7.3. Profitability analysis, pricing guidelines, and compensation

All banks we interviewed conduct internal profitability analysis, but in some cases the analysis does not incorporate internal ratings. Analysis at other banks involves expected loss costs and perhaps costs of allocated capital that vary by internal rating. The higher such costs, the lower the measured profitability of a business unit or individual transaction. Rating-sensitive profitability analysis thus has significant implications for the design and operation of internal rating systems.

To implement such analysis, the bank must estimate expected losses for assets in each grade as well as the amount of economic capital to allocate (if it allocates capital). For purposes of discussion, we define economic capital for the bank as a whole as that needed to maintain the bank's solvency in the face of unexpectedly large losses. The process of estimating the additional economic capital needed as a result of making any given loan is complex but, as a practical matter, the loan's internal rating is a primary (if not the sole) operational determinant of the capital allocations imposed by current risk-sensitive profitability models.

The measured profitability of business units enters importantly into management decisions about which units should grow or shrink. When risk-sensitive profitability is appraised at the individual loan or relationship level, unprofitable loans are either not made or unprofitable relationships are eventually dropped. At a growing number of banks, employee compensation is formally tied to measured profitability.

Upon introduction of risk-sensitive profitability analysis, pressure to rate loans favorably arises because expected losses and capital allocations are lower for lower-risk loans. Some banks found that many loans were upgraded shortly after the introduction of profitability analysis, although the overall degree of the shift was modest. Many noted that the number of disagreements in which relationship managers pressed for more favorable ratings increased once such systems were put into place.

In addition to pressure for more favorable ratings, rating-sensitive profitability analysis also creates pressure to increase the number of rating categories. Such pressure, which comes both from business line staff and the profitability analysis unit itself, arises because some of the loans in any given grade are less risky than others and thus should bear smaller credit costs. Creation of more grades allows for better recognition of such differences in risk. Interviewees remarked that the pressure to increase the number of grades has become more pronounced as competitive forces have compressed loan spreads in recent years, noting that a reduction in expected loss factors by a few basis points, or a small decrease in the amount of capital allocated to a loan, may be the difference between a transaction that meets internal profitability "hurdles" and one that does not.

These stresses place increased pressure on the loan review unit to maintain discipline and enforce consistency, stability, and accuracy. As the number of grades on the scale increases and the distinctions of risk become finer, reasonable people will naturally disagree more frequently about ratings, and thus the control of biases becomes more difficult. The difficulty seems likely to be greatest just after the number of grades is increased because the loan review staff must enforce (and if necessary, develop) new cultural definitions for the grades. The latter task is somewhat easier at banks that use external referents in assigning or reviewing ratings, such as default probability models and agency ratings of borrowers, because such referents give loan reviewers objective benchmarks to use in identifying problems and communicating with staff. Rating scale redesigns that split existing grades into smaller compartments are also easier to implement because the existing cultural definitions can be refined rather than replaced.

Risk-sensitive profitability analysis also introduces new demands for internal loss experience data and for mappings to external referents because such analysis demands relatively precise quantification of the risk characteristics of each grade. However, such analysis can also have the effect of making existing data and mappings less useful, at least in the short run, because rating pressures or changes in architecture may, to some extent, change the effective meaning of grades.

7.4. Using ratings to trigger administrative actions

As noted, many banks include an internal Watch grade on their scales in addition to the regulatory problem-asset grades (formally, the Watch grade would be counted among the Pass grades). Reassignment of a loan to Watch or regulatory grades typically triggers a process of quarterly (or even monthly) reporting and formal reviews of the loan. At banks where the main uses of ratings are for monitoring and regulatory reporting, raters' main interest is to avoid getting caught assigning ratings that are *not risky enough*. Such an offense can harm the rater's career, and thus raters have an incentive to assign credits to the riskiest Pass grade that is not Watch. For example, some banks are especially likely to penalize raters when a loan review reassigns a credit from one of the less risky grades into the Watch grade or a regulatory grade. Thus, in the absence of carefully designed controls, the use of grades for administrative purposes can tend to reduce the accuracy of Pass grade assignments. This sort of bias is less likely at the largest banks because the countervailing incentives of rating-sensitive profitability analysis are most likely to operate there.

However, incentives associated with rating-sensitive profitability analysis can reduce the effectiveness of administrative management of problem loans. Staff may delay assigning credits to Watch or regulatory grades because such reassignments may reduce measured profitability. Thus, there is a tension in the

simultaneous use of rating systems for administrative purposes and for profitability analysis. Such tension can be overcome with proper oversight, the implementation of which represents another burden on loan reviewers.

7.5. Potential uses of internal ratings by external entities

Internal ratings are a potential source of information for bank investors and regulators. For example, disclosure of the profile of a bank's loans across its internal rating categories might enhance the ability of shareholders and analysts to assess bank risk. Information about the internal ratings of assets underlying asset-backed securities originated by banks, especially securitizations of traditional commercial loans, might allow investors to better appraise the risk of the securities. Some banks are reportedly already considering using internal ratings in structuring securitizations. For example, when loans in a pool are paid off, the new loans replacing them may be required to be drawn from those with a particular internal grade.

External validation of internal ratings is likely to become a prerequisite for such applications because investors (or rating agencies) must understand the loss characteristics of each internal grade and have confidence that such characteristics will remain stable over time. Currently, such validation appears quite difficult not so much because each bank's rating scale is different, but because loss concepts are ambiguous and the rating criteria are largely embedded in bank culture rather than written down. Moreover, as noted, most banks do not have sufficient historical data on loss experience by internal grade to support objective measurement of loss characteristics.

US bank regulatory agencies are already beginning to make greater use of internal ratings in supervision and regulation. This is part of a continuing emphasis on the importance of strong risk management within banks and a shift of examination focus toward the adequacy of internal risk management and somewhat away from testing of individual transactions. [22] Supervisors are reviewing the adequacy of internal rating systems, monitoring distributions of loans across individual institutions' internal grades, and where possible using mappings of internal grades to the S&P or Moody's scales in order to make comparisons of risk profiles and trends in profiles. Because ratings are forward-looking indicators of credit risk, supervisory use of internal ratings helps provide a concrete basis for discussions between banks and supervisors about credit risk posture.

[22] The shift in focus in part recognizes that detailed examination of the contents of a portfolio at a point in time are less useful in a rapidly changing world, and also makes supervision and regulation less intrusive and less restrictive of innovation.

Internal ratings might become one consideration in scaling regulatory capital requirements more closely to the riskiness of bank portfolios. The current risk-based capital regime (based on the 1988 Basle Accord) provides for lower risk-weights on certain low-risk assets (for example, those that are government-issued or guaranteed), but applies the same 8% capital requirement to essentially all loans to private borrowers regardless of underlying risk. This distortion in the regulatory "pricing" of risk has motivated banks to invent a variety of schemes for regulatory capital arbitrage, that is, for effectively circumventing the capital requirement (Jones, 2000). Although no formal proposals have been released at the time of this writing, senior regulators from a variety of nations have been discussing various means by which regulatory capital requirements might be made more sensitive to risk, and several suggestions rely heavily on banks' internal ratings.

Greater reliance on internal ratings for supervision and regulation would require that supervisors be confident of the rigor and integrity of internal rating systems. Hitherto, examiners have sought to validate rating assignments only as they relate to the regulatory problem-asset grades. Supervisory reliance on Pass grade information implies that some validation and testing of assignments to Pass grades may be necessary.

External use of internal ratings would introduce new stresses on internal rating systems. In some respects, the stresses would be similar to those associated with internal rating-sensitive profitability analysis. Incentives would arise to grade optimistically and to alter the rating system to produce more fine-grained distinctions of risk. However, incentive conflicts would be between outsiders and the bank as a whole rather than internal to the bank. Such new conflicts could overwhelm existing checks and balances currently provided by internal review functions. Even in the absence of such incentive conflicts, external users might demand a greater degree of accuracy or consistency in rating assignments than required internally. For both reasons, external reviews and validation of the rating system might be necessary. In addition, banks and external parties should both be aware that any additional stresses of external uses, if not properly controlled, could impair the effectiveness of internal rating systems as a tool for managing the bank's credit risk. [23]

[23] In the early 1990s, the National Association of Insurance Commissioners (NAIC) introduced a system of risk-based capital requirements for insurance companies in which requirements vary with the ratings of assets. Although such ratings are assigned by the NAIC's Securities Valuation Office (SVO), the SVO does take into account any ratings of an assets published by major rating agencies. In the wake of this and other developments in the insurance industry, the rating agencies experienced substantial pressure from both issuers and investors (insurance companies) to assign favorable ratings to some assets. This was a new and difficult development for the agencies in that issuers and investors had traditionally applied opposing pressures.

8. Some recommendations for bank and regulators

Although this article's primary goal is illumination of the relationships between the forms and functions of large banks' internal rating systems, in performing the research we have come to believe there are (at least) 11 specific internal rating system characteristics that should be promoted but that are not currently very widely implemented. By altering its rating system to have these characteristics, a bank not only will be better able to manage credit risk internally, it will be in a better position to take advantage of opportunities that arise as investors and regulators begin to make more use of internal rating information. The last three recommendations are directed equally at banks and bank regulators. These recommendations are our own opinions and do not represent the policy of the Federal Reserve.

1. Internal ratings should be two-dimensional and should embody clear loss concepts, with one rating scale reflecting obligors' probability of default and the other scale reflecting the LIED of facilities. Such clarity and separation is helpful not only when ratings are used as inputs to analytical applications like internal profitability analysis, but also aids review of rating assignments because PD and LIED are basic building blocks of risk and are clearly displayed rather than being mixed together in a single facility grade measuring expected loss.
2. The number of grades on each scale and their dispersion across the spectrum of risk should achieve usefully fine distinctions of risk that are appropriate to the bank's business mix. Different scales are appropriate for different banks, and the exact number of grades that is optimal will vary. For example, three or four Pass grades may not provide sufficient differentiation where internal ratings are used in pricing or profitability analysis. It seems equally clear that there is a point beyond which more fine-grained distinctions become counterproductive. For example, on a scale with 100 grades, it is likely that neither raters nor reviewers of ratings would be able to reliably distinguish risks just one or two grades apart.
3. Rating criteria and rating decisions should be written in more detail. Although written definitions that cover every eventuality may be too bulky, detail greater than the current norm of a few sentences would promote consistency of ratings within a bank and is likely to be a practical necessity if external validation of internal rating systems by investors, public rating agencies, or regulators becomes the norm. Similarly, written documentation of rating decisions in relatively standardized form is likely to help ensure that rating definitions are followed and will aid both data warehousing and external validation.
4. Independent credit staff should assign ratings to the extent economically feasible. Although it may be necessary that relationship managers assign

ratings to small exposures, using independent credit staff to rate large exposures will limit bias in rating assignments.

5. Simultaneous review of rating assignments for all obligors of a given type (for example, in a given industry or a given country) appears to be especially effective in revealing inconsistencies in grade assignments because the rating criteria are comparable. However, some provision for ensuring that ratings are consistent across obligor and facility types is also necessary.
6. Banks should make serious attempts to quantify the loss characteristics for each internal grade, and such quantification should involve a variety of stakeholders within the bank, not just the "quants" in the risk modeling group. Even if the process of developing estimates of PD, LIED, or EL for each grade does not yield precise estimates, the process is likely to reveal disagreements about rating criteria and thus will promote consistency.
7. Banks should use as many external referents as possible in assigning and reviewing ratings and in quantifying the loss characteristics of grades, but such tools should be used with care. A variety of statistical models of borrower default probability are commercially available, and each has strengths and weaknesses that should be understood as they are applied. Internal ratings should also be mapped to agency grades, but with careful attention to the pitfalls of mapping described previously.
8. Banks should collect and warehouse loss experience data for their own portfolios and should use such data to tune their rating criteria and adjust quantitative estimates of loss characteristics as appropriate. Such data should identify exposures which performed and those which did not and the loss that was experienced for the latter, and should include complete rating histories as well as characteristics of obligors and facilities. Although such characteristics are likely to represent large volumes of data, without characteristics it will be difficult to evaluate and tune rating criteria across asset types and over time.
9. Neither banks nor regulators should forget that most internal rating systems are typically used to guide loan origination and monitoring activities as well as to measure risk. Rating system changes (or changes in uses that affect the incentives of raters) that degrade the quality of the bank's monitoring or its underwriting controls are obviously undesirable, and steps should be taken to offset or avoid any such degradation.
10. Both banks and regulators should be sensitive to the fact that increasing use of internal ratings in supervision and regulation will introduce new stresses that, if left unchecked, might impair the bank's ability to manage credit risk.
11. Any mandate that all banks use a single, standardized internal rating scale is inadvisable. Internal rating system form should follow function, and functions will continue to differ significantly across banks. A standardized system, especially one imposed by regulators, might discourage innovation and hinder some banks' ability to manage risk efficiently.

9. Concluding remarks

It is our impression that at most banks, internal rating systems were first introduced mainly to support loan approval and loan monitoring processes and to support regulatory requirements for identification and tracking of problem assets. A close reading of Udell (1987) implies that as recently as a decade ago, it was common for bank internal scales to have three Pass grades. Most Pass loans probably were in the middle grade, with the lowest Pass grade being an internal Watch grade that triggered extra monitoring and the top grade being for very low-risk loans that required less monitoring and for which underwriting decisions might be streamlined and loan limits increased.

The uses of internal ratings have multiplied over the past 10 years and promise to continue to grow, and thus a bank's decisions about its internal rating system can have an increasingly important effect on its ability to manage credit risk. At the same time, development of appropriate internal rating system architectures and operating designs is becoming an increasingly complex task. The central role of human judgment in the rating process and the variety of possible uses for ratings mean that internal incentives can influence rating decisions. Thus, careful design of controls and internal review procedures is a crucial consideration in aligning form with function.

No single internal rating system is best for all banks. Banks' systems vary widely largely because of differences in business mix and in the uses to which ratings are put. Among variations in business mix, the share of large corporate or institutional loans in a bank's portfolio has the largest implications for its internal rating system. Among the current uses of internal ratings, profitability analysis and product pricing models have the most significant implications for the rating system because they give bank staff with a personal interest in transactions an incentive to rate too favorably. Thus, careful attention to review and control procedures that limit biases in ratings is becoming increasingly important to the accuracy and consistency of internal ratings. As outside investors and regulators make more use of banks' internal ratings, it is likely that new and powerful stresses on the rating systems will be introduced. Incentive conflicts that pit banks' interests against those of external entities will compound existing internal tensions, and it is likely that some sort of external validation of banks' rating systems will become necessary. Such validation, whether by regulators or other entities like public rating agencies, will be difficult to achieve and will lead to pressures for greater clarity and rigor in rating system architecture and operation.

By their nature, banks' credit cultures typically adapt slowly to changes in conditions. The rapid pace of change in risk management practice has been increasing the stresses on credit cultures in general and internal rating systems in particular. Careful attention to the many considerations noted in this article,

including the recommendations made in the preceding section, can help accelerate the process of adjustment and thus the easing of stresses.

References

Altman, E.I., Saunders, A., 1997. Credit risk measurement: Developments over the last 20 years. Journal of Banking and Finance 21, 1721–1742.

Altman, E.I., Suggitt, H.J., 2000. Default rates in the syndicated bank loan market: A mortality analysis. Journal of Banking and Finance 24 (1/2), 229–253, this issue.

Asarnow, E., Edwards, D., 1995. Measuring loss on defaulted bank loans: A 24-year study. The Journal of Commercial Lending 11–23.

Brady, T.F., English, W.B., Nelson, W.R., 1998. Recent changes to the Federal Reserve's survey of terms of business lending. In: Federal Reserve Bulletin. vol. 84, Washington, DC, pp. 604–615.

Cantor, R. Packer, F., 1994. The credit rating industry. In: Federal Reserve Bank of New York Quarterly Review, Summer/Fall, pp. 1–26.

Carty, L.V., Lieberman, D., 1996. Special report: Defaulted bank loan recoveries. Moody's Investors Service, New York.

Jones, D.S., 2000. Emerging problems with the Accord: Regulatory capital arbitrage and related issues. Journal of Banking and Finance 24 (1/2), 35–58, this issue.

Ederington, L.H., Yawitz, J.B., 1987. The bond rating process. In: E. Altman, (Ed.). Handbook of Financial Markets and Institutions, 6th ed. Wiley, New York.

English, W.B., Nelson, W.R., 1998. Bank risk rating of business loans. Working paper, Board of Governors of the Federal Reserve System. Washington, DC.

Machauer, A., Weber, M., 1998. Bank behavior based on internal credit ratings of borrowers. Journal of Banking and Finance 22, 1355–1383.

Mingo, J.J., 2000. Policy implications of the Federal Reserve study of credit risk models at major US banking institutions. Journal of Banking and Finance 24 (1/2), 15–33, this issue.

Moody's Investors Service, 1991. Global Credit Analysis. IFR Publishing, New York.

Robert Morris Associates and First Manhattan Consulting Group, 1997. Winning the credit cycle game: A roadmap for adding shareholding value through credit portfolio management. Philadelphia, PA.

Santomero, A.M., 1997. Commercial bank risk management: An analysis of the process. Journal of Financial Services Research 12, 83–115.

Society of Actuaries, 1998. 1986–1994 Credit Risk Loss Experience Study: Private Placement Bonds. Schaumburg, IL.

Standard and Poor's, 1998. Corporate ratings criteria. New York.

Treacy, W.F., Carey, M., 1998. Credit risk rating at large US banks. In: Federal Reserve Bulletin 84. Washington, DC, pp. 897–921.

Udell, G.F., 1987. Designing the optimal loan review policy: An analysis of loan review in Midwestern banks. Herbert V. Prochnow Educational Foundation, Madison, WI.

Udell, G.F., 1989. Loan quality, commercial loan review, and loan officer contracting. Journal of Banking and Finance 13, 367–382.