

The Relevance of Peer Groups In Early Warning Analysis

During the past decade, many researchers have studied the possibility of developing a statistical relationship that would provide early warning of severe financial deterioration or severe weakness at a bank.¹ These efforts have relied mainly on regularly reported financial data to evaluate a bank's potential strength or weakness.

Some of this research has been conducted within the Federal Reserve System. Its results contributed to the development of an early warning screen implemented during the mid-1970s. This screen uses regularly reported financial data and thus provides continuous quarterly information on member banks to Federal Reserve supervisory personnel. In general, such information has proved to be helpful in focusing supervisory resources where they are most likely to be needed. In more recent years, experience suggests a need to improve the efficiency of screening programs. Three important sets of issues have been raised.

1. The early warning screens used by the Federal Reserve and the comptroller compare banks with their peers to arrive at an evaluation. This raises several issues. For purposes of early warning analysis, how broad and diverse can a peer group be for the banks in the group to be considered essentially in the same type of business, with comparable risks? How should differences in the scope and nature of the banking services involved be taken into account? Are size groupings an adequate method

of dealing with differences in banks' operating characteristics?

2. How can potentially vulnerable banks be identified without also seeming to target a sizable number of institutions that turn out **not** to be vulnerable to future weakness?

3. Regulators need to develop additional financial variables to sharpen the focus of early warning analysis, especially in light of new data being reported by banks.

Most financial analysts have resolved such questions in a practical way, using their knowledge and experience to define peer bank groups, usually based on specific size classifications. The answer is not so clear, however, for determining a bank's potential vulnerability to future financial weakness, since serious financial problems have occurred among banks of all size classes, with no single class more or less prone to severe financial problems.

This article will report on an effort to refine peer groupings, to improve the early warning screen's efficiency and to provide a more realistic appraisal of bank vulnerability in light of each bank's business orientation and overall market exposure. This analysis relies on the same financial ratios that have been at the core of the program for a number of years, leaving for later the task of developing additional or revised financial variables.

Results

The research we will describe appears to have yielded promising results in suggesting how identifying potential problem banks could be made more efficient. The following general findings are worth pointing out:

- (1) The five key financial ratios employed in early warning analysis at the Federal Reserve Bank of New York (see Exhibit 1) continue to give good results, especially when used in

¹Research most directly relevant can be found in the following works. See Leon Korobow, David P. Stuhr and Daniel Martin, "A Nationwide Test of Early Warning Research in Banking," Federal Reserve Bank of New York *Quarterly Review*, Autumn 1977, and references cited therein, for a discussion of the early warning concepts employed in this research and a report of the evidence underlying those concepts. See Daniel Martin, "Early Warning of Bank Failure," *Journal of Banking and Finance*, 1977, pp. 249-276 for a detailed analysis of modern early warning techniques and their applicability to the anticipation of bank failure. See also Leon Korobow "Measuring and Managing Bank Risk: Some Recent History," paper delivered to the International Symposium on Forecasting, June 5, 1983, Philadelphia.

Exhibit 1. Key Surveillance Ratios

1.
$$\frac{\text{Loans \& Leases}}{\text{Total Sources of Funds}}$$
2.
$$\frac{\text{Equity Capital}}{\text{Risk Assets}}$$
3.
$$\frac{\text{Operating Expenses}}{\text{Operating Revenues}}$$
4.
$$\frac{\text{Gross Loan Losses}}{\text{Net Operating Income \& Provision for Loan Losses}}$$
5.
$$\frac{\text{Commercial \& Industrial Loans}}{\text{Total Loans}}$$

conjunction with a revised system of peer groupings.

(2) Participation by banks in both domestic and foreign markets appears to be a good common-risk criterion for developing improved peer groups.

(3) Compared to earlier studies, the initial refinement in peer groups produced a considerably increased concentration of potentially vulnerable banks in the lower rankings.

(4) Further research yielded a new type of "merged" bank ranking, developed from data for a selected peer group within the broad class of banks that operate offices both domestically and abroad. An overall bank ranking was formed by integrating the early warning scores computed separately from the subgroup's financial ratio data with the scores obtained for the remaining banks in the sample.

(5) The merged ranking approach showed promise of improving efficiency. Approximately three of four banks in the weakest percentiles of the ranking developed from 1979 ratio data received low supervisory ratings in 1981. Translated into probabilities, this means the 1979 data implied a future weakness probability of 78 percent, on average, for the banks that subsequently received low supervisory ratings in 1981. That compares favorably with a range of comparable probabilities of 43 to 63 percent found in earlier studies.

(6) The subgroup scoring approach establishes a rationale for selecting peer groups. The

subgroups are designed to produce bank scores that can be integrated into a single bank rating in which the scores have no significant relationship with bank size. This technique eliminates bias for or against particular size classes in the calculation of bank scores.

(7) Experiments with several large bank peer groups—for example, the top 40 banks with at least one foreign office or banks with \$10 billion or more in total assets—produced scores in which size had little or no influence on the overall bank ranking formed from the subgroup scores and those obtained for all other banks having at least one foreign office.

(8) The merged rankings obtained from these subgroups, and the probabilities derived from them, showed several interesting properties:

- (a) the rankings' ability to highlight exceptionally strong or exceptionally weak banks was maintained and strengthened, even as the subgroup was narrowed to the largest multinational organizations.
- (b) Some banks' early warning indicators were highly sensitive to the peer group in which they were placed.

(9) These results arose because the variability of some key ratios (such as equity and loss charge-offs) declined significantly as the subgroups were narrowed, amplifying the sensitivity of the score to deviations from the means of those ratios. This suggests it is important to place banks in peer groups consistent with their respective business orientation and management skills. Moreover, this decision should not rest on size alone. One approach would be to review with special care banks whose early warning scores are highly sensitive to placement in a particular peer group.

(10) While the subgroup scoring approach promises to be a useful tool in a more objective study of peer group classifications, our research suggests that supervisory discretion remains important in placing a bank in an appropriate classification.

Early Warning Analysis: The Fed's Program

The Federal Reserve's present screening program focuses on two main categories of member banks: (1) those having total assets of \$300 million or more and (2) smaller banks. The System's approach is to screen the large bank

Exhibit 2. Current Federal Reserve Surveillance Ratios

- | | |
|---|---|
| 1. Loans & Leases
Total Sources of Funds | 7. Operating Expenses
Operating Revenue |
| 2. Liquid Assets
Total Sources of Funds | 8. Dividends
Net Income |
| 3. Interest-Sensitive Funds
Total Sources of Funds | 9. Gross Loan Losses
Net Operating Income &
Provision for Loan Losses |
| 4. Primary Capital
Risk Assets | 10. Noninterest Expenses
Total Operating
Income - Interest Expense |
| 5. Total Capital
Risk Assets | 11. Commercial & Industrial Loans
Total Loans |
| 6. Net Income
Total Assets - Cash Items | |

**Exhibit 3. Components of the Composite Score:
The Current System Program****Ratios:**

1. Loans & Leases/Total Sources of Funds
3. Interest-Sensitive Funds/Total Sources of Funds
4. Equity Capital/Adjusted Risk Assets
7. Total Operating Expenses/Total
Operating Revenue
9. Gross Loan Losses/Net Operating Income +
Provisions for Loan Losses
10. Noninterest Expenses/Total Operating Income-
Interest Expense
11. Commercial & Industrial Loans/Total
Loans, Gross

group on the basis of nationwide data and the smaller banks by Reserve District.

The \$300 million and over class was selected in part because these banks constitute a large, nationally representative group. For several years this was the only group submitting comprehensive quarterly condition and income reports. A related factor was a reluctance to segment banks into many size classifications that might obscure weakening trends within an entire size class. In addition, research had produced little evidence to suggest that peer groupings by size contributed to the accuracy of early warning analysis. Nonetheless, there remained a strong underlying view that significant differences among classes of banks—such as multinational, regional and community banks—require separate treatment to assess operating and risk characteristics correctly.

The System's screening program is a multistage process involving 11 financial ratios (see Exhibit 2). Seven ratios (Exhibit 3) are combined to form a "composite" bank score. A bank is treated as an exception if it falls below a predetermined cutoff level of the composite score. In addition, separate cutoff levels are established for each of the 11 financial ratios; banks below the individual ratio cutoffs are also treated as exceptions. This multistage process appears to have been successful in screening out banks that subsequently developed serious supervisory problems. However, the initial list of exceptions generated by the cutoff levels usually is far larger than the final list, leaving the challenge of attempting to isolate potential problem banks within as small an initial list of exceptions as possible.

The composite score is an important tool in the screening process. This score is developed from

a combination of key financial variables designed to capture a bank's overall strength or weakness. Selection of the relevant list of financial ratios and the methodology used to combine them as an early warning indicator have drawn on research conducted at the New York Fed. That research has focused on the efficiency and accuracy with which key financial variables could identify potential problem banks years in advance of severe deterioration. The System's early warning program incorporates a somewhat larger group of financial ratios than can be supported by historical research, in part because some data for these ratios became available only in the latter part of the 1970s. The core of the composite scoring approach, however, rests on five key financial ratios identified from research over a number of years. These five ratios (see Exhibit 1) appear to provide early warning results as good or better than many other combinations tested during several years of research.

It will be helpful to review briefly how the composite score is developed and the results it has produced. The first step is to determine peer group baseline averages for each of the five ratios, using the data of an appropriate group of banks for a selected base year. To calculate a bank's score on a ratio, the bank's deviation from the baseline average is divided by the standard deviation of the peer group's average of that ratio. A score above the average is assigned a negative sign for all the ratios except the equity capital ratio. The resulting scores for each of the

bank's ratios are added algebraically to form a composite score that captures the cumulative effects of strength or weakness in all the key characteristics. Banks with high positive scores are considered strong, banks with large negative scores are considered weak.

The use of a baseline average for a selected group of banks rests on the assumption that average behavior within an appropriately representative group constitutes a "norm" against which all banks in the group can be measured. Clearly, this concept must be used carefully, lest a pattern of weakness characterizing an entire group be adopted as an acceptable norm. Moreover, the particular group for which means and standard deviations are calculated has a significant effect on how banks of varying size are treated by the scoring system.

The research supporting this system found that: 1) Poor scores often preceded low supervisory ratings by several years. 2) The incidence of low supervisory ratings was greatest among banks with especially poor scores. 3) Banks that undertook substantially more risk than industry norms tended to have a substantially above-average chance of experiencing future weakness. These results allowed us to develop rules of thumb to isolate banks that appear particularly vulnerable—the weakest 10 percent, 20 percent or 30 percent of the ranking.

This simple approach has been the basis for a more sophisticated early warning model in our further research. A statistical early warning function can be computed with the values of each of the key ratios for each bank in the sample and a measure of the incidence among those banks of low supervisory ratings in two or three subsequent years. The form of the statistical function is shown below. The explanatory variables are the five key ratios. The dependent variable is the probability of a bank's receiving a low supervisory rating in the future.²

$$P = .5 + \frac{1}{\pi} \arctan (a_0 + a_1 LL.TS + a_2 EQ.RA + a_3 EXP.OP + a_4 GCO.NI + a_5 CI.LN),$$

where

P = Probability that a bank will receive a low supervisory rating;

LL.TS = Loans and leases/deposits and related sources of funds;

EQ.RA = Equity capital/risk assets;

EXP.OP = Operating expenses/operating revenues;

GCO.NI = Gross charge-offs/net operating income + provision for loan losses;

CI.LN = Commercial and industrial loans/total loans; and

a_0 = a constant term.

The function provides a measure of the estimated probability of low supervisory ratings computed from financial data for a particular current period, where the coefficients of the function are developed from a prior historical period.

The most recent test of this early warning function was in 1980 when we investigated the potential for weakness of all large member banks over the period 1976-1978. We used a function whose parameters were estimated from financial ratio data of 1972 and supervisory ratings over the period 1973-75.³ We entered financial ratio data for 1975 for all member banks having \$300 million or more in total assets into the latter function to compute each bank's likelihood of a poor supervisory rating. Of 298 member banks in the study group, 76 actually had low supervisory ratings during the period 1976-1978.

A high percentage of the low-rated banks were in the weakest percentiles of the ranking and, therefore, were accorded high probabilities of receiving a low rating. The average probability of the 76 low-rated banks for 1976-1978 was 63 percent. The average probability of the 202 member banks that did not receive low ratings in 1976-1978, was 32 percent. The function probabilities differed significantly from the average probability of 25 percent, assuming a uniform distribution of low ratings among all the banks in the ranking.

If these results appear to be reasonably good, any conclusions must be tempered by the possibility that the results may have been influenced

²In fitting the function to actual data, the historical values of P, the dependent variable, must be established for each bank. This is accomplished through a computer program in which the observed values of the dependent variable are represented by a variable coded as 1 if the bank received a low rating in the estimation period and 0 if it did not. The computer program fits probabilities over the estimation period as closely as possible to these observed values while preserving the functional form indicated.

³To produce forecast probabilities for one or more banks, the values of the five key ratios for a base year two or three years prior to the forecast period would be entered into the function. The research evidence on the significance of the functional relationship between the five key ratings, and on the accuracy of forecasts developed from these historical relationships, is described in detail in the work cited in note 1.

by a somewhat high incidence of low supervisory ratings in 1976-1978. Further, many of the 76 low-rated banks of that period already were in a problem status by 1976.

While the results produced thus far are encouraging, they leave substantial room for improvement from the standpoint of efficient early warning analysis. For example, in the forecast just described, it would take a cut-off point that included about half of the 298 sample banks to ensure that all but 10 of the low-rated banks were captured. At least 60 banks would have to be reviewed to capture 42 low-rated banks. Clearly, it would be desirable to reduce the segment of the ranking likely to contain a high percentage of banks that are seriously vulnerable.

Peer Group Alternatives

One promising way of addressing the model's efficiency is to refine its peer grouping. The present rankings for banks having \$300 million or more in total assets are influenced by bank size as well as bank soundness. Larger banks are somewhat more likely to get lower composite scores purely because they are larger. The observed relationship can be traced to the tendency of large banks to have lower equity capital ratios and higher ratios of commercial loans, expenses and loan losses than regional and local banks.

Does the ratio profile of large banks mean they tend to be riskier and, therefore, more likely to develop serious supervisory problems than smaller institutions? This is possible, but other inferences can be drawn. For example, large banks may tend to be more diversified and better managed than smaller institutions, and may not be as risky as financial ratios alone might suggest.

In any case, the present treatment of very large banks has certain practical pluses and minuses. The tendency toward an adverse score reduces the chances of missing a very large weak bank. At the same time, very large banks might look artificially weaker and regional banks relatively stronger than they otherwise would if scores were calculated from a less diverse group than the present large-bank grouping. If users of the rankings know this, early warning signals about large banks may be discounted. Just as important, the surveillance program may fail to detect emerging weakness at large regional organizations.

There are no simple or clear-cut approaches to developing more refined peer groupings. Market

observers and financial analysts have long employed size groupings to separate banks into classes having common risk and operating characteristics. However, by themselves, size groupings do not indicate that institutions share common operating characteristics.

We began a revision of the large-bank group by selecting, as a possible "common risk" group among the present \$300 million and over class, all member banks operating in the U.S. market that have at least one foreign office. The classification is based on an assumption that banks that operate both domestically and abroad must deal with the broadest type of market risk and perhaps the most critical form of market discipline. On this point, they could be considered significantly more homogeneous with regard to risk than the present large-bank grouping. This new grouping divides the present large-bank peer group of approximately 350 member banks having \$300 million or more in total assets in 1979 into two segments of roughly equal numbers of institutions—banks that operate at least one foreign office, and banks whose offices are all in the United States.

The next step was to determine how this classification performed in isolating potential problem banks. We obtained the names of member banks of \$300 million or more in assets that had weak or marginal supervisory ratings in 1981. Five of these banks had weak ratings and eight banks were on the borderline.

Interestingly, 12 of the 13 weak or marginal banks and four of the five weak banks had at least one foreign office. Thus, the group of banks that operated both domestically and abroad in 1979 contained practically all those accorded low supervisory ratings in 1981. This grouping, therefore, could be expected to produce a higher concentration of low-rated banks in the weakest segments of a 1979 ranking (based on 1979 ratio data) than that produced by a broader ranking. That turned out to be the case. The average probability of future weakness computed for the 12 banks that had both domestic and foreign offices and had actually received low ratings in 1981 was 65.5 percent, compared to 28.6 percent when these 12 were compared to the total group of 352 member banks.

While this first pass at finding a more effective peer grouping provided promising results, an analysis of the relationship of bank scores to asset size continued to show a small but significant negative relationship between composite

Exhibit 4. Comparison of Average Probabilities of Low-Rated Banks on Early Warning Screen Selected Peer Groups

Average Probability In Percent	I. 1979 Ratio Data, 1981 Supervisory Ratings				
	Peer Group				
	(1) \$300 million or more in total assets +	(2) At least one foreign office*	(3) 10% or more in foreign deposits*	(4) Top 40 banks in asset size*	(5) \$10 billion or more in total assets*
1. Low rated banks:	28.6%	65.5%	67.1%	70.1%	73.4%
2. All banks in the screen:***	3.4	7.7	7.7	7.7	7.7
	II. 1981 Ratio Data, 1981 Supervisory Ratings				
1. Low-rated banks:	—	41.8%	39.0%	46.3%	48.2%
2. All banks in the screen:***	—	7.5	7.5	7.5	7.5

+ A total of 352 member banks

* Approximately 160 banks, all at least \$300 million or larger in total assets.

** Subgroups of banks noted above. Merged ranking produced from scores calculated separately from each subgroup and from the respective group of remaining banks.

*** The probability of a low rating if all banks are equally likely to receive a low supervisory rating.

Note: The probabilities in the table have been obtained by means of a cumulative interval calculation. This procedure is similar but not identical to the procedure used in developing probability estimates in the early warning function described in the text.

scores and bank size. To eliminate this source of bias, the 160 banks with at least one foreign office were divided into a further subgroup of banks deemed to be more homogeneous in operating characteristics than the 160-bank group as a whole. We experimented with several subgroups from among the multinational organizations: (1) banks with at least 10 percent of their deposits from abroad, (2) the largest 40 banks in the group, and (3) all banks having \$10 billion or more in total assets.

The results are summarized in Exhibit 4, Section I. The average frequencies of the low-rated banks are noticeably higher in each of the three rankings developed from subgroup scores than the average frequencies obtained from either the 352 bank group ranking (column (1)) or the 160-bank group ranking (column (2)). The increased probability for low-rated banks moving from column (1) to (2) may be interpreted as a

measure of the concentrating effect of the "foreign office" classification. The average probability of low-rated banks rose to 70.1 percent (column (4)) for the ranking in which the top 40 banks were a subgroup and to 73.4 percent (column (5)) for the ranking in which banks with \$10 billion or more in total assets were a subgroup. Further, the negative relationship between bank scores and total assets was barely significant for the 40-bank subgroup, and nonsignificant for the \$10 billion and over subgroup.

Section 2 of Exhibit 4 indicates that the 12 banks that had low ratings in 1981 were identified reasonably well by the revised rankings.

The effect of the subgroup scoring procedure is shown in more detail in Exhibit 5, which indicates the various bank rankings' ability to distinguish weak banks from those on the borderline. As noted earlier, when the 160 banks with at least one foreign office were the sample group,

Exhibit 5. Effect of Peer Grouping on Early Warning Probabilities of Selected Member Banks*

	Average Probabilities Based on 1979 Ratio Data, 1981 Supervisory Ratings*			
	Peer Group			
	(1) 352 Member Banks: Assets \$300 Million Or Greater	(2) At least One Foreign Office	(3) Top 40 Banks**	(4) \$10 billion Or More In Total Assets**
I. Low-Rated Banks				
(a) Four banks, 1981 CAMEL rating of 4 or 5 :	38.8%	67.9%	77.8%	77.8%
(b) Eight banks, 1981 CAMEL rating of 3 :	23.5	65.6	66.2	71.3
II. Selected Multinational Banks				
(a) Selected strong banks	10.8%	17.4%	12.4%	9.2%
(b) Selected weak banks	37.8	80.4	73.9	66.6
(c) Selected large regionals	17.7	27.7	14.6	38.1

* Probabilities are approximated through a cumulative interval calculation. The results may not correspond exactly with probabilities obtained from the early warning function described in the text.
 ** Scores and probabilities obtained from a merged ranking.

the result was a considerable increase in the probability accorded the weakest banks—to nearly 68 percent (see columns (1) and (2)). The use of the subgroup scoring procedure raised that probability to about 78 percent (column (4)). Moreover, the probabilities accorded borderline banks behaved consistently and increased in parallel with those of the weakest banks, but were lower than those of the weakest banks.

The second portion of Exhibit 5 indicates the effect of the various peer groupings and subgroup rankings on a selected number of large multinational banks grouped into three categories: strong, weak or borderline and large

regionals. Narrowing the peer group to banks with at least one foreign office raised all the banks' probabilities of future weakness, as might be expected.

Nonetheless, the probability levels for strong banks remained far below probability levels that past experience has indicated is a danger zone. In contrast, the probabilities of the weaker institutions, which were high to begin with, rose sharply to very high levels. Other interesting results arose from the refined peer group rankings and affected the early warning signals of a number of larger banks in the sample. The probabilities of weakness of the strongest institutions declined.

This increased strength emerged because several key financial ratios (the equity capital ratio and the gross chargeoff ratio) showed significantly less variability in the relatively small peer groups of the largest banks than in the more diverse groupings. Consequently, exceptional strength or weakness in key financial ratios tended to be amplified by the subgroup bank scores. At the same time, the probabilities for some of the weaker banks declined in the \$10 billion and over subgroup ranking simply because they were being measured against others whose operating characteristics were more like their own. Still, these banks' probabilities of future weakness remained high.

The behavior of the respective probabilities accorded several large regional organizations points out some interesting aspects of the decision process that must be involved in peer group analysis. When grouped with the top 40 banks, which includes practically all the nation's major multinational banks, the selected regionals look relatively strong. When the scores were obtained from a \$10 billion and over subgroup, these regional banks were placed in the "all other" bank group. The financial ratios of these regional banks then looked substantially weaker in relation to a group consisting of large regional, regional and community banks than they did with the top 40 banks. When the narrowest subgroup was used, the average probability of weakness for the regional banks rose sharply to 38.1 percent, a fairly high level. The selection of the appropriate peer group clearly is important in this illustration and hinges on a careful assessment of a bank's ability to handle various kinds of risks.

Conclusions

In general, these results suggest that some form of peer grouping along the lines developed in this report might be considered for use within early warning programs. This approach requires some confidence that bank scores should be calculated in such a way that bank size itself should not affect composite scores. In any case, the subgroup scoring approach permits this influence to be controlled and provides a more objective means of defining reasonably homogenous peer classes than perhaps has been available thus far.

Supervisory judgments will continue to be important, since a bank's management philosophy can't be captured easily by financial ratios. A change in management philosophy or objectives may become apparent to supervisors before it affects financial ratios. Under such circumstances, the bank's peer group classification might appropriately be reviewed and the effects of a shift in the bank's class assessed. Used in this manner, peer group classes may hold the potential for enhancing supervisory insight into future problem situations.

—Leon Korobow and
David P. Stuhr

Leon Korobow is an assistant vice president of the Federal Reserve Bank of New York, David P. Stuhr is an economist in the Banking Studies Department of the New York Fed. The authors wish to thank Daniel Martin for his advice in computer programming, Christopher Long for his expert handling of software and Robert Wasky for his capable research assistance. The authors, nonetheless, accept full responsibility for this paper.