The SCOR System of Off-Site Monitoring: Its Objectives, Functioning, and Performance

*by Charles Collier, Sean Forbush, Daniel A. Nuxoll, and John O'Keefe**

The Federal Deposit Insurance Corporation (FDIC) and other bank supervisors have developed a number of tools with which to monitor the health of individual banks as well as the health of the industry as a whole.[1] One tool is on-site examinations: each bank is examined every 12 to 18 months and is assigned a CAMELS rating.[2] These examinations provide the most complete and reliable information about banks' financial health, and supervisors regard CAMELS ratings as the best single indicator of banks' condition. However, between examinations a bank's financial condition may change so that the CAMELS rating is no longer accurate. Therefore, the FDIC and other bank supervisors have developed other tools: off-site systems to monitor insured institutions between examinations.

The FDIC's major off-site monitoring tool is the **S**tatistical **C**AMELS **O**ff-site **R**ating (SCOR) system. The system was designed to help the FDIC identify institutions that have experienced noticeable financial deterioration. This article discusses that objective and the data and method used to meet it. The article then discusses the performance of SCOR in terms of that objective, as well as some auxiliary features that make the system more useful. Two appendices address key technical issues that arose during the development of SCOR.

Objectives of the Project

The SCOR system was developed in the late 1990s to detect banks whose financial condition had substantially deteriorated since their last on-site examination. As its name indicates, the model is an off-site system that is meant to supplement the current system of on-site examinations.

* The authors are all on the staff of the Federal Deposit Insurance Corporation (FDIC): Charles Collier and Sean Forbush are in the Division of Supervision and Consumer Protection (Collier as chief of the Information Management Section, Forbush as a senior financial analyst), and Daniel Nuxoll and John O'Keefe are in the Division of Insurance and Research (Nuxoll as a senior economist, O'Keefe as chief of the Financial Risk Measurement Section). The multiyear development of SCOR involved many of the authors' colleagues, a number of whom have contributed to this article.
The opinions expressed here are the authors' and do not necessarily reflect the views of the FDIC.

[1] Throughout this article, the term *banks* includes all insured financial institutions–commercial banks, savings banks, and thrifts.

[2] CAMELS is an acronym for **C**apital, **A**sset quality, **M**anagement, **E**arnings, **L**iquidity, and market **S**ensitivity. Examiners have rated sensitivity only since 1998. Strictly speaking, examination ratings before that year are CAMEL ratings, but we ignore this distinction and use "CAMELS" throughout, except in Appendix 1.

After an examination, examiners assign the bank a composite CAMELS rating—a rating that reflects the bank's overall financial condition. The ratings range from 1 to 5, with 1 the best and 5 the worst (the meanings of the ratings are summarized in table 1). Banks with a rating of 4 or 5 are considered problem banks. Examiners also rate each of the six CAMELS components, again on a scale of 1 to 5. The meanings of the component ratings parallel those of the composite rating.

Table 1

Definitions of the CAMELS Ratings

Rating	Characteristic
1	"sound in every respect"
2	"fundamentally sound"
3	"exhibit some degree of supervisory concern"
4	"generally exhibit unsafe or unsound practices or condition"
5	"exhibit extremely unsafe or unsound practices or condition"

Source: FDIC Manual of Examination Policies.

Off-site monitoring at the FDIC attempts to identify institutions that received a rating of 1 or 2 on the last examination but might well receive a rating of 3 or worse at the next examination. According to the definitions in table 1, institutions with a rating of 1 or 2 are sound, whereas those with a rating of 3 or worse have some significant problems; once an institution is rated 3 or worse, it has been identified as a concern, and the FDIC monitors it intensively. Consequently, only the likely passage from 1 or 2 to 3 or worse is of interest in off-site monitoring. Identifying 3- or 4-rated institutions that are likely to receive a worse rating at the next examination is not particularly useful from the supervisory perspective.

The difference between a rating of 2 and a rating of 3 has a number of practical implications. Institutions with a rating of 3 or worse are examined more frequently, generally receive closer supervision, pay higher deposit insurance premiums, and may face some legal restrictions on their activities. (Supervisors often take either formal or informal enforcement actions against these banks, and enforcement actions generally restrict an institution's activities or commit it to remedying an identified problem in its operations.)[3]

Consequently, the major objective of the SCOR project was to identify correctly the 1- and 2-rated institutions that were in danger of being downgraded to 3 or worse. The accuracy of the proposed system was analyzed in terms of two types of error, conventionally called Type I and Type II errors. Type I errors consist of false negatives or, more colloquially, "freeing the guilty." In our context, a false negative is failing to detect a downgrade before it occurs, so the level of Type I errors is the percentage of downgraded banks that the model did not identify as problems. Conversely, Type II errors consist of false positives, or "convicting the innocent." The level of Type II errors is the percentage of banks that are identified by the model, yet are found to be sound by a subsequent examination.[4]

There is a trade-off between Type I and Type II errors. Anyone can achieve 0 percent Type I error without a model simply by identifying all banks as likely to be downgraded. By identifying all banks, one has certainly identified all banks that will actually be downgraded. However, one has also identified as problems all of the banks not actually downgraded, so Type II error is 100 percent. Conversely, one can easily attain 0 percent Type II error by identifying no banks; however, this results in 100 percent Type I error. Generally, the more banks identified by a model, the lower the Type I error and the higher the Type II error.

Ideally, the users of a model determine the acceptable trade-off of Type I and Type II errors in terms of the relative costs of the two types of error. At

[3] See Curry et al. (1999) for a discussion of the effectiveness of enforcement actions.

[4] Actually, the relevant Type I and Type II errors are not those discussed in the text. For the FDIC's purposes, the critical question is whether the regional office is aware that a bank might present a supervisory concern, but because that awareness cannot be established retrospectively, all backtesting uses examination ratings. Because case managers have information besides examination ratings, the regional office is often aware of potential downgrades before they occur, but the backtests assume that the regional office is not aware of problems until an examination has begun. Thus, the backtests overstate the model's ability to identify banks that present a concern.

the FDIC, each bank is assigned a case manager at the appropriate regional office. After a bank has been identified by SCOR as likely to be downgraded, the bank's case manager reviews the information available about the bank and determines whether further action is warranted. If the review causes sufficient concern, the FDIC manager can schedule an examination and can allocate resources to supervise the bank more closely. In the context of off-site monitoring, therefore, the cost of Type I error is slow reaction to problems at a bank—that is, a delay in increasing the supervision of the bank. On the other hand, the cost of Type II error is the waste of staff time spent conducting unnecessary reviews. In addition, Type II error undermines the credibility of the system, so case managers have little reason to be conscientious about reviews.

For ease of presentation, this article discusses Type I and Type II accuracy instead of error. Type I accuracy is the percentage of actual downgrades that were identified in advance by the model. Type II accuracy is defined analogously as the percentage of identified banks that are in fact subsequently downgraded.

For the designers of SCOR, accuracy was the major objective, and the benchmark for accuracy was CAEL, the off-site monitoring system developed at the FDIC during the mid-1980s. CAEL was an expert system that used basic ratios from the Call Reports (the quarterly financial reports filed by banks) to rate **C**apital, **A**sset quality, **E**arnings, and **L**iquidity (hence the name **CAEL**);[5] CAEL did not produce management ratings because the quality of management cannot readily be identified with any financial ratio. The ratings of the four components were combined by means of a complicated system of weights to produce a composite rating, which was used to identify institutions for off-site review.

CAEL rated institutions on a scale of 0.5 to 5.5. Conceptually, CAEL ratings are easily mapped to CAMELS ratings: a CAEL rating between 0.5 and 1.5 corresponds to a CAMELS ratings of 1, a CAEL rating of 1.5 to 2.5 corresponds to a CAMELS rating of 2, and so forth.[6]

SCOR was intended to produce ratings comparable to CAEL's while also being easier to analyze. CAEL's use of a complicated system of weights to derive a final composite rating made it difficult for examiners to understand which financial ratios were responsible for the poor ratings an institution received. Thus, although CAEL informed examiners which institutions had problems, it was not always informative about the nature of the problems. Consequently, a secondary objective for the designers of SCOR was to develop a method of analyzing ratings in terms of the underlying ratios.

Development and Functioning of SCOR

In contrast to the expert-system approach of CAEL, SCOR uses a statistical model. It compares examination ratings with the financial ratios of a year earlier. SCOR identifies which financial ratios were most closely related to examination ratings and uses that relationship to forecast future ratings.[7] For example, to predict ratings on the basis of the June 2003 Call Report, SCOR compares data from the Call Report of June 2002 with actual examination ratings from the period July

[5] This expert system was designed by a group of experienced examiners, who decided which ratios were the best precursors of future problems. Updating the system would involve convoking another group of experienced examiners to deliberate about the model. For more information on CAEL, see FDIC (1997), 507ff.

[6] Actually, the mapping is not quite that simple because CAEL was built with a bias toward downgrading institutions. Without any bias, an institution receiving a CAEL rating of 2.5 would be as likely to receive a 2 at the next examination as a 3. The bias, however, means that an institution with a rating of 2.5 will in fact be more likely to receive a 2 than a 3. Because of the bias, CAEL identifies more banks as possible problems, thus increasing the Type II errors while decreasing the Type I errors.

[7] The SCOR model is very similar to the SEER rating model, originally called FIMS, developed by the Federal Reserve System. Both SEER and SCOR draw on a long history of models of bank failure and distress. Demirgüç-Kunt (1989) reviews pre-FIMS developments, and Gilbert, Meyer, and Vaughn (1999) explain the rationale behind such models. For a discussion of the SEER system, see Cole, Cornyn, and Gunther (1995). SEER and SCOR differ in one important respect: SCOR does not use past CAMELS ratings to forecast future ratings. For a discussion of the issue of using past ratings to forecast future ratings, see Appendix 1.

2002 to June 2003. This procedure identifies the Call Report data that were the best indicators of ratings over the past year and uses that relationship to forecast ratings based on June 2003 data. The assumption is that the data that were the best indicators of ratings over the year just past will also be the best indicator over the year to come. The SCOR method, by identifying which ratios are consistently related to examination ratings, attempts to identify which ratios examiners consider the most significant and therefore could be interpreted as an attempt to read examiners' minds.

If the relationship between examination ratings and financial ratios changes, that change will be reflected in the model, generally through a change in coefficients, but only after a delay. For example, if examiners find that intangible factors (such as underwriting) have on average deteriorated and if they therefore assign poorer ratings, then the average SCOR rating will also worsen, even if the deterioration will not yet have affected the basic financial ratios. But because the model is estimated with examination ratings from the past year, the changes in the relationship between ratings and ratios will not be incorporated into the model until the next year.[8]

It is also important to note that SCOR is estimated every quarter and that therefore the ratings for June 2003 (for example) do not depend on any data before June 2002. The estimated relationship between ratings and ratios depends only on very recent data and changes slightly from quarter to quarter. Consequently, even if the Call Report ratios were identical, the ratings for June 2003 could be very different from those for June 1993—in principle. In practice, however, banks similar to those that had poor ratings in 1993 would also have poor ratings in 2003.

SCOR uses a stepwise estimation procedure that eliminates ratios whose relationship with examination ratings is not consistent (that is, ratios that are not statistically significant). In general, the stepwise procedure drops relatively few variables.

SCOR uses only two peer groups: banks and thrifts. Experimentation has indicated that additional peer groups do not improve the model's forecasting power.[9]

The model was developed with a somewhat conservative bias to avoid the problem of excessive data mining. This problem occurs because one can always find a complete coincidence that is statistically significant if one looks at enough data. For example, one might find that banks with a disproportionate number of left-handed tellers had poor CAMELS ratings. Clearly, one would be foolish to use this information to forecast ratings because there is no plausible connection between these two phenomena.[10]

One can avoid this pitfall by choosing variables that actually do cause problems in banks. Choosing such variables necessarily involves using informed judgment. The original specification for SCOR was chosen after both a review of the literature on bank failures and discussions with bank examiners.[11] Discussions with examiners were particularly germane because examiners actually assign the ratings that the model is attempting to forecast. Alternative specifications were tested, and if testing demonstrated that a specification clearly improved the model's ability to detect downgrades of 1- and 2-rated institutions to 3 or worse, changes were made.

[8] With regard to underwriting, at the conclusion of each exam FDIC examiners evaluate separately the quality of the institution's underwriting practices. The FDIC is currently researching whether these ratings can be used to forecast future examination ratings.

[9] In two different experiments, credit card banks and large banks were eliminated from the model. In both cases, the model's forecasting power was worse. Homogeneity is the enemy of statistical models.

[10] The FDIC does not collect data on left-handed tellers, but it does collect a vast amount of data on banks. It would be truly remarkable if some of these data were not correlated with CAMELS ratings. Statisticians are well aware that statistics can demonstrate correlation but not causation.

[11] The variables discussed in Cole, Cornyn, and Gunther (1995) are typical of those used in failure models. See also Hooks (1995) and Demirgüç-Kunt (1989).

The final SCOR model uses 12 variables; all are financial data from the Call Report, expressed as a percentage of assets. Table 2 lists the variables and some ratios for a completely hypothetical bank.[12]

Table 2

SCOR Variables and Ratios for a Hypothetical Bank

SCOR Variable	Percentage of Total Assets
Equity	13.59
Loan-Loss Reserves	1.31
Loans Past Due 30–89 Days	2.23
Loans Past Due 90+ Days	0.89
Nonaccrual Loans	1.51
Other Real Estate	0.45
Charge-offs*	1.18
Provisions for Loan Losses and Transfer Risk*	1.28
Income before Taxes and Extraordinary Charges*	0.10
Volatile Liabilities	25.31
Liquid Assets	28.16
Loans and Long-Term Securities	68.79

*Flow variables. These variables are lagging 12-month totals and have been adjusted for mergers.

Tests of statistical significance show that all the variables are closely related to CAMELS ratings. In addition to the variables in table 2, we experimented with other variables, such as loan growth and average employee salaries. We also experimented with the definitions of some of the variables. For example, we experimented with using Tier-1 capital instead of simple equity, and with using average total assets instead of total assets in the denominators of the ratios. We did not find any other specification that produced consistently better forecasts than the model currently embodied in SCOR.

In table 2, the variables marked with asterisks are items from the income statement (flows), in contrast to the unmarked variables, which are from the balance sheet (stocks). Stocks are measured at a point in time; SCOR uses the end-of-quarter figures from the Call Report. Flows are measured over a period of time; SCOR uses trailing four-quarter totals, instead of the year-to-date numbers found on Call Reports.

Four-quarter totals can be significantly affected by mergers. To eliminate these effects, SCOR uses merger-adjusted data. If banks merge, SCOR does a pro forma merger of the data from pre-merger quarters. Although certainly not ideal, this method eliminates a major distortion due to mergers.[13]

The model forecasts the probability that a bank will receive a specific rating. An example of ratings for a completely hypothetical bank can be found in table 3. According to SCOR, this completely hypothetical bank has approximately a 3 percent chance of receiving a rating of 1, a 55 percent chance of receiving a rating of 2, and so forth.

Table 3

Sample SCOR Output for a Hypothetical Bank

Rating	Probability
1	3.2
2	55.0
3	36.5
4	4.9
5	0.4
Probability of Downgrade	41.8
SCOR Rating	2.44

[12] An earlier specification of SCOR used 13 variables. Dividends were included, and net income was used instead of income before taxes. Pretax income is now used because of the increasing number of banks that are sub-chapter S corporations and do not pay corporate income tax, and dividends were dropped because supervisors commonly restrict dividends at troubled institutions. Thus, dividends are necessarily low at an institution after supervisors have identified it as troubled (it is important to remember, however, that low dividends do not necessarily signal that a sound institution is having trouble). Both changes–dropping dividends, and replacing pretax income with net income–demonstrate the importance of using informed judgment when selecting variables. The current version of SCOR is at least as accurate as the older version.

[13] However, this method might introduce another distortion. Suppose major portions of the disappearing bank (for example, branches or a credit card portfolio) were sold within 12 months of the merger. SCOR's method of adjusting for mergers would include income from operations that were not part of the merged entity. Although examples of this sort of distortion can be found, they are relatively uncommon.

The SCOR model also estimates the probability of receiving a downgrade. If our hypothetical bank is currently rated 2 or better, that probability is defined as its chance of receiving a rating of 3 or worse (36.5% + 4.9% + 0.4% = 41.8%).[14] Associated with these probabilities is a SCOR rating that equals the expected rating [(1 x 3.2%) + (2 x 55.0%) + (3 x 36.5%) + (4 x 4.9%) + (5 x 0.4%)].[15]

The FDIC flags any bank with a downgrade probability of 35 percent or greater. Flagging means a bank must be reviewed by its case manager, and 35 percent was chosen because case managers have only a limited amount of time for reviewing banks. SCOR flags approximately as many banks as CAEL, but during the 1991–1992 recession the SCOR system would have flagged many more banks than CAEL. If SCOR flags so many banks that the review process overwhelms regional analysts—which could happen, for example, during a recession—the flag can be easily changed.[16]

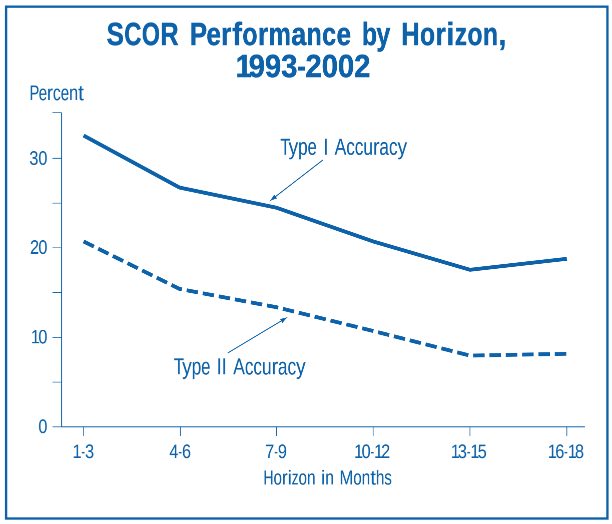
Results

The previous section refers to various experiments that were done while SCOR was being developed. The success of these experiments was evaluated in terms of the objective of the model: whether the modifications produced a model able to correctly identify banks that were subsequently downgraded. This section reports on the results of the final model and demonstrates the type of testing that was repeatedly done during the course of this project, and the type of testing that demonstrated SCOR's superiority to its predecessor, CAEL.

Although the forecasts were evaluated at a variety of time horizons, testing focused on downgrades that occurred four to six months after a given Call Report date. The rationale for this emphasis is that the Call Report data are finalized 60 days after the Call Report date. Consequently, forecasts are not available to bank supervisors until 60 days after the Call Report date.[17]

Figure 1 shows the accuracy of the model at various time horizons. These results include only the first examination after the Call Report was filed. Clearly, accuracy decreases as the forecast horizon lengthens. However, SCOR has some success even at horizons of 16–18 months. Even at this time horizon, SCOR is at least seven times better than a random guess.[18]

Figure 1



[14] If the bank had a rating of 3, the probability of a downgrade would equal 5.3 percent (4.9% + 0.4%). If it had a rating of 4, the probability of a downgrade would equal 0.4 percent. By definition, the downgrade probability for 5-rated banks is zero.

[15] In practice, most banks have a high probability of receiving one or two specific ratings and almost no probability of receiving the other three ratings. The example shown in table 3 is typical. In these cases, we can closely approximate the downgrade probability by dropping the integer part of the SCOR rating. For the hypothetical bank in table 3, the approximate probability of being downgraded to a 3 is 42 percent. This approximation is exact if three of the five probabilities are zero.

[16] When SCOR was adopted in 1999, a 30 percent downgrade probability was used to flag banks for review. By 2001, when the weakening economy undoubtedly affected the financials at banks and caused more poor SCOR ratings and more reviews, that flag was resulting in too many reviews. Accordingly, the higher probability was adopted.

[17] Another reason for focusing on the four- to six-month horizon is that after an examination, the Call Reports filed immediately before the examination almost always (but not absolutely always) are revised. Revisions to the Call Report that immediately precede the exam will bias the backtest because the SCOR model will have access to the corrected data instead of the data that were actually available to supervisors before the examination. This bias is minimized by the use of forecasts based on the Call Report filed four to six months before the examination–a Call Report less likely to be revised.

[18] Banks that have problems are more likely to be examined, and the reported results include only the first examination after the Call Report that provided the data for the SCOR rating. Consequently, the results for a 16- to 18-month horizon include only the strongest banks, and only 2.6 percent of these were downgraded. In contrast, 18.8 percent of the banks identified by SCOR are downgraded 16 to 18 months later.

Figure 2 shows, by Call Report year, the Type I and Type II accuracy achieved under the SCOR system. (The data for figure 2 are found in table 4.) Accuracy is assessed at a four- to six-month horizon, which corresponds closely to the period when the forecasts would be available to supervisors.

Figure 2

Performance of SCOR over Time, Four- to Six-Month Horizon

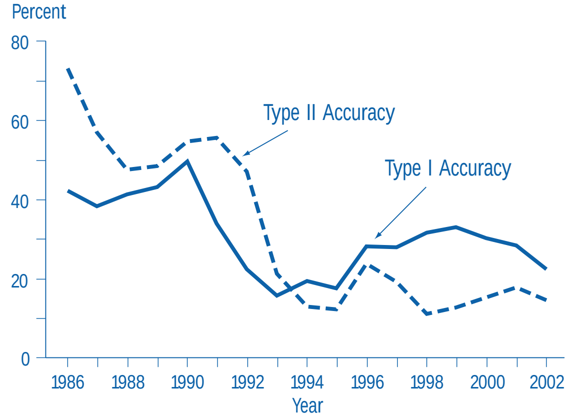


Table 4

Performance of SCOR at Four- to Six-Month Horizon

	Examined	Downgraded	Flagged	Correct
1986	6,465	1,038	1,800	760
1987	6,990	691	1,027	394
1988	6,655	669	767	318
1989	7,236	691	776	335
1990	7,098	894	988	490
1991	7,740	714	1,170	397
1992	9,403	348	727	164
1993	9,911	187	253	40
1994	9,444	191	129	25
1995	8,961	153	108	19
1996	8,279	167	142	40
1997	7,321	156	107	30
1998	6,805	232	82	26
1999	7,020	302	118	39
2000	6,676	274	139	42
2001	6,623	312	196	56
2002	3,878	143	93	21
Total	126,505	7,162	8,622	3,196

Note: This table reports on the results of examinations conducted four to six months after the Call Report that is the basis for a SCOR rating. Downgrades are from ratings of 1 or 2 to 3 or worse.

Clearly the accuracy of the model has declined substantially, and performance has been especially weak since 1993. Since 1993, SCOR has identified approximately 16 percent of the banks that were subsequently downgraded (Type I accuracy), and approximately 27 percent of the banks identified by SCOR were downgraded (Type II accuracy). It must be noted, however, that although the SCOR model is not extremely accurate, it is informative. While Type II accuracy of 27 percent is low, it is approximately nine times better than a random guess. The model does produce valuable information, distinguishing banks that are likely to be downgraded from those that are not.[19] SCOR was adopted to replace CAEL because it had higher levels of Type I and Type II accuracy for almost all time periods.

The low level of accuracy might be expected inasmuch as SCOR relies completely on financial ratios. Any such model will probably be more accurate when the reasons for downgrades are financial, and less accurate when the reasons have to do with some aspect of bank operations that does not affect the bank's financial ratios. For example, examiners may downgrade a bank because they discover that it has significantly weakened its underwriting standards or has weak internal controls—but as long as the more risky loans have not become past due, problems might not have made their way to the financial statements. Consequently, one might reasonably expect that SCOR would be less accurate over the last decade.[20]

The reliance on financial data has several other effects on SCOR's performance. For one thing, it means that SCOR is completely dependent on the accurate reporting of financial information. But in two of the more spectacular bank failures of the last few years—BestBank and the First National Bank of Keystone—the bank's condition had been substantially misstated; consequently, SCOR gave extremely good ratings to both banks.

[19] The reviews done by case managers almost inevitably indicate that banks flagged by SCOR have noticeable weaknesses, even though the weaknesses might not warrant an examination or closer supervision.

[20] The recession of 2001 affected mostly the larger banks and had minimal effects on the rest of the industry.

These problems with SCOR demonstrate that it can never be a substitute for full-scope examinations. Examinations can detect unsafe practices before they affect the bank's financial condition; examinations can also detect misstated financial reports.[21] As we have said before and will say again, SCOR is a complement to bank examinations, not a substitute for them.

Additional Features

A secondary objective of the SCOR project was to produce ratings that were easier to understand and analyze than CAEL ratings. Several features were added to the model to help users of the ratings understand the reasons SCOR identifies a particular institution. First, the SCOR system produces component ratings that help identify specific areas of weakness in a bank. The most controversial of the component ratings has been the management rating because the conventional wisdom is that a model that uses financial ratios cannot identify weaknesses in management. Nonetheless, the SCOR management rating does indicate which banks are at risk of being downgraded.

The second auxiliary tool is a system of weights that indicate which variables are causing poor ratings. The operation of these weights is discussed in this section, while the more technical explanation is relegated to an appendix.

In addition to producing ratings that are more easily analyzed than CAEL ratings, SCOR has also proved useful for tracking trends in the industry. This ability is an extension of the more traditional off-site monitoring.

The Component Ratings

The SCOR model produces a forecasted rating not only for the CAMELS composite but also for each of the six CAMELS components. Case managers and examiners find these ratings useful for identifying the weaknesses in banks.[22]

The component ratings are produced by exactly the same method that is used to produce the composite rating. Most notably, the same variables are used for all the component ratings. But although all the variables in table 2 are relevant to the composite rating, some are more relevant to one or another of the six components. For example, the equity-asset ratio is obviously relevant to the capital component of CAMELS but is less important to the earnings component. SCOR, however, uses all the variables to forecast all the components and, by means of the stepwise procedure mentioned above, selects the variables that are more relevant to explaining the observed component.

The results indicate that examiners do not rate the components in isolation. Consider the capital component. Although the equity-asset ratio is critical for the rating of this component, other variables, too, are used to forecast it. For example, high levels of loans past due 30–89 days are consistently related to poor capital ratings. The reason SCOR uses this variable for this component may not be obvious, but the capital rating is determined by the adequacy of the bank's capital in relationship to its need for capital, and banks with high levels of past-due loans are likely to experience more losses in the future and are therefore likely to need more capital to absorb those losses. Consequently, if two banks have the same equity-asset ratio but one of them has a very high level of past-due loans, that one would receive a worse capital rating.[23]

Although the component ratings are widely used, several financial analysts have raised questions about using SCOR to forecast the management rating. In contrast to the other components, this one is not obviously directly related to any finan-

21 The FDIC's Division of Supervision and Consumer Protection rightly insists that bank examiners are not a substitute for adequate internal and external auditing. However, it was examinations that uncovered the fraud at both BestBank and First National Bank of Keystone. See Berger and Davies (1998) for a discussion of the auditing function of examinations.

22 As mentioned above, CAEL produced ratings for capital, asset quality, earnings, and liquidity, which were then combined into a composite rating.

23 CAEL captured this type of relationship by treating some ratios as primary causes of ratings and others as secondary causes.

cial ratios;[24] internal controls and underwriting standards, for example, cannot be readily reduced to such ratios. In other words, many of the factors behind management ratings are intangible, and a statistical model cannot consider factors that cannot be reduced to accounting.

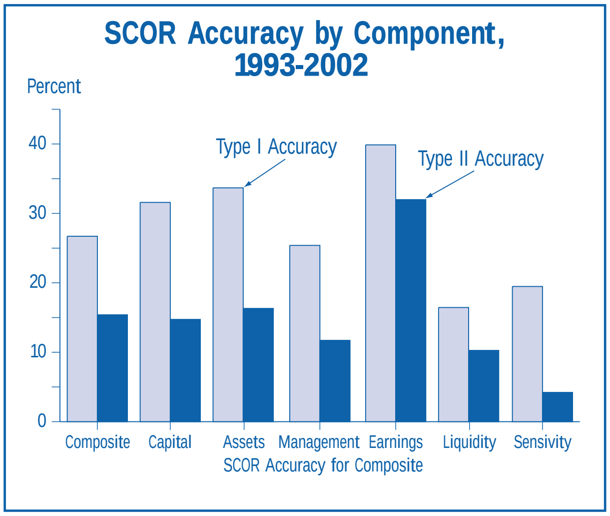
However, all the data in the Call Report can be viewed as indicators of the quality of a bank's management. Obviously factors such as economic conditions affect a bank's financial health, but the quality of management is always a critical factor as well. In the case of loans past due 30–89 days, for example, a high level of such loans implies that the bank has a problem with the quality of its assets and is more likely to have a poor asset rating at the next examination. However, that same level of loans past due 30–89 days might also mean that the bank's management has done a poor job underwriting the loan portfolio and that the bank is more likely to have a poor management rating. Other factors besides underwriting standards affect past-due ratios, so management ratings and past-due ratios do not move in lockstep.

Moreover, the management rating is not alone in involving factors that do not appear on the Call Report. All the other components also involve such factors. For example, the asset rating depends on the level of classified loans, but no data on loan classifications are available until after the examination is actually complete. Thus, asset ratings cannot be assigned only on the basis of information from the Call Report. Similarly, capital ratings depend on the level of classifications as well as on qualitative assessments of the risk because the fundamental question is whether the available capital is adequate for the level of risk.

In short, the management rating is much like the other ratings. SCOR forecasts management ratings by using the same technique it uses for the other ratings: it examines the characteristics of banks to which examiners have recently assigned poor management ratings. SCOR has found that examiners give poor management ratings to banks with low earnings, low reserves for loan losses, and high levels of past-due and problem loans.

Most importantly, SCOR can produce reasonably accurate forecasts of management ratings. Figure 3 shows the accuracy of the component (and composite) forecasts. Although management forecasts are less accurate than some others, SCOR can still use relevant Call Report data to identify institutions likely to have management problems.[25]

Figure 3



Weights of the Call Report Data

Besides producing forecasted composite (and component) ratings, SCOR produces a system of weights that highlights which aspects of a bank's data are responsible for poor ratings.[26] Each ratio is assigned a weight that indicates the contribution that that ratio made to the poor SCOR rating. By indicating which aspects of a bank's operations account for the subpar rating, these weights give case managers and others a starting point for analyzing ratings.

[24] Recall that this is the reason CAEL excluded the management component.

[25] Some other aspects of the forecasted management rating are worth noticing. First, the accuracy of the forecasted management rating deteriorates less over time, so at long horizons it is among the more accurate of the component forecasts. Second, the forecasted management rating does help signal downgrades in the composite rating.

[26] The mathematics behind the "weighting" system can be found in Appendix 2.

In order to define a poor rating, SCOR needs some standard for a good rating. SCOR uses the typical 2-rated bank as the benchmark because, by definition, 2-rated banks are sound institutions with some minor weaknesses.[27] In contrast, 1-rated banks are very strong institutions, and banks with 3, 4, and 5 ratings have weaknesses severe enough that the institutions warrant close supervision. SCOR was designed to identify those banks that are in danger of receiving examination ratings worse than 2, so the 2-rated banks are the obvious standard of comparison.

SCOR considers the "median-2" bank to be the typical 2-rated bank. The median-2 bank is constructed from the median financial data for all the banks that were rated 2 in on-site examinations over the previous year. Thus, the capital-asset ratio is the median ratio for all the banks that received a 2-rating in the previous year. The median-2 bank does not actually exist; it is a statistical construct.[28]

The median-2 bank does not necessarily have a SCOR rating of exactly 2. If the typical 2-rated bank is a strong 2 (more like a 1-rated bank than a 3-rated bank), then the median 2 would probably have a SCOR rating of better than 2. In fact, at present the industry is very healthy. As a result, the median-2 bank has a SCOR rating of approximately 1.6.

Table 5 reports a hypothetical example of the SCOR weighting system. The weights indicate that the problems in the hypothetical bank are due primarily to poor-quality assets and low earnings. Income has a weight of approximately 29 percent, and nonaccrual loans have a weight of 28 percent. This means that the difference in income ratios accounts for approximately 29 percent of the difference between the SCOR rating of the median-2 bank and the rating of the hypothetical bank.

Table 5

SCOR Weights for a Hypothetical Bank

	Median-2 Bank	Hypothetical Bank	
Variable	Ratio	Ratio	Weight
Equity	9.31	13.59	–7.16
Loan-Loss Reserves	0.80	1.31	0.00
Loans Past Due 30–89 Days	0.75	2.23	18.56
Loans Past Due 90+ Days	0.10	0.89	10.27
Nonaccrual Loans	0.23	1.51	28.05
Other Real Estate	0.00	0.45	9.31
Charge-offs	0.13	1.18	4.66
Provisions for Loan Losses and Transfer Risk	0.18	1.28	0.00
Income before Taxes and Extraordinary Charges	1.38	0.10	29.13
Volatile Liabilities	14.13	25.31	5.24
Liquid Assets	32.34	28.16	1.95
Loans and Long-Term Securities	71.29	68.79	0.00
SCOR Rating	1.60	2.44	

Loans past due 30–89 days and loans past due 90+ days also have high weights.

Weights can be negative or zero. Weights are used to explain poor ratings, and those variables that would actually contribute to a better rating receive negative weights. For example, in table 5 the bank actually has more capital than the median-2 bank, so equity has a negative weight. This ratio is better, not worse, than that of the median 2, so it would tend to be a reason for a better, not a worse, rating.

Zero weights occur when there is no consistent relationship between a ratio and the examination ratings. For example, in table 5 loan-loss reserves have a zero weight. This could occur if some banks with high loan-loss reserves were being conservative and providing for any possible losses whereas other banks with high loan-loss reserves had asset-quality problems. In such a case, some banks with high reserves would have good ratings and some would have poor ratings, and SCOR would not find a consistent relationship.[29] The

[27] See table 1. Peer groups could easily be used for this analysis but currently are not.

[28] The median-2 bank is used as the basis of comparison instead of the "mean 2" because outliers tend to increase mean financial ratios. A simple example illustrates the point. If 99 banks have capital-asset ratios of 9 percent and 1 has a capital-asset ratio of 90 percent, the mean capital-asset ratio is 9.81 percent. The median is 9 percent, which is more representative.

[29] There are other possible reasons that a stepwise procedure might eliminate a variable. For example, if two of the explanatory variables were highly correlated, the stepwise procedure would choose the one most closely related to CAMELS ratings and would ignore the other.

In practice, the coefficients used by SCOR are very stable from one period to the next, and the stepwise procedure adds or drops only marginally important variables. In historical tests, SCOR uses almost all the variables each quarter to forecast either a composite or a component rating.

stepwise procedure assigned that variable a zero coefficient.[30]

The weights in table 5 are typical of the banks that are identified as potential concerns. In general, these banks have either asset problems (high levels of loans past due, of nonaccrual loans, or of other real estate) or poor earnings. High levels of noncore funding or lack of liquid assets are also occasional contributing factors.

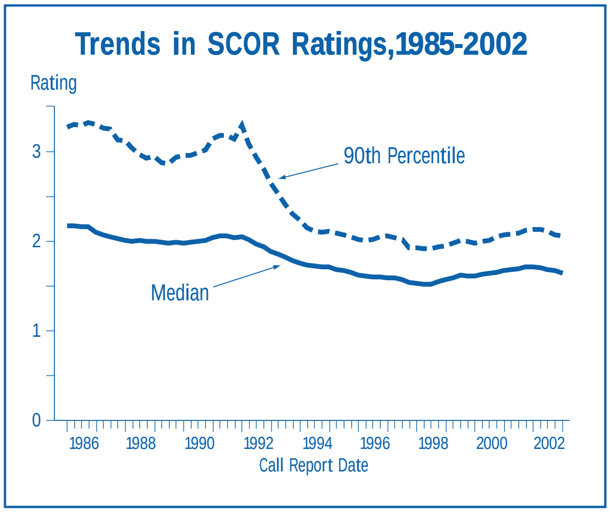
The weights are a starting point for analysis. They do not diagnose the problem, but they do indicate which factors are of special concern and which are not particularly important.

Trends in the Industry

Although SCOR was developed to identify specific institutions, trends in SCOR ratings can also be used to identify changes in the overall health of the banking industry. Figure 4 shows the trends in the median SCOR composite rating and in the 90th percentile. By definition, 50 percent of the banks have ratings better than the median, while 90 percent have ratings better than the 90th percentile and 10 percent have worse ratings. The median can be interpreted as the rating of the typical bank, whereas the 90th percentile indicates trends among the 10 percent of the banks that have the worst ratings. These banks, of course, are the ones of particular concern to supervisors.

The banking problems of the late 1980s and early 1990s are apparent in the data presented by figure 4. The figure also indicates that the banking industry's health peaked in 1998, when the median SCOR rating was 1.52. By the end of 2001, the median rating was 1.71. During 2002, ratings improved.[31]

Figure 4



Concluding Comments

SCOR permits the FDIC to track industry trends and helps identify the institutions that are especially weak. The SCOR output also helps the FDIC identify which financial ratios contribute to poor ratings. However, in periods of economic prosperity, SCOR forecasts are wrong more often than they are right, and since 1993 the model has missed approximately 80 percent of the downgrades, and its forecasts of a downgrade have been incorrect about 75 percent of the time. In contrast, when data are used from the early 1990s—a period when recession was causing financial problems for many banks—SCOR produces more accurate forecasts. Although this single piece of evidence is not conclusive, it does suggest that SCOR will become even more useful if economic troubles again begin affecting the banking industry. SCOR could then help the FDIC focus its limited resources on the institutions that need closer supervision.

[30] Because these weights are calculated with a Taylor first-order approximation, they necessarily sum to 100 percent.

[31] As discussed above, SCOR is reestimated each quarter, so the coefficients change slightly over time. As also discussed above, changes in coefficients would occur if, for example, examiners found that underwriting standards had changed. However, by using the 1998 coefficients to rate banks in 2001 and using 2001 coefficients to rate banks in 1998, one can determine whether the change in SCOR ratings is driven by the underlying financial ratios or by the change in coefficients. This exercise indicates that the change in coefficients accounts for approximately half the trend between 1998 and 2001, while changes in the ratios account for the other half. The change in the model could be interpreted as reflecting examiners' growing concern about aspects of bank operation (for example, underwriting) that are not measured by the ratios.

The model identifies systematic financial strength or weakness but does not consider intangible factors. However, intangibles are too important to ignore because during periods of economic prosperity, poor ratings are more likely to be the result of poor policies and procedures—that is, intangible factors—than of financial weakness. Consequently, the accuracy of SCOR will be lower during periods of prosperity, as it is during the current period. Thus off-site monitoring, with its dependence on financial ratios, cannot replace on-site monitoring. The SCOR model and other systems of off-site monitoring are an aid to examiners but should never be allowed to replace regular examinations.

REFERENCES

Berger, Allen, and Sally M. Davies. 1998. The Information Content of Bank Examinations. *Journal of Financial Services Research* 14:117–44.

Cole, Rebel A., and Jeffery W. Gunther. 1998. Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems. *Journal of Financial Services Research* 13:103–17.

Cole, Rebel A., Barbara G. Cornyn, and Jeffery W. Gunther. 1995. FIMS: A New Monitoring System for Banking Institutions. *Federal Reserve Bulletin* (January): 1–15.

Curry, Timothy J., John P. O'Keefe, Jane Coburn, and Lynne Montgomery. 1999. Financially Distressed Banks: How Effective Are Enforcement Actions in the Supervision Process? *FDIC Banking Review* 12, no. 2:1–18.

Demirgüç-Kunt, Asli. 1989. Deposit-Institution Failures: A Review of the Empirical Literature. Federal Reserve Bank of Cleveland *Economic Review* (Quarter 4): 2–18.

Federal Deposit Insurance Corporation (FDIC). 1997. *History of the Eighties—Lessons for the Future*. Vol. 1, *An Examination of the Banking Crises of the 1980s and Early 1990s*. FDIC.

Gilbert, R. Alton, Andrew P. Meyer, and Mark D. Vaughan. 1999. The Role of Supervisory Screens and Econometric Models in Off-Site Surveillance. Federal Reserve Bank of St. Louis *Review* (November–December): 31–56.

Hooks, Linda M. 1995. Bank Asset Risk—Evidence from Early Warning Models. *Contemporary Economic Policy* 13, no. 4:36–50.

APPENDIX 1

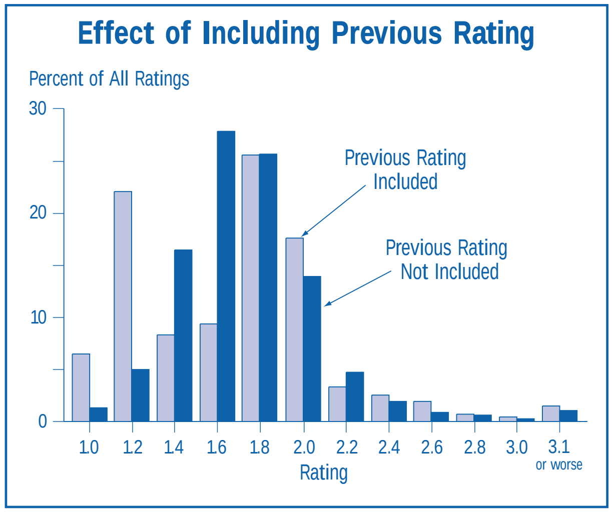
Exclusion of Current CAMELS Ratings

The SCOR model does not use current CAMELS ratings as an explanatory variable, for several reasons. First, the models that use current ratings produce forecasts that tend to cluster around the integers of 1, 2, 3, 4, or 5. For example, ratings near 2 (say, 2.05) are common, but ratings further from 2 (say, 2.40) are rare. This clustering suggests that most banks are really one of five identifiable types. If 2-rated banks are in fact substantially different from other banks, most 2-rated banks will have actual ratings close to 2 (say, 2.05), and only the odd institution that does not really fit one of the established types will have an intermediate rating (say, 2.40).

However, CAMELS ratings are undoubtedly approximate measures of financial strength, and 2-rated banks are not an identifiable type as much as they are a group of banks whose "true" financial strength might be rated somewhere between 1.5 and 2.5. The category of 2-rated banks includes both "strong 2s" (with "true" ratings of 1.6) and "weak 2s" (with "true" ratings of 2.4).

SCOR ratings tend to be more uniformly distributed than ratings produced by models that incorporate prior examination ratings; the distribution of SCOR ratings probably reflects the actual distribution of financial strength among banks. Figure A.1 illustrates the difference in the distributions of the two types of rating systems. The screened bars show the distribution of SCOR ratings based on December 1996 data. The solid bars show the distribution of ratings from an otherwise identical model that includes the CAMEL rating as of December 1996. The forecasts that use the CAMEL ratings are clearly clustered, whereas the distribution of SCOR ratings is smoother.[32]

Figure A-1



The second reason the SCOR model does not use current CAMELS ratings is that examiners wanted a system that used only financial data: they were suspicious of any model that forecasted future ratings in terms of current ratings, especially when the model said that ratings tend not to change. A model that exhibits inertia might miss changes in a bank's condition. There is some evidence that information in CAMELS ratings does become dated, so older CAMELS ratings might well be misleading.[33]

Finally and most importantly, the historical data did produce some evidence confirming examiners' concerns. Models that use CAMELS ratings are marginally worse than SCOR at forecasting the ratings of those banks of most interest to the FDIC—formerly sound banks that are currently experiencing difficulties. Over the past couple of years, the SCOR model has produced better (albeit only slightly better) forecasts of downgrades than models that use prior examination ratings.[34] Consequently, the SCOR model uses only financial data.

[32] The distribution of SCOR ratings also resembles the distribution of the average of the six component ratings. It should be noted that this average is not meaningful because examiners would almost certainly assign higher weights to some components than to others. Moreover, SCOR ratings are distributed much like CAEL ratings, though CAEL ratings tend to be lower. (As explained above, CAEL was intended to be "biased" toward downgrading banks, and SCOR is not biased.)

[33] See Cole and Gunther (1998).

[34] The differences are not statistically significant.

On the other hand, including current ratings would have some advantages, and some prototypes of SCOR did use this approach.[35] First, CAMELS ratings include information not available on the balance sheet. When examiners rate banks, they consider many intangible factors, such as the quality of internal controls, and these intangible factors tend to persist over time. A model that uses only financial data ignores this extra information.

Second, models that include current ratings are more accurate in distinguishing between 1- and 2-rated banks.[36] SCOR cannot differentiate between these banks, apparently because 1-rated banks are financially very similar to 2-rated banks. Conventional wisdom holds that most of the difference between 1-rated and 2-rated banks lies in intangible factors.

Third, models that use current CAMELS ratings tend to produce forecasted ratings that differ only slightly from the current examination ratings, and in fact the best single predictor of future ratings is the current rating. Almost all banks that have a 2 rating before an examination receive a 2 rating after it.

[35] The Federal Reserve Board's contemporaneous SEER model includes management ratings.

[36] As discussed in the main body of the text, however, bank supervisors are relatively unconcerned about distinguishing between 1- and 2-rated banks.

APPENDIX 2

Calculation of the SCOR Weights

The method used to calculate the SCOR weights takes advantage of the linear portion of the logit model. Ignoring the intercept terms, the linear portion is a weighted sum of the bank's financial data, which can be denoted βx which equals $\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{12} x_{12}$.

If the weights are computed for the composite CAMELS rating, this sum can be considered a measure of the bank's general financial strength. If the weights are computed for the capital rating, βx can be considered the measure of the bank's capital adequacy.

The ratings of two banks can be readily compared. Consider two institutions: Bank A (with financial data $x^A = x_1^A, x_2^A, \ldots, x_{12}^A$) and Bank B (with financial data $x^B = x_1^B, x_2^B, \ldots, x_{12}^B$). The difference in the measure of financial strength of the two banks is $\beta x^A - \beta x^B = \beta\,(x^A - x^B)$. The first variable accounts for $\beta_1\,(x_1^A - x_1^B)$ of this difference, or, in percentage terms:

$$100 * \frac{\beta_1 \left(x_1^A - x_1^B\right)}{\beta \left(x^A - x^B\right)}$$

This percentage would indicate the importance of the capital-asset ratio, for example, in explaining the difference in financial strength of the two banks. These percentages (for variables x_1, x_2, and so forth) necessarily sum to 100. The percentages can be negative; a negative percentage could occur if Bank A were stronger, on the whole, than Bank B but had a lower (weaker) capital-asset ratio.

It might be noted that this method is closely related to a Taylor expansion of the logit model. The first derivative of the logistic function equals K β_i where K is a number that depends on the point at which the derivative is evaluated. However, K is the same for all variables. Thus, the first term in a Taylor expansion about the point x^B is $K\,\beta_1\,(x_1^A - x_1^B)$, and the total is $K\,\beta\,(x^A - x^B)$. Of course, the intercept terms will not enter the Taylor expansion because they are constants. If the individual terms are expressed as percentages of the total, then K cancels from both numerator and denominator, and the result is identical to the formula above.