R. Alton Gilbert is vice president and banking advisor at the Federal Reserve Bank of St. Louis. Andrew P. Meyer is an economist at the Federal Reserve Bank of St. Louis. Mark D. Vaughan is senior manager and economist at the Federal Reserve Bank of St. Louis. The authors thank John Block and Michael DeClue for suggesting this topic. They also thank Bob Avery, Kevin Bertsch, Don Conner, Joan Cronin, Tom Fitzgerald, Bill Francis, Bill Gavin, Mike Gordy, Jeff Gunther, Jim Harvey, Jim Houpt, Gene Knopik, Ellen Lamb, Jose Lopez, Kim Nelson, Frank Schmid, and Dave Wheelock along with seminar participants at the Federal Reserve System Surveillance Conference, the annual meeting of the Federal Reserve System Committee on Financial Structure and Regulation, the Federal Reserve Bank of St. Louis, and the Federal Reserve Bank of Kansas City for helpful comments on earlier drafts. All remaining errors and omissions are our own. Boyd Anderson, Thomas King, and Judith Hoffman provided excellent research assistance.

The Role of Supervisory Screens and Econometric Models in Off-Site Surveillance

R. Alton Gilbert, Andrew P. Meyer, and Mark D. Vaughan

Banking is one of the more closely supervised industries in the United States, reflecting the view that bank failures have stronger adverse effects on economic activity than other business failures. Bank failures can disrupt the flow of credit to local communities (Gilbert and Kochin, 1989), interfere with the operation of the payments system (Gilbert and Dwyer, 1989), and reduce the money supply (Friedman and Schwartz, 1963). Bank failures also can have lingering effects on the real economy. Indeed, a growing body of literature blames the length of the Great Depression on the disruption of credit relationships that followed the wave of bank failures during the early 1930s (Bernanke, 1983; Bernanke, 1995; and Bernanke and James, 1991).

The existence of unfairly priced deposit insurance bolsters the case for bank supervision. Without insurance, depositors have strong incentives to monitor and discipline risky institutions by withdrawing funds or demanding higher interest rates. Insured depositors, in contrast, have little incentive to monitor and discipline risk (Flannery, 1982). Moreover, deposit insurance premiums established under the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) do not appear to punish risk adequately. The spread between the premiums paid by the riskiest and safest banks is only 27 basis points, and just 562 of the 10,486 FDIC-insured institutions paid any premiums during the first half of 1999 (Barancik, 1999). As a result, bank supervisors must act as agents of the taxpayers to limit risk. Supervisory limits on bank risk reduce the likelihood that failures will exhaust the deposit insurance fund and impose direct costs on the taxpayers.[1]

Bank supervisors use on-site examination and off-site surveillance to identify banks likely to fail. Supervisors then can take steps to reduce the likelihood that these institutions will fail. The most useful tool for identifying problem institutions is on-site examination, in which examiners travel to a bank and review all aspects of its safety and soundness. On-site examination is, however, both costly and burdensome: costly to supervisors because of its labor-intensive nature and burdensome to bankers because of the intrusion into day-to-day operations. As a result, supervisors also monitor bank condition off-site. Off-site surveillance yields an ongoing picture of bank condition, enabling supervisors to schedule and plan exams efficiently. Off-site surveillance also provides banks with incentives to maintain safety and soundness between on-site visits.

In off-site surveillance, supervisors rely primarily on two analytical tools: supervisory screens and econometric models. Supervisory screens are combinations of financial ratios, derived from bank balance sheets and income statements, that have, in the past, given forewarning of safety-and-soundness problems. Supervisors

[1] See White (1991) for a discussion of the role of lax government supervision in the thrift debacle of the 1980s.

draw on their experience to weigh the information content of these ratios. Econometric models also combine information from bank financial ratios. These models, however, rely on a computer rather than judgement to combine ratios, boiling the information about bank condition in the financial statements down to one number. In some models this number represents the likelihood that a bank will fail. In others, the number represents the supervisory rating that would be awarded if the bank were examined today.

In past statistical comparisons, econometric models have outperformed supervisory screens, yet screens continue to enjoy considerable popularity in the surveillance community. Cole, Cornyn, and Gunther (1995) demonstrated that the Federal Reserve's econometric model, the System for Estimating Examination Ratings (SEER), outperformed a surveillance approach based on screens (the Uniform Bank Surveillance System or UBSS), both as a predictor of failures and as an identifier of troubled institutions. Nonetheless, analysts at the Board of Governors and in each of the Reserve Banks continue to generate a variety of screens to aid in exam scheduling and scoping. To economists who are not involved in day-to-day surveillance, the continuing popularity of screens is somewhat puzzling.

We explore two possible explanations for the popularity of screens: (1) perhaps the extra precision of econometric models is not worth the added cost, or (2) perhaps the flexibility of screens makes them particularly attractive in today's dynamic banking environment. Although models can tease information out of bank financials that the human eye might overlook, they are more costly to operate than screens, requiring surveillance analysts to learn to interpret complex statistical output. If models only marginally outperform screens in flagging banks headed for problems, then the marginal benefit of the extra precision might not exceed the marginal learning costs. Another possible explanation for the attachment to screens is the ease with which they can be adapted to new environments. The last 15 years have witnessed remarkable change in the banking industry. In such a fluid environment, screens can be adapted to reflect changes in the sources of safety-and-soundness problems faster than econometric models.

We demonstrate that econometric models still significantly outperform supervisory screens in statistical horse races, implying that the marginal benefit of using models does indeed outweigh any marginal learning costs. Specifically, we use data from the 1980s and 1990s to compare the performance of supervisory screens and econometric models as tools for predicting failures 12 to 24 months in the future. We highlight the resource savings associated with using each approach rather than random examination. We also estimate an econometric model designed to predict the likelihood that a bank, currently considered safe and sound, will suffer a significant slip in its supervisory rating in 12 to 24 months. Finally, we demonstrate how econometric models can be used to pinpoint the source of developing problems.

Despite the statistical advantages of using econometric models, screens can still add tremendous value in off-site surveillance. In today's fast-changing world of banking, supervisors can modify screens well before econometric models can be re-estimated. Moreover, experience with new screens then can inform the respecification of econometric models. In short, supervisory screens and econometric models play important complementary roles in allocating examination resources.

ON-SITE AND OFF-SITE SURVEILLANCE: A CLOSER LOOK

To appreciate the roles of models and screens in off-site surveillance, it is important to first place these tools in the overall framework of bank supervision. Bank supervisors rely principally on regular on-site examinations to maintain bank safety and soundness. Examinations ensure the integrity of bank financial statements and identify banks that should be subject to supervisory sanctions.[2] During a routine exam, examiners assess six components of safety and soundness—capital protection (C), asset quality (A), management competence (M), earnings strength (E), liquidity risk (L) and market risk (S)—and assign a grade of 1 (best) through 5 (worst) to each component. Examiners then use these six scores to award a composite rating, also expressed on a 1 through 5 scale.[3] At present, most banks boast 1 or 2 CAMELS composites. Indeed, at year-end 1998, only 285 of 8,264 U.S. banks carried 3, 4, or 5 composite ratings.

Although on-site examination is the most effective tool for constraining bank risk, it is both costly to supervisors and burdensome to bankers. As a result, supervisors face continuous pressure to limit exam frequency. During the 1980s, supervisors yielded to this pressure, and many banks escaped yearly examination (Reidhill and O'Keefe, 1997). In 1991, however, the Federal Deposit Insurance Corporation Improvement Act (FDICIA) required annual examinations for all but a handful of small, well-capitalized, highly rated banks, and even these institutions must be examined every 18 months. This new mandate reflected the lessons learned from the wave of failure during the late 1980s, namely that more frequent exams, though likely to increase the up-front costs of supervision, reduce the down-the-road costs of resolving failures by revealing problems at an early stage.

Although recent changes in public policy have mandated greater exam frequency, supervisors still can use off-site surveillance tools to flag banks for accelerated exams and to plan regularly scheduled, as well as accelerated exams. Bank condition can deteriorate rapidly between on-site visits (Cole and Gunther, 1998). In addition, the Federal Reserve now employs a "risk-focused" approach to exams, in which supervisors allocate on-site resources according to the risk exposures of the bank (Board of Governors, 1996). Off-site surveillance helps supervisors allocate on-site resources efficiently by identifying institutions that need immediate attention and by pinpointing risk exposures for regularly scheduled as well as accelerated exams. For these reasons, an interagency body of bank and thrift supervisors—the Federal Financial Institutions Examinations Council (FFIEC)—requires banks to submit quarterly Reports of Condition and Income, often referred to as call reports. Surveillance analysts then use call report data to conduct financial statement analysis between exams.

Using their field experience as a guide, supervisors have developed rules of thumb for exam scheduling and scoping with call report data.[4] These rules of thumb are called supervisory screens. To give an example of the use of screens, supervisors might flag a bank for an accelerated examination (or plan to allocate more resources to a given area on a scheduled exam) if a certain financial ratio, like a risk-based capital

Table 1

How to Interpret CAMELS Composite Ratings

CAMELS Composite Rating	Description
1	Financial institutions with a composite-1 rating are sound in every respect and generally have individual component ratings of 1 or 2.
2	Financial institutions with a composite-2 rating are fundamentally sound. In general, a 2-rated institution will have no individual component ratings weaker than 3.
3	Financial institutions with a composite-3 rating exhibit some degree of supervisory concern in one or more of the component areas.
4	Financial institutions with a composite-4 rating generally exhibit unsafe and unsound practices or conditions. They have serious financial or managerial deficiencies that result in unsatisfactory performance.
5	Financial institutions with a composite-5 rating generally exhibit extremely unsafe and unsound practices or conditions. Institutions in this group pose a significant risk for the deposit insurance fund and their failure is highly probable.

Source: *Federal Reserve Commercial Bank Examination Manual*

[2] See Flannery and Houston (1999) for evidence that holding company inspections help ensure the integrity of financial statements. See Gilbert and Vaughan (1998) for a discussion of the sanctions available to bank supervisors.

[3] See Hall, King, Meyer, and Vaughan (1999) for a discussion of the factors used to assign individual and composite ratings.

[4] See Putnam (1983) for a description of the use of supervisory screens in off-site surveillance during the late 1970s and early 1980s.

ratio, is suspect. Another example might be a rule that flags a bank if 10 out of 15 ratios either exceed or fall short of desired levels. This approach offers two advantages: simplicity and flexibility. An experienced supervisor can detect emerging problems easily, as well as the sources of these problems, without sophisticated statistical analysis. An experienced supervisor also can easily modify the screens in changing banking environments. On the negative side, supervisors who rely only on subjective judgment to "screen" might miss subtle but important interactions among financial ratios.

Econometric models offer a more systematic way to combine call report data for scheduling and scoping. A common type of model used in surveillance estimates the marginal impact of a change in a financial ratio on the probability that a bank will fail, holding all other ratios constant. These models can examine ratios simultaneously, capturing subtle but important interactions. The Federal Reserve uses two different models in off-site surveillance. One model combines financial ratios to estimate the probability that each Fed-supervised bank will fail within the next two years. Another model estimates the CAMELS rating that would be awarded based on the bank's latest financial statements. Every quarter, economists at the Board of Governors feed the latest call report data into these models and forward the results to each of the 12 Reserve Banks. Surveillance analysts in the Reserve Banks then investigate the institutions that the models flag as "exceptions."

SPECIFYING REPRESENTATIVE VERSIONS OF SUPERVISORY SCREENS AND ECONOMETRIC MODELS

To compare the performance of supervisory screens and econometric models, we first specified a representative version of each surveillance tool. To specify a set of supervisory screens, we interviewed safety-and-soundness officers and examiners in the Eighth Federal Reserve District. To specify an econometric model, we reviewed the academic literature. After conducting interviews and reviewing literature, we identified a set of financial ratios common to both approaches. We included only these common ratios in our representative screens and models to facilitate a comparison of relative performance. The financial ratios common to both the screens and the models reflect the individual components of bank condition in the CAMEL framework. (Bank regulators added the "S" to the CAMEL framework on January 1, 1997. During our sample period, however, examiners explicitly graded only five aspects of safety and soundness.) Although our screens and models are representative of the screens and models regularly used in off-site surveillance, they are not identical to the tools currently used by the Board of Governors or the individual Reserve Banks.

In both the screens and models, we used the ratio of total equity to total assets (EQUITY) to assess capital adequacy. Higher levels of capital protection provide a larger buffer against losses and increase the owners' stake in the bank. We expect, therefore, that higher levels of capital will reduce the likelihood of safety-and-soundness problems. A safety-and-soundness problem first is defined as an outright failure; later in the paper we define a safety-and-soundness problem as a downgrade from a CAMEL-1 or CAMEL-2 rating to a CAMEL-3, CAMEL-4, or CAMEL-5 rating.

We gauged asset quality with three different measures: the ratio of nonperforming loans to total loans (BAD-LOANS), the ratio of consumer loans to total assets (CONSUMER), and the ratio of other real estate owned to total loans (OREO). Nonperforming loans are loans that are 90 or more days past due or in nonaccrual status. (In bank accounting, loans are either classified as accrual or nonaccrual. As long as a loan is classified as accrual, the interest due is counted as current revenue, even if the borrower falls behind on interest payments.) We used the nonperforming loan ratio as a measure of asset quality because banks ultimately charge off relatively high percentages of nonperforming loans. We used the consumer loan ratio because the charge-off rate for consumer loans has been higher historically than for other types of loans. For example, nationwide, the average charge-off rate for all types of bank loans from 1990 through 1997 was 0.86 percent; for consumer loans, the average was 2.08 percent. Finally, we included "other real estate owned" because the term generally applies to collateral seized after loan defaults; banks with higher OREO ratios tend to have more credit risk exposure. We expect that banks with higher values of these ratios will experience more safety-and-soundness problems.

As proxies for managerial competence, we used noninterest expense as a percentage of total revenue (OVERHEAD), insider loans as a percentage of total assets (INSIDER), and occupancy expense as a percentage of average assets (OCCUPANCY). Because well-managed banks hold down overhead costs, avoid excessive lending to insiders, and pay reasonable amounts for office space, we expect that banks with higher values of these ratios will suffer more safety-and-soundness problems.

We measured earnings strength with the ratio of net income to total assets (return on assets, or ROA), and the ratio of interest income accrued, but not collected, to total loans (UNCOLLECTED). All other things being equal, higher earnings provide a greater cushion for withstanding adverse economic shocks. We expect, therefore, that higher returns on assets will reduce the likelihood of safety-and-soundness problems. Banks with high levels of interest income accrued but not collected are vulnerable to large restatements of earnings and capital because the loans generating accrued-interest-that-has-not-been-collected could be reclassified as nonaccrual. We expect, therefore, that higher levels of uncollected interest income point to future safety-and-soundness problems.

We gauged liquidity risk with three measures: liquid assets (cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets (LIQUID), large time deposits as a percentage of total assets (LARGE-TIME), and core deposits as a percentage of total assets (CORE). A larger stock of liquid assets indicates greater ability to meet unexpected liquidity needs. Larger stocks of liquid assets, therefore, should translate into fewer safety-and-soundness problems. Liquidity risk also depends on the division of bank liabilities between volatile and core funding. Large time deposits represent a volatile source of funding because they are not fully insured by the FDIC; a sudden jump in market interest rates or a sudden deterioration in bank condition could raise funding costs dramatically. All other things being equal, greater reliance on large time deposits implies a greater likelihood of safety-and-soundness problems. Similarly, the smaller a bank's volume of nonvolatile or core deposits, the greater the likelihood of safety-and-soundness problems.

Finally, we included control variables for bank size and holding company affiliation in the representative versions of the screens and models. We added the natural logarithm of total assets (SIZE) because larger banks should be better able to diversify across product lines and geographic regions and, therefore, avoid safety-and-soundness problems. We also added a control variable to capture the effect of holding company affiliation. This variable, BHCRATIO, equaled the ratio of total assets in the sample bank to total assets in all banks in the parent-holding company. Because holding companies are better able to serve as a source of strength for their smaller members, we expect that lower values of BHCRATIO imply fewer safety-and-soundness problems in the future. (The shaded insert discusses the holding company control variable in more detail.) Table 2 presents a complete list of the variables used in this article as the supervisory screens and as independent variables in the econometric models. The table also includes a positive or negative sign indicating the hypothesized relationship between each variable and the likelihood of outright failure or CAMEL downgrade from CAMEL 1 or 2 to CAMEL 3, 4, or 5.

WHY CONTROL FOR HOLDING COMPANY MEMBERSHIP?

It may seem curious that we included a variable related to holding company membership in the supervisory screens and the econometric model. We included this variable because theory and evidence suggest that small banks belonging to large holding companies are less likely to fail or suffer supervisory downgrades.

To see why small banks belonging to large holding companies are less likely to encounter safety-and-soundness problems, suppose that such a bank is facing serious asset quality problems. The owners of the holding company must confront a trade-off when deciding whether to inject equity into this subsidiary. On the one hand, alternative investments are likely to offer higher returns because loan losses will absorb some of the injections. On the other hand, not injecting equity into the troubled subsidiary could lead to a failure, which, in turn might taint the reputation of the holding company in the eyes of financial markets or bank supervisors. Because the bank is small, the injection is more likely to prevent a failure and the attendant reputational damage. In short, when a subsidiary bank is relatively small, the holding company is better able to serve as a source of strength.

For this reason, we added BHCRATIO, the assets of the sample bank divided by the total assets of all bank subsidiaries of its holding company, to the list of screens and explanatory variables. BHCRATIO assumed a value of unity when the sample bank did not belong to a holding company or was the only bank in the holding company. All other things being equal, the smaller the assets of the sample bank relative to the assets of the holding company, the smaller the value of BHCRATIO. We expect to observe a positive relationship between BHCRATIO and future safety-and-soundness problems (failures or downgrades of CAMEL ratings to problem status).

Empirical studies confirm that BHCRATIO helps explain both bank failures and capital injections into troubled holding company subsidiaries. Belongia and Gilbert (1990) found that a variable constructed like BHCRATIO enhanced the explanatory power of a model of agricultural bank failures: the smaller the agricultural banks relative to the size of their parent organizations, the lower their probabilities of failure. Gilbert (1991) also found that a variable constructed like BHCRATIO helped explain equity injections into undercapitalized banks; the smaller the undercapitalized banks relative to the size of their parent organizations, the larger the equity injections into the undercapitalized banks.

Taken together, our empirical evidence supports the hypothesis that BHCRATIO is positively related to both failures and CAMEL downgrades. When used as a screen, the means differed in the hypothesized direction in two of the three failure samples (1988 and 1989) and six of the seven downgrade samples. When used in the econometric model, the coefficient on BHCRATIO was positive and significant in only one of the three failure prediction models (1987), but it was positive and statistically significant in all the CAMEL downgrade equations. The lack of supporting evidence from the failure prediction screens and models may be the result of the Texas bank failures of the late 1980s. In several prominent cases, regulators shut down entire holding companies even when many of the subsidiary banks were safe and sound. See Cannella, et. al. (1995) for additional discussion of the closure of these holding companies and banks.

Table 2

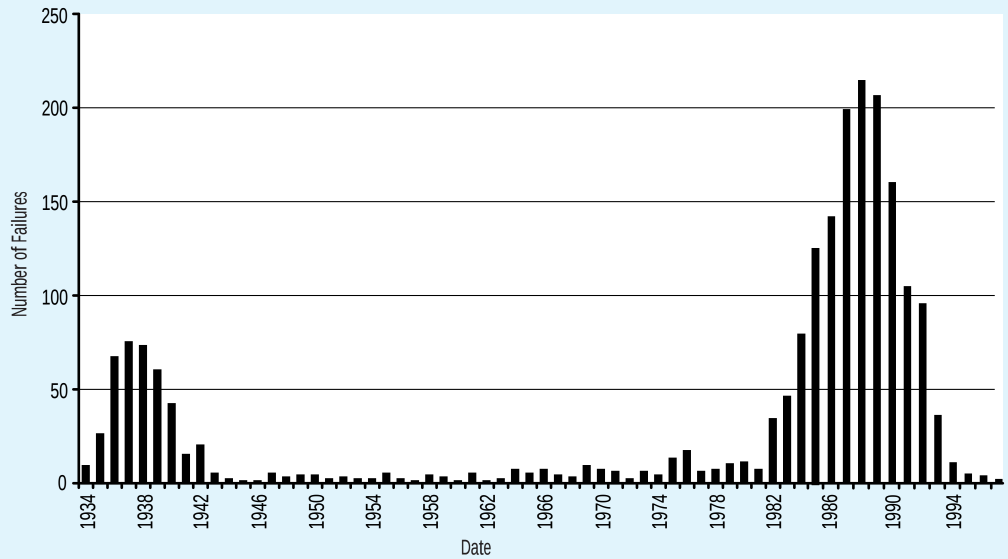
What Variables Help Predict Bank Failures or CAMEL Downgrades?

This table lists the single-variable screens and independent variables used in our econometric models. The sign indicates the hypothesized relationship between the variable and the likelihood of a safety-and-soundness problem. For example, the negative sign for the equity-to-assets ratio indicates that a higher capital ratio would reduce the likelihood of a failure or CAMEL downgrade.

Symbol	Description	Hypothesis about sign of coefficient for predicting failure or CAMEL downgrades (positive sign indicates positive correlation with probability of failure or rating downgrade).
EQUITY	Equity as a percentage of total assets.	–
BAD-LOANS	Nonperforming loans as a percentage of total loans.	+
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.	+
CONSUMER	Consumer loans as a percentage of total assets.	+
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.	+
OVERHEAD	Noninterest expense as a percentage of total revenue.	+
OCCUPANCY	Occupancy expense as a percentage of average assets.	+
ROA	Net income as a percentage of total assets.	–
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.	+
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.	–
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.	+
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.	–
SIZE	Natural logarithm of total assets, in thousands of dollars.	–
BHCRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO $\equiv$ 1.	+

Figure 1

Number of Commercial Bank Failures by Year 1934-97



This figure shows that U.S. commercial bank failures peaked in 1988 and dropped precipitously during the 1990s.

GAUGING SUPERVISORY SCREENS AND ECONOMETRIC MODELS AS PREDICTORS OF BANK FAILURE

We began by using the representative supervisory screens on historical data to gauge how well they would have predicted bank failures during 1989, 1990, and 1991. To conduct these tests, we partitioned a list of all U.S. banks during those years into failures and survivors for each year. The sample ended in 1991 because so few banks failed after the early 1990s (see Figure 1). We then used 1987, 1988, and 1989 call report data to generate screen values for the sample banks two years before the observation of failure or survival. An individual screen would provide early warning if the mean value of the screen for the failed banks differed significantly from the mean value for the survivor banks in the direction hypothesized. The capital screen, for example, would meet this condition if the mean equity-to-asset (EQUITY) ratio for the failed banks was significantly below the mean ratio for the surviving banks two years before the observation of failure or survival. Table 3 presents the means and standard deviations of the screen ratios for both banks that failed and banks that survived.

Overall, the individual screens would have done a good job predicting bank failures during 1989, 1990, and 1991. For 11 of the 14 variables, the average screen values for the failed and surviving banks differed significantly in the hypothesized direction across all three years. Indeed, only the consumer loans screen, the core deposit screen, and the size control variable failed to correlate consistently with future failures. The capital screen clearly illustrates the signaling value of individual supervisory screens. In all three years, the differences in means were economically large and statistically significant—banks with weaker capital ratios were more likely to fail. For example, the fourth-quarter 1987 equity-to-asset ratio for banks that would fail during 1989 (4.30 percent) was well below the ratio for banks that would survive that year (8.50 percent).

Table 3

How Well Do the Individual Screens Predict Bank Failures?

This table presents evidence about the failure prediction record of individual supervisory screens. The far-left and right columns for each year contain the mean values of the screens; standard deviations appear in parentheses below the means. An asterisk indicates a significant difference (at the 5-percent level) between the means for failed and survivor banks. Shading highlights screens with significant predictive power in all three years. The center column for each year (‡) shows the number of survivor banks with screen values worse than those of the average failed bank; the larger this number, the worse the performance of the screen. Taken together, this evidence shows that screens warn of potential failures but also can lead to many unnecessary exams.

	Data as of 1987:4 for:			Data as of 1988:4 for:			Data as of 1989:4 for:		
	149 banks that failed in 1989	‡	11,838 banks that survived 1989	115 banks that failed in 1990	‡	11,446 banks that survived 1990	82 banks that failed in 1991	‡	11,246 banks that survived 1991
EQUITY	4.30*	359	8.50	3.38*	180	8.58	4.24*	273	8.69
	(2.23)		(3.09)	(3.82)		(3.22)	(2.36)		(3.38)
BAD-LOANS	8.19*	612	2.54	8.22*	386	2.16	6.79*	493	2.02
	(6.02)		(2.95)	(5.20)		(2.52)	(4.03)		(2.56)
OREO	6.85*	360	1.28	7.36*	317	1.24	5.24*	631	1.22
	(9.71)		(2.26)	(7.29)		(2.32)	(5.68)		(2.46)
CONSUMER	10.54	4,817	10.79	12.72*	3,361	10.74	12.63	3,380	10.71
	(8.68)		(7.82)	(10.33)		(7.98)	(12.03)		(7.97)
INSIDER	1.09*	1,724	0.52	1.51*	1,074	0.54	1.00*	1,850	0.53
	(2.20)		(0.92)	(2.35)		(1.10)	(1.12)		(0.96)
OVERHEAD	46.62*	1,277	35.04	49.79*	741	34.11	41.36*	1,423	32.05
	(26.12)		(22.33)	(16.59)		(10.50)	(12.33)		(9.89)
OCCUPANCY	0.66*	2,276	0.49	0.80*	1,185	0.49	0.76*	1,383	0.48
	(0.39)		(0.31)	(0.42)		(0.30)	(0.41)		(0.31)
ROA	-2.16*	358	0.67	-2.55*	177	0.80	-1.28*	336	0.87
	(3.19)		(1.16)	(2.73)		(1.11)	(1.67)		(1.03)
UNCOLLECTED	0.96*	2,037	0.67	0.94*	2,418	0.71	0.97*	2,594	0.76
	(0.62)		(0.39)	(0.47)		(0.40)	(0.42)		(0.43)
LIQUID	32.99*	2,702	45.36	32.76*	2,776	44.43	27.82*	1,469	43.87
	(13.86)		(15.24)	(11.84)		(15.14)	(10.33)		(14.84)
LARGE-TIME	22.98*	757	9.27	17.07*	1,666	9.65	14.77*	2,390	10.06
	(13.04)		(7.90)	(8.25)		(7.41)	(7.83)		(7.30)
CORE	69.42*	1,466	79.41	77.33	3,796	78.90	77.23	3,904	78.38
	(14.32)		(9.73)	(9.22)		(9.53)	(10.98)		(9.42)
SIZE	10.98	7,374	10.79	10.72	5,876	10.84	11.26*	7,717	10.89
	(1.35)		(1.24)	(1.09)		(1.26)	(1.55)		(1.27)
BHCRATIO	0.62*	8,517	0.75	0.83*	7,849	0.75	0.92*	7,571	0.75
	(0.44)		(0.39)	(0.31)		(0.39)	(0.23)		(0.39)

Table 4

What Were the CAMEL Ratings of Banks that Failed in
1989, 1990, and 1991?

This table shows that supervisors already were aware of problems in most of the banks that failed in 1989, 1990, and 1991. Shading highlights the failure record of problem banks (CAMEL 3, 4, or 5). Supervisors recognize that these banks are significant failure risks and, therefore, monitor them closely. CAMEL-1 or -2 banks rarely fail, so they are not monitored as closely.

Rate of Bank Failure by Prior CAMEL Rating

Date of Rating (Calendar Year of Failure)	CAMEL Rating	Number of Banks	Number of Failures	Percentage Failed
March 1988 (1989)	1	1,908	0	0.00%
	2	5,029	6	0.12
	3	1,493	30	2.01
	4	643	52	8.09
	5	115	27	23.48
March 1989 (1990)	1	2,409	0	0.00
	2	6,130	10	0.16
	3	1,585	19	1.20
	4	673	48	7.13
	5	139	36	25.90
March 1990 (1991)	1	2,573	0	0.00
	2	6,423	9	0.14
	3	1,474	14	0.95
	4	629	31	4.93
	5	158	27	17.09

A better measure of the value added by individual screens, however, is their record in identifying failure candidates that were not already on supervisors' watch lists. Suppose, for example, that it is March 1988, and supervisors are scheduling and staffing exams for the rest of the year. Most of the banks with CAMEL composite ratings below 2 already are under scrutiny, so supervisors would like to use the latest call report data (year-end 1987) to identify CAMEL 1 or 2-rated banks that are significant failure risks in 1989. A tool that accurately predicted the 1989 failures of CAMEL 3, 4, and 5-rated banks, but did a poor job predicting the failures of CAMEL 1 or 2-rated banks, would not add much value in off-site surveillance because it would give supervisors little new information.

With this standard in mind, we looked again at the failure prediction record of the single-variable screens for 1989, 1990, and 1991. First, we identified all the CAMEL-2 banks as of March 1988, 1989, and 1990 and partitioned that set into banks that failed and banks that did not fail the following calendar year. We then generated the corresponding screen values using call report data from the previous December. Finally, we calculated the percentage of CAMEL-2 banks that would have to be examined, using each screen as a guide, to flag one-half of the CAMEL-2 banks that failed the next year. We selected one-half of the failures as a threshold because catching all of the CAMEL-2 failures would require, in some cases, examining most of the CAMEL-2 banks. We looked at only CAMEL-2 banks because no banks rated CAMEL 1 as of March 1988, March 1989, or March 1990 failed during the following calendar year. Table 4 puts the CAMEL-2 failure numbers in perspective by showing the failure rates for each CAMEL cohort, while Table 5 shows the percentage of CAMEL-2 banks that must be examined, using each screen, to catch one-half of the failures the next year.

The evidence for 1989, 1990, and 1991 failures shows that single-variable screens would have improved significantly over random examination of CAMEL-2 banks. In each of the years, several screens were particularly informative. The large-time-deposits-to-total-assets ratio, for example, outperformed the other 13 screens as a tool for identifying 1989 failures. Had supervisors used the fourth-quarter 1987 value of this ratio as a guide, they would have caught one-half of the 1989 failures after examining only 1.7 percent of the CAMEL-2 banks. For 1990 failures, the return-on-asset screen was dominant; had supervisors scheduled exams using fourth-quarter 1988 values of this screen they would have caught one-half of 1990's failures after visiting only 0.9 percent of the CAMEL-2 banks. Finally, for 1991 failures, the non-performing loan screen turned in the best performance. Supervisors could have identified one-half of that year's failures by examining only 2.2 percent of CAMEL-2 banks. To put these numbers in perspective, if supervisors scheduled examinations randomly, on the average examiners would have had to visit 50 percent of CAMEL-2 banks to catch one-half of those that failed during the next 12 to 24 months. The average three-year performance of every single-variable screen except the consumer loan screen and the size control variable was well below 50 percent.

Next, we fit an econometric model to the data on bank failures and the measures used as screens to gauge how well it would have predicted failures. Again, we partitioned U.S. banks into failures and survivors for each year, assigning a "1" to banks that failed and a "0" to banks that survived. This binary observation served as the dependent variable in the model. As independent variables, we used the two-year lagged screen values, including the size and holding company control variables. We estimated a logit model—a specific type of econometric model used when the dependent variable is a "0" or "1"—year by year; that is, we fit the model to 1985 screen values and 1987 failure observations, then to 1986 screen values and 1988 failure observations, and finally to 1987 screen values and 1989 failure observations. Table 6 presents the estimation results.

The econometric model would also have done a good job identifying failures in 1987, 1988, and 1989. For all three years, we could reject the hypothesis that the model had no explanatory power. Moreover, six individual coefficients differed statistically from zero with the hypothesized signs across all three equations. Specifically, low capital ratios (EQUITY), low liquid-asset ratios (LIQUID), high nonperforming-loan ratios (BAD-LOANS), high other-real-estate-owned ratios (OREO), high interest-accrued-but-not-collected ratios (UNCOLLECTED), and high large-time-deposit ratios (LARGE-TIME) correlated strongly with future failures. Overall, the econometric model implies that capital protection, asset quality, and liquidity positions are the most important determinants of failure risk.

Next, we used the econometric model to identify failure candidates that were not already on supervisors' watch lists. The evidence from 1989, 1990, and 1991 (which appears in Table 7) shows that the econometric model also would have improved significantly over random examination. Specifically, if the sample banks had been examined from the highest to the lowest estimated probability of failure (based on year-end 1987 data), 55 banks would have had to be examined to catch three of the six that would fail in 1989. To flag five of the 10 banks that would fail in 1990, 51 examinations would have been necessary. To identify five of the nine failures in 1991, 155 banks would have had to be examined. At first glance these numbers might seem high, but 55 banks represented only 1.1 percent of all CAMEL-2 banks in 1988; 51 represented only 0.8 percent of CAMEL-2 banks in 1989; and 155 represented a mere 2.4 percent of all CAMEL-2 banks in 1990. In short, the econometric model improves significantly on the random examination of CAMEL-2 banks.

Table 5

Do Individual Supervisory Screens Improve Over Random Examination of CAMEL-2 Banks?

This table demonstrates that individual supervisory screens improve over random examination of CAMEL-2 banks. To catch one-half of the following year's failures using a random examination strategy, supervisors would have to order, on average, visits to one-half of the CAMEL-2 banks. Only the consumer loan screen and the size control variable had average performance ratios above 50 percent.

Note, however, the considerable variance in the performance of individual supervisory screens. The performance ranking of individual screens changed significantly from year to year. Shading highlights screens that placed among the top five predictors in all three years. Only two screens placed consistently among the top five predictors.

Single-variable screen	**For each year, the first column shows the percentage of CAMEL-2 banks that must be examined to include one-half of the banks that failed in the following calendar year. The second column indicates the rank of each screen from best (1) to worst (14).**					
	Banks that Failed in:					
	1989		**1990**		**1991**	
	Percent based on 1987:4 data	Rank of screen	Percent based on 1988:4 data	Rank of screen	Percent based on 1989:4 data	Rank of screen
EQUITY	4.6	3	2.3	2	4.0	2
BAD-LOANS	16.9	6	7.3	4	2.2*	1
OREO	8.6	5	21.6	8	17.6	6
CONSUMER	26.9	10	37.0	11	86.1	14
INSIDER	44.8	12	9.3	6	37.1	11
OVERHEAD	22.6	7	5.6	3	56.7	12
OCCUPANCY	69.6	14	8.7	5	14.2	5
ROA	5.3	4	0.9*	1	4.7	3
UNCOLLECTED	25.5	8	37.2	12	31.1	10
LIQUID	25.9	9	15.9	7	5.0	4
LARGE-TIME	1.7*	1	23.7	10	29.1	8
CORE	3.9	2	46.6	14	30.7	9
SIZE	42.0	11	41.7	13	70.4	13
BHCRATIO	55.9	13	21.9	9	20.0	7

*Lowest among the screens.

EQUITY	Equity as a percentage of total assets.
BAD-LOANS	Nonperforming loans as a percentage of total loans.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.
CONSUMER	Consumer loans as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.
OCCUPANCY	Occupancy expense as a percentage of average assets.
ROA	Net income as a percentage of total assets.
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
SIZE	Natural logarithm of total assets, in thousands of dollars.
BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.

Table 6

How Well Does the Econometric Model Fit the Bank Failure Data?

This table presents the estimated regression coefficients for the failure prediction logit. The model predicts in-sample failures ("1" represents failure; "0" denotes survivor) for calendar year *t* with year *t*-2 call report data. Standard errors appear in parentheses below each coefficient. Three asterisks denote significance at the 1 percent level; two asterisks denote significance at the 5-percent level. Shading highlights coefficients that were significant with the correct sign in all three years. Overall, the evidence in this table suggests that the econometric model predicted in-sample failures well.

	Banks that Failed or Survived in:		
Independent Variables	**1987**	**1988**	**1989**
Intercept	-0.994 (2.801)	-2.588 (2.525)	-6.479 (3.499)
EQUITY	-0.303*** (0.055)	-0.314*** (0.056)	-0.285*** (0.051)
BAD-LOANS	0.107*** (0.018)	0.099*** (0.020)	0.095*** (0.023)
OREO	0.097*** (0.031)	0.047** (0.024)	0.122*** (0.019)
CONSUMER	0.007 (0.012)	0.002 (0.012)	-0.018 (0.012)
INSIDER	0.041 (0.023)	0.084 (0.048)	0.102 (0.054)
OVERHEAD	-0.014 (0.014)	-0.012 (0.013)	0.001 (0.002)
OCCUPANCY	0.710** (0.314)	0.450 (0.374)	-0.069 (0.308)
ROA	-0.061 (0.052)	0.007 (0.065)	0.007 (0.050)
UNCOLLECTED	0.935*** (0.132)	0.608*** (0.160)	0.828*** (0.215)
LIQUID	-0.041*** (0.010)	-0.019** (0.008)	-0.033*** (0.009)
LARGE-TIME	0.072*** (0.021)	0.074*** (0.016)	0.115*** (0.026)
CORE	0.003 (0.022)	0.007 (0.018)	0.034 (0.025)
SIZE	-0.356*** (0.111)	-0.120 (0.101)	-0.011 (0.116)
BHCRATIO	1.075*** (0.340)	-0.119 (0.236)	-0.348 (0.260)
Number of Observations	12,645	12,345	11,987
Pseudo-R^2	0.375	0.275	0.403
-2 log likelihood testing whether all coefficients (except the intercept) = 0	633.108***	453.035***	645.996***

EQUITY	Equity as a percentage of total assets.
BAD-LOANS	Nonperforming loans as a percentage of total loans.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.
CONSUMER	Consumer loans as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.
OCCUPANCY	Occupancy expense as a percentage of average assets.
ROA	Net income as a percentage of total assets.
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
SIZE	Natural logarithm of total assets, in thousands of dollars.
BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.

Table 7

How Well Does the Econometric Model Identify CAMEL-2 Failure Candidates?

This table quantifies the supervisory value added by the econometric model. Specifically, it shows how many CAMEL-2 banks must be examined in each year, based on logit probability estimates using data from the previous year, to catch each potential failure. For example, in 1988, supervisors would have had to examine 18 (or 0.4 percent) of the 2-rated banks to catch one of the 1989 failures. Catching one-half of the 1989 failures would have required examining 55 (or 1.1 percent) of the 2-rated banks. To catch all six failures, supervisors would have had to examine 650 (or 12.9 percent) of the 2-rated banks. Shading highlights the number of banks that must be examined to catch one-half of the failures in each year. Overall, the evidence suggests that the econometric model improved significantly on random examinations of CAMEL-2 banks.

Among the CAMEL-2 rated banks, rank based on probability of failure:		Estimated probability of failure	Percentage of CAMEL-2 rated banks that must be examined to include this failed bank
Among those that failed	**Among all CAMEL-2 rated banks**		
Among banks rated CAMEL 2 as of March 1988, six that failed during 1989:			
1	18	5.2%	0.4%
2	20	4.9	0.4
3	55	2.9	1.1
4	82	2.2	1.6
5	547	0.6	10.9
6	650	0.5	12.9
Among banks rated CAMEL 2 as of March 1989, 10 that failed during 1990:			
1	4	33.8	0.1
2	8	12.5	0.1
3	34	5.1	0.6
4	43	4.8	0.7
5	51	4.4	0.8
6	58	4.1	0.9
7	206	2.1	3.4
8	544	1.2	8.9
9	1,324	0.7	21.6
10	3,488	0.3	56.9
Among banks rated CAMEL 2 as of March 1990, nine that failed during 1991:			
1	34	4.7	0.5
2	72	3.4	1.1
3	101	2.9	1.6
4	141	2.5	2.2
5	155	2.3	2.4
6	212	2.0	3.3
7	523	1.1	8.1
8	1,913	0.4	29.8
9	5,774	0.0	89.9

At first glance, the resource-savings benchmark—the number of CAMEL-2 banks that must be examined to catch one-half of the following year's failures—appears to suggest that the screens and the model would have been comparable tools for allocating on-site examination resources. The comparison appears in Table 8, which combines data from Tables 5 and 7. In each year, the performance of the dominant screen is relatively close to the performance of the econometric model. For example, using the econometric model as a guide, supervisors would have had to examine 1.1 percent of all CAMEL-2 banks (as of March 1988) to catch one-half of the 1989 failures. If supervisors had used the dominant screen instead—the large-time-deposit ratio—they would have had to examine 1.7 percent of the CAMEL-2 banks. For 1990, the econometric model would have identified one-half of the failures after 0.8 percent of 2-rated banks had been examined; the comparable figure for the dominant screen (return on assets) was 0.9 percent. Finally, for 1991 failures, the dominant screen outperformed the econometric model. The nonperforming-loan screen identified one-half of the failures after examining 2.2 percent of the CAMEL-2 banks; the figure for the econometric model was 2.4 percent.

A closer look, however, reveals that the screens and the model would not have been equally effective surveillance tools. Although during each year the performance of the dominant screen is close to that of the econometric model, the dominant screens vary from year to year. Moreover, only two screens ranked among the top five in all three years, and in only one of those six cases (two screens, three years) did a screen beat the model. On average during the three-year period, the model significantly outperformed all of the individual screens. On average, supervisors could have caught one-half of the surprise failures by examining only 1.4 percent of the CAMEL-2 banks. The lowest average for the supervisory screens—the return-on-asset screen and the equity screen—was 3.6 percent. To put this evidence in perspective, suppose supervisors decided on the basis of 1989 screen performance to use the large-time-deposits-to-total-assets ratio as a guide for predicting 1990 failures. With such a guide, they would have had to examine 23.7 percent of the banks rated CAMEL-2 as of March 1989 to catch one-half of the failures. The comparable percentage using the econometric model is 0.8 percent. In summary, for single-variable screens to be as effective as the model, supervisors would have to know at the beginning of each year which screen would perform relatively well—an unrealistic information requirement.

It also is important to compare the performance of the screens and the model for a broader range of type-1 and type-2 errors. Put another way, the resource savings benchmark, while intuitively appealing, represents only one possible type-1/type-2 error trade-off. Type-1 errors, in this context, are missed failures; these errors impose unexpected costs on the deposit insurance fund and the real economy. Type-2 errors are missed survivors; these errors waste scarce examination resources and impose undue burdens on banks. Consider a concrete example of type-2 error using the individual capital screen. Suppose bank supervisors scheduled 1989 exams for all banks (CAMEL 1 through 5) using only fourth quarter 1987 values of the capital screen. Because the distributions of capital screen values for the failed and survivor banks overlap considerably (see Figure 2), this approach would lead to a large number of type-2 errors. For example, 359 survivor banks had weaker equity ratios than the average ratio for all the failed banks (see Table 3).

The evidence from a broader range of type-1/type-2 error trade-offs confirms the statistical dominance of the econometric model. An econometric model would dominate a set of screens as devices for identifying failures if it produced fewer type-2 errors (missed survivors) for any desired level of type-1 errors (missed failures). In pictures, meeting this condition implies that a curve tracing the trade-off between the two types of errors for the econometric model lies completely below

Table 8

How Do the Individual Supervisory Screens and the Econometric Model Compare as Tools for Allocating On-Site Examination Resources?

This table illustrates the superior performance of the econometric model as a tool for allocating on-site examination resources. It combines data from Tables 5 and 7. The columns show the percentage of banks that must be examined, using either the econometric model or a specific supervisory screen as a guide, to catch one-half of the banks that will fail that year. In each year, the dominant screen comes close to the model's performance, but the dominant screen varies year to year. Moreover, the three-year average for the model is well below the averages for the single variable screens.

Method of ranking banks by probability of failure.	**Among banks rated CAMEL 2, the percentage that must be examined to include one-half of the banks that failed in the following calendar year.**			
	Banks that failed in:			
	1989	1990	1991	Mean Percentage
Model	1.1%	0.8%	2.4%	1.4%
Screens				
EQUITY	4.6	2.3	4.0	3.6
BAD-LOANS	16.9	7.3	2.2*	8.8
OREO	8.6	21.6	17.6	15.9
CONSUMER	26.9	37.0	86.1	50.0
INSIDER	44.8	9.3	37.1	30.4
OVERHEAD	22.6	5.6	56.7	28.3
OCCUPANCY	69.6	8.7	14.2	30.8
ROA	5.3	0.9*	4.7	3.6
UNCOLLECTED	25.5	37.2	31.1	31.3
LIQUID	25.9	15.9	5.0	15.6
LARGE-TIME	1.7*	23.7	29.1	18.2
CORE	3.9	46.6	30.7	27.1
SIZE	42.0	41.7	70.4	51.4
BHCRATIO	55.9	21.9	20.0	32.6

*Lowest among the screens for that year.

EQUITY	Equity as a percentage of total assets.
BAD-LOANS	Nonperforming loans as a percentage of total loans.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.
CONSUMER	Consumer loans as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.
OCCUPANCY	Occupancy expense as a percentage of average assets.
ROA	Net income as a percentage of total assets.
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
SIZE	Natural logarithm of total assets, in thousands of dollars.
BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.

the trade-off curves for every single-variable screen.[5] Figure 3 presents the 1990 failure trade-off curve for the econometric model and the four best single variable screens, using the sample of CAMEL-2 banks. With only two exceptions, the trade-off curve for the econometric model does indeed lie below the curves for the individual screens. For small ranges of values, trade-off curves for the return-on-assets and the capital screens dip below the curve for the econometric model. Similar curves for 1989 and 1991 failures reveal similar patterns—the trade-off curve for the econometric lies below the curves for the individual screens with only a few exceptions. In those cases, one or two screens outperform the model for a small range of type-1/type-2 error trade-offs, but no one screen consistently outperforms the model. In summary, only by correctly guessing which screen will dominate at the beginning of the year and by preselecting a desired type-1 error rate from a small range of values can a supervisor beat the econometric model with a single-variable individual screen. These conditions are clearly difficult to meet.

GAUGING SUPERVISORY SCREENS AND ECONOMETRIC MODELS AS PREDICTORS OF CAMEL DOWNGRADES

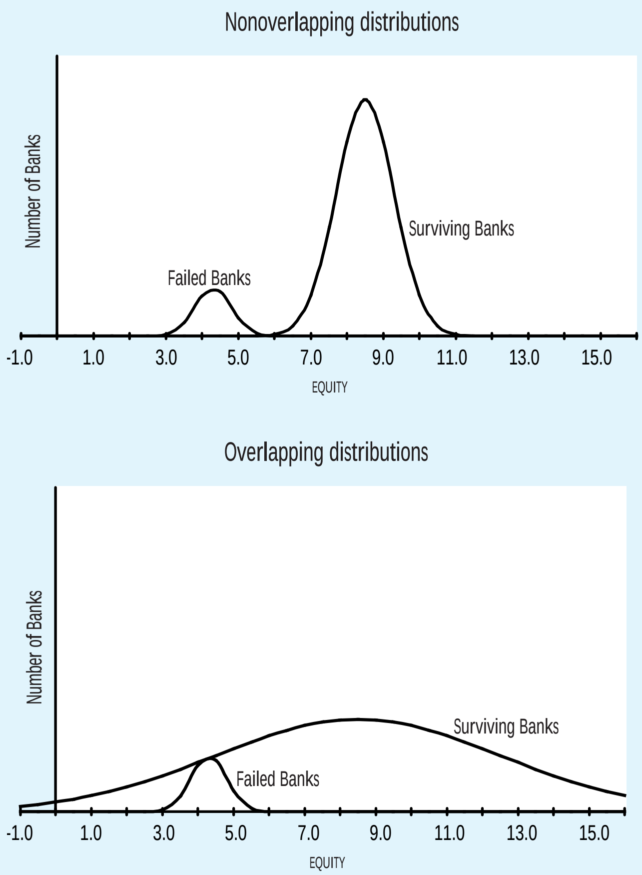
Because failures have fallen off sharply since the early 1990s, supervisors have become interested in developing tools for flagging safe-and-sound banks that will develop problems. For this reason, we estimated an econometric model designed to capture the likelihood that a bank's CAMEL rating will be downgraded from CAMEL 1 or 2 to CAMEL 3, 4, or 5. Because such downgrades remained relatively common through 1997, we have a large enough sample to conduct a meaningful comparison of the resource savings obtained with the screens and the econometric model. (Figure 4 and Table 9 provide data on the frequency of these downgrades.)

To estimate a downgrade model, we changed the definition of a safety-and-soundness problem and the sample selection criteria. Now, in the econometric model, we assigned a "1" to banks that suffered a downgrade from safe-and-sound status (CAMEL 1 or 2) to problem status (CAMEL 3, 4, or 5) and a "0" to all other

Figure 2

Hypothetical Distributions of Equity Ratios

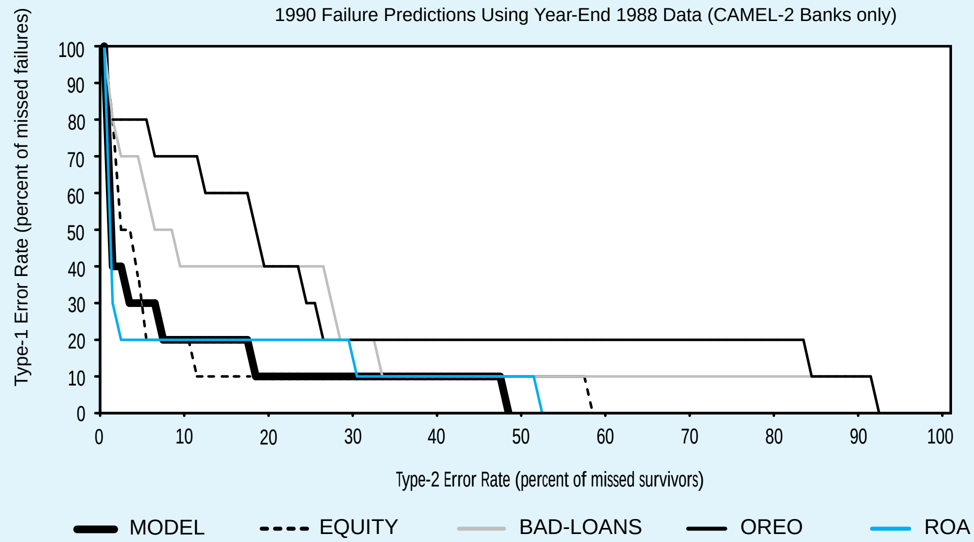
This figure demonstrates type-2 error (the problem of missed survivors) using supervisory screens. When the distributions of the screen ratios for failing and surviving banks overlap considerably, supervisors who rely only on screens to schedule exams will devote a large quantity of on-site resources to banks unlikely to fail. Suppose that the figures below are capital screen (equity-to-total-asset ratio) distributions. In the top panel, the distribution for failures lies completely to the left of the distribution for survivors. If the actual distributions looked like this, supervisors could allocate on-site resources efficiently by examining banks with the lowest capital ratios. Unfortunately, the actual distributions are more like those in the bottom panel. For example, in late 1987, 359 of the 11,838 banks that would survive through 1989 had equity-to-asset ratios below the mean for the 149 banks that would fail that year. If supervisors flagged banks with the lowest equity-to-asset ratios in late 1987, their watch list would have included many more survivor banks than failed banks.



[5] Our graphical analysis of error trade-offs follows the approach used by Cole, Cornyn, and Gunther (1995).

Figure 3

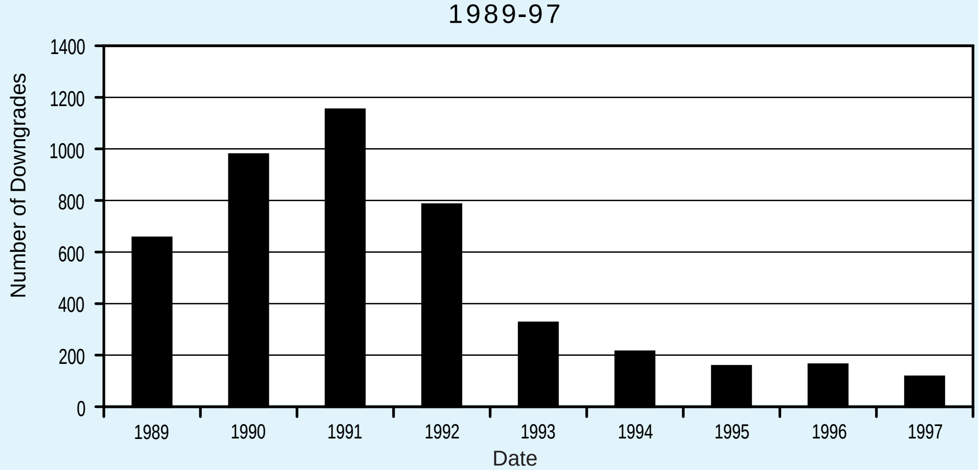
What is the Trade-Off Between False Negatives and False Positives in the Failure-Prediction Model Compared to the Individual Screens?



This figure shows the trade-off between the type-1 error rate (missed failures) and the type-2 error rate (missed survivors). The type-1 error rate is the percentage of banks rated CAMEL 2 that subsequently failed but were not identified by the model (or screen). The type-2 error rate is the percentage of banks rated CAMEL 2 that did not subsequently fail but were misidentified by the model (or screen) as a failure risk. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than most individual screens. Moreover, even in years when individual screens dominate the logit model over some ranges of the type-1 versus -2 trade-off, the dominant screens are not consistently the same. (For clarity, only the four best screens are shown.)

Figure 4

Number of Commercial Bank Downgrades by Year 1989-97



This figure shows that downgrades to problem status (CAMEL 3,4, or 5) are still relatively common, although the absolute number has declined since the early 1990s.

Table 9

How Many CAMEL-1 and CAMEL-2 Banks Suffered Downgrades to CAMEL 3, 4, or 5 from 1991 to 1997?

This table shows the number of our sample banks that were downgraded to problem status in each year. We excluded banks receiving downgrades to problem status the same year as the CAMEL 1 or 2 observation from the sample to avoid biasing comparisons against supervisory screens. Note: As overall banking performance improved in the 1990s, the percentage of banks suffering downgrades fell, but downgrades were still much more common than failures.

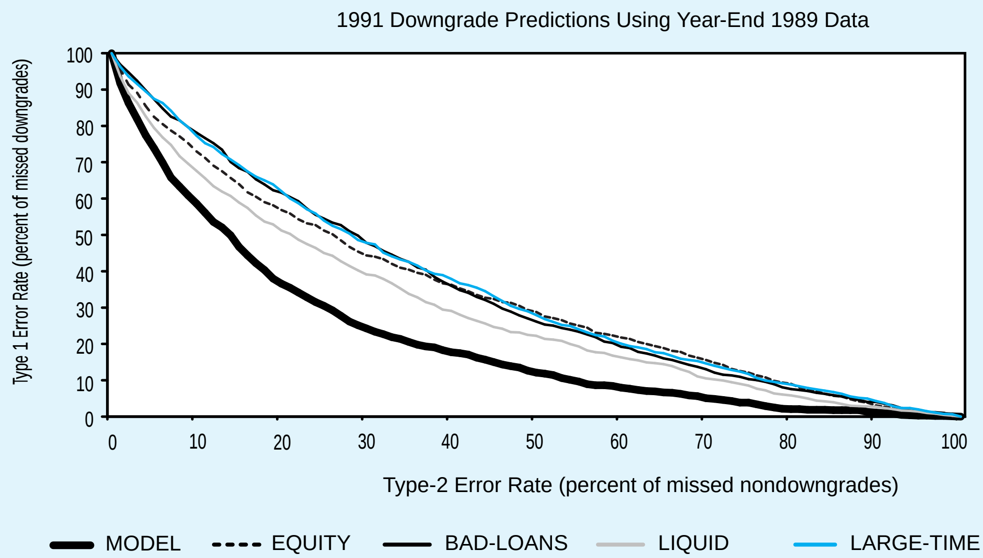
Date of Rating (Year of Downgrade)	CAMEL Rating	Number of Banks	Number of Banks Downgraded	Percentage Downgraded
March 1990 (1991)	1	2,057	79	3.84%
	2	5,036	987	19.60
March 1991 (1992)	1	1,956	51	2.61
	2	4,985	670	13.44
March 1992 (1993)	1	1,972	17	0.86
	2	5,212	292	5.60
March 1993 (1994)	1	2,041	14	0.69
	2	5,030	185	3.68
March 1994 (1995)	1	2,359	13	0.55
	2	4,446	127	2.86
March 1995 (1996)	1	2,583	13	0.50
	2	3,940	135	3.43
March 1996 (1997)	1	1,931	9	0.47
	2	2,420	103	4.26

banks. All of the sample banks were examined during the year of the downgrade. We excluded banks receiving downgrades the same year as the CAMEL 1 or 2 observation. Without this exclusion, the predictive power of both the screens and the model would be seriously weakened. A simple example illustrates the problem. Suppose we are selecting sample banks for 1990. We begin with all CAMEL-1 and CAMEL-2 banks as of March 1990. If we did not exclude banks receiving downgrades during the remainder of 1990, the predictive power of 1989 screens for 1991 downgrades would be weakened because banks reclassified as problems in 1990 would not be in the set of 1991 downgrades. Apart from the change in the dependent variable and the sample selection criteria, the empirical tests were identical to those conducted on failures. Our dataset, however, now includes CAMEL downgrades from 1991 through 1997 and the corresponding lagged call report ratios.

The supervisory screens and the econometric model would each have done a good job predicting CAMEL downgrades. (Due to space constraints, the tables containing means and coefficient values can be found on the Research Department website of the Federal Reserve Bank of St. Louis <www.stls. frb.org/publications>.) For seven of the 14 individual screens, the differences between the means for the downgraded and non-downgraded banks were statistically significant with the hypothesized sign across all seven years. At the same time, the hypothesis that the econometric model had no explanatory value could be soundly rejected for all seven years. More specifically, across the seven equations, coefficients on five of the 14 independent variables were consistently significant with the hypothesized sign. In each approach, several

Figure 5

What is the Trade-Off Between False Negatives and False Positives in the Downgrade-Prediction Model Compared to the Individual Screens?



This figure shows the trade-off between the type-1 error rate (missed downgrades) and the type-2 error rate (missed nondowngrades). The type 1-error rate is the percentage of banks rated CAMEL-1 or -2 that were subsequently downgraded by supervisors but were not identified by the model (or screen). The type-2 error misidentified by the model (or screen) as a downgrade risk. A desirable early-warning system minimizes the increase in type-2 errors for any given decrease in type-1 errors. This graph shows that for any level of type-1 error rate tolerated by supervisors, the econometric model (in bold) leads to fewer type-2 errors than any individual screen. For clarity, only the four best screens are shown.

additional variables also were factors in downgrades during most of the years. Looking at the evidence from the screens and the model, the credit and liquidity risk variables appear most closely correlated with future downgrades.

Next, we directly compared the performance of the screens and the model using the resource benchmark and the error trade-off benchmark. Table 10 contains the comparison for CAMEL-1 banks that will be downgraded, while Table 11 contains the comparison for CAMEL-2 banks. Figure 5 illustrates the type-1 versus type-2 error trade-offs for 1991 downgrades. (Due to space constraints, the error trade-off figures for 1992 through 1997 are available on the Research Department website of the Federal Reserve Bank of St. Louis.)

By the resource savings benchmark, the model would have outperformed the screens for both 1- and 2-rated institutions. For CAMEL-1 banks, the econometric model posted lower exam percentages than any of the screens during four of the seven years. Moreover, as was the case for failure predictions, the rankings of the screens varied considerably from year to year. Finally, to catch one-half of the downgrades during the seven-year sample, supervisors would have had to examine only 16.9 percent of the CAMEL-1 banks using the econometric model. The lowest average for the supervisory screens—shared by the nonperforming-loan screen and the uncollected-interest-income screen—was 27.1 percent. For the CAMEL-2 banks, the results were even stronger: The econometric model outperformed the dominant screen every year. Again, the screen rankings varied considerably from year to year, and the dominant screen one year was not necessarily dominant the next. On average, supervisors could have caught one-half of

Table 10

How Does the Econometric Model Compare with the Single-Variable Screens as a Tool for Predicting CAMEL-1 Downgrades?

This table compares the econometric model and the individual screens as tools for predicting which CAMEL-1 banks will be downgraded to problem status. The columns show the percentage of banks that must be examined, using either the econometric model or a specific supervisory screen as a guide, to catch one-half of the downgrades the following year. In each year, the dominant screen comes close to the model's performance, but the dominant screen varies year to year. Moreover, on average, the model is clearly superior. The evidence suggests that the econometric model is the better tool for allocating on-site resources.

Method of Ranking Banks by Probability of Downgrade	**Among banks rated CAMEL 1, the percentage of banks that must be examined to include one-half of the banks that were downgraded in the following calendar year.**							
	Banks that were downgraded in:							
	1991	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**Mean Percentage**
Model	12%	11%	9%	23%	31%	23%	9%	16.9%
EQUITY	29	46	51	31	49	34	24	37.7
BAD-LOANS	31	17*	25	22	23	35	37	27.1
OREO	50	42	39	46	74	52	67	52.9
CONSUMER	54	54	59	36	48	27	17	42.1
INSIDER	46	45	33	17	43	56	42	40.3
OVERHEAD	35	24	22	14*	45	28	64	33.1
OCCUPANCY	34	31	39	31	55	48	39	39.6
ROA	41	37	30	32	24	21	92	39.6
UNCOLLECTED	37	42	30	32	21*	17*	11*	27.1
LIQUID	19*	29	17*	64	59	37	15	34.3
LARGE-TIME	30	25	24	16	39	23	35	27.4
CORE	34	30	35	42	49	38	53	40.1
SIZE	73	52	40	21	32	26	54	42.6
BHCRATIO	56	44	19	27	36	36	58	39.4

*Lowest number among single-variable screens that year.

EQUITY	Equity as a percentage of total assets.
BAD-LOANS	Nonperforming loans as a percentage of total loans.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.
CONSUMER	Consumer loans as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.
OCCUPANCY	Occupancy expense as a percentage of average assets.
ROA	Net income as a percentage of total assets.
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
SIZE	Natural logarithm of total assets, in thousands of dollars.
BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO ≡ 1.

Table 11

How Does the Econometric Model Compare with the Single-Variable Screens as a Tool for Predicting CAMEL-2 Downgrades?

This table compares the econometric model and the individual screens as tools for predicting which CAMEL-2 banks will be downgraded to problem status. The columns show the percentage of banks that must be examined, using either the econometric model or a specific supervisory screen as a guide, to catch one-half of the downgrades the following year. In each year, the dominant screen comes close to the model's performance, but the dominant screen varies year to year. Moreover, on average, the model is clearly superior. The evidence in this table suggests that the econometric model is the better tool for allocating on-site resources.

Method of Ranking Banks by Probability of Downgrade	**Among banks rated CAMEL 2, the percentage of banks that must be examined to include one-half of the banks that were downgraded in the following calendar year.**							
	Banks that were downgraded in:							
	1991	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**Mean Percentage**
Model	24%	18%	13%	19%	21%	15%	16%	18.0%
EQUITY	35	39	45	48	51	52	42	44.6
BAD-LOANS	37	35	26	38	33	35	30	33.4
OREO	45	44	39	44	39	43	50	43.4
CONSUMER	50	47	47	53	45	45	36	46.1
INSIDER	51	46	42	42	44	47	45	45.3
OVERHEAD	42	38	24*	39	35	42	43	37.6
OCCUPANCY	38	34	27	35	34	39	40	35.3
ROA	40	39	26	32*	37	41	32	35.3
UNCOLLECTED	47	40	44	43	32	29*	34	38.4
LIQUID	29*	25*	25	34	38	35	28*	30.6
LARGE-TIME	34	34	32	41	31*	36	35	34.7
CORE	37	39	38	45	36	41	40	39.4
SIZE	62	58	51	39	37	33	36	45.1
BHCRATIO	49	42	36	32	28	33	40	37.1

*Lowest number among single-variable screens that year.

EQUITY	Equity as a percentage of total assets.
BAD-LOANS	Nonperforming loans as a percentage of total loans.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.
CONSUMER	Consumer loans as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.
OCCUPANCY	Occupancy expense as a percentage of average assets.
ROA	Net income as a percentage of total assets.
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
SIZE	Natural logarithm of total assets, in thousands of dollars.
BHCRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO = 1.

the CAMEL-2 downgrades by examining only 18 percent of the CAMEL-2 banks. The lowest average for the supervisory screens—the liquid asset screen—was 30.6 percent.

Broadening the desired range of type-1 errors to other values besides 50 percent confirms the dominance of the econometric model. Figure 5 contains the 1991 error trade-off curves for the model and the individual supervisory screens, based on a pooled sample of CAMEL-1 and CAMEL-2 banks. For all ranges of type-1 errors, the econometric trade-off curve model lies below the curves for the individual supervisory screens. The curves for 1992 through 1997 reveal a similar pattern. In one year, the trade-off curves for the return-on-asset screen and the holding company control variable dipped below the econometric-model curve for a small range of values. Again, to beat the model with a single screen, supervisors would have had to guess correctly which screen would turn in a superior performance and the appropriate level of type-1 error. In summary, by the resource savings benchmark or the error trade-off benchmark, the econometric model clearly outperforms individual supervisory screens as a tool for predicting CAMEL downgrades.

RISK-SCOPING WITH ECONOMETRIC MODELS

To be useful in risk-focused supervision, an off-site surveillance tool must go beyond identifying institutions that are likely to develop safety-and-soundness problems and pinpoint the source of the developing problems. Armed with this information, supervisors then can determine the appropriate size and experience level of the examination team. Screens are attractive for risk-scoping because the specific financial ratios are designed to conform to the CAMELS framework. With some minor tweaking, however, supervisors also can use the output from econometric models to scope exams.

Table 12 demonstrates how the downgrade model can reveal the source of developing safety-and-soundness problems for a randomly selected bank in our sample. (Currently, the Board of Governors provides similar information to each Reserve Bank to support SEER. This information is contained in the Risk Profile Analysis Report.) The table presents the actual values of the regression variables for a sample bank with a sizable downgrade probability (column 2), along with average values for all the sample banks (column 3). Overall, this bank has an 11.31 percent chance of suffering a downgrade to problem status during the next 12 to 24 months, roughly three times the average downgrade probability for the sample. In addition, the actual values for the regression variables at this bank are weaker than the sample average in every case except uncollected revenue and core deposits.

Asset quality and management competence appear to be the principal sources of weakness at this bank. We isolated these sources of weakness by calculating the downgrade probability that we would obtain for each independent variable if it were set equal to the peer average and all the other independent variables remained at their actual values. These numbers appear in column four of the table. Column five of Table 12 then shows the difference between this hypothetical probability and the overall probability of a downgrade. A large positive number in column five indicates that the screen value makes a relatively large contribution to the downgrade probability. For example, OREO is the largest single contributor to risk for this bank: The ratio is 4.10 compared with a 0.23 average figure for the sample. If that OREO ratio were set equal to the average ratio for the sample, the overall downgrade probability for the bank would fall 5.07 percentage points, from 11.31 percent to 6.24 percent. Viewed another way, the high OREO ratio at this bank accounts for nearly one-half of its overall downgrade probability. The nonperforming loan ratio and the overhead expense ratio also contribute substantially to the downgrade probability. Supervisors risk-scoping this exam would assign more examiners to loan review and discussions with management.

Table 12

What Does the Econometric Model Tell Us About the Factors Contributing to a Downgrade?

This table shows how the econometric model can be used to isolate the variables most responsible for a likely downgrade. Column one lists the explanatory variables in the model. The second column gives the value of each variable for a sample bank with an 11.31 percent downgrade probability. The third column shows the average value of each variable among all the sample banks. Column four shows what the predicted downgrade probability would be if the selected variable were set equal to the sample peer average and all the other variables were kept at their actual values. The final column shows the difference between this hypothetical probability and the actual downgrade probability (11.31 percent). A large positive number in column 5 indicates that the given variable makes a significant contribution to the bank's risk. For example, the largest single contributor to risk at this bank is the OREO ratio (4.10 compared with the peer average of 0.23). In contrast, favorable core deposit and uncollected interest income rates, relative to peer, improve the bank's standing by 0.34 and 0.13 percentage points.

Random Bank from the Downgrade Regression Sample

Downgrade Probability: 11.31 percent

Regression Variable (1)	Most recent value of bank's ratio (in %) (2)	Average sample value for variable (in %) (3)	Downgrade probability with variable set to sample average (4)	Difference from bank's actual downgrade probability (5)
EQUITY	8.55	9.94	10.64	0.67
BAD-LOANS	1.75	0.75	9.32	1.99
OREO	4.10	0.23	6.24	5.07
CONSUMER	10.93	9.00	11.11	0.20
INSIDER	0.32	1.27	11.25	0.06
OVERHEAD	49.73	35.49	8.05	3.26
OCCUPANCY	0.56	0.39	11.23	0.08
ROA	0.75	1.10	10.61	0.70
UNCOLLECTED	0.57	0.59	11.44	-0.13
LIQUID	28.62	38.55	8.60	2.71
LARGE-TIME	9.05	8.50	11.00	0.32
CORE	80.47	77.67	11.65	-0.34
SIZE	10.15	11.27	8.60	2.71
BHCRATIO	0.99	0.65	6.94	4.37

EQUITY	Equity as a percentage of total assets.
BAD-LOANS	Nonperforming loans as a percentage of total loans.
OREO	Other real estate owned (real estate other than bank premises) as a percentage of total loans.
CONSUMER	Consumer loans as a percentage of total assets.
INSIDER	The value of loans to insiders (officers and directors of the bank) as a percentage of total assets.
OVERHEAD	Noninterest expense as a percentage of total revenue.
OCCUPANCY	Occupancy expense as a percentage of average assets.
ROA	Net income as a percentage of total assets.
UNCOLLECTED	Interest accrued as revenue but not collected as a percentage of total loans.
LIQUID	Liquid assets (sum of cash, securities, federal funds sold, and reverse repurchase agreements) as a percentage of total assets.
LARGE-TIME	Large denomination time deposit liabilities as a percentage of total assets.
CORE	Core deposits (transactions, savings and small time deposits) as a percentage of total assets.
SIZE	Natural logarithm of total assets, in thousands of dollars.
BCHRATIO	The ratio of each bank's total assets to the total assets of its holding company. Banks without holding companies have BHCRATIO $\equiv$ 1.

Supervisors also can use information provided by the control variables in exam planning. For example, both of the control variables, SIZE and BHCRATIO, make significant contributions to the downgrade probability for this bank. Recall that all other things being equal, both large banks and small banks that are members of large holding companies are less likely to encounter safety-and-soundness problems. In the example, the large values for SIZE and BHCRATIO imply that the management and loan quality problems demand more examiner attention because this bank is not a large, well-diversified institution and cannot rely on a parent company as a source of strength.

SUPERVISORY SCREENS AND ECONOMETRIC MODELS AS COMPLEMENTS

Our statistical evidence does not, however, imply that the screens currently employed by supervisors add no value in off-site surveillance. First, as noted earlier, our screens and models are not the actual screens and models currently used by the surveillance community. Second, our tests are biased in favor of econometric models. Finally, our tests do not measure the potential value of screens in a rapidly changing banking environment.

Our comparisons contain several biases against screens. As noted, in practice, supervisory screens typically are weighted averages of financial ratios. Our representative screens, in contrast, are single-variable screens. In addition, supervisors modify their screens regularly based on feedback from field examiners. Our approach implicitly assumes that supervisors used the same single-variable screens throughout the entire sample. A better approach would rely on a time series of the actual multiple-variable screens used, but unfortunately, no such series exists. Finally, it is possible that successful use of the screens weakened their predictive power. Suppose supervisors ignored the output of econometric models, relying exclusively on screens to identify the banks that were most vulnerable to failure or ratings downgrades. Suppose further that supervisors then intervened to prevent these failures or downgrades. From a statistical standpoint, the more successful the use of screens, the weaker their predictive power would be.

Our simple statistical horse races also fail to capture the value that supervisory screens can add in a dynamic banking environment. The agricultural bank problems of the 1980s demonstrate this value. Before the 1980s, the agricultural-loan-to-total-loan ratio would not have correlated positively with bank failures. That changed with the sharp declines in farm income and prices after 1981 (Belongia and Gilbert, 1990). By 1982, examination reports revealed that banks top-heavy with agricultural loans were significant failure risks. Failures did not rise sharply, however, until the second half of 1984, after declines in farm income and prices had absorbed the net worth of farmers and their banks (Kliesen and Gilbert, 1996). Because of the need to re-estimate coefficients and conduct new performance tests, new econometric models would not have been available to warn of agricultural bank vulnerability until 1985 or perhaps 1986. In short, supervisors could have developed screens for predicting agricultural bank failures long before econometric models would have signaled a rise in failure probabilities.

CONCLUSION

Off-site surveillance involves using accounting data to identify banks likely to develop safety-and-soundness problems. Early intervention, based on this information, can limit losses to the deposit insurance fund and the real economy. Supervisors rely heavily on two tools to flag developing problems: supervisory screens and econometric models. We used data from the 1980s and 1990s to compare, once again, the performance of these two approaches to off-site surveillance. As in earlier comparisons, the econometric models outperformed the supervisory screens. These results do not, however, suggest that screens should be dropped from the surveillance toolbox.

When abrupt changes in the causes of bank failures and CAMEL downgrades occur, supervisors can use their first-hand knowledge to modify screens long before models can be revised to reflect the new conditions. In short, the flexibility of supervisory screens makes them an important complement for econometric models in off-site surveillance.

REFERENCES

Barancik, Scott. "Few Banks and Thrifts Pay FDIC Premiums," *American Banker*, April 21, 1999, page 26.

Belongia, Michael T., and R. Alton Gilbert. "The Effects of Management Decisions on Agricultural Bank Failures," *American Journal of Agricultural Economics* (November 1990), pp. 901-10.

Bernanke, Ben S. "The Macroeconomics of the Great Depression: A Comparative Approach," *Journal of Money, Credit, and Banking* (February 1995), pp. 1-28.

________. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression," *American Economic Review* (June 1983), pp. 257-76.

________ and Harold James. "The Gold Standard, Deflation, and Financial Crisis in the Great Depression: An International Comparison," in *Financial Markets and Financial Crises*, R. Glenn Hubbard, ed., University of Chicago Press, 1991, pp. 33-68.

Board of Governors of the Federal Reserve System. "Risk-Focused Safety and Soundness Examinations and Inspections," SR 96-14, May 24, 1996.

Cannella, Albert A. Jr., Douglas R. Fraser, and D. Scott Lee. "Firm Failure and Managerial Labor Markets: Evidence from Texas Banking," *Journal of Financial Economics* (June 1995), pp. 185-210.

Cole, Rebel A., and Jeffrey W. Gunther. "A Comparison of On- and Off-Site Monitoring Systems," *Journal of Financial Services Research* (April 1998), pp. 103-17.

________, Barbara G. Cornyn, and Jeffrey W. Gunther. "FIMS: A New Monitoring System for Banking Institutions," *Federal Reserve Bulletin* (January 1995), pp. 1-15.

Flannery, Mark J. "Deposit Insurance Creates a Need for Bank Regulation," Federal Reserve Bank of Philadelphia *Business Review* (January/February 1982), pp. 17-27.

________ and Joel F. Houston. "The Value of a Government Monitor for U.S. Banking Firms," *Journal of Money, Credit, and Banking* (February 1999), pp. 14-34.

Friedman, Milton, and Anna Jacobson Schwartz. *A Monetary History of the United States: 1867-1960*, Princeton University Press, 1963.

Gilbert, R. Alton. "Do Bank Holding Companies Act as Sources of Strength for Their Bank Subsidiaries," this *Review* (January/February 1991), pp. 3-18.

________ and Gerald P. Dwyer, Jr. "Bank Runs and Private Remedies," this *Review* (May/June 1989), pp. 43-61.

________ and Levis A. Kochin. "Local Economic Effects of Bank Failures," *Journal of Financial Services Research* (December 1989), pp. 333-45.

________ and Mark D. Vaughan. "Does the Publication of Supervisory Enforcement Actions Add to Market Discipline?" *Research in Financial Services Public and Private Policy* (1998), pp. 259-80.

Hall, John R., Thomas B. King, Andrew P. Meyer, and Mark D. Vaughan. "Do Certificate of Deposit Holders and Supervisors View Bank Risk Similarly? A Comparison of the Factors Affecting CD Yields and CAMEL Composites," Supervisory Policy Analysis Working Paper, October 1999.

Kliesen, Kevin L., and R. Alton Gilbert. "Are Some Agricultural Banks Too Agricultural?" this *Review* (January/February 1996), pp. 23-35.

Putnam, Barron H. "Early Warning Systems and Financial Analysis in Bank Monitoring: Concepts of Financial Monitoring," Federal Reserve Bank of Atlanta *Economic Review* (November 1983), pp. 6-13.

Reidhill, Jack, and John O'Keefe. "Off-Site Surveillance Systems," in *History of the Eighties: Lessons for the Future, Volume I*. Washington: Federal Deposit Insurance Corporation, 1997.

White, Lawrence J. *The S&L Debacle: Public Policy Lessons for Bank and Thrift Regulation*, Oxford University Press, 1991.