Could a CAMELS Downgrade Model Improve Off-Site Surveillance?

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

The cornerstone of bank supervision is a regular schedule of thorough, on-site examinations. Under rules set forth in the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA), most U.S. banks must submit to a full-scope federal or state examination every 12 months; small, well-capitalized banks must be examined every 18 months. These examinations focus on six components of bank safety and soundness: capital protection (C), asset quality (A), management competence (M), earnings strength (E), liquidity risk exposure (L), and market risk sensitivity (S). At the close of each exam, examiners award a grade of one (best) through five (worst) to each component. Supervisors then draw on these six component ratings to assign a composite CAMELS rating, which is also expressed on a scale of one through five. (See the insert for a detailed description of the composite ratings.) In general, banks with composite ratings of one or two are considered safe and sound, whereas banks with ratings of three, four, or five are considered unsatisfactory. As of March 31, 2000, nearly 94 percent of U.S. banks posted composite CAMELS ratings of one or two.

Bank supervisors support on-site examinations with off-site surveillance. Off-site surveillance uses quarterly financial data and anecdotal evidence to schedule and plan on-site exams. Although on-site examination is the most effective tool for spotting safety-and-soundness problems, it is costly and burdensome. On-site examination is costly to supervisors because of the examiner resources required and burdensome to bankers because of the intrusion into daily operations. Off-site surveillance reduces the need for unscheduled exams. Off-site surveillance also helps supervisors plan exams by highlighting risk exposures at specific institutions.[1] For example, if pre-exam surveillance reports indicate that a bank has significant exposure to interest rate fluctuations, then supervisors will add interest-rate-risk specialists to the exam team.

The two most common surveillance tools are supervisory screens and econometric models. Supervisory screens are combinations of financial ratios, derived from quarterly bank balance sheets and income statements, that have given warning in the past about the development of safety-and-soundness problems. Supervisors draw on their experience to weigh the information content of these ratios. Econometric models also combine information from bank financial ratios. These models rely on statistical tests rather than human judgment to combine ratios, boiling the information from financial statements down to an index number that summarizes bank condition. In past comparisons, econometric models have outperformed supervisory screens as early warning tools (Gilbert, Meyer, and Vaughan, 1999; Cole, Cornyn, and Gunther 1995). Nonetheless, screens still play an important role in off-site surveillance. Supervisors can add screens quickly to monitor emerging sources of risk; econometric models can be modified only after new risks have produced a sufficient number of safety-and-soundness problems to allow re-specification and out-of-sample testing.

At the Federal Reserve, the off-site surveillance toolbox includes two distinct econometric models that are collectively known as SEER—the System for Estimating Examination Ratings. One model, the SEER risk rank model, uses the latest quarterly financial data to estimate the probability that each Fed-supervised bank will fail within the next two years. The other model, the SEER rating model, uses the latest financial data to produce a "shadow" CAMELS rating for each supervised institution. That is, the model estimates the CAMELS rating that examiners would have assigned had the bank been examined using the most recent set of financial

R. Alton Gilbert is a vice president and banking advisor, Andrew P. Meyer is an economist, and Mark D. Vaughan is a supervisory policy officer and economist at the Federal Reserve Bank of St. Louis. The authors thank economists Robert Avery, Jeffrey Gunther, James Harvey, Tom King, Jose Lopez, Don Morgan, Chris Neely, and David Wheelock; bank supervisors Carl Anderson, Kevin Bertsch, and Kim Nelson; and seminar participants at the meetings of the SEER Technical Working Group and the Western Economics Association for their comments. Judith Hazen provided research assistance.

© 2002, The Federal Reserve Bank of St. Louis.

[1] See Board of Governors (1996) for a description of risk-focused examination.

WHAT ARE CAMELS RATINGS?

CAMELS composite rating	Description
Safe and sound	
1	Financial institutions with a composite one rating are sound in every respect and generally have individual component ratings of one or two.
2	Financial institutions with a composite two rating are fundamentally sound. In general, a two-rated institution will have no individual component ratings weaker than three.
Unsatisfactory	
3	Financial institutions with a composite three rating exhibit some degree of supervisory concern in one or more of the component areas.
4	Financial institutions with a composite four 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 five rating generally exhibit extremely unsafe and unsound practices or conditions. Institutions in this group pose a significant risk to the deposit insurance fund and their failure is highly probable.

NOTE: CAMELS is an acronym for six components of bank safety and soundness: capital protection (C), asset quality (A), management competence (M), earnings strength (E), liquidity risk exposure (L), and market risk sensitivity (S). Examiners assign a grade of one (best) through five (worst) to each component. They also use these six scores to award a composite rating, also expressed on a one-through-five scale. As a rule, banks with composite ratings of one or two are considered safe and sound while banks with ratings of three, four, or five are considered unsatisfactory.

SOURCE: *Federal Reserve Commercial Bank Examination Manual*.

statements and the previous CAMELS rating. Every quarter, analysts in the surveillance section at the Board of Governors feed the latest call report data into these models and forward the results to the 12 Reserve Banks. The Federal Deposit Insurance Corporation (FDIC) and the Office of the Comptroller of the Currency (OCC) also use statistical models in the off-site surveillance of the banks they supervise.[2]

The Federal Reserve employs two distinct models in off-site surveillance to accomplish two distinct objectives. One objective, embodied in the SEER risk rank model, is to identify a core set of financial variables that consistently foreshadows failure. Due to the paucity of bank failures since the early 1990s, the coefficients of the risk rank model were last estimated on data ending in 1991. A fixed-coefficient model, such as the risk rank model, allows surveillance analysts to gauge how much of any change in failure probabilities over time is due to changes in the values of these core financial variables. The second objective is to allow for changes over time in the relationship between financial performance today and bank condition tomorrow. The second half of the SEER framework, the SEER rating model, meets this objective by allowing analysts to reestimate the relationship quarterly, adjusting for any changes in the factors that produce safety-and-soundness problems.

Identifying banks with composite CAMELS ratings of one or two that are at risk of downgrade to a composite rating of three, four, or five is another important objective of the SEER framework, although this relationship is not directly estimated in either SEER model. Supervisors view a downgrade from safe-and-sound condition to unsatisfactory condition as serious because three-, four-, and five-rated banks are much more likely to fail. For example, Curry (1997) found that 74 percent of the banks that failed from 1980 through 1994 held three, four, or five composite CAMELS ratings two years prior to failure. Table 1 contains an update of Curry's

[2] See Reidhill and O'Keefe (1997) for a history of the off-site surveillance systems at the Federal Reserve, FDIC, and OCC.

figures, indicating that 53 of the 58 banks (91 percent) that failed in the years 1993 through 1998 held unsatisfactory ratings at least one year prior to failure. Because of their high failure risk, banks in unsatisfactory condition receive constant supervisory attention. An econometric model designed to flag safe-and-sound banks at risk of downgrade could help allocate supervisory resources not already devoted to troubled institutions. Such a model might also yield even earlier warning of emerging financial distress—warning that could reduce the likelihood of eventual failure by allowing earlier supervisory intervention. Although SEER failure probabilities and "shadow" CAMELS ratings for one- and two-rated banks certainly provide clues about downgrade risks, these index numbers are not the product of a model estimated specifically to flag downgrade candidates.

Even so, the SEER models may produce "watch lists" of one- and two-rated banks that differ little from watch lists produced by a downgrade-prediction model. The CAMELS downgrade model, the SEER risk rank model, and the SEER rating model generate ordinal rankings of banks based on risk. The models differ by the specific measure of overall risk—the risk of failure (SEER risk rank model), the risk of receiving a poor current CAMELS rating (SEER rating model), or the risk of moving from satisfactory to unsatisfactory condition in the near future (downgrade model). The models also differ by the sample of banks used for estimation—the SEER models are estimated on all commercial banks, whereas a downgrade model is estimated only on one- and two-rated institutions. But if the financial factors that explain CAMELS downgrades differ little from the financial factors that explain failures or CAMELS ratings, then all three models will produce similar risk rankings and, hence, similar watch lists of one- and two-rated banks. Only formal empirical tests can determine the potential contribution of a downgrade-prediction model to off-site surveillance at the Federal Reserve.

To answer our title question—could a CAMELS downgrade model improve off-site surveillance—we compare the out-of-sample performance of a downgrade-prediction model and the SEER models using 1990s data. We find only slight differences in the ability of the three models to spot emerging financial distress among safe-and-sound banks. Specifically, in out-of-sample tests for 1992 through 1998, the watch lists produced by the downgrade-prediction model outperform the watch lists produced by the SEER models by only a small margin. We conclude that, in relatively tranquil banking environments like the 1990s, a downgrade model adds little value in off-site surveillance. We caution, however, that a downgrade-prediction model might prove useful in more turbulent banking times.

THE RESEARCH STRATEGY

Our downgrade-prediction model is a probit regression that uses bank financial data to estimate the probability each sample bank will tumble from a composite CAMELS rating of one or two to a composite CAMELS rating of three, four, or five. Specifically, the dependent variable takes a value of one for any bank whose CAMELS rating falls from satisfactory to unsatisfactory in the 24 months following the quarter of the financial data; the dependent variable is zero if the bank is examined but not downgraded in the 24-month window. Although bank failure declined dramatically in the 1990s, CAMELS downgrades were still common, thereby allowing frequent reestimation of the model. (See Table 2 for data on CAMELS downgrades in the 1990s.) The SEER risk rank model is also a probit model, using financial data to estimate the probability that a Fed-supervised bank will fail or see its tangible capital fall below 2 percent of total assets in the next 24 months. The SEER rating model is a multinomial logit regression that uses financial data to estimate a "shadow" CAMELS rating—the composite rating that examiners would have awarded had the bank been examined that quarter. A multinomial logit differs from a standard logit by predicting a range of discrete values (in this case CAMELS composite ratings, which range from one to five) rather than two discrete values (failure/no failure or downgrade/no downgrade).

The explanatory variables for the downgrade-prediction model include a set of financial performance ratios and a bank size variable that all appear in the SEER risk rank model, as well as two additional CAMELS-related variables. Table 3 describes the explanatory variables and the expected relationship between each variable and the likelihood of a future downgrade. The financial performance ratios capture the impact of leverage risk, credit risk, and liquidity risk—three risks that have consistently produced financial distress in commercial banks (Putnam, 1983; Cole and Gunther, 1998). The bank size and CAMELS-related variables capture the impact of other factors that may affect downgrade risk.

The downgrade-prediction model captures leverage risk with total equity minus goodwill as a

Table 1

How Often Did Unsatisfactory Banks Fail in the 1990s?

Year of failure	CAMELS rating at least one year prior to failure	Number of banks in each CAMELS cohort	Number of failures in each CAMELS cohort	Percentage failed in each CAMELS cohort	Percentage of all failures with CAMELS ratings of 3, 4, or 5 one year in advance
1993	1	2,396	1	0.04	91.7
	2	6,549	2	0.03	
	3	**1,877**	**4**	**0.21**	
	4	**762**	**14**	**1.84**	
	5	**218**	**15**	**6.88**	
1994	1	2,508	0	0.00	90.9
	2	6,693	1	0.01	
	3	**1,578**	**0**	**0.00**	
	4	**562**	**5**	**0.89**	
	5	**124**	**5**	**4.03**	
1995	1	3,299	0	0.00	100
	2	6,469	0	0.00	
	3	**916**	**0**	**0.00**	
	4	**303**	**2**	**0.66**	
	5	**56**	**3**	**5.36**	
1996	1	3,759	0	0.00	75.0
	2	5,995	1	0.02	
	3	**587**	**1**	**0.17**	
	4	**158**	**1**	**0.63**	
	5	**39**	**1**	**2.56**	
1997	1	4,041	0	0.00	100
	2	5,472	0	0.00	
	3	**400**	**0**	**0.00**	
	4	**91**	**1**	**1.10**	
	5	**23**	**0**	**0.00**	
1998	1	4,328	0	0.00	100
	2	4,941	0	0.00	
	3	**329**	**0**	**0.00**	
	4	**57**	**1**	**1.75**	
	5	**15**	**0**	**0.00**	

NOTE: This Table shows that banks with composite CAMELS ratings of one or two were less likely to fail in the 1990s than were banks with composite ratings of three, four, or five. The number of failed banks that were classified as unsatisfactory banks (CAMELS three, four, or five composite ratings) at least one year prior to failure are shown in bold. Supervisors recognized that these banks were significant failure risks and, therefore, monitored them closely. Because supervisors do not monitor CAMELS one- and two-rated banks as closely, they are interested in a tool that can identify which of these institutions is most likely to encounter financial distress.

Table 2

How Common Were CAMELS Downgrades in the 1990s?

Year of downgrade	CAMELS rating at beginning of year	Number of banks	Number of banks downgraded to unsatisfactory status	Percentage of banks downgraded to unsatisfactory status	Total number of downgrades to unsatisfactory status
1990	1	2,182	38	1.74	728
	2	5,572	690	12.38	
1991	1	2,189	34	1.55	698
	2	5,475	664	12.13	
1992	1	1,959	22	1.12	424
	2	5,275	402	7.62	
1993	1	2,289	7	0.31	182
	2	5,976	175	2.93	
1994	1	2,910	9	0.31	162
	2	5,717	153	2.68	
1995	1	3,091	8	0.26	102
	2	4,885	94	1.92	
1996	1	3,260	10	0.31	126
	2	4,487	116	2.59	
1997	1	3,223	7	0.22	123
	2	3,719	116	3.12	
1998	1	3,006	19	0.63	153
	2	3,090	134	4.34	

NOTE: This Table demonstrates that downgrades from safe-and-sound to unsatisfactory status were common in the 1990s, thereby making it possible to reestimate a downgrade-prediction model on a yearly basis. Specifically, the far right column shows the number of sample banks rated as safe and sound (CAMELS one or two) at each year-end that were downgraded to unsatisfactory status (CAMELS three, four, or five) within the following year. Note that two-rated banks were much more likely to slip into unsatisfactory status than one-rated banks. Note also that the percentage of banks suffering downgrades to unsatisfactory status fell as overall banking performance improved in the mid-1990s, but the trend reversed in the late 1990s.

percentage of total assets (NET WORTH) and net income as a percentage of total assets (or, return on assets [ROA]). Leverage risk is the risk that losses will exceed capital, rendering a bank insolvent. We expect higher levels of capital (lower leverage risk) to reduce the likelihood of CAMELS downgrades. We include ROA in the leverage risk category because retained earnings are an important source of additional capital for many banks and because higher earnings provide a greater cushion for withstanding adverse economic shocks (Berger, 1995). We expect that higher earnings reduce the risk of a future downgrade.

The downgrade-prediction model captures credit risk with the ratio of loans 30 to 89 days past due to total assets (PAST-DUE 30), the ratio of loans over 89 days past due to total assets (PAST-DUE 90), the ratio of loans in nonaccrual status to total assets (NONACCRUING), the ratio of other real estate owned to total assets (OREO), the ratio of commercial and

Table 3

What Factors Help Predict Downgrades to Unsatisfactory Condition (CAMELS Three, Four, or Five)?

Independent variables (risk proxies)	Symbol	Hypothesized relationship
Leverage risk		
Total net worth (equity capital minus goodwill) as a percentage of total assets	NET WORTH	–
Net income as a percentage of average assets (return on average assets)	ROA	–
Credit risk		
Loans past due 30-89 days as a percentage of total assets	PAST-DUE 30	+
Loans past due 90+ days as a percentage of total assets	PAST-DUE 90	+
Nonaccrual loans as a percentage of total assets	NONACCRUING	+
Other real estate owned as a percentage of total assets	OREO	+
Commercial and industrial loans as a percentage of total assets	COMMERCIAL LOANS	+
Residential real estate loans as a percentage of total assets	RESIDENTIAL LOANS	–
Liquidity risk		
Book value of securities as a percentage of total assets	SECURITIES	–
Deposits >$100M (jumbo CDs) as a percentage of total assets	LARGE TIME DEPOSITS	+
Non-financial variables		
Natural logarithm of total assets, in thousands of dollars	SIZE	?
Dummy variable equal to 1 if bank has a CAMELS rating of 2	CAMELS-2	+
Dummy variable equal to 1 if the bank's management rating is worse than its composite CAMELS rating	BAD MANAGE	+

NOTE: This Table lists the independent variables used in the downgrade-prediction model. The signs indicate the hypothesized relationship between each variable and the likelihood of a downgrade from satisfactory status (a CAMELS one or two composite rating) to unsatisfactory status (a CAMELS three, four, or five rating). For example, the negative sign for the net worth ratio indicates that, other things equal, higher net worth today reduces the likelihood of a downgrade to unsatisfactory status tomorrow.

industrial loans to total assets (COMMERCIAL LOANS), and the ratio of residential real estate loans to total assets (RESIDENTIAL LOANS). Credit risk is the risk that borrowers will fail to make promised interest and principal payments. The model contains six measures of credit risk because this risk was the driving force behind bank failures in the late 1980s and early 1990s (Hanc, 1997). We include the past-due and nonaccruing loan ratios because banks charge off higher percentages of these loans than loans whose payments are current.[3] We include other real estate owned, which consists primarily of collateral seized after loan defaults, because a high OREO ratio often signals poor credit risk management—either because a bank has had to foreclose on a large number of loans or because it has had trouble disposing of seized collateral. PAST-DUE 30, PAST-DUE 90, NONACCRUING, and OREO are backward-looking because they register asset quality problems that have already emerged (Morgan and Stiroh, 2001). To give the model a forward-looking dimension, we add the commercial-and-industrial-loan ratio because, historically, the charge-off rate for these loans has been higher than for other types of loans. We also employ the residential real estate ratio because, historically, losses on these loans have been relatively low. With the exception of the residential loan ratio, we expect a positive relationship between the credit risk measures and downgrade probability.

The downgrade-prediction model captures liquidity risk with investment securities as a per-

[3] In bank accounting, loans are classified as either 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.

centage of total assets (SECURITIES) and jumbo certificates of deposit (CDs) as a percentage of total assets (LARGE TIME DEPOSITS). Liquidity risk is the risk that a bank will be unable to fund loan commitments or meet withdrawal demands at a reasonable cost. A larger stock of liquid assets—such as investment securities—indicates a greater ability to meet unexpected liquidity needs and should, therefore, translate into a lower downgrade probability. Liquidity risk also depends on a bank's reliance on non-core funding. Core funding—which includes checking accounts, savings accounts, and small time deposits—is relatively insensitive to the difference between the interest rate paid by the bank and the market rate. Non-core funding—which includes jumbo CDs—can be quite sensitive to interest rate differentials. All other things equal, greater reliance on jumbo CDs implies a greater likelihood of a funding runoff or an interest expense shock and, hence, a future CAMELS downgrade.

The downgrade-prediction model also includes variables that capture the impact of asset size, bank heterogeneity, and management competence on downgrade risk. We add the natural logarithm of total assets (SIZE) because large banks can reduce risk by diversifying across product lines and geographic regions. As Demsetz and Strahan (1997) have noted, however, geographic diversification relaxes a constraint, enabling bankers to assume more risk, so we make no prediction about the relationship between size and downgrade probability. We include a dummy variable equal to one if a bank's composite CAMELS rating is two; we do this because two-rated banks tumble into unsatisfactory status more often than one-rated banks. (See Table 2 for data on the downgrade rates for one- and two-rated institutions.) Finally we employ a dummy variable (BAD MANAGE) equal to one if the management component of the CAMELS rating is higher (weaker) than the composite rating. In these cases, examiners have registered concerns about the quality of bank management, even though these problems have yet to produce financial consequences.

After estimating the downgrade-prediction model, we use all three models to produce rank orderings, or "watch lists," of one- and two-rated banks. With the downgrade model, the list ranks safe-and-sound banks from the highest probability of tumbling into unsatisfactory condition to the lowest. With the SEER risk rank model, the list ranks safe-and-sound banks from the highest probability of failing to the lowest. With the SEER rating model, the list ranks safe-and-sound banks from the highest (weakest) shadow CAMELS rating to the lowest. Although each model produces a different index number, they all may produce similar ordinal rankings. Supervisors could use the SEER framework to monitor safe-and-sound banks by focusing on the riskiest one- or two-rated banks as identified by either the rating or failure-prediction model. Again, only a formal test of out-of-sample performance can gauge the value added by a customized downgrade-prediction model. Out-of-sample tests—which use an evaluation period subsequent to the estimation period—are crucial because supervisors use econometric models this way in practice.

We compare out-of-sample performance of the watch lists by examining the type-one and type-two error rates associated with each list. Type-one errors are sometimes called false negatives; type-two errors are false positives. Each type of error is costly to supervisors. A missed downgrade—a type-one error—is costly because an accurate downgrade prediction gives supervisors more warning about emerging financial distress, and early intervention reduces the likelihood of failure. A type-two error occurs when a predicted downgrade does not materialize. An over-predicted downgrade is costly because it wastes scarce supervisory resources on a healthy bank. Type-two errors also impose unnecessary costs on healthy banks because on-site examinations disrupt day-to-day operations.

Following Cole, Cornyn, and Gunther (1995), we generate power curves for the three watch lists that indicate the minimum achievable type-one error rates for any desired type-two error rate. (These curves are illustrated in Figures 1 and 2.) Power curves allow comparison of each list's ability to reduce false negatives and false positives simultaneously. A more theoretically appealing approach would minimize a loss function that places an explicit weight on the benefits of early warning about financial distress and the costs of wasted examination resources and unnecessary disruption of bank activities. The relative performance of the watch lists could then be assessed for the optimal type-one (or type-two) error rate. Unfortunately, the data necessary to pursue such an approach are unavailable. Without concrete data about supervisor loss functions, we opt for power curves that make no assumptions about the weights that should be placed on type-one and type-two errors. This approach also allows supervisors to use our results to compare model performance over any desired range of error rates.

For example, the SEER risk rank power curve shows the type-one and type-two error rates when an ordinal ranking based on failure probability is interpreted as a rank ordering of downgrade risk. We trace out the curve by starting with the assumption that no one- or two-rated bank is a downgrade risk. This assumption implies that all subsequent downgrades are surprises, making the type-one error rate 100 percent. In this case, the type-two error rate is zero because no banks are incorrectly classified as downgrade risks. We obtain the next point by selecting the one- or two-rated bank with the highest failure probability. If the selected bank suffers a subsequent downgrade, then the type-one error rate for the SEER risk rank watch list decreases slightly. The type-two error rate remains at zero because, again, no institutions are incorrectly classified as downgrade risks. If the selected bank does not suffer a downgrade, then the type-one error rate remains at 100 percent and the type-two error rate increases slightly. By selecting banks in order of their failure probability and recalculating type-one and type-two error rates, we can trace out a power curve. At the lower right extreme of the curve, the entire failure probability rank ordering is considered at risk of a downgrade. At this extreme, the SEER risk rank watch list posts a type-one error rate equal to zero percent and a type-two error rate equal to 100 percent.

The area under the power curves provides a basis for comparing the out-of-sample performance of each watch list. A smaller area implies a lower overall type-one and type-two error rate and a more accurate model. We express the area for each watch list as a percentage of the total area in the box. A useful benchmark is the case in which downgrade risks are selected at random. Random selection of one- and two-rated banks, over a large number of trials, produces power curves with an average slope of –1. The area under a "random" watch list power curve equals, on average, 50 percent of the area of the entire box.

THE DATA

We exploit two data sources for our analysis—the Federal Financial Institutions Examination Council (FFIEC) and the National Information Center of the Federal Reserve System (NIC). We use income and balance sheet data from the Reports of Condition and Income (the call reports), which are collected under the auspices of the FFIEC. The FFIEC requires all commercial banks to submit quarterly call reports to their principal supervisors; most call report items are available to the public. We rely on CAMELS composite and management ratings from the NIC database. This database is available to examiners and analysts in the banking supervision function of the Federal Reserve System but not to the public. We also draw on the NIC database for the SEER failure probabilities and "shadow" CAMELS ratings.

To ensure an unbiased comparison of the models, we exclude any bank with an operating history under five years from the estimation sample for the downgrade-prediction model. The financial ratios of these start-up, or *de novo*, banks often take extreme values that do not signal safety-and-soundness problems (DeYoung, 1999). For example, *de novos* often lose money in their early years, so their earnings ratios are poor. These extreme values distort model coefficients and could compromise the relative performance of the downgrade-prediction model. Another reason for excluding *de novos* is that supervisors already monitor these banks closely. The Federal Reserve conducts a full-scope on-site examination every six months for a newly chartered state-member bank.[4] Full-scope exams continue on this schedule until the *de novo* earns a one or two composite CAMELS rating for two consecutive exams.

As an additional safeguard, we use a timing convention for estimating the downgrade-prediction model that corresponds to the timing convention used to estimate the SEER risk rank model. Specifically, we estimate the downgrade model six times—each time using financial data for one- and two-rated institutions in the fourth quarter of year t and downgrade status (1 = downgrade, 0 = no downgrade) in years $t + 1$ and $t + 2$. For example, to produce the first downgrade equation (reported as the "1990-91" equation in Table 4), we use a sample of banks rated CAMELS one or two as of December 31, 1989. We then regress downgrade status during 1990 and 1991 on fourth quarter 1989 data. A bank that is examined but maintains a one or two rating during the entire two-year period is classified as "no downgrade." A bank that is examined and suffers a downgrade to a three, four, or five composite rating anytime in the two-year period is classified as "downgrade."

Finally, when comparing out-of-sample performance of the models, we note biases that result from

[4] The Federal Reserve supervises bank holding companies and state-chartered banks that belong to the Federal Reserve System. The FDIC supervises state-chartered banks that do not belong to the Federal Reserve System. The OCC supervises banks chartered by the federal government.

using revised call report data rather than originally submitted call report data. Supervisors sometimes require banks to revise their call report data after an on-site examination. Indeed, some economists have argued that this auditing function is the principal value of examinations (Berger and Davies, 1998; Flannery and Houston, 1999). Revisions of fourth quarter data tend to be particularly large because banks strive to make their year-end financial reports look as healthy as possible (Allen and Saunders, 1992). Gunther and Moore (2000) have found that early warning models estimated on revised data outperform models estimated on originally submitted data. Because of this evidence, estimation and simulation of an early warning model with the original data, rather than the revised data, would provide a more appropriate test of the value of a model for surveillance. The original data, however, are not available for all banks and all periods. Hence, we estimate the downgrade model on revised rather than original call report data. The coefficients of the SEER risk rank model were estimated using revised call report data, and we apply these coefficients to revised call report data to generate failure probability rankings. Because the SEER risk rank model and the downgrade-prediction model are estimated with revised data, our performance comparisons do not favor either model ex ante. But because the SEER rating model was estimated on originally submitted call report data, out-of-sample comparisons favor the downgrade-prediction model over the rating model. Data limitations do not allow us to correct for this bias, so we bear it in mind as we interpret the power curve evidence for these two models.

IN-SAMPLE FIT OF THE DOWNGRADE-PREDICTION MODEL

As noted, we estimate the downgrade-prediction model six times—first regressing downgrade status in 1990 and 1991 on fourth quarter 1989 financial data, then regressing downgrade status in 1991 and 1992 on fourth quarter 1990 data, and so on, up through regressing downgrade status in 1995 and 1996 on fourth quarter 1994 data. The results of these regressions appear in Table 4.

Overall, the downgrade-prediction model fits the data relatively well in-sample. For each of the six regressions, the log-likelihood test statistic allows rejection of the hypothesis that all model coefficients equal zero at the 1 percent level of significance. The pseudo-R^2, which indicates the approximate proportion of the variance of downgrade/no downgrade status explained by the model, ranges from a low of 14.9 percent for the 1993-94 equation to a high of 22.4 percent for the 1991-92 equation. These pseudo-R^2 numbers may seem low, particularly when viewed against the figures for failure-prediction models—the pseudo-R^2 for the SEER risk rank model is 63.2 percent—but CAMELS downgrades are less severe than outright failures and, therefore, much more difficult to forecast. In this light, the pseudo-R^2 figures look more respectable. The estimated coefficients on eight explanatory variables—the jumbo-CD-to-total-asset ratio, the net-worth-to-total-asset ratio, the past-due and nonaccruing loan ratios, the net-income-to-total-asset ratio, and the two CAMELS dummy variables—are statistically significant with the expected sign in all six equations. The coefficient on the size variable has a mixed-sign pattern, which is not surprising, given the theoretical ambiguity in the relationship between bank size and risk. The coefficients on the other four explanatory variables are statistically significant with the expected sign in at least three of the six equations.

The in-sample fit of the downgrade-prediction model does deteriorate slightly through time. The log-likelihood statistic declines monotonically from the 1991-92 equation through the 1995-96 equation. Indeed, the psuedo-R^2 averages 20.7 percent for the first three equations (1990-91, 1991-92, 1992-93) and 16.5 percent for the last three equations (1993-94, 1994-95, 1995-96). The number of statistically significant coefficients with expected signs also declines slightly over the estimation years. For instance, the coefficients on the commercial-and-industrial-loan-to-total-asset ratio are statistically significant with the expected sign in the first three equations but in only one of the last three equations (1995-96). The monotonic deterioration in model fit reflects the decline in the number of downgrades. In the first three regressions, the average number of downgrades per year was 500; in the last three regressions, the average dropped to 127 downgrades per year.

OUT-OF-SAMPLE PERFORMANCE COMPARISONS OF THE SEER RISK RANK MODEL, THE SEER RATING MODEL, AND THE DOWNGRADE-PREDICTION MODEL

With a timing convention that mimics the way supervisors use econometric models in surveillance,

Table 4

How Well Did the CAMELS Downgrade-Prediction Model Perform In-Sample?

	Years of downgrades in CAMELS ratings		
Explanatory variables	1990-91	1991-92	1992-93
Intercept	–2.053*** (0.232)	–0.923*** (0.249)	–0.284 (0.290)
COMMERCIAL LOANS	0.010*** (0.003)	0.013*** (0.003)	0.012*** (0.003)
RESIDENTIAL LOANS	–0.005** (0.002)	–0.003 (0.002)	–0.004 (0.003)
LARGE TIME DEPOSITS	0.017*** (0.003)	0.018*** (0.003)	0.014*** (0.004)
NET WORTH	–0.053*** (0.008)	–0.050*** (0.010)	–0.049*** (0.011)
PAST-DUE 90	0.396*** (0.038)	0.304*** (0.039)	0.232*** (0.045)
PAST-DUE 30	0.100*** (0.021)	0.136*** (0.021)	0.151*** (0.025)
NONACCRUING	0.227*** (0.027)	0.201*** (0.030)	0.188*** (0.035)
ROA	–0.242*** (0.031)	–0.330*** (0.038)	–0.104*** (0.038)
SECURITIES	–0.015*** (0.002)	–0.017*** (0.002)	–0.014*** (0.002)
OREO	0.212*** (0.030)	0.210*** (0.032)	0.021 (0.033)
SIZE	0.076*** (0.016)	–0.029* (0.017)	–0.128*** (0.022)
CAMELS-2	0.622*** (0.060)	0.542*** (0.067)	0.577*** (0.081)
BAD MANAGE	0.488*** (0.050)	0.405*** (0.053)	0.429*** (0.058)
Number of observations	8,927	8,636	8,361
Pseudo-R^2	0.218	0.224	0.179
–2 log likelihood testing whether all coefficients (except the intercept) = 0	5,909.617***	5,020.667***	3,476.658***

NOTE: This Table contains the estimated regression coefficients for the downgrade-prediction model. The model regresses downgrade status (1 for a downgrade and 0 for no downgrade) in calendar years $t+1$ and $t+2$ on explanatory variables from the fourth quarter of year t. See Table 3 for the definitions of the explanatory variables. Standard errors appear in parentheses next to the coefficients. One asterisk denotes significance at the 10 percent level, two asterisks denote significance at the 5 percent level, and three asterisks denote significance at the 1 percent level. Shading highlights coefficients that were significant with the expected sign in all six years. The pseudo-R^2 gives the approximate proportion of the total variance of downgrade status explained by the model. Overall, the downgrade-prediction model predicts in-sample downgrades well. Eight of the 13 regression variables are significant with the predicted sign in all six years, and all of the variables are significant in at least some years. Note that, by most measures of in-sample fit, the model declines in power over time, primarily due to the decrease in the number of downgrades.

we conduct six separate tests of the out-of-sample performance of the downgrade-prediction model. As noted, the first downgrade-prediction model regresses downgrade status in 1990 and 1991 on year-end 1989 financial data. By the end of 1991, supervisors would have had coefficient estimates from that regression. Our first out-of-sample test applies those coefficients to year-end 1991 financial ratios to compute downgrade probabilities for each sample bank. We then use the ranking of downgrade probabilities to construct power curves for type-one and type-two errors over the 1992-93 test window. To ensure compatibility between the in-sample and out-of-sample data, we limit the first out-of-sample test to banks with five-year operating histories, with CAMELS ratings of one or two as of year-end 1991, and with at least one full-scope examination in 1992 or 1993. The next five out-of-sample tests of the downgrade-prediction model—for the 1993-94, 1994-95, 1995-96, 1996-97, and 1997-98 windows—employ the same timing convention and the same sample restrictions.

Our out-of-sample tests of the SEER risk rank and the SEER rating models use the same timing convention as the out-of-sample tests of the downgrade-prediction model. Specifically, we apply the fixed SEER risk rank coefficients to year-end 1991 data and rank the one- and two-rated banks by their estimated probabilities of failure. We then derive a power curve reflecting the type-one and type-two

Table 4 cont'd

How Well Did the CAMELS Downgrade-Prediction Model Perform In-Sample?

	Years of downgrades in CAMELS ratings		
Explanatory variables	1993-94	1994-95	1995-96
Intercept	0.340 (0.358)	–0.809** (0.379)	0.069 (0.425)
COMMERCIAL LOANS	0.005 (0.005)	0.007 (0.005)	0.013** (0.005)
RESIDENTIAL LOANS	–0.005 (0.003)	–0.002 (0.003)	–0.013*** (0.004)
LARGE TIME DEPOSITS	0.018*** (0.005)	0.023*** (0.005)	0.021*** (0.005)
NET WORTH	–0.094*** (0.014)	–0.025* (0.013)	–0.034*** (0.012)
PAST-DUE 90	0.329*** (0.058)	0.286*** (0.063)	0.324*** (0.073)
PAST-DUE 30	0.169*** (0.032)	0.113*** (0.034)	0.162*** (0.035)
NONACCRUING	0.148*** (0.046)	0.183*** (0.045)	0.146*** (0.050)
ROA	–0.137*** (0.040)	–0.252*** (0.050)	–0.162*** (0.038)
SECURITIES	–0.007*** (0.002)	–0.003 (0.003)	–0.010*** (0.003)
OREO	0.080* (0.041)	0.193*** (0.044)	0.154*** (0.052)
SIZE	–0.171*** (0.027)	–0.149*** (0.030)	–0.210*** (0.034)
CAMELS-2	0.444*** (0.095)	0.625*** (0.103)	0.590*** (0.102)
BAD MANAGE	0.453*** (0.066)	0.406*** (0.073)	0.515*** (0.078)
Number of observations	8,600	9,169	9,200
Pseudo-R^2	0.149	0.153	0.193
–2 log likelihood testing whether all coefficients (except the intercept) = 0	2,248.122***	1,911.719***	1,628.444***

errors of this ordinal ranking, assuming that a higher failure probability at year-end 1991 indicates a higher downgrade probability in 1992 and 1993. For each year of the sample, we repeat this procedure, applying the fixed SEER risk rank model coefficients to the end-of-year call report data for one- and two-rated banks. Because SEER rating model estimates are not available for 1991 and 1992, we start out-of-sample testing of this model with "shadow" CAMELS ratings based on year-end 1993 data. We derive a power curve for the ordinal ranking of shadow CAMELS ratings based on the assumption that higher (weaker) estimated ratings indicate higher downgrade risk in 1994 and 1995. We use the same timing convention for the remaining three out-of-sample tests of the rating model (1995-96, 1996-97, and 1997-98).

Using any of the three models to flag downgrade candidates markedly improves the results compared with randomly selecting one- and two-rated banks. Panel A of Table 5 presents the results of the out-of-sample performance tests of the downgrade-prediction model and the two SEER models. Figures 1 and 2 offer the same information in visual form. Over the four test windows that include both SEER models—1994-95 through 1997-98—the average area under the power curves for the three models is 20.78 percent, substantially less than the 50 percent area under the power curve for random selection. Over all six test windows—1992-93 through 1997-98—the average of the area under the downgrade-prediction power curve and the SEER risk rank power curve equals 21.41 percent. Across individual models and individual years, the areas range from a high of 26.59 percent for the SEER risk rank model in flagging 1994-95 downgrades to a low of 15.14 percent for the downgrade-prediction model in flagging 1996-97 downgrades.

Overall, the downgrade-prediction model slightly outperforms the two SEER models in the out-of-sample performance comparisons. Over four tests covering the years 1994 through 1998, the downgrade-prediction model produces an average power curve area of 18.48 percent, whereas the two SEER models, on average, produce an area of 21.93 percent. Over six tests covering 1992 through 1998,

Table 5

How Did the Out-of-Sample Performance of the Downgrade-Prediction Model and the SEER Models Compare?

Panel A: Area under power curves

Downgrade years	Downgrade model (%)	SEER risk rank model (%)	SEER rating model (%)
1992-93	21.01	22.06	NA
1993-94	22.64	25.54	NA
1994-95	22.31	26.59	21.88
1995-96	17.40	21.45	22.13
1996-97	15.14	19.09	19.35
1997-98	19.08	24.62	20.30
Mean over all years	**19.60**	**23.23**	**20.92**

Panel B: Area under power curves below 20 percent type-two error rate

Downgrade years	Downgrade model (%)	SEER risk rank model (%)	SEER rating model (%)
1992-93	12.03	12.32	NA
1993-94	12.14	12.61	NA
1994-95	11.87	12.70	11.95
1995-96	10.72	11.66	11.82
1996-97	10.28	11.24	11.52
1997-98	10.92	12.28	11.73
Mean over all years	**11.33**	**12.14**	**11.76**

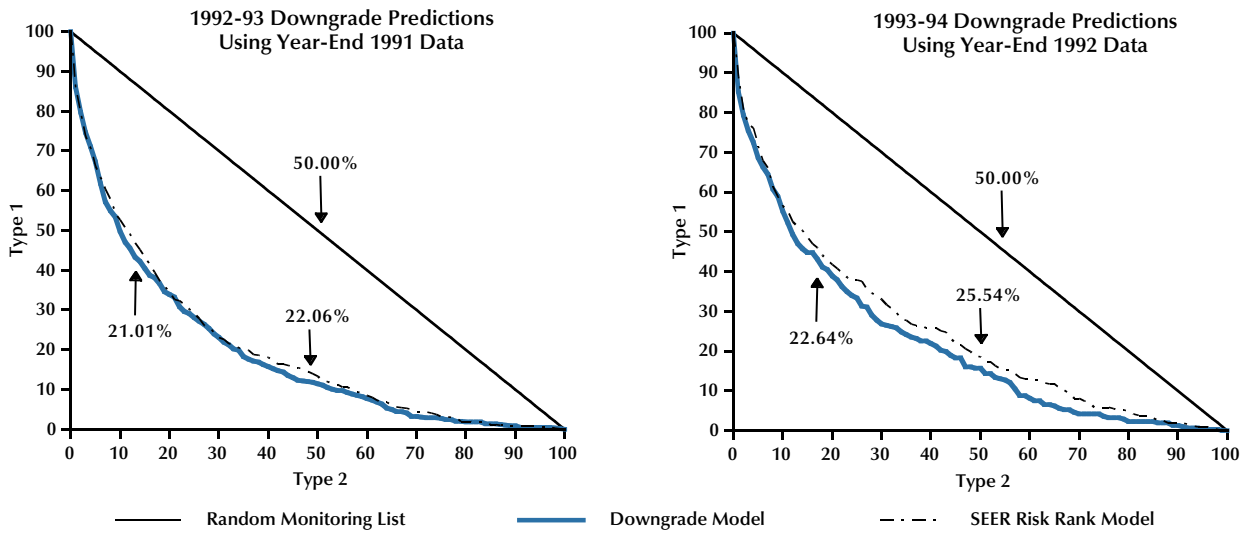
NOTE: This Table contains the areas under each model's power curve for each two-year test window. Each power curve reveals the trade-offs between type-one errors (missed downgrades) and type-two errors (over-predicted downgrades) for a particular model. We assess relative performance by comparing areas under the curves; smaller is better because smaller areas imply simultaneous reduction of both types of errors. The SEER rating model data were not available before 1993, so the Table contains no shadow CAMELS areas for the 1992-93 and 1993-94 test windows. When comparing areas, we bear in mind that the area produced by a randomly generated watch list equals, on average, 50 percent. Although all three models improve considerably over random selection of downgrade candidates, the downgrade-prediction model does not materially outperform the two SEER models (Panel A). When the maximum allowable type-two error rate is 20 percent, the results are virtually identical (Panel B). We use this cut-off as representative of model comparisons when supervisors insist on small watch lists.

the downgrade-prediction model generates an average area of 19.60 percent; the SEER risk rank model generates an average area of 23.23 percent. In each of the six test windows, the downgrade-prediction model outperforms the SEER risk rank model, the difference in area ranging from 1.05 percentage points for the 1992-93 window to 5.54 percentage points for the 1997-98 window. The downgrade-prediction model outperforms the SEER rating model in three of the four test windows, with area differentials ranging from 1.22 percentage points for the 1997-98 test to 4.73 percentage points for the 1995-96 test. Only in the 1994-95 test did the SEER rating model outperform the downgrade model—by an area differential of 0.43 percentage points.

Still, the difference between the out-of-sample performance of the downgrade-prediction model and the SEER models is quite small. On average over all tests, the downgrade model outperforms the SEER models by an area differential of just 2.48 percentage points. Over the restricted area (with less than a 20 percent type-two error), the area differential is even smaller—only 0.62 percentage points. At the same time, on average, the two SEER models outperform random selection by an area differential of 27.93 percentage points. Moreover, the out-of-sample tests are biased in favor of the downgrade-prediction model in two respects: the

Figure 1

How Well Do the Models Predict Out-of-Sample CAMELS Downgrades?



NOTE: This Figure shows that the SEER risk rank model and the downgrade model have similar type-one vs. type-two tradeoffs for most of the range of errors for 1992-93 and 1993-94 downgrades. The downgrade model slightly edges out the SEER failure model by 21.01 percent to 22.06 percent for the 1992-93 downgrades, and by 22.64 percent to 25.54 percent for the 1993-94 downgrades. The SEER rating model numbers were not available before 1993, so a SEER rating model power curve does not appear in the Figure.

This Figure depicts the trade-off between the type-one error rate and the type-two error rate for the SEER risk rank model, SEER rating model, and the downgrade model. The type-one error rate is the number of missed downgrades (false negatives) divided by the total number of CAMELS one- and two-rated banks; the type-two error rate is the number of incorrectly flagged downgrades (false positives) divided by the total number of CAMELS one- and two-rated banks. The area under each curve, divided by the total area in the box, offers a convenient way to compare the performance of each model. Smaller areas imply lower levels of *both* types of errors and, hence, better model performance. The 50 percent line indicates the type-one and type-two error rates associated with random selection of one- and two-rated banks.

coefficients on the SEER risk rank model have been fixed since 1991, and the SEER rating model is estimated on originally submitted call report data. The small difference in performance, particularly when viewed in light of these potential biases, suggests that the SEER models and our customized downgrade-prediction model flag downgrade candidates equally well.

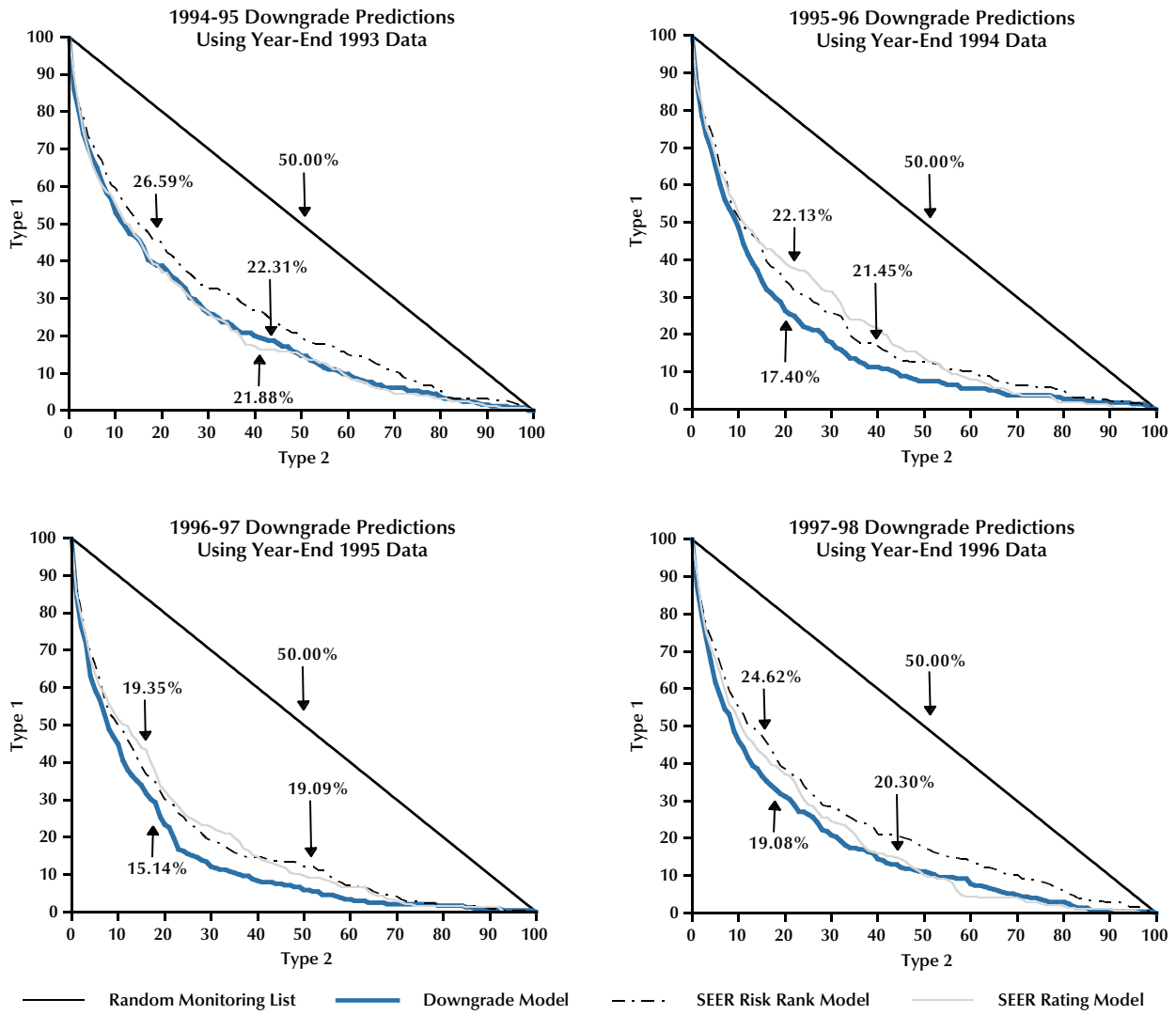
Analyzing a region with low type-two error rates confirms that the out-of-sample performances of the downgrade-prediction model and those of the SEER models are comparable. If monitoring healthy banks were costless, then supervisors would want a watch list long enough to catch all downgrade risks—a list that produced a zero type-one error rate. But because monitoring healthy banks is costly, supervisors would prefer a watch list that is reasonably sized. Panel B of Table 5 contains the areas under the power curves for all three models when the maximum allowable type-two error rate is 20 percent. Over this restricted region, the difference in the performance of the downgrade-prediction model and the two SEER models—again, expressed in terms of average areas under power curves over multiple tests—is less than 1 percentage point. Although the 20 percent threshold is arbitrary, it conveys a larger message—that the small difference between the performance of the downgrade model and the SEER models becomes even smaller when the comparison focuses on regions where supervisors are likely to operate.

ROBUSTNESS CHECKS

To check the robustness of our findings, we experimented with a "fresh" set of explanatory variables for each of the six downgrade-prediction

Figure 2

How Well Do the Models Predict Out-of-Sample CAMELS Downgrades?



NOTE: This Figure shows that the downgrade model, the SEER risk rank model, and the SEER rating model produce similar type-one vs. type-two tradeoffs for most of the range of errors for 1994-95 downgrades. The downgrade and SEER rating models do slightly outperform the SEER risk rank model, largely because the coefficients of the risk rank model are fixed. The downgrade model slightly outperforms both the SEER risk rank model and the SEER rating model as a tool for flagging downgrade candidates in 1995-96, 1996-97, and 1997-98.

This Figure depicts the trade-off between the type-one error rate and the type-two error rate for the SEER risk rank model, SEER rating model, and the downgrade model. The type-one error rate is the number of missed downgrades (false negatives) divided by the total number of CAMELS one- and two-rated banks; the type-two error rate is the number of incorrectly flagged downgrades (false positives) divided by the total number of CAMELS one- and two-rated banks. The area under each curve, divided by the total area in the box, offers a convenient way to compare the performance of each model. Smaller areas imply lower levels of *both* types of errors and, hence, better model performance. The 50 percent line indicates the type-one and type-two error rates associated with random selection of one- and two-rated banks.

regressions. If the factors driving downgrades change through time, then the out-of-sample performance of a model with a fixed set of explanatory variables should decay over a sequence of tests, even if new coefficients for the fixed set of variables are obtained each year. To combat this bias, we compiled an expanded list of candidate variables based on a review of the early warning literature.[5] Next, we identified the best subset of explanatory variables in each year based on in-sample fit of the model.[6] Specifically, our variable selection technique resembled stepwise and backward-elimination variable selection but improved upon these methods by considering all possible combinations, rather than adding or subtracting explanatory variables sequentially. Because our technique is most effective when the explanatory variables are not highly correlated, we started by grouping candidates into clusters based on correlation.[7] For example, we grouped all the nonperforming asset ratios in one problem-loan cluster. Then, for each year, we identified the variable in each cluster that was least correlated with the variables in the other clusters. Finally, we added this variable to the final set of explanatory variables for that year's downgrade-prediction model.

As an additional robustness check, we shortened the forecast horizon for all three models. In our previous analysis, we compared each model's ability to forecast downgrades two years into the future. Because the SEER rating model regresses this quarter's ratings on last quarter's financial data (i.e., uses a one-quarter lag), out-of-sample performance comparisons over a two-year horizon may be biased against the shadow CAMELS. To correct for this potential bias, we regressed downgrade status in the first quarter of year t on financial data from the fourth quarter of year $t-1$. Then, we applied the coefficients from that model to financial data from the fourth quarter of year $t+1$ to generate downgrade probabilities. Next, we traced out power curves for the downgrade-prediction model—and for the two SEER models—using the first quarter of year $t+2$ as a test window. Finally, we compared the areas under each model's power curve four times with all three models (first quarter 1994 through first quarter 1997) and six times for the SEER risk rank model and the downgrade-prediction model (first quarter of 1992 through the first quarter of 1997).

Both robustness checks confirmed our principal empirical result—that the downgrade-prediction model does not improve significantly over the SEER models as a tool for flagging downgrade candidates. In the first robustness check, we found that respecifying the CAMELS downgrade model annually did not improve its out-of-sample accuracy. Indeed, the resulting power curves were nearly identical to those obtained with the original downgrade-prediction model. In the second robustness check, we found that shortening the forecast horizon did improve the out-of-sample performance of all three models, presumably because predicting near events is easier than predicting more distant ones. For example, the average area under the downgrade-prediction power curve improved 4.32 percentage points (six tests), the average area under the SEER risk rank power curve improved 3.22 percentage points (six tests), and the average area under the SEER rating model power curve improved by 4.55 percentage points (four tests). Still, average areas produced by each model were fairly close: 15.28 percent for the downgrade-prediction model, 20.01 percent for the SEER risk rank model, and 16.37 percent for the SEER rating model. When viewed against the random selection benchmark, these performance differences seem economically insignificant.

CONCLUSION

The Federal Reserve's off-site surveillance system includes two econometric models that are collectively known as the System for Estimating Examination Ratings (SEER). One model, the SEER risk rank model, uses the latest financial statements to estimate the probability that each Fed-supervised bank will fail within the next two years. The other model, the SEER rating model, uses the latest financial statements to produce a "shadow" CAMELS rating for each supervised bank. Banks identified as risky by either model receive closer supervisory scrutiny than other Fed-supervised banks.

Because many of the banks flagged by the SEER models have already tumbled into poor condition and, hence, receive considerable supervisory attention, we developed an alternative model to identify safe-and-sound banks headed for financial distress. Such a model could help supervisors allocate scarce

[5] In addition to the papers we already cited, we drew on Cole and Gunther (1995), Hooks (1995), Wheelock and Wilson (2000), and Estrella, Park, and Peristiani (2000).

[6] See Lawless and Singhal (1978) for details.

[7] See Jackson (1991).

on- and off-site resources by pointing to banks not currently under scrutiny that need watching. Specifically, we estimated a model to flag banks with composite CAMELS ratings of one and two that are likely to receive downgrades to composite ratings of three, four, or five in the next two years. We then compared the out-of-sample performance of the model and the SEER models as tools for identifying downgrade candidates.

Over a range of two-year test windows in the 1990s, we found that the CAMELS downgrade model outperformed the SEER models by only a small margin. Our evidence suggests that, during relatively tranquil banking times such as the 1990s, a downgrade-prediction model contributes little to the Federal Reserve's off-site surveillance framework. Our evidence also indirectly validates the performance of the current SEER framework as a tool for supporting on-site examinations by the Federal Reserve.

Our evidence does not imply, however, that downgrade-prediction models have no role to play in off-site surveillance. Our sample period is marked by the longest economic expansion in U.S. history. During this period, the U.S. banking industry enjoyed robust profitability and healthy asset quality. Indeed, downgrades to unsatisfactory status as well as outright failures dropped off considerably in the 1990s relative to the 1980s. A possible interpretation of our findings is that one early warning model is as good as another when financial distress in the banking industry is relatively rare. The downgrade-prediction model could materially outperform the SEER models in a different economic climate—for example, the early stages of a contraction in which downgrades are frequent but failures still relatively rare. Only a series of out-of-sample tests that span the business cycle can conclusively determine the value added by a CAMELS downgrade model.

REFERENCES

Allen, Linda and Saunders, Anthony. "Bank Window Dressing: Theory and Evidence." *Journal of Banking and Finance*, June 1992, *16*(3), pp. 585-623.

Berger, Allen N. "The Relationship Between Capital and Earnings in Banking." *Journal of Money, Credit, and Banking*, May 1995, *27*(2), pp. 432-56.

_____ and Davies, Sally M. "The Information Content of Bank Examinations." *Journal of Financial Services Research*, October 1998, *14*(2), pp. 117-44.

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

Cole, Rebel A.; Cornyn, Barbara G. and Gunther, Jeffrey W. "FIMS: A New Monitoring System for Banking Institutions." *Federal Reserve Bulletin*, January 1995, *81*(1), pp. 1-15.

_____ and Gunther, Jeffrey W. "Separating the Likelihood and Timing of Bank Failure." *Journal of Banking and Finance*, September 1995, *19*(6), pp. 1073-89.

_____ and _____. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems." *Journal of Financial Services Research*, April 1998, *13*(2), pp. 103-17.

Curry, Timothy. "Bank Examination and Enforcement," in *History of the Eighties: Lessons for the Future*, Vol. 1. Washington, DC: Federal Deposit Insurance Corporation, 1997, pp. 421-75.

Demsetz, Rebecca S. and Strahan, Philip E. "Diversification, Size, and Risk at Bank Holding Companies." *Journal of Money, Credit, and Banking*, August 1997, *29*(3), pp. 300-13.

DeYoung, Robert. "Birth, Growth, and Life or Death of Newly Chartered Banks." Federal Reserve Bank of Chicago *Economic Perspectives*, Third Quarter 1999, *23*(3), pp. 18-35.

Estrella, Arturo; Park, Sangkyun and Peristiani, Stavros. "Capital Ratios as Predictors of Bank Failure." Federal Reserve Bank of New York *Economic Policy Review*, July 2000, *6*(2), pp. 33-52.

Flannery, Mark J. and Houston, Joel F. "The Value of a Government Monitor for U.S. Banking Firms." *Journal of Money, Credit, and Banking*, February 1999, *31*(1), pp. 14-34.

Gilbert, R. Alton; Meyer, Andrew P. and Vaughan, Mark D. "The Role of Supervisory Screens and Econometric Models in Off-Site Surveillance." Federal Reserve Bank of St. Louis *Review*, November/December 1999, *81*(6), pp. 31-56.

Gunther, Jeffrey W. and Moore, Robert R. "Early Warning Models in Real Time." Financial Industry Studies Working Paper No. 1-00, Federal Reserve Bank of Dallas, October 2000.

Hanc, George. "The Banking Crises of the 1980s and Early

1990s: Summary and Implications," in *History of the Eighties: Lessons for the Future*, Vol. 1. Washington, DC: Federal Deposit Insurance Corporation, 1997, pp. 3-85.

Hooks, Linda M. "Bank Asset Risk: Evidence from Early-Warning Models." *Contemporary Economic Policy*, October 1995, *13*(4), pp. 36-50.

Jackson, J. Edward. *A Users Guide to Principal Components*. New York: Wiley, 1991.

Lawless, J.F. and Singhal, K. "Efficient Screening of Nonnormal Regression Models." *Biometrics*, 1978, *34*, pp. 318-27.

Morgan, Donald P. and Stiroh, Kevin J. "Market Discipline of Banks: The Asset Test." Working Paper, Research Department, Federal Reserve Bank of New York, 2001.

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 O'Keefe, John. "Off-Site Surveillance Systems," in *History of the Eighties: Lessons for the Future*, Vol. 1. Washington, DC: Federal Deposit Insurance Corporation, 1997, pp. 477-520.

Wheelock, David C. and Wilson, Paul W. "Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions." *Review of Economics and Statistics*, February 2000, *82*(1), pp. 127-38.