FEDERAL RESERVE BANK OF ST. LOUIS
SUPERVISORY POLICY ANALYSIS WORKING PAPER

Working Paper No. 2002-08

Can Feedback from the Jumbo-CD Market Improve Off-Site Surveillance of Community Banks?

R. Alton Gilbert
Research Department
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-8559
gilbert@stls.frb.org

Andrew P. Meyer
Banking Supervision and Regulation
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442
ameyer@stls.frb.org

Mark D. Vaughan
Banking Supervision and Regulation
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442
mark.vaughan@stls.frb.org
314.444.8859

June 2002

Can Feedback from the Jumbo-CD Market Improve Off-Site Surveillance of Community Banks?

R. Alton Gilbert
Research Department
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-8559
gilbert@stls.frb.org

Andrew P. Meyer
Banking Supervision and Regulation
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-4647
ameyer@stls.frb.org

Mark D. Vaughan*
Banking Supervision and Regulation
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
314-444-8859
mark.vaughan@stls.frb.org

*corresponding author

June 2002

JEL Codes: G21, G28
Keywords: Bank Supervision, Early Warning Models, Surveillance

Critical feedback from a number of sources greatly improved this work. Specifically, we would like to thank the following bank supervisors—Carl Anderson, John Block, Joan Cronin, and Kim Nelson—and the following economists—Rosalind Bennett, Doug Evanoff, Mark Flannery, John Jordan, John Hall, Jim Harvey, Tom King, John Krainer, Bill Lang, Jose Lopez, Dan Nuxoll, Evren Ors, Jim Thomson, Sherrill Shaffer, and Larry Wall—for helpful comments and discussions. We also profited from exchanges with seminar participants at the Federal Reserve System Surveillance Conference, the Federal Reserve System Committee on Financial Structure meetings, the Financial Management Association meetings, the Office of the Comptroller of the Currency, and Washington University in St. Louis. Any remaining errors and omissions are ours alone. The views expressed in this paper do not represent official positions of the Federal Reserve Bank of St. Louis, the Board of Governors, or the Federal Reserve System.

Can Feedback from the Jumbo-CD Market Improve Off-Site Surveillance of Community Banks?

Abstract

We examine the value of feedback from the jumbo-certificate-of-deposit (CD) market in the off-site surveillance of community banks. Using accounting data, we construct proxies for default premiums on jumbo CDs. Then, we produce rank orderings of community banks—defined as institutions holding less than $500 million in assets (constant 1999 dollars)—based on these proxies. Next, we use an econometric surveillance model to generate rank orderings based on the probability of encountering financial distress. Finally, we compare these rank orderings as tools for flagging emerging problems. Our comparisons include eight out-of-sample test windows during the 1990s.

We find that feedback from the jumbo-CD market would have added little value in community-bank surveillance during our sample period. Specifically, rank orderings based on output from the econometric model significantly outperformed rank orderings based on jumbo-CD default premiums. More important, the jumbo-CD orderings improved little over a random ordering. Other attempts to extract risk signals from the jumbo-CD data yielded similar results. Taken together, our findings validate current surveillance practices. We conclude by arguing that the robust economic environment of the 1990s probably plays a large role in our results.

JEL Codes: G21, G28

Keywords: Bank Supervision, Early Warning Models, Surveillance

1. Introduction

In recent years, bank supervisors around the developed world have explored strategies for harnessing market pressure to contain bank risk. Indeed, the new Basel capital accord counts market discipline as an explicit pillar of bank supervision—along with supervisory review and capital requirements. In the United States, one popular proposal for enhancing market discipline involves requiring large banks to issue a standardized form of subordinated debt (Board of Governors, 2000; Board of Governors, 1999; Meyer, 1999). Advocates of this proposal argue that high-powered performance incentives in the sub-debt market will lead to accurate assessments of bank risk. And, in turn, these assessments—expressed for high-risk institutions as rising yields or difficulties rolling over maturing debt—will pressure bank managers to maintain safety and soundness (Lang and Robertson, 2000).

Even if financial markets provide little direct pressure to contain bank risk, market-generated risk assessments could still contribute to one component of supervisory review—off-site surveillance. Off-site surveillance involves the use of accounting data and anecdotal evidence to schedule on-site exams and to monitor bank progress in addressing previously identified deficiencies. Risk assessments from financial markets could contribute to surveillance in two ways. First, market signals about risk might flag troubled banks missed by conventional surveillance tools. Second, market signals might point to emerging problems before conventional surveillance tools. In the U.S., off-site surveillance relies heavily on financial statements that banks must file quarterly with their principal supervisor. Markets, in contrast, render risk verdicts at a much higher frequency, in some cases even daily.

To date, discussions about harnessing financial markets for supervisory purposes have centered on large, complex banking organizations. Discussions have centered on larger banks because the supervisory benefits would be highest and the compliance costs lowest for these institutions (Emmons, Gilbert, and Vaughan, 2001). The benefits would be highest for large banks because of their complexity; these institutions engage in non-bank activities frequently and use derivative instruments heavily. Large banks also account for the lion's share of U.S. banking assets, making the stability of the financial system dependent on their safety and soundness. The compliance cost would be lowest because most of these institutions already tap national financial markets routinely. For example, at the end of 2000, 44 of the largest 50 commercial banks, and 46 of the 50 largest bank holding companies, had subordinated debt outstanding.

Financial markets could play a useful role in the supervision of community banks. Community banks do fund most of their operations with core deposits (Bassett and Brady, 2001). And the prices and quantities of these deposits—typically checking accounts, passbook savings deposits, and small time deposits—tend to be "sticky" (Flannery, 1982). Still, community banks face one form of market discipline—they compete in the national market for jumbo certificates of deposit, that is time deposits with balances exceeding the $100,000 ceiling for deposit insurance coverage. At year-end 2000, banks holding less than $500 million in assets—the definition of community bank set forth in the Financial Modernization Act of 1999—funded 13.2 percent of their assets with jumbo CDs, up from 9.1 percent at year-end 1990. Academic research has repeatedly confirmed that jumbo-CD holders react to bank-specific risk. [See Hall, King, Meyer, and Vaughan (2002) for recent evidence as well as a review of previous research.]

This reaction could translate into pressure on community bankers to contain risk or signals to community-bank supervisors to take action.

Supplementary information from the jumbo-CD market might prove particularly useful in community-bank surveillance. Many community banks operate under extended exam schedules, with up to 18 months elapsing between full-scope examinations. This extended schedule may reduce the information content of community bank financial statements, thereby reducing the effectiveness of off-site supervisory monitoring. Verification of financials is one important source of value created by examinations (Berger and Davies, 1998; Flannery and Houston, 1999); indeed, recent research has documented large adjustments in asset quality measures following examinations, particularly for institutions with emerging problems (Gunther and Moore, 2000). It is possible that the holders of a community bank's jumbo-CDs, because their own money is at stake, supplement published income statements and balance sheets with independent research. Another possibility is that information about safety and soundness leaks to uninsured depositors through the bank's board of directors, a body that typically includes prominent local businesspeople. Thus, sudden changes in jumbo-CD yields or withdrawal patterns might signal impending trouble more quickly or more reliably than surveillance tools based on quarterly financial statements.

Another reason to consult the jumbo-CD market for help in community-bank surveillance is that these banks represent a disproportionate threat to the deposit insurance fund. Community banks fail more often than larger banks. From 1984 through 1998, for example, the average failure rate for banks with less than $500 million in assets

(1998 constant dollars) was 0.73 percent; the failure rate for banks holding over $500 million in assets was 0.33 percent. In addition, community banks have accounted for a disproportionately large share of assets at risk in failures. At year-end 2000, the U.S. banking sector contained 7,570 banks holding less than $500 million in assets. These banks accounted for 91.7 percent of U.S. banks and 12.9 percent of U.S. banking assets. Between 1984 and 1998, community banks accounted for 40 percent of the assets of failed U.S. banks. Finally, losses to the FDIC from community-bank failures, per dollar of assets, have exceeded losses from large-bank failures. Losses have been higher because community banks are top-heavy with assets that are expensive to liquidate—fewer securities and more fixed assets—and because community banks have fewer uninsured and unsecured claimants to absorb failure costs (James, 1991). More timely or more accurate warning about emerging problems would help supervisors reduce the costs of community bank failures to the FDIC.

Although the jumbo-CD market could, in theory, provide supervisors with useful information, available data permit only the construction of crude proxies for the desired signals about community-bank risk. Just a handful of large banks issue jumbo-CDs that are actively traded in secondary markets, so real-time, market-generated yields are not available for small institutions (Morris and Walter, 1993). It is possible, however, to use quarterly financial data to construct average jumbo-CD yields for almost every bank in the country. These yields can then be combined with data from the Treasury market to produce proxies for default premiums on jumbo CDs. Other researchers have successfully used this approach to test hypotheses about bank risk (for example, James, 1988; Keeley 1990; and, more recently, Martinez-Peria and Schmukler, 2001). Still, proxies based on these yields suffer from two, related types of measurement error. First, they are average rather than marginal measures and, therefore, somewhat backward looking. Second, they are quarterly accounting measures rather than real-time economic measures. How much value default-premium proxies add in community-bank surveillance is ultimately an empirical question. Answering this question is important as a validation exercise for current surveillance practices. Answering this question is also important because proxies for jumbo-CD default premiums offer the only real hope for integrating market feedback into community-bank surveillance.

Assessing the supervisory value of feedback from the jumbo-CD market requires establishing a benchmark with output from current surveillance tools. It is not enough to note that jumbo-CD measures react to bank risk contemporaneously or forecast emerging safety-and-soundness problems successfully because supervisors already have systems in place for these purposes. The acid test of the jumbo-CD measures, or of any potential off-site tool for that matter, is whether it improves materially upon current practice. Three recent papers evaluate market signals with a "best-practices-in-surveillance" benchmark. Evanoff and Wall (2001) compare regulatory-capital ratios and subordinated-debt yields as predictors of supervisory ratings; they find that sub-debt yields modestly outperform capital ratios in one-quarter-ahead tests. Gunther, Levonian, and Moore (2001) add estimated default frequencies from the KMV model—a risk measure drawn from the equity market—to an econometric model designed to predict holding-company supervisory ratings with accounting data. They find that the KMV data improve in-sample fit. Krainer and Lopez (2001) also include equity-market variables—in this case, cumulative abnormal stock returns and KMV default probabilities—in a model of holding-company ratings. Unlike Gunther, Levonian, and Moore, they assess the value added by equity-market variables by measuring improvement in one-quarter-ahead forecasts. Like Evanoff and Wall, they find that market variables provide only a modest boost to out-of-sample performance.

We compare risk rankings based on jumbo-CD signals with risk rankings based on the output of an econometric surveillance model as tools for flagging deteriorating community banks. Specifically, we combine our yield measure with Treasury yields to obtain proxies for jumbo-CD default premiums. We then compile rank orderings of the sample banks based on these proxies. Next, we estimate a probit model with financial and supervisory data to predict the likelihood that each sample bank will encounter financial distress in the next eight quarters. We then compile a rank ordering based on estimated distress probabilities. Finally, we track the out-of-sample prediction records of the default-premium orderings and the distress-probability orderings over an eight-quarter horizon. Because uninsured depositors may react to risk by withdrawing their money rather than demanding a higher yield, we also examine the performance of rank orderings based on jumbo-CD runoff. As robustness checks, we evaluate the predictive power of jumbo-CD orderings extracted with a more sophisticated technique and with different cuts of the sample banks. We also assess orderings obtained when jumbo-CD default premiums are added to the econometric surveillance model. All these tests are designed build on the emerging literature on the value of market signals in bank surveillance.

Our empirical tests indicate that feedback from the jumbo-CD market would have contributed little to community-bank surveillance in the 1990s. In all eight out-of-sample test windows, rank orderings based on output from the econometric surveillance model significantly outperform rank orderings based on jumbo-CD default premiums or runoffs. More important, the jumbo-CD orderings improve little on random orderings. These findings prove robust to different extraction techniques and sample cuts. Finally, including jumbo-CD default premiums in the early warning model does not enhance its out-of-sample performance. Collectively, our findings validate current surveillance practices. We argue that the robust economic environment of the 1990s probably plays a large role in our results.

2. A Primer on Off-Site Surveillance

The cornerstone of supervisory review is thorough, regularly scheduled, 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 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 one through five scale. In general, banks with "one" or "two" composite ratings are considered safe and sound while banks with "three," "four," or "five" ratings are considered unsatisfactory. As of March 31, 2000, supervisors classified just over six percent of U.S. banks as unsatisfactory.

Bank supervisors support on-site examinations with off-site surveillance. Off-site surveillance refers to the use of quarterly accounting data and anecdotal evidence to monitor the condition of supervised banks between scheduled 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. 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.

Two commonly used off-site tools are supervisory screens and econometric models. Supervisory screens are combinations of financial ratios, derived from quarterly balance sheets and income statements, that have given warning in the past about developing safety-and-soundness problems. Supervisors draw on their experience to weigh the joint information content of these ratios. Econometric models also combine information from financial ratios. These models rely on statistical tests rather than human judgment to combine ratios, boiling 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 develop 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.

The Federal Reserve uses two econometric models in off-site surveillance. These models are collectively known as SEER, the System for Estimating Examination Ratings. One model, the SEER risk-rank model, uses the latest quarterly accounting 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 data to produce a "shadow" CAMELS rating for each supervised institution—that is, the rating that would have been assigned had the bank been examined using its most recent financial statements. Every quarter, analysts at the Board of Governors feed the latest financial data into the SEER models and forward the results to the 12 Reserve Banks. The surveillance sections at each Reserve Bank, in turn, follows up on all "red-flagged" institutions. The Federal Deposit Insurance Corporation (FDIC) and the Office of the Comptroller of the Currency (OCC) also use econometric models as a part of the off-site surveillance regimen for the banks they supervise (Reidhill and O'Keefe, 1997).

To assess the marginal value of feedback from the jumbo-CD market, we use a downgrade-prediction model to "benchmark" the performance of econometric surveillance models. A downgrade-prediction model is designed to flag banks headed for financial distress. More specifically, it estimates the likelihood of tumbling from "satisfactory" condition (CAMELS "one" or "two ratings) into "unsatisfactory" condition (composite CAMELS "three," "four," or "five") in the coming quarters. In theory, a model designed to predict downgrades could improve upon the SEER framework. Downgrades to unsatisfactory condition were common throughout the 1990s (see Table

1), whereas few banks failed during this period. Thus, a downgrade-prediction model can be re-estimated quarterly. The SEER risk-rank model, in contrast, was estimated during the years 1985-1991, and model coefficients have been "frozen" since the estimation period. More important, a downgrade-prediction model may flag banks not currently under close scrutiny that need watching. Institutions with “three,” “four,” or “five” composite ratings fail at a much higher rate than institutions with “one” or “two” ratings, so “unsatisfactory” institutions already receive scrutiny between exams. Recent research suggests that a downgrade-prediction model would have improved slightly over the SEER framework in the 1990s in flagging emerging distress (Gilbert, Meyer, and Vaughan, 2002). Because of the downgrade-prediction model’s theoretical appeal and slight performance edge on 1990s data, we use it to produce risk rankings that will benchmark current surveillance practices.

In essence, we compare the performance of a single supervisory screen with the performance of an econometric model as early warning tools. The supervisory screen used here—measures of jumbo-CD default premiums or runoffs—differs from other screens by summarizing overall bank risk, not just one type of exposure such as liquidity or credit risk. Previous evidence from the jumbo-CD market confirms risk pricing. (See Table 2 for a summary of published findings). The effect should be stronger in the wake of FDICIA—an act designed in part to shift failure costs from the FDIC to uninsured depositors (Benston and Kaufman, 1998). As a check, we regressed our measures of jumbo-CD yields and runoffs on SEER failure probabilities and a host of control variables. These results—reported in table 3—confirm responsiveness to bank-specific risk, though this response is economically small. The economically small response suggests that jumbo-CD default premiums and runoffs put little direct pressure on banks to contain risk. It does not suggest, however, that ordinal rankings based on these measures lack information content. Again, only out-of-sample performance tests can determine whether these rankings improve upon rankings obtained with current surveillance tools.

3. The Data

Our data set includes quarterly accounting data for most U.S. community banks as well as confidential supervisory assessments of these banks, beginning with the late 1980s and continuing through the 1990s. Specifically, we use accounting data from the fourth quarter of each year, beginning in 1989 and ending in 1998; we use confidential supervisory assessments from 1990 through 2000. Following the approach used in the SEER framework, we estimate the downgrade-prediction model on all CAMELS one- and two-rated U.S. banks and then use model coefficients to out-of-sample downgrade-probability estimates for community banks. The number of observations underlying the downgrade-prediction model range from 7,836 (1992-93 regression) to 8,666 (1995-96 regression).

The data come from 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 that all U.S. commercial banks submit quarterly call reports to their principal supervisors; most reported items are available to the public. We also rely on CAMELS composite and management ratings from the National Information Center (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.

To reduce bias in the performance tests, we exclude from the sample any bank with an operating history of less than five years. Financial ratios for these start-up, or *de novo*, banks often take extreme values that do not imply 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 can 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. For example, the Federal Reserve examines each of its newly chartered banks every six months. Full-scope exams continue on this schedule until the *de novo* earns a composite CAMELS rating of “one” or “two” in consecutive exams.

Our calculation of default premiums begins with three items taken from each sample bank’s quarterly call report submissions. These items are interest expense paid on jumbo-CDs, dollar volume of jumbo CDs outstanding, and the distribution of remaining maturities on jumbo CDs. To generate quarterly yield observations for each bank, we divide jumbo-CD interest expense by average jumbo-CD balances. We use distributions of remaining maturities to control for term-structure effects. From 1989 through 1996, the FFIEC required banks to slot jumbo CDs in one of four buckets based on time remaining to maturity—“less than three months remaining,” “three months to one year remaining,” “one year to five years remaining,” and “over five years remaining.” In 1997, the two longest maturity buckets changed to “one year to three years remaining,” and "over three years remaining." These reporting conventions allow only a crude adjustment for maturity. For example, jumbo CDs in "less than three months remaining" bucket may include long-term instruments issued several years ago that are about to mature. Nevertheless, these data offer the only available means of controlling for the impact of maturity on yields.

We use two distinct methods to turn these call-report items into default-risk premiums—a yield-spread method and a simple-residuals method. We use two methods to insure that subsequent performance tests are not biased by reliance on only one, possible poor, proxy for default premiums. Admittedly, both methods control only for term-to-maturity. Later in Section Six, we try to isolate default premiums with a more sophisticated approach that controls for influences on yields besides maturity. We begin with simple methods because the resulting proxies are similar to screens commonly used by surveillance analysts. The default-premium measures obtained from the two methods are highly correlated; the correlation coefficient over all the sample years is 0.92.

The yield-spread method generates default-premiums by subtracting risk-free yields from jumbo-CD yields. Specifically, we start with the dollar volume of jumbo CDs in each maturity class for each sample bank. We then multiply each of these dollar figures by that quarter's rate of return on Treasury issues with comparable-maturity. The sum of these resulting values, divided by the total jumbo-CD balances, approximates each bank's risk-free yield. Next, we obtain a default-premium series by subtracting—for every quarter in the sample—each bank's risk-free yield from its average jumbo-CD yield.

The simple-residuals method proxies default premiums with residuals from year-by-year yield regressions. The dependent variable in each regression is fourth-quarter jumbo-CD interest expense divided by average fourth-quarter jumbo-CD balances. The independent variables include an intercept term and the percentage of the jumbo-CD portfolio in each of the four maturity buckets. Tables 4a and 4b contain the regression results. Overall model fit is poor—perhaps because other, non-maturity factors play important roles in determining yields—but the combined explanatory power of the maturity variables is statistically significant for each year's regression. The estimated intercepts—which approximate the yield on "less than three months to maturity" instruments—change in concert with market rates on three-month jumbo CDs, though adjustments take place with a lag. For instance, the intercept for 1991 equation (5.72 percent) is well above the secondary-market yield on three-month CDs in the fourth quarter of 1991 (4.91 percent). Because interest rates were lower that quarter (1991:IV) than in prior quarters, jumbo CDs about to mature included higher-yield instruments issued in prior quarters. Still, for each year, coefficient signs and magnitudes for the maturity variables reveal that yields rose with time-to-maturity, implying that the equation residuals may serve as acceptable proxies for default premiums.

4. The Surveillance Benchmark—a CAMELS-Downgrade-Prediction Model

The CAMELS-downgrade-prediction model transforms available accounting and supervisory data into distress probabilities. Specifically, the model is a probit regression. The dependent variable equals one for any sample bank whose composite CAMELS rating slips from one/two status to three/four/five status in the following eight quarters;

the variable equals zero for banks that are examined but not downgraded in the eight-quarter window. The explanatory variables include the financial-performance ratios and bank size measure used in the SEER risk-rank model, as well as two additional CAMELS-related variables. Table 5 describes the explanatory variables and the expected relationship between each variable and the likelihood of a downgrade. The financial-performance ratios are designed to measure 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 relies on six measures of credit risk, the risk that borrowers will fail to make promised interest and principal payments. The credit-risk measures include the ratio of loans 30-89 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 non-accrual status to total assets (NONACCRUING), the ratio of other-real-estate-owned to total assets (OREO), the ratio of commercial-and-industrial loans to total assets (COMMERCIAL-LOANS), and the ratio of residential-real-estate loans to total assets (RESIDENTIAL-LOANS). The model relies on 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 non-accruing loan ratios because banks tend to charge off higher percentages of these loans than loans in current status. We include "other real estate owned" (OREO), which consists primarily of collateral seized after loan defaults, because a high OREO ratio often signals poor credit-risk management—either because a bank had to foreclose on a large number of loans or because it 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 loans. Similarly, we include 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 model contains two measures of leverage risk, the risk that losses will exceed capital and produce insolvency. The leverage-risk measures include total equity minus goodwill as a percentage of total assets (NET-WORTH) and net income as a percentage of total assets (or, return on assets, ROA). We expect higher levels of capital (lower leverage risk) to reduce the likelihood of CAMELS downgrades. Return on assets bears on leverage risk because retained earnings are an important source of additional capital for many banks and because higher earnings provide a larger cushion for withstanding adverse economic shocks (Berger, 1995). We expect higher earnings to reduce the risk of downgrade.

The downgrade-prediction model uses two ratios to capture liquidity risk, the risk that loan commitments cannot be funded or withdrawal demands met at a reasonable cost. The liquidity-risk measures include investment securities as a percentage of total assets (SECURITIES) and jumbo-CD balances as a percentage of total assets (LARGE-TIME-DEPOSITS). A larger stock of liquid assets—such as investment securities—indicates a greater ability to meet unexpected funding needs and should, therefore, translate into a lower downgrade probability. Liquidity risk also depends on a bank's reliance on non-core funding. Non-core funding—which includes jumbo CDs—can be quite sensitive to the difference between the interest rate paid by the bank and the market rate. 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 CAMELS downgrade.

In addition to financial ratios, the model includes variables designed to capture the impact of asset size, CAMELS differences, 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 a "two" because two-rated banks tumble into unsatisfactory status more often than one-rated banks. (See table 1 for downgrade rates for one- and two-rated institutions in the 1990s.) 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 management's competence, even though these problems have yet to produce financial consequences.

We estimate the downgrade-prediction model for eight separate sample windows. In each equation, downgrade status (1 = downgrade, 0 = no downgrade) in years $t+1$ and $t+2$ is regressed on accounting and supervisory data for banks rated CAMELS "one" or "two" in the fourth quarter of year *t*. For example, to produce the first downgrade equation (reported as the "1990-91" equation in Table 6), we use all non-*de novo,* U.S. community banks with one or two composite ratings as of December 31, 1989. We then regress downgrade status during 1990 and 1991 on fourth quarter 1989 (1989:IV) data. We continue with this timing convention through a regression of downgrade status in 1997 and 1998 on 1996:IV data.

The downgrade-prediction model fits the data relatively well. (Table 6 contains these regression results.) For all eight regressions, the log-likelihood test statistic allows rejection of the hypothesis that the model coefficients jointly equal zero at the one-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 15.0 percent for the 1994-95 equation to a high of 22.6 percent for the 1991-1992 equation. 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 non-accruing loan ratios, the net income-to-total asset ratio, and the two CAMELS dummy variables—are statistically significant with the expected sign in all eight equations. The coefficient on the size variable has a mixed-sign pattern, which is not unexpected, given the theoretical ambiguity about the relationship between size and risk. The coefficients on the other five explanatory variables are statistically significant with the expected sign in at least three of the eight equations. Although model performance for the decade is good overall, in-sample fit does deteriorate slightly in later years due to the decline in downgrade frequency.

5. Out-of-Sample Performance Tests—Jumbo-CD vs. the Downgrade Rankings

Next, we conduct performance tests of the risk rankings drawn from downgrade probabilities and jumbo-CD default premiums. For each year, we use the probit model to estimate the likelihood that each sample bank will suffer a downgrade in the next eight quarters and then rank banks from highest downgrade probability to lowest. Similarly, for each year we rank the banks from highest to lowest default premium. We obtain separate rankings for the yield-spread measure and the simple-residuals measure. Although each approach produces a different index number, they may all lead to similar ordinal rankings. Only out-of-sample testing can determine whether the default-premium rankings differ from the downgrade-probability rankings and whether differences in rankings favor default premiums as a surveillance tool. Out-of-sample tests—which use an evaluation period subsequent to the estimation period—are crucial because they mimic the way supervisors actually conduct off-site surveillance. Also, as has been demonstrated in the "profitability-of-trading rules" literature, superior in-sample performance often fails to translate into superior out-of-sample performance (Roll, 1994; Sullivan, Timmerson, and White, 1999).

We assess out-of-sample performance using the type-one and type-two error rates for each risk ranking. Each type of forecast error is costly. A missed downgrade—a type-one error—is costly because accurate downgrade predictions give supervisors more warning about emerging problems, and early intervention reduces the likelihood of failure. A predicted downgrade that does not materialize—a type-two error—is costly because it wastes scarce supervisory resources and imposes unnecessary regulatory burdens. A trade-off exists between the two types of error; supervisors can completely eliminate over-predicted downgrades, for example, by assuming that no banks are downgrade risks.

For each risk ranking, power curves can be drawn to indicate the minimum achievable type-one error rates for any desired type-two error rate (Cole, Cornyn, and Gunther, 1995). For example, the yield-spread power curve shows the type-one and type-two error rates when an ordinal ranking based on spread-over-Treasuries is interpreted as a rank ordering of downgrade risk. We trace out the curve by starting with the assumption that no sample bank is a downgrade risk. This assumption implies that all subsequent downgrades are surprises—a 100 percent type-one error rate. Because no banks are incorrectly classified as downgrade risks, the type-two error rate is zero. We obtain the next point on the curve by selecting the bank with the highest spread. If that bank suffers a downgrade in the following eight quarters, then the type-one error rate for the yield-spread ordering 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 yield spread and re-calculating type-one and type-two error rates, we can trace out a power curve. At the lower right extreme of the curve, all banks are considered at risk of a downgrade. At this extreme, the type-one error rate is zero percent, and the type-two error rate is 100 percent. Figure 1 shows 1992-93 power curves for each risk ranking.

The areas under the power curves provide a basis for comparing the forecast accuracy of each risk ranking. Smaller areas imply lower overall rates of type-one and type-two errors and, hence, more accurate risk rankings. For each ranking, we express the area under the curve as a percentage of the total area in the box. A useful benchmark for evaluating the economic significance of differences in forecast accuracy is the area produced when risk rankings are randomly selected. Random selection, over a large number of trials, produces power curves with an average slope of negative one. Thus, the area under a random-ordering power curve equals, on average, 50 percent of the area of the entire box.

This method of evaluating relative performance—though somewhat atheoretic—does make the best use of the existing data. A more theoretically appealing approach would start by minimizing 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 rankings 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 performance over any desired range of error rates.

Our out-of-sample tests follow a timing convention that reflects the way supervisors actually conduct surveillance. To test the downgrade-model rankings, we start by regressing 1990-91 downgrade status on financial and supervisory data from the fourth quarter of 1989. By the end of 1991, supervisors would have possessed coefficient estimates from this regression. We then apply these coefficients to fourth-quarter 1991 data for each sample bank to obtain a downgrade probability for 1992 and 1993. Finally, we rank banks by these probabilities and use the ranking—together with actual downgrade incidence in 1992 and 1993—to construct a power curve. Downgrade-model curves for the remaining seven test windows reflect the same timing convention. To test the jumbo-CD rankings, we first order banks by their year-end 1991 yield spreads. Next, we derive a power curve for this ranking, assuming that high spreads map into high downgrade probabilities for the 1992-93 window. We follow this procedure seven more times, each time deriving a new ranking when new call-report data would have become available and drawing a new yield-spread curve for this ranking. To complete the jumbo-CD tests, we rank banks by residuals from the 1991 "yield-on-maturity" regression and derive power curves assuming that a high residual equals a high downgrade risk in the 1992-93 window. Once again, we repeat this procedure seven times, producing eight simple-residuals power curves in all.

The evidence overwhelmingly points to the downgrade-prediction model as the superior surveillance tool. Indeed, the default-premium rankings barely improve on random rankings. In the first test window (1992-93), for example, the area under the yield-spread power curve (48.16 percent) and the area under the simple-residuals power curve (47.28 percent) are close to the random-selection benchmark of 50 percent. In sharp contrast, the area under the downgrade-model curve is just 19.46 percent. The power-curve patterns over the next seven test windows are remarkably consistent with the patterns in the first test window. (Figure 1 contains the power curves for the 1992-93 test window. Because the power curves for the other test windows are so similar, they were omitted to save space. Table 7 presents power-curve areas for each surveillance tool and each test window, as well as the average area for each tool over all eight tests.) Over all eight tests the average area under the downgrade-model curves is 19.64 percent, the average area under the yield-spread curves is 46.99 percent, and the average area under the simple-residuals curves is 46.88 percent. In the individual test windows, the downgrade-model areas range from 15.61 percent (1996-97 test window) to 23.19 percent (1994-95). Meanwhile, the areas under the yield-spread rankings range from 43.09 (1995-96) percent to 51.01 percent (1994-95), and the areas under the simple-residuals rankings range from 43.32 percent (1995-96) to 50.49 percent (1994-95).

6. Robustness Checks

Proxies for default premiums appear to have little value in community-bank surveillance. It is possible, however, that the poor performance of our proxies stems from a poor method of extracting signals from jumbo-CD data. It is also possible that jumbo-CD screens would have proved useful in monitoring a different cut of community banks. Finally, it is possible that jumbo-CD screens would have improved the performance of the downgrade-prediction model. In short, before ruling out a role for jumbo-CD screens in surveillance, we must check our results for robustness.

One potential explanation for the poor performance of our default-premium measures is that risky banks substitute insured for uninsured deposits (Billet, Garfinkel, and O Neal, 1998). If such a "substitution effect" were important, then signals about changing risk exposures would show up in jumbo-CD runoffs rather than default premiums. To explore this possibility, we compute the year-over-year percentage change in outstanding jumbo CDs for each sample bank. Then, we rank banks each year from largest to smallest percentage runoff, assuming that high runoff percentages map into high downgrade probabilities. Finally, as before, we use these rankings, together with observed downgrades, to generate power curves for all eight out-of-sample windows. Column 6 of Table 7, labeled "simple runoffs," contains the power-curve areas for the runoff rankings. The average area over all the out-of-sample tests is 46.72 percent, on par with the 46.99 percent average for the yield-spread rankings and 46.88 percent average for the simple-residuals rankings. This evidence suggests that runoff screens, like the default-premium screens, would contribute little to community bank surveillance.

Our relatively simple approach to computing default premiums and runoff percentages may also account, in part, for our findings. Accordingly, we test jumbo-CD rankings obtained by a more sophisticated method. This method, which we term the "complex residuals" method, involves first regressing yields on a richer set of non-financial explanatory variables, then ranking banks by the residuals, and finally generating power curves with these rankings. We repeat this procedure for runoff percentages. The richer set of explanatory variables was identified through an extensive specification search in related research. [See Hall, King, Meyer, and Vaughan (2002) for a detailed discussion of this search and the control variables.] These explanatory variables include controls for interest-rate levels, regional economic conditions, idiosyncratic aspects of the bank's jumbo-CD portfolio, and idiosyncratic aspects of demand and supply for the bank's jumbo CDs. These controls should reduce noise in the default-premium and runoff measures and, thereby, increase the forecasting power of risk rankings based on these measures. Out-of-sample performance of risk rankings based on "complex" yield residuals (column 7 of Table 7) is slightly better than the performance of the other default-premium rankings—the average area under the eight power curves dips about three percentage points to 43.69 percent. At the same time, the performance of the risk rankings based on runoff residuals ("complex residuals-runoff" in column 8 of Table 7) worsens slightly—the average area rises nearly one percentage point to 47.62 percent. More important, these areas are still close to random-selection areas and far from the downgrade-model areas. The evidence implies that even jumbo-CD rankings based on sophisticated measures of default premiums and runoffs contain little information content.

Although our default-premium and runoff measures turn in poor forecasting performances as single-variable screens, they may add value as regressors in the downgrade-prediction model. Indeed, previous research has identified single-variable surveillance screens with this property (Gilbert, Meyer and Vaughan, 1999). Pursuing this angle, we estimated "enhanced" downgrade-prediction models that include each default-premium proxy and each runoff measure. As before, we rank the sample banks using downgrade probabilities generated by the enhanced model. Column 3 in Table 7 contains the results of out-of-sample tests of the model enhanced by simple yield spreads. In seven of eight test windows, the baseline downgrade model outperforms the enhanced downgrade model. On average, the power-curve area under the enhanced model's risk rankings is 20.49 percent—slightly worse than the 19.64 percent area under for the rankings from the baseline model (column 2). The results for the other enhanced downgrade models are similar. In short, proxies for jumbo-CD default premiums or runoffs do not improve surveillance as independent variables in the downgrade-prediction model.

Another possibility is that we define community banks too rigidly. Raising the asset cutoffs, for example, might lead to more informative risk rankings because jumbo CDs at small banks act like core deposits. Also, jumbo-CD holders at larger institutions might be more sophisticated and, hence, more likely to monitor bank condition. To test for an "asset threshold" effect, we conduct all the out-of-sample performance comparisons (default-premium rankings vs. runoff rankings vs. downgrade-model rankings) on non-*de novo*, U.S. banks holding less than $1 billion (1999 dollars) in assets. Raising the size threshold to $1 billion—or to a variety of thresholds between $500 million and $1 billion for that matter—does not alter our findings. In out-of-sample tests based on a variety of asset thresholds, the downgrade-prediction model still emerges as the superior surveillance tool.

7. Conclusions

We find that feedback from the jumbo-CD market would have added little value in community-bank surveillance during the 1990s. Specifically, rank orderings produced by a downgrade-prediction model—a model chosen to benchmark current surveillance practices—would have significantly outperformed rank orderings based on jumbo-CD default premiums and runoffs throughout the decade. And jumbo-CD orderings would have improved little over random orderings. Finally, adding jumbo-CD screens to the downgrade-prediction model did not improve its out-of-sample forecasting record. We see this evidence as a partial validation of current surveillance practices.

Problems with the jumbo-CD data or frictions in the jumbo-CD market could explain the poor performance of the screens. As noted, all of our default-premium proxies are based on a noisy measure of yields. Large degrees of measurement error, due to marginal vs. average problems or accounting data vs. market data problems, would explain the lack of information content in the default-premium rankings. Another possibility is that jumbo-CD rates posted by community banks "cluster" around integers and even fractions (Kahn, Pennacchi, and Sopranzetti, 1999); such clustering would make rates less responsive to changes in bank risk and risk rankings based on those rates less informative. Still another possibility is that jumbo-CDs in community banks are relatively free of default risk—either because balances barely exceed $100,000 or because balances belong mostly to municipalities (and are therefore, by law, collateralized with blue-chip securities). A final possibility is that holders of community-bank jumbo CDs are simply noise traders, giving little thought to default-risk when deciding where to place their funds.

We are not persuaded that data problems and market frictions account solely for our findings; the long business-cycle expansion of the 1990s, no doubt, also plays a role. Over this period, bank profitability ratios, capital ratios, and reserves-to-problem loan ratios soared to near record highs, while failure rates plummeted to near record lows. In such an environment, jumbo-CD screens—no matter how accurately measured or precisely determined—are unlikely to convey information about risk because the benefits of monitoring are so low. Such an explanation would account for the successful use of average yields in bank-risk studies a decade ago—a time when financial distress was fairly common and failures were sharply rising. Such an explanation would also account for the inability of other researchers to find significant supervisory value in real-time, economic data from the 1990s. In short, the peculiarity of the sample period may compromise attempts to assess the supervisory value of any market-based signal. Assessing the supervisory potential of market feedback in bank surveillance may have to await data from all phases of the business cycle.

REFERENCES

Baer, Herbert and Elijah Brewer III. " Uninsured Deposits as a Source of Market Discipline: Some New Evidence." Federal Reserve Bank of Chicago *Economic Perspectives* (September/October 1986): 23-31.

Bassett, William F. and Thomas F. Brady. "The Economic Performance Small Banks, 1985-2000." *Federal Reserve Bulletin* (November 2001): 719-728.

Benston, George J. and George G. Kaufman. "Deposit Insurance Reform in the FDIC Improvement Act: The Experience to Date." The Federal Reserve Bank of Chicago, *Economic Perspectives* (Second Quarter 1998): 2-20.

Berger, Allen N. "The Relationship between Capital and Earnings in Banking." *Journal of Money, Credit and Banking* 27 (1995): 432-456.

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

Billett, Matthew T.; Jon A. Garfinkel; and Edward S. O'Neal. "The Cost of Market versus Regulatory Discipline in Banking," *Journal of Financial Economics* 48 (1998): 333-358.

Board of Governors of the Federal Reserve System and the United States Department of the Treasury. *The Feasibility and Desirability of Mandatory Subordinated Debt*, December 2000.

Board of Governors of the Federal Reserve System. "Using Subordinated Debt as an Instrument of Market Discipline." Staff Study 172, December 1999.

Brewer, Elijah III and Thomas H. Mondschean. "An Empirical Test of the Incentive Effects of Deposit Insurance." *Journal of Money, Credit, and Banking* 26 (1994): 146-64.

Cargill, Thomas F. "CAMEL Ratings and the CD Market." *Journal of Financial Services Research* 3 (1989): 347-358.

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

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

Cook, Douglas O. and Lewis J. Spellman. "Repudiation Risk and Restitution Costs: Toward an Understanding of Premiums on Insured Deposits." *Journal of Money, Credit and Banking* 26 (1994): 439-459.

Crane, Dwight B. "A Study of Interest Rate Spreads in the 1974 CD Market." *Journal of Bank Research* 7 (1976): 213-24.

Demsetz, Rebecca S. and Philip E Strahan. "Diversification, Size, and Risk at Bank Holding Companies." *Journal of Money, Credit, and Banking* 29 (1997): 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.

Ellis, David M., and Mark J. Flannery. "Does the Debt Market Assess Large Banks' Risk? Time Series Evidence from Money Center CDs," *Journal of Monetary Economics* 30 (1992): 481-502.

Emmons, William R.; R. Alton Gilbert; and Mark D. Vaughan. "Controlling Risk by Big Banks: Can Markets Help?" Federal Reserve Bank of St. Louis *Regional Economist* (October 2001): 5-9.

Evanoff, Douglas D. and Larry D. Wall. "Sub-debt Yield Spreads as Bank Risk Measures." *Journal of Financial Services Research* 20 (2001): 121-45.

Flannery, Mark J. "Retail Bank Deposits as Quasi-Fixed Factors of Production." *American Economic Review* 72 (1982): 527-536.

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

Gilbert , R. Alton, Andrew P. Meyer and Mark D. Vaughan. "Could a CAMELS Downgrade Model Improve Off-Site Surveillance?" Federal Reserve Bank of St. Louis *Review* (January/February 2002): 47-63.

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

Gunther, Jeffery W. and Robert R. Moore. "Financial Statements and Reality: Do Troubled Banks Tell All?" Federal Reserve Bank of Dallas *Economic and Financial Review* (Third Quarter 2000): 30-35

Gunther, Jeffery W.; Mark E. Levonian; and Robert R. Moore. "Can the Stock Market Tell Bank Supervisors Anything They Don't Already Know?" Federal Reserve Bank of Dallas *Economic and Financial Review* (Second Quarter, 2001): 2-9.

Hall, John R., Thomas King, Andrew P. Meyer, and Mark D. Vaughan. "Did FDICIA Improve Market Discipline? A Look at Evidence from the Jumbo-CD Market." Supervisory Policy Analysis working paper, Federal Reserve Bank of St. Louis, June 2002.

Hanc, George. "The Banking Crises of the 1980s and Early 1990s: Summary and Implications." In *History of the Eighties: Lessons for the Future*, volume 1. Washington, DC: Federal Deposit Insurance Corporation, 1997.

Hannan, Timothy and Gerald A. Hanweck. "Bank Insolvency Risk and the Market for Large Certificates of Deposit." *Journal of Money, Credit, and Banking* 20 (1988): 203-211.

James, Christopher. "Heterogeneous Creditors and LDC Lending." *Journal of Monetary Economics* 25 (1990): 325-346.

James, Christopher. "The Losses Realized in Bank Failures." *Journal of Finance* 46 (1991): 1223-1242.

James, Christopher. "The Use of Loan Sales and Standby Letters of Credit by Commercial Banks." *Journal of Monetary Economics* 22 (1988): 395-422.

Kahn, Charles; George Pennacchi; and Ben Sopranzetti. "Bank Deposit Rate Clustering: Theory and Empirical Evidence." *Journal of Finance* 56 (1999): 2185-2214.

Keeley, Michael C. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80 (1990): 1183-1198.

Krainer, John and Jose A. Lopez. "Incorporating Equity Market Information into Supervisory Monitoring Models." Working Paper, Research Department, Federal Reserve Bank of San Francisco, August 30, 2001.

Lang, William W. and Douglas Robertson. "Analysis of Proposals for a Minimum Subordinated Debt Requirement." Comptroller of the Currency, Economic and Policy Analysis Working Paper #2000-4, March 2000.

Martinez-Peria, Maria Soledad and Sergio L. Schmukler. "Do Depositors Punish Banks for Bad Behavior? Market Discipline, Deposit Insurance, and Banking Crises." *Journal of Finance* 56 (2001): 1029-1051.

Meyer, Laurence H. "Market Discipline as a Complement to Bank Supervision and Regulation," Speech before the Conference on Reforming Bank Capital Standards, Council on Foreign Relations, New York, June 14, 1999.

Morgan, Donald P. and Kevin J. Stiroh. "Market Discipline of Banks: The Asset Test." *Journal of Financial Services Research* 20 (2001): 195-208.

Morris, Marc D. and John R. Walter. "Large Negotiable Certificates of Deposit." In *Instruments of the Money Market.* Richmond, VA: Federal Reserve Bank of Richmond, 1993.

Park, Sangkyun. "Market Discipline by Depositors: Evidence from Reduced-Form Equations." *Quarterly Review of Economics and Finance* 35 (1995): 497-514.

Park, Sangkyun and Stavros Peristiani. "Market Discipline by Thrift Depositors." *Journal of Money, Credit, and Banking* 30 (1998): 347-364.

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: 6-13.

Reidhill, Jack and John O'Keefe. "Off-Site Surveillance Systems." In *History of the Eighties: Lessons for the Future*, volume 1. Washington, DC: Federal Deposit Insurance Corporation, 1997.

Roll, Richard. "What Every CFO Should Know About Scientific Progress in Financial Economics: What is Known, and What Remains to be Resolved." *Financial Management* 23, no. 2 (1994): 69-75.

Sullivan, Ryan; Allan Timmerman; and Halbert White. "Data-Snooping, Technical Trading Rule Performance and the Bootstrap." *Journal of Finance* 54 (1999): 1647-1691.

Table 1

How common were downgrades to "unsatisfactory" condition in the 1990s?

This table demonstrates that downgrades from safe-and-sound to unsatisfactory status were common in the 1990s, thereby allowing re-estimation of a downgrade-prediction model on a yearly basis. Specifically, the far right column shows the number of sample banks rated as safe and sound (composite CAMELS of "one" or "two") at each year-end that were downgraded to unsatisfactory status (composite CAMELS of "three," "four," or "five") in the following 12 months. The data reveal that two-rated banks were much more likely to tumble into unsatisfactory condition than one-rated banks. Although downgrades to unsatisfactory condition were common throughout the decade, they became less frequent in the mid-1990s as overall banking conditions improved.

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	2,190	34	1.55	699
	2	5,482	665	12.13	
1991	1	1,959	22	1.12	424
	2	5,277	402	7.62	
1992	1	2,291	7	0.31	182
	2	5,980	175	2.93	
1993	1	2,911	9	0.31	162
	2	5,726	153	2.67	
1994	1	3,091	8	0.26	102
	2	4,885	94	1.92	
1995	1	3,284	10	0.30	127
	2	4,522	117	2.59	
1996	1	3,242	7	0.22	125
	2	3,741	118	3.15	
1997	1	3,022	19	0.63	152
	2	3,105	133	4.28	
1998	1	3,067	19	0.62	201
	2	3,047	182	5.97	
1999	1	3,088	12	0.39	190
	2	3,320	178	5.36	
2000	1	3,226	3	0.09	49
	2	3,684	46	1.25	

Table 2

Does published evidence point to "risk pricing" by U.S. jumbo-CD holders?

This table summarizes research on risk pricing by jumbo-CD holders in the United States. (We use the term "bank" to refer bank holding companies as well as banks.) These studies employed both cross-section and time-series techniques and used a variety of risk proxies and control variables. Overall, the evidence suggests that jumbo-CD holders price bank-specific risk.

Authors	Issuer of Jumbo Certificate of Deposit	Sample Dates	Was Bank-Specific Risk Priced?
Crane (1976)	Bank	1974	Somewhat
Baer and Brewer (1986)	Bank	1979-82	Yes
Hannan and Hanweck (1988)	Bank	1985	Yes
James (1988)	Bank	1984-86	Yes
Cargill (1989)	Bank	1984-86	Yes
James (1990)	Bank	1986-87	Yes
Keeley (1990)	Bank	1984-86	Yes
Ellis and Flannery (1992)	Bank	1982-88	Yes
Cook and Spellman (1994)	Thrift	1987-88	Yes
Brewer and Mondschean (1994)	Thrift	1987-89	Yes
Park (1995)	Bank	1985-92	Yes
Park and Peristiani (1998)	Thrift	1987-91	Yes
Hall, King, Meyer, and Vaughan (2002)	Bank	1988-90, 1993-95	Yes

Table 3

Do our measures of jumbo-CD yields and runoffs “price” bank-specific risk?

This table displays the results of regressions of jumbo-CD yields and runoffs on failure probability and controls. The failure-probability measure—produced by the SEER risk-rank model—serves as a summary measure of bank-specific risk. The other explanatory variables control for influences on yields and runoffs besides bank-specific risk. These influences include the percentage of the bank’s jumbo-CD portfolio with fixed rates, the average maturity of the bank’s jumbo-CD portfolio, the bank’s asset size, the bank’s relationship with a holding company (“1” if a member), the bank’s propensity to tap national funding markets (“1” if a user of brokered deposits), and the bank’s degree of market power (“1” if in an MSA). State dummies (not reported) control for differences in tax laws and regional economic conditions. Time dummies and maturity-adjusted Treasury yields control for business-cycle effects. The regressions use pooled end-of-year data for 1988 through 1996.

The SEER failure probability has a statistically significant but economically small impact on yields and runoffs. For example, the coefficient on SEER failure probability is significant at the one-percent level in each equation. At the same time, these coefficients imply that jumping from the sample-mean failure probability of 0.545 percent to 3.522 percent (one standard deviation higher) would have depressed the average sample bank's return on assets by only one-quarter basis point—from 1.1898 percent to 1.1873 percent.

The economically small risk effect suggests that the jumbo-CD market puts little direct pressure on banks to maintain safety and soundness. It does not suggest, however, that ordinal rankings based on jumbo-CD screens lack information content. Again, only out-of-sample performance tests can determine whether these rankings improve upon rankings obtained with current surveillance tools.

		Jumbo-CD Yield Regression			**Jumbo-CD Runoff Regression**		
	Independent Variable	**Coef.**	**Std. Err.**		**Coef.**	**Std. Err.**	
	Intercept	-0.4774	0.1337	***	14.9297	1.7430	***
Bank-Specific Risk Variable	SEER failure probability	0.0073	0.0014	***	-0.1568	0.0180	***
Control Variables	Fixed-rate percentage	-0.0024	0.0006	***	0.0212	0.0083	**
	Mty-weighted Treasury yield	1.1126	0.0203	***	-1.9718	0.2643	***
	Maturity	-0.0758	0.0141	***	-0.6689	0.1841	***
	Maturity-Treasury interactive	-0.1915	0.0136	***	-0.2069	0.1769	
	Holding-company dummy	-0.0087	0.0104		-0.0490	0.1357	
	Brokered-deposit dummy	0.1408	0.0159	***	1.1418	0.2067	***
	MSA dummy	-0.0514	0.0088	***	0.2401	0.1143	**
Regression Statistics	Number of observations	109,635			109,635		
	R^2	0.5747			0.0344		

Table 4a

Are residuals from the "jumbo-CD yield on remaining maturity" regressions potentially good proxies for default premiums?

As one measure of default premiums, we use residuals from year-by-year regressions of jumbo-CD yields on remaining maturities. This table displays the results for the 1991 through 1996 regressions. Regression coefficients appear on top; standard errors appear below in parentheses. Three stars indicate statistical significance at the one-percent level; two stars indicate significance at the five-percent level; and one star indicates significance at the 10-percent level. The intercept estimates approximate the average fourth-quarter yields on jumbo CDs maturing in less than three months. The intercept estimate plus the coefficient estimate on "3 TO 12" approximates the average yield on instruments maturing in three months to one year. The intercept estimate plus the coefficient estimate on "12 TO 60" approximates the yield on jumbo CDs maturing in the next 12 months to five years. The coefficient on OVER 60 has a comparable interpretation. Coefficient signs and magnitudes for the maturity variables reveal that in each year yields rose with time-to-maturity, implying that the equation residuals may serve as good proxies for default premiums.

	Fourth quarter of:					
Independent variables	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**
Intercept	0.0572*** (0.0006)	0.0367*** (0.0005)	0.0332*** (0.0004)	0.0433*** (0.0004)	0.0538*** (0.0006)	0.0512*** (0.0005)
3 TO 12	0.0090*** (0.0013)	0.0122*** (0.0009)	0.0060*** (0.0008)	0.0012 (0.0008)	0.0051*** (0.0010)	0.0041*** (0.0009)
12 TO 60	0.0256*** (0.0021)	0.0276*** (0.0012)	0.0253*** (0.0009)	0.0117*** (0.0009)	0.0091*** (0.0012)	0.0107*** (0.0011)
OVER 60	0.0137 (0.0113)	0.0470*** (0.0065)	0.0290*** (0.0045)	0.0085* (0.0047)	0.0182** (0.0082)	0.0024 (0.0078)
Adjusted R^2	0.0149	0.0531	0.0690	0.0164	0.0070	0.0099
F-statistic for significance of the three independent variables	59.338***	208.245***	264.166***	57.65***	23.68***	30.91***

Percentage of deposits with remaining maturity:	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**
Within 3 months	53.17%	48.93%	45.90%	45.01%	45.62%	43.55%
3 to 12 months	29.43	30.00	30.38	31.35	33.8	40.92
12 to 60 months	15.99	18.94	21.32	21.88	19.11	15.09
Over 60 months	1.41	2.13	2.41	1.76	1.47	0.44
Yield on three-month CDs in the secondary market	4.91%	3.44%	3.28%	5.86%	5.72%	5.41%

Table 4b

Are residuals from the "jumbo-CD yield on remaining maturity" regressions potentially good proxies for default premiums?

This table displays the results from the 1997, 1998, and 1999 regressions. The long-term maturity buckets for these three years differed from the buckets for 1991 through 1996. Again, regression coefficients appear on top; standard errors appear below in parentheses. Three stars indicate statistical significance at the one-percent level; two stars indicate significance at the five-percent level; and one star indicates significance at the 10-percent level. Once again, coefficient signs and magnitudes for the maturity variables reveal that in each year yields rose with time-to-maturity, implying that the equation residuals may serve as good proxies for default premiums.

	Fourth quarter of:		
Independent variables	**1997**	**1998**	**1999**
Intercept	0.0531*** (0.0008)	0.0493*** (0.0007)	0.0495*** (0.0004)
3 TO 12	0.0044*** (0.0015)	0.0074*** (0.0011)	0.0031*** (0.0007)
12 TO 36	0.0084*** (0.0019)	0.0127*** (0.0015)	0.0078*** (0.0010)
OVER 36	0.0072 (0.0047)	0.0154*** (0.0031)	0.0135*** (0.0020)
Adjusted R^2	0.0026	0.0127	0.0134
F-statistic for significance of the three independent variables	8.83***	38.11***	39.78***

Percentage of deposits with remaining maturity:	**1997**	**1998**	**1999**
Within 3 months	40.44%	39.69%	38.95%
3 to 12 months	42.80	43.60	45.42
12 to 36 months	13.90	13.36	12.71
Over 36 months	2.85	3.35	2.92
Yield on three-month CDs in the secondary market	5.73%	5.20%	6.06%

Table 5

Which factors help predict downgrades to unsatisfactory status?

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 reduces the likelihood of a downgrade over the next two years.

	Independent Variables	Symbol	Hypothesized Relationship
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	–
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	–
Liquidity Risk	Book value of investment securities as a percentage of total assets.	SECURITIES	–
	Deposits > $100M (jumbo CDs) as a percentage of total assets.	LARGE-TIME-DEPOSITS	+
Control 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	+

Table 6

How well does the downgrade-prediction model fit the data?

This table presents the results of probit regressions of downgrade status on financial performance ratios and control variables. The dependent variable equals "1" for a downgrade and "0" for no downgrade in calendar year *t+1* and *t+2*. Values for independent variables are taken from the fourth quarter of year *t*. Standard errors appear in parentheses below the coefficients. One asterisk denotes statistical significance at the 10-percent level, two asterisks at the five-percent level, and three at the one-percent level. The pseudo-R^2 indicates the approximate proportion of the variance in downgrade status that is explained by the model.

Overall, the downgrade-prediction model fit the data relatively well. For each of the eight regressions, the log-likelihood test statistic allows rejection of the hypothesis that all model coefficients equal zero. In addition, eight of the 13 regression variables are significant with the predicted sign in all eight years, and all of the variables were significant in at least some years. Although model performance for the decade is good overall, in-sample fit does deteriorate slightly in later years due to the decline in downgrade frequency.

		Period of Downgrade in CAMELS rating			
	Independent Variable	**1990-1991**	**1991-1992**	**1992-1993**	**1993-1994**
	Intercept	-2.087*** (0.246)	-0.957*** (0.264)	-0.081 (0.318)	0.048 (0.375)
Credit Risk	PAST-DUE-30	0.112*** (0.021)	0.150*** (0.022)	0.136*** (0.026)	0.174*** (0.033)
	PAST-DUE-90	0.376*** (0.039)	0.328*** (0.040)	0.239*** (0.047)	0.304*** (0.060)
	NONACCRUING	0.235*** (0.029)	0.199*** (0.030)	0.291*** (0.036)	0.178*** (0.045)
	OREO	0.220*** (0.030)	0.216*** (0.032)	0.145*** (0.031)	0.167*** (0.043)
	COMMERCIAL-LOANS	0.009*** (0.003)	0.013*** (0.003)	0.009** (0.004)	0.002 (0.005)
	RESIDENTIAL-LOANS	-0.005** (0.002)	-0.004 (0.002)	-0.004 (0.003)	-0.005 (0.003)
Leverage Risk	NET-WORTH	-0.054*** (0.010)	-0.048*** (0.011)	-0.073*** (0.013)	-0.074*** (0.013)
	ROA	-0.241*** (0.035)	-0.318*** (0.039)	-0.200*** (0.043)	-0.263*** (0.051)
Liquidity Risk	SECURITIES	-0.016*** (0.002)	-0.017*** (0.002)	-0.013*** (0.002)	-0.009*** (0.003)
	LARGE-TIME-DEPOSITS	0.017*** (0.003)	0.019*** (0.003)	0.015*** (0.004)	0.017*** (0.005)
Control Variables	SIZE	0.079*** (0.017)	-0.029 (0.019)	-0.125*** (0.024)	-0.147*** (0.030)
	CAMELS-2	0.633*** (0.062)	0.517*** (0.068)	0.509*** (0.087)	0.432*** (0.102)
	BAD-MANAGE	0.488*** (0.051)	0.401*** (0.054)	0.478*** (0.061)	0.466*** (0.069)
	Number of Observations	8,494	8,065	7,836	8,058
	Pseudo-R^2	0.219	0.226	0.208	0.161
	-2 log likelihood testing whether all coefficients (except the intercept) = 0	7211.785***	6053.423***	3946.413***	2444.832***

Table 6 (Continued)
How well does the downgrade-prediction model fit the data?

		Period of Downgrade in CAMELS rating			
	Independent Variable	**1994-1995**	**1995-1996**	**1996-1997**	**1997-1998**
	Intercept	-0.780* (0.402)	-0.011 (0.436)	-0.162 (0.415)	-1.371*** (0.388)
Credit Risk	PAST-DUE-30	0.119*** (0.035)	0.164*** (0.035)	0.189*** (0.033)	0.093*** (0.029)
	PAST-DUE-90	0.296*** (0.064)	0.322*** (0.074)	0.399*** (0.064)	0.347*** (0.057)
	NONACCRUING	0.192*** (0.046)	0.145*** (0.051)	0.157*** (0.046)	0.187*** (0.044)
	OREO	0.192*** (0.044)	0.153*** (0.052)	0.091 (0.059)	0.156** (0.067)
	COMMERCIAL-LOANS	0.007 (0.005)	0.013*** (0.005)	0.010** (0.005)	0.005 (0.005)
	RESIDENTIAL-LOANS	-0.002 (0.004)	-0.013*** (0.004)	-0.009*** (0.003)	0.000 (0.003)
Leverage Risk	NET-WORTH	-0.032** (0.014)	-0.034*** (0.013)	-0.036*** (0.014)	-0.020* (0.012)
	ROA	-0.229*** (0.052)	-0.164*** (0.038)	-0.393*** (0.063)	-0.110** (0.044)
Liquidity Risk	SECURITIES	-0.002 (0.003)	-0.010*** (0.003)	-0.015*** (0.003)	-0.011*** (0.003)
	LARGE-TIME-DEPOSITS	0.024*** (0.005)	0.020*** (0.005)	0.023*** (0.005)	0.019*** (0.004)
Control Variables	SIZE	-0.150*** (0.033)	-0.202*** (0.035)	-0.150*** (0.032)	-0.101*** (0.030)
	CAMELS-2	0.594*** (0.104)	0.589*** (0.103)	0.501*** (0.099)	0.760*** (0.093)
	BAD-MANAGE	0.389*** (0.075)	0.510*** (0.078)	0.406*** (0.083)	0.535*** (0.081)
	Number of Observations	8,664	8,666	8,574	8,306
	Pseudo-R^2	0.150	0.188	0.223	0.182
	-2 log likelihood testing whether all coefficients (except the intercept) = 0	2130.311***	1977.293***	2189.627***	2347.306***

PAST-DUE-30	Loans over 30 days past due as a percentage of total loans	NET-WORTH	Equity less goodwill as a percentage of total assets
PAST-DUE-90	Loans over 90 days past due as a percentage of total loans	ROA	Net income as a percentage of total assets.
		SECURITIES	Book value of securities as a percentage of total assets
NONACCRUING	Loans on nonaccrual status as a percentage of total loans	LARGE-TIME-DEPOSITS	Large denomination time deposit liabilities as a percentage of total assets.
OREO	Other real estate owned as a percentage of total assets	SIZE	Natural logarithm of total assets, in thousands of dollars.
COMMERCIAL-LOANS	Commercial and industrial loans as a percentage of total assets	CAMELS-2	Dummy variable equal to 1 if bank has a CAMELS rating of 2.
RESIDENTIAL-LOANS	Residential real-estate loans as a percentage of total assets	BAD-MANAGE	Dummy variable equal to 1 if the bank's Management rating is worse than composite CAMELS rating.

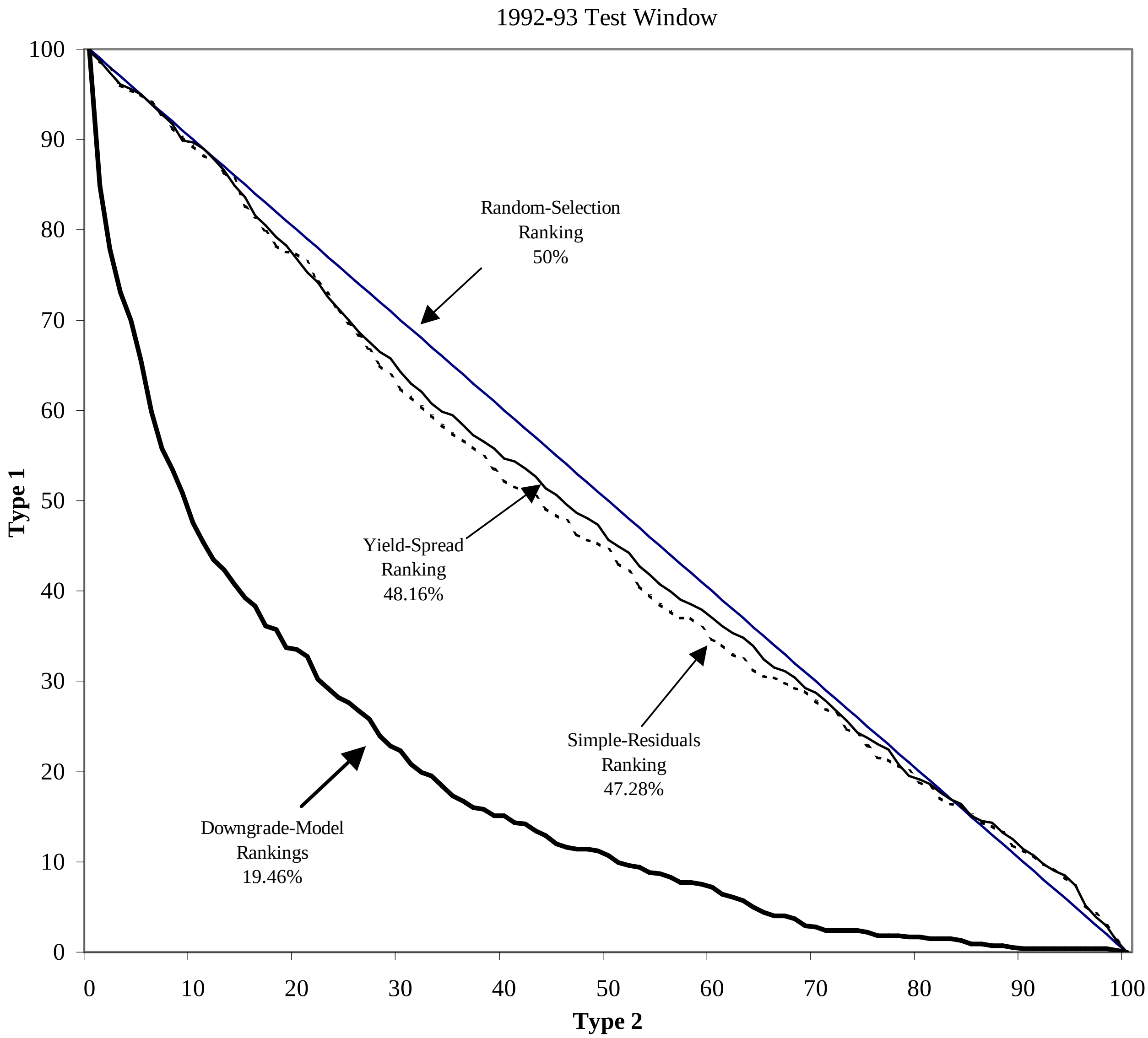
Table 7
Do measures of jumbo-CD default premiums and runoffs add value in community-bank surveillance?

This table summarizes the out-of-sample performance of risk rankings obtained from the downgrade-prediction model and the jumbo-CD data. Performance comparisons are based on areas under power curves. Power curves illustrate the trade off between type-one errors (the percentage of missed downgrades) and type-two errors (the percentage of over-predicted downgrades) for the risk rankings produced by each surveillance tool. A smaller area under the curve implies a lower rate of both types of errors and, thus, a better surveillance tool. Each cell in columns two through eight contains the area under the power curve for a specific risk ranking over a specific test window.

The evidence suggests that jumbo-CD rankings add little value in community-bank surveillance. Risk rankings produced by the downgrade-prediction model (column two) perform considerably better than a random ranking (area of 50 percent). At the same time, risk rankings based on jumbo-CD default premiums or runoffs (columns four through eight) perform little better than random risk rankings. Moreover, adding jumbo-CD default premiums to the downgrade-prediction model (column three) did not improve its out-of-sample performance.

Downgrade Years (1)	Downgrade Model (2)	Downgrade Model with Yield Spread (3)	Yield-Spread (4)	Simple Residuals (Yields) (5)	Simple Runoff (6)	Complex Yield Residuals (7)	Complex Runoff Residuals (8)
1992-93	19.46%	20.42%	48.16%	47.28%	49.54%	48.36%	52.06%
1993-94	21.87%	23.34%	47.23%	46.88%	47.77%	39.75%	52.22%
1994-95	23.19%	23.55%	51.01%	50.49%	45.53%	48.27%	47.25%
1995-96	17.85%	18.68%	43.09%	43.32%	47.88%	38.90%	46.12%
1996-97	15.61%	18.03%	47.45%	47.46%	45.87%	42.11%	41.55%
1997-98	18.84%	19.47%	45.61%	46.26%	45.23%	42.31%	44.50%
1998-99	21.12%	21.59%	47.25%	47.21%	47.81%	43.50%	49.28%
1999-00	19.14%	18.80%	46.10%	46.13%	44.09%	46.35%	48.00%
Mean All Years	**19.64%**	**20.49%**	**46.99%**	**46.88%**	**46.72%**	**43.69%**	**47.62%**

Figure 1
Do measures of jumbo-CD default premiums and runoffs add value in community-bank surveillance?



This figure contains power curves corresponding to 1992-93 out-of-sample test of three risk rankings--the ranking produced by the downgrade-prediction model and the rankings produced by the two proxies for default premiums (yield spreads and simple residuals) Each curve portrays rates of type-one ("missed" downgrades) and type-two errors ("over-predicted" downgrades) for a specific risk ranking. Because simultaneous reduction of both errors is desirable, we use the area under a power curve, expressed as a percentage of the total area in the box, to compare the out-of-sample performance of each ranking. The 50 percent line depicts the power curve produced by random risk rankings. This line offers a benchmark for judging the economic significance of performance differences. The power curves from the 1992-93 test show that the downgrade-prediction model significantly outperforms both jumbo-CD default premiums as a tool for generating risk rankings. Indeed, rankings based on default premiums hardly improve over rankings compiled through random selection.