WORKING PAPER SERIES

Can Feedback from the Jumbo-CD Market Improve Bank Surveillance?

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

Working Paper 2003-041A
http://research.stlouisfed.org/wp/2003/2003-041.pdf

December 2003

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

Can Feedback from the Jumbo-CD Market Improve Bank Surveillance?

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

May 2003

JEL Codes: G21, G28
Keywords: Market Discipline, 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—as well as the following economists—Rosalind Bennett, Mark Carey, Bill Emmons, Doug Evanoff, Mark Flannery, John Jordan, John Hall, Jim Harvey, Tom King, John Krainer, Bill Lang, Jose Lopez, Dan Nuxoll, Evren Ors, Jeremy Piger, Jim Thomson, Sherrill Shaffer, Scott Smart, and Larry Wall—for helpful comments and discussions. We also profited from exchanges with seminar participants at Baylor University, the Federal Deposit Insurance Corporation, 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 (Department of Economics and the Olin School of Business). 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 Bank Surveillance?

R. Alton Gilbert, Mark D. Vaughan, and Andrew P. Meyer
Federal Reserve Bank of St. Louis

Abstract

We examine the value of jumbo certificate-of-deposit (CD) signals in bank surveillance. To do so, we first construct proxies for default premiums and deposit runoffs and then rank banks based on these risk proxies. Next, we rank banks based on the output of a logit model typical of the econometric models used in off-site surveillance. Finally, we compare jumbo-CD rankings and surveillance-model rankings as tools for predicting financial distress. Our comparisons include eight out-of-sample test windows during the 1990s. We find that rankings obtained from jumbo-CD data would not have improved on rankings obtained from conventional surveillance tools. More importantly, we find that jumbo-CD rankings would not have improved materially over random rankings of the sample banks. These findings validate current surveillance practices and, when viewed with other recent empirical tests, raise questions about the value of market signals in bank surveillance.

JEL Codes: G21, G28
Keywords: Market Discipline, 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 (Meyer, 2001; Board of Governors, 2000; and Board of Governors, 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 risky institutions in rising yields or difficulties rolling over maturing debt—will pressure bank managers to maintain safety and soundness (Lang and Robertson, 2002).

Even if the subordinated-debt market—or any other market for bank claims—applies little direct pressure on bank managers, market-generated risk assessments could still contribute to one component of supervisory review—off-site surveillance.[1] Off-site surveillance involves the use of accounting data and anecdotal evidence to schedule on-site examinations and to monitor bank progress in addressing previously identified deficiencies. Market-generated risk assessments could contribute to surveillance in three ways. First, market signals might flag problem banks missed by conventional surveillance tools. Second, market signals might uncover emerging problems before conventional surveillance tools. Third, market signals might increase confidence about risk assessments produced by conventional surveillance tools (Flannery, 2001).

[1] Bliss and Flannery (2001) looked for evidence that managers of bank holding companies respond to market pressure to contain risk. They found none, though Rajan (2001) questioned the ability of their framework to unearth strong evidence of managerial responses.

To date, discussions about harnessing sub-debt or other market signals for supervisory purposes have centered on large complex banking organizations. Discussions have centered on large banks because the supervisory benefits are thought to be the highest and the compliance costs lowest for these institutions (Emmons, Gilbert, and Vaughan, 2001). The benefits are perceived as the 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 costs are thought to be the lowest because most of these institutions already tap national financial markets routinely. For example, at year-end 2002, 41 of the 50 largest commercial banks, and 48 of the 50 largest bank holding companies, had subordinated debt outstanding.

But before forcing all large banking organizations to issue subordinated debt in a standardized form, supervisors should make sure that existing securities do not produce useful risk signals. A mandatory security issue is an implicit capital-structure tax. We know of no evidence suggesting that the welfare loss from such a tax is negligible. That most large banks currently issue sub-debt does not imply a negligible loss. Voluntary issuance varies considerably over time with market conditions. For example, issuance of subordinated debt by the top 50 banking organizations rose from less than 10 per year during 1988-1990, to almost 86 per year during 1995-98, only to fall to 42 during 1999 (Covitz, Hancock, and Kwast, 2002). At any given time, the banks with no outstanding sub-debt may be just those institutions for which issuance is the most costly and risk signals the most valuable. Moreover, those banks now issuing sub-debt may not be choosing maturity structures likely to produce the most valuable supervisory signals, so even they would face an implicit tax. The uncertain impact of a sub-debt tax—together with the lack of conclusive evidence of the tax's supervisory value—suggest that supervisors should first try to extract useful signals about safety and soundness from claims banks already issue.[2]

A logical place to look for useful risk signals is the market for jumbo certificates of deposit (CDs)—time deposits with balances exceeding the $100,000 ceiling for deposit-insurance coverage. As noted, only the very largest banks and bank holding companies now issue subordinated debt. Similarly, only about 700 of the largest bank holding companies have publicly traded stock. Even though these large holding companies are the most important economically, the focus of off-site surveillance—indeed of prudential supervision in the U.S.—is at the bank level. And a negative risk signal from holding company claims would not, by itself, help supervisors identify troubled subsidiary banks. Unlike sub-debt or public equity, jumbo CDs are an important part of the capital structure of all commercial banks. At year-end 2002, U.S. commercial banks on average funded 12.7 percent of their assets with jumbo CDs (unweighted mean). For banks holding more than $500 million in assets, the year-end 2002 jumbo-CD-to-total-asset ratio was 12.8 percent; for banks holding less than $500 million in assets, the ratio was 12.0 percent. Research over the past 25 years has repeatedly confirmed that jumbo-CD holders perceive and price default risk.[3]

On top of offering a potential improvement in large-bank surveillance, signals from the jumbo-CD market could prove useful in the off-site monitoring of community banks. Community banks are relatively small institutions, specializing in making loans to and taking deposits from distinct regions such as small towns or city suburbs. Many of these banks operate under extended exam schedules, with up to 18 months elapsing between full-scope examinations. This extended schedule diminishes the information content of community-bank financial statements, thereby reducing the effectiveness of off-site

---

[2] Available evidence suggests that holders of bank-issued subordinated-debt do price default risk. Flannery and Sorescu (1996), for example, document increases in the risk sensitivity of sub-debt yields as the U.S. government retreated from "too-big-to-fail" guarantees. This evidence does not, however, imply that sub-debt signals have significant supervisory value. Bliss (2001) argues that the poor microstructure of the sub-debt market renders the risk signals from default spreads unreliable. Evanoff and Wall (2001) provide supporting evidence—finding that sub-debt yields barely outperform regulatory capital ratios—themselves poor proxies for overall supervisory assessments—as predictors of financial distress.

[3] See Table 1 for a survey of published research.

supervisory monitoring.[4] It is possible that the holders of community-bank jumbo CDs, because their own money is at stake, supplement published income statements and balance sheets with independent research. It is also possible that information about safety and soundness leaks to uninsured depositors from bank boards of directors, bodies that typically include 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 financial statements.

Another reason to consult the jumbo-CD market for help in community-bank surveillance is that these banks represent a non-trivial 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. Moreover, 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 expensive to liquidate—more fixed assets and fewer securities—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, improve large- as well as community-bank surveillance, available data permit only the construction of crude proxies for the desired risk signals. 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 most 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

[4] Verification of financials is one important source of value created by bank 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).

for default premiums. 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.

Measurement-error problems do not, by themselves, imply that accounting-based jumbo-CD signals are valueless in off-site surveillance. Jumbo-CD runoff—a quantity response to changes in bank risk—can be measured relatively error-free with accounting data. Moreover, financial-distress models based on accounting data have been a cornerstone of regulatory and private-sector surveillance for decades (Altman and Saunders, 1997). Indeed, bank surveillance models give heavy weight to book-value measures of credit risk and capital protection, both of which are known to suffer from serious measurement error (Barth, Beaver, and Landsman, 1996). Finally, and most importantly, the supervisory value of jumbo-CD signals—or any market signal for that matter—depends not on the power of the signal alone, but rather on the power of the signal adjusted for the cost of extracting it. Suppose, for example, that a sub-debt signal would generate $21 worth of supervisory value, expressed in terms of the present value of saved failure-resolutions costs. Further suppose, that the cost of extracting this signal—in terms of the welfare loss of the capital structure tax and the cost of revising surveillance practices to incorporate sub-debt signals—is $16. In contrast, suppose an accounting-based jumbo-CD signal would generate only $7 worth of supervisory value—one-third of the value of a pure market signal—because of measurement error. But suppose that the cost of obtaining the signal is only $1 because banks already report jumbo-CD data on their financial statements and because additional accounting-based signals can easily be integrated into current surveillance practices. In this hypothetical scenario, the flawed jumbo-CD signals contribute more value at the margin ($6 compared with $5 for the sub-debt signal). In short, the value of jumbo-CD data in bank surveillance is ultimately an empirical issue.

Properly assessing the supervisory value of market signals requires the use of a benchmark for current surveillance practices. It is not enough to note that a signal reacts to bank risk contemporaneously or forecasts emerging safety-and-soundness problems successfully because supervisors already have systems in place for these purposes. The acid test of market signals is whether they improve materially upon current practice. Four recent papers assess the supervisory value of market signals with a current-practices benchmark. Evanoff and Wall (2001) compare regulatory-capital ratios and subordinated-debt yields as predictors of supervisory ratings, finding that sub-debt yields modestly outperform capital ratios in one-quarter-ahead tests. Gunther, Levonian, and Moore (2001) add estimated KMV default probabilities—a risk measure drawn from the equity market—to an econometric model designed to predict holding-company supervisory ratings with accounting data. They note an improvement in in-sample fit. Krainer and Lopez (forthcoming) 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 value added by measuring improvement in one-quarter-ahead forecasts. Like Evanoff and Wall, they find that market data provide only a modest boost to out-of-sample performance. Finally, Curry, Elmer and Fissel (2001) add other signals derived from the market for bank equity to an econometric surveillance model designed to predict four-quarter-ahead supervisory ratings. They find some evidence of an improvement in both in- and out-of-sample performance. Their out-of-sample test, however, relies on a contemporaneous holdout sample rather than a period-ahead sample that would mimic current surveillance practices.

Building on the emerging literature on market signals in bank surveillance, we compare the performance of jumbo-CD signals and econometric models in the off-site surveillance of commercial banks. Specifically, we combine our yield measure with Treasury yields to obtain proxies for jumbo-CD default premiums. We then compile 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 two years. We then order sample banks by these 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 performance of orderings based on jumbo-CD runoff. As robustness checks, we evaluate the predictive power of jumbo-CD orderings obtained with a more sophisticated technique as well as orderings of different cuts of the sample banks. We also assess orderings obtained when default premiums and deposit runoff are added to the benchmark surveillance model. Unlike other research, we assess value added with a surveillance model actually used by a Federal bank supervisor, and we employ out-of-sample timing conventions that mimic current surveillance practices among Federal bank supervisors. Unlike most of the other research, we conduct all our tests on bank data rather than holding company data—an important distinction because the principal focus of off-site surveillance among Federal supervisors is the bank, not the holding company.

Our empirical tests indicate that feedback from the jumbo-CD market would have contributed nothing to large-bank or community-bank surveillance in the 1990s. In all eight out-of-sample test windows, orderings based on output from the econometric surveillance model significantly outperform orderings based on jumbo-CD default premiums or runoffs. More important, the jumbo-CD orderings improve little on random orderings. These findings are robust to different extraction techniques and sample cuts. Finally, including measures of jumbo-CD default premiums or runoff in the econometric surveillance model does not enhance its out-of-sample performance. Taken together, our evidence suggests that jumbo-CD signals would not have flagged problem institutions missed by current surveillance tools, would not have flagged problem institutions before current surveillance tools, and would not have reduced uncertainty about problem institutions flagged by current surveillance tools. These findings validate current surveillance practices and, when viewed with other recent empirical tests, raise questions about the value of market signals in bank surveillance.

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 (L), and market-risk sensitivity (S).[5] At the close of each exam, a grade of one (best) through five (worst) is awarded 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 December 31, 2001, supervisors classified just over 6.41 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 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 increase the number of market-risk specialists on the exam team.

---

5 The "S" component was introduced in January 1997. Before that time, examiner assessments of market risk were embedded in the other five components. In our empirical work, we use CAMELS composites for post-1996 tests and CAMEL composites for pre-1997 tests.

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 surveillance models also combine information from financial ratios. These models rely on statistical tests rather than human judgment to boil financial statements down to an index number summarizing bank condition. In past comparisons, econometric surveillance 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 System 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 section at each Reserve Bank, in turn, follows up on each "red-flagged" institutions. The Federal Deposit Insurance Corporation (FDIC) and the Office of the Comptroller of the Currency (OCC) use similar econometric models as a part of the off-site surveillance regimen for the banks they supervise (Reidhill and O'Keefe, 1997).

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 that a satisfactory bank (CAMELS one or two composite rating) will tumble into unsatisfactory condition (composite CAMELS three, four, or five) in the coming eight quarters. In theory, a model designed to predict downgrades could improve upon the SEER framework. Although few banks failed in the 1990s, many banks suffered downgrades to unsatisfactory status, so a downgrade-prediction model can be re-estimated quarterly. (See Table 2 for data on downgrade frequency.) The SEER risk-rank model, in contrast, was estimated on 1985-1991 data, and model coefficients have been frozen since the original estimation. More important, a downgrade-prediction model may flag banks not currently under close scrutiny. Institutions with three, four, or five composite ratings fail at much higher rates than institutions with one or two ratings, so unsatisfactory institutions already receive close 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). In addition, a downgrade-prediction model is used to supplement SEER output in one Federal Reserve District. 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 surveillance model. 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 leverage risk or credit risk. As noted, previous evidence from the jumbo-CD market confirms risk pricing. 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 (Kroszner and Strahan, 2001; Benston and Kaufman, 1998). As a check, we regressed the simple measures of jumbo-CD yields and runoffs on the financial ratios used in downgrade-prediction model and several control variables suggested by recent research.[6] These results confirm responsiveness to bank risk, though this response is economically small. The economically small response suggests that jumbo-CD default premiums and runoffs put little direct pressure on bank managers. It does not imply, however, that rankings based on these measures lack meaningful information about bank condition. Again, only out-of-sample performance tests can determine whether jumbo-CD rankings improve upon rankings obtained with current surveillance tools.

3. The Data

Our data set includes quarterly accounting data for all U.S. commercial 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. The income and balance sheet data come from the Reports of Condition and Income (the call reports), which are collected under the auspices of the Federal Financial Institutions Examination Council (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 only to examiners, analysts, and economists in the banking supervision function of the Federal Reserve System.

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. Extreme values distort model coefficients and could compromise the relative performance of orderings from the downgrade-prediction model. Another reason for excluding *de novos* is that supervisors already monitor these banks closely.

---

[6] Following Hall, King, Meyer and Vaughan (2002), we include controls for influences on yields and runoffs that do not explicitly affect the probability of financial distress. An example of such a factor is the maturity

For example, the Federal Reserve examines each of its newly chartered banks every six months. Full-scope exams follow this schedule until the *de novo* earns a composite CAMELS rating of one or two in consecutive exams.

We generate two distinct proxies for jumbo-CD default premiums, starting with two items from each sample bank's call reports: interest expense on jumbo CDs and dollar volume of jumbo CDs outstanding. First, we divide jumbo-CD interest expense by average jumbo-CD balances for each sample bank for each quarter. We then use two different methods to convert these yield measures into default-risk premiums—a yield-spread method and a yield-residuals method. The use of two methods reduces the chance that performance tests will be biased by reliance on one, possibly poor, proxy for default premiums. The yield-spread method controls only for term-to-maturity, but the yield-residuals method controls for other influences on yields besides maturity. We begin with a simple method because the resulting proxy is similar to screens commonly used by surveillance analysts. The default-premium measures obtained from the two methods are highly correlated; the average year-by-year correlation coefficient over all the sample years is 0.98.

To control for the impact of term-structure yields, we use a third item taken from each sample bank's call reports: distributions of remaining maturities. 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 are crude—jumbo CDs in the "less than three months remaining" bucket could include currently maturing instruments that were issued several years ago—but the resulting data offer the only means of controlling for maturity. In the yield-spread method, we multiply the dollar volume of jumbo CDs in each maturity class for each sample bank by that quarter's yield on Treasury issues with comparable maturity. The sum of these resulting values,

---

of the bank's jumbo-CD portfolio.

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. In the yield-residuals method, for each sample year we regress each bank's average jumbo-CD yield on a broad set of control variables suggested by recent literature, examiner interviews, and specification tests.[7] Then, we use the residuals from these regressions as proxies for default premiums. We control for term-to-maturity with each bank's average jumbo-CD maturity as well as its maturity-weighted Treasury yield in each sample quarter. To obtain the maturity-weighted Treasury yield, we multiply the proportion of each bank's CDs in each maturity pool by that quarter's yield on a comparable-maturity Treasury issue. Tables 3a and 3b contain the regression results. The combined explanatory power of the maturity variables is statistically significant for each year as is the combined explanatory power of the non-maturity control variables.

In addition to proxies for default premiums, we also use jumbo-CD runoff to rank risky institutions. Park and Peristiani (1998) note that yields completely summarize default risk only when the jumbo-CD market is frictionless. When transaction or information frictions are present, jumbo-CD holders may choose to withdraw their funds from high-risk institutions. In addition, Jordan (2000) and Billett, Garfinkel, and O'Neal (1998) have documented a tendency for risky banking organizations to substitute insured for uninsured deposits to escape higher default premiums. If such substitution effects are important, then signals about changing risk exposures may show up in jumbo-CD runoffs rather than default premiums. To explore this possibility, we compute the quarterly percentage change in outstanding jumbo CDs for each sample bank. Then, we conduct out-of-sample performance tests with rankings based on these runoff percentages. We also regress percentage jumbo-CD runoff on the explanatory variables used to obtain the yield residuals and then use these runoff residuals to generate a risk ranking.

---

7 We include controls for regional influences (state dummies), economic conditions (time dummies), idiosyncratic aspects of the bank's jumbo-CD portfolio (holding company affiliation as proxied by a dummy variable for BHC membership) and willingness to tap national funding markets (as proxied by a dummy variable for use of brokered deposits), and idiosyncratic aspects of demand and supply for the bank's jumbo CDs (as proxied by a dummy for operation in an MSA).

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

The CAMELS-downgrade-prediction model transforms accounting and supervisory data into financial-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 (financial distress) 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 a bank-size measure used in the SEER risk-rank model, as well as two additional CAMELS-related variables. Table 4 describes the explanatory variables and the expected relationship between each variable and the likelihood of a downgrade. The financial-performance ratios are designed to capture 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 a CAMELS downgrade. 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 downgrade risk.

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 offered by the bank and the market. 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 two because two-rated banks tumble into unsatisfactory status more often than one-rated banks. (See Table 2 for supporting evidence.) 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 concerns about management competence, even though this problem has yet to produce financial consequences.

We estimate the downgrade-prediction model for eight overlapping 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 5), we use all non-*de novo*, U.S. commercial 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. Following the approach used in the SEER framework, we estimate the downgrade-prediction model using all CAMELS one- and two-rated U.S. banks and later use estimated coefficients to generate downgrade-probability estimates for each sample bank. The number of observations underlying each estimation of the downgrade-prediction model ranges from 7,836 (1992-93 regression) to 8,666 (1995-96 regression).

The downgrade-prediction model fits the data relatively well in sample. (Table 5 contains these regression results.) For all eight regressions, the log-likelihood test statistic allows rejection of the hypothesis that 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 for 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 signs in all eight equations. The coefficient on the size variable has a mixed-sign pattern, which is not unexpected given the 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 with the decline in downgrade frequency.

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

Next, we conduct performance tests of the risk rankings based on downgrade probabilities and jumbo-CD signals. For each year, we use the probit model to estimate the likelihood that each sample bank will suffer a downgrade (encounter financial distress) in the next eight quarters, and then rank banks from highest downgrade probability to lowest. Similarly, for each year we rank sample banks from highest to lowest default premium and from largest to smallest deposit runoff. We obtain separate rankings for yield spreads, yield residuals, percentage runoffs, and runoff residuals. Although each approach produces different specific numbers, they may all lead to similar rankings. Only out-of-sample testing can determine whether the jumbo-CD rankings differ from the downgrade-probability rankings and whether differences in rankings favor jumbo-CD signals as a surveillance tool. Period-ahead out-of-sample tests—which use an evaluation period subsequent to the estimation period rather than a contemporaneous holdout sample—are crucial because they mimic the way supervisors actually conduct off-site surveillance. Also, as has been demonstrated in the empirical literature on technical stock-market analysis, superior in-sample performance often fails to translate into superior out-of-sample performance (Malkiel, 1999; Roll, 1994).

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 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 rate 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 ordering based on spread-over-Treasuries is interpreted as an ordering of downgrade risk. We trace out the curve 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 downgrade. At this extreme, the type-one error rate is zero percent, and the type-two error rate is 100 percent.

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 random. 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. Evaluating relative performance with power-curve areas—though somewhat atheoretic—does make the best use of existing data. A more theoretically appealing approach would start by minimizing a loss function that explicitly weighs the benefits of early warning about financial distress over against 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 the results to compare performance over any desired range of error rates.

As noted, the out-of-sample tests follow a timing convention that reflects the way supervisors actually conduct surveillance. We start by regressing 1990-91 downgrade status on financial and supervisory data from 1989:IV. By the end of 1991, supervisors would have possessed coefficient estimates from this regression. We then apply these coefficients to 1991:IV data for each sample bank to obtain downgrade probabilities 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 follow the same timing convention. To test the yield-spread 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. We use the same procedures to produce yearly rankings based on percentage runoff. To test the yield-residuals rankings, we order banks by residuals from the 1991 regression of CD yields on the explanatory variables in Table 3a 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 more times for each model. We use the same procedure to produce rankings based on runoff residuals.

The out-of-sample evidence indicates that rankings based on output from the downgrade-prediction model outperform rankings based on jumbo-CD default premiums. Indeed, the default-premium rankings barely improve on random rankings. In the first test window (1992-93), the area under the yield-spread power curve (46.67 percent) and the area under the yield-residuals power curve (47.97 percent) are close to the random-selection benchmark of 50 percent. The power-curve patterns over the next seven test windows are 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 are omitted to save space. Table 6 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 19.83 percent, the average area under the yield-spread curves is 44.43 percent, and the average area under the yield-residuals curves is 44.02 percent. In the individual test windows, the downgrade-model areas range from 15.39 percent (1996-97) to 21.58 percent (1992-93). Meanwhile, the areas under the yield-spread rankings range from 40.02 (1995-96) percent to 49.72 percent (1994-95), and the areas under the yield-residuals rankings range from 39.68 percent (1995-96) to 50.11 percent (1994-95). The poor performance of these jumbo-CD rankings relative to the downgrade-model rankings suggests that default premiums would not flag banks missed by conventional surveillance tools. The poor performance relative to the random-selection benchmark suggests that default-premiums would not increase supervisor confidence about rankings produced by the downgrade-prediction model.

The out-of-sample performance of the runoff rankings mirrors the performance of default-premium rankings. In the first test window (1992-93), the area under the percentage-runoff power curve (49.62 percent) and the area under the runoff-residuals power curve (48.36 percent) are close to the random-selection benchmark. And, once again, the patterns over the other test windows are consistent. Over all eight tests, the average area under the percentage-runoff curves is 45.80 percent, and the average area under the runoff-residuals curves is 52.19 percent. In the individual test windows, the areas under the percentage-runoff rankings range from 40.06 (1996-97) percent to 50.84 percent (1994-95), and the areas under the runoff-residuals rankings range from 48.36 percent (1992-93) to 57.41 percent (1996-97). The consistently poor performance of the runoff rankings suggests that they do not contribute to surveillance—either by flagging missed banks or by increasing supervisor confidence about flagged banks.

6. Robustness Checks

6.1 Robustness Checks for Jumbo-CD Signals as Univariate Screens

Although jumbo-CD screens would have contributed little to surveillance in the tests presented thus far, they might add value for different sample cuts or for different forecasting horizons. For example, jumbo-CD screens may provide useful risk signals for banks whose jumbo-CD portfolios have short maturities, for banks with a large volume of assets, or for banks with no foreign depositors. Still another possibility is that jumbo-CD risk screens flag banks headed for financial distress before conventional surveillance tools.

The marginal-average problem noted earlier could explain the weak performance of the default-premium proxies. As an arithmetic matter, today's average yield will be more representative of today's risk levels if maturities are short. In addition, banks planning to embark on risky strategies may issue long-maturity CDs to escape market discipline (Flannery, 1986). To explore these possibilities, we replicated the out-of-sample tests for the sub-sample of banks with weighted-average portfolio maturities of less than six months. The results did not change; for example, the average area under the yield-spread power curves increased slightly (worsened) from the 44.02 percent for the all-bank sample to 44.88 percent for the short-maturity sub-sample. Put simply, the evidence indicates that long maturities do not account for the poor performance of the default-premium CD screens.

Jumbo-CD data might contain useful risk signals only for large, complex banking organizations. Jumbo CDs at community-banks may be more like core deposits than money-market instruments. And because prices and quantities of core deposits are known to be sticky (Flannery, 1982), yields and runoffs of community-bank CDs could respond sluggishly to changes in risk no matter how short the maturity of the portfolio. Also, jumbo-CD signals for large banking organizations may be stronger because the cost of monitoring these institutions—most of which have publicly traded debt and equity—is lower. To test for an asset-threshold effect, we reproduced the out-of-sample tests for a sub-sample of banks holding more over $500 million in assets (1999 dollars). We used this threshold because the Financial Modernization Act of 1999 established this upper bound on community-bank size for regulatory purposes. The out-of-sample tests on the large-bank sub-sample were qualitatively similar to the full-sample results. For example, the average area under the yield-spread power curves worsened slightly from 44.43 percent for the full sample to 44.90 percent for the large-bank subsample. We repeated the out-of-sample tests for banks holding more than $1 billion in assets and for banks with SEC registrations. Each time we compared the results for the large bank sub-sample with the results for the remaining sub-sample (i.e., banks holding less $500 million in assets, banks holding less than $1 billion in assets, and banks with no SEC registration), looking for a difference in performance across size cohorts. The size-split evidence was consistent—for large-complex banking organizations as well as for community banks, the downgrade-prediction model proved the far superior surveillance tool, and jumbo-CD signals barely improved over randomly generated signals.

Jumbo-CD signals might contain surveillance information for banks with no foreign deposits. The National Depositor Preference Act of 1993 elevated the claims of domestic depositors over the claims of foreign depositors, reducing expected losses for jumbo-CD holders (Marino and Bennett, 1999). Domestic jumbo CD holders in banks with foreign offices may have perceived no default-risk exposure because they knew that the foreign depositors would provide a financial cushion. To test for a depositor-preference effect, we screened out banks with foreign deposits and replicated the out-of-sample tests. Again, the results mirrored the full-sample results; for example, the average power-curve area for the yield-spread screens on the no-foreign-deposits sub-sample (43.87 percent) improved only slightly over the area for the all-banks sample (44.43 percent). Even for banks with no foreign-deposit cushion, jumbo-CD signals contained no useful supervisory information.

As final robustness checks on univariate jumbo-CD screens, we experimented with different out-of-sample forecasting horizons. Jumbo-CD signals might, for example, perform well at one-year forecasting horizons but not at two-year horizons. We replicated the out-of-sample tests for shorter as well as for longer test windows. Varying forecasting horizons did not alter the results—jumbo-CD signals would not have improved upon current surveillance practices or even random signals and would have barely improved over random signals. For example, the average power-curve area for the yield-spread screens using one-year-ahead forecasts was 43.84 percent, only slightly better than the 44.43 percent area for the two-year-ahead specification. This evidence goes to the timeliness of jumbo-CD screens. As noted, jumbo-CD screens add supervisory value if they flag problem banks missed by current surveillance tools, flag problem banks before current surveillance tools, or reduce uncertainty about banks flagged by current surveillance tools. Evidence from the baseline tests and the sub-sample tests indicates that jumbo-CD screens would not have flagged missed institutions or increased confidence about flagged institutions at a 24-month horizon. The additional timing evidence indicates that screens would not have improved over random selection at any forecasting horizon.

6.2 Jumbo-CD Signals as Regressors in the Downgrade-Prediction Model.

Although the default-premium and runoff measures perform poorly as univariate screens, they may add value as regressors in the downgrade-prediction model. Indeed, previous research has identified surveillance screens with this property (Gilbert, Meyer and Vaughan, 1999). To pursue this angle, we estimated an "enhanced" downgrade-prediction model that included the default-premium and runoff screens—singly and jointly. As before we assessed out-of-sample performance in terms of impact on power-curve areas—first by adding the screens to the baseline model and then by dropping them from the enhanced model. As an added check we assessed performance with the Quadratic Probability Score (QPS)—a probit analogue for the Root Mean Square Error (RMSE) statistic (Estrella, 1998 and Estrella and Mishkin, 1998).[8] If jumbo-CD screens contribute to the downgrade-prediction model, they will lower the QPS. Column 7 of Table 6 contains the power curve areas for the enhanced downgrade-prediction model. Column 2 of Panel A in Table 7 notes the impact of the jumbo-CD screens on the power curve areas of the enhanced downgrade model; Column 2 of Panel B in Table 7 shows the impact the jumbo-CD screens on the QPS of the downgrade model. To facilitate interpretation of these numbers, columns 3 through 6 in Panel A and B of Table 7 note the impact of other variable blocks—such as the leverage-risk

---

8 Formally, $QPS = 1/T\sum_{t=1}^{T} 2(\Pr(R_t = 1 | z_1 \cdots z_K) - R_t)^2$

In words, QPS is obtained by computing the downgrade probability for each sample bank with the downgrade-prediction model. Then, R_t—a binary variable equal to one if the bank is downgraded in the out-of-sample test window and zero if not—is subtracted from the downgrade-probability estimate. This difference is then squared, multiplied by two and averaged across all the sample banks. An ideal model generates probabilities close to unity for banks with subsequent downgrades and probabilities close to zero for non-downgrades, so higher QPS figures imply weaker out-of-sample performance—just as higher power curve areas imply weaker performance.

variables (equity-to-asset ratio and return on assets)—on QPS and power-curve areas. In Table 7 changes in QPS and power-curve areas are expressed in percentage-change terms to permit direct comparison.

By the power-curve and QPS metrics, jumbo-CD screens weakened the out-of-sample performance of the downgrade-prediction model. Adding the yield-spread and percentage runoff screens to the downgrade-prediction model increased the average power-curve area by 2.17 percent (0.43 percentage points, from 19.83 pecent for the baseline model to 20.26 percent). Similarly, removing the yield-spread and percentage-runoff screens from the enhanced model reduced the average power curve area (improved performance) by 1.06 percent and the QPS by 0.24 percent. In contrast, dropping the leverage-risk variables from the enhanced model increased the power curve area (worsened performance) by an average of 7.12 percent over all eight tests and increased the QPS by an average of 2.17 percent. The results for the residual-based screens were qualitatively similar. These results held up when we conducted the tests on the sample cuts and forecasting horizons described in 6.1.

Finally, we estimated "parallel" downgrade models with only default-premium and runoff screens as regressors. The downgrade-prediction model may be such a good surveillance tool that no additional explanatory variable could significantly improve its performance. At the same time, the jumbo-CD screens might jointly contain the information embedded in the explanatory variables of the downgrade model. In this case, a model with only screens as regressors could add value—even if it turned in a poor absolute performance—by reducing supervisor uncertainty about some of the banks flagged by the baseline model. To test this possibility, we estimated a model on yield spreads and percentage runoff only and a model on yield and runoff residuals only. We then tested the out-of-sample performance for these parallel models—for the full sample at an eight-quarter horizon as well as for all the various sample cuts and forecasting horizons.

The out-of-sample evidence, once again, indicated that the jumbo-CD screens would have added no value in surveillance. The average power-curve area for yield-spread/percentage runoff model across all eight test windows was 47.63 percent—once again near the random-selection benchmark. Including the dummy variables for banks with two composite ratings and with weak management ratings boosted this average to 30.27 percent. But a model with these two dummies alone produced an area of 30.07 percent. The results for the residuals model were similar—across every sample cut and forecasting horizon. In short, as regressors in the downgrade model, jumbo-CD screens did not help flag problem banks and did not increase supervisor confidence about flagged banks.

7. Conclusions

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

Problems with the jumbo-CD data or frictions in the jumbo-CD market could explain the poor performance of jumbo-CD signals. As noted, runoff is measured without error, but default premium proxies are extracted from noisy measures of yields. Another possibility is that posted jumbo-CD rates "cluster" around integers and even fractions (Kahn, Pennacchi, and Sopranzetti, 1999); such clustering would make rates less responsive to changes in bank risk—and rankings based on those rates less informative. Similarly, jumbo-CD holders may receive other services—commercial loans and checking services, for example—from the bank and, thus, price the relationship comprehensively rather than CDs individually. Such pricing practices would further weaken the link between bank risk and jumbo-CD yields. Still another possibility is that many jumbo CDs are relatively risk free—either because deposits barely exceed $100,000 or because depositors are mostly municipalities.[9] A final possibility is that jumbo-CD holders are noise traders, giving little thought to default-risk when choosing a bank.

We are not persuaded that data problems and market frictions account solely for the findings. As noted, several recent studies that used actual debt and equity market data rather than accounting proxies have also found little surveillance value in market signals. Rather, we believe that economic conditions in the 1990s play an important role. Over this period, bank profitability ratios soared to near record highs and failure rates plummeted to near record lows—largely as a result of an unprecedented economic boom that was enjoyed by virtually every region in the country (Berger, et al, 2000).[10] In such an environment, jumbo-CD signals—no matter how accurately measured or precisely determined—would convey little 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 evidence in Martinez-Peria and Schmukler (2001). With a dataset and research strategy similar to ours, they studied the impact of banking crises on market discipline in Argentina, Chile and Mexico, finding little discipline before but significant discipline after the crises. Interpreted in this light, our findings suggest that future policy and research work on market signals should focus on identifying the specific bank claims that yield the most surveillance value in each state of the business cycle.[11]

---

9 In most states, Treasury or agency securities must back municipal deposits. Such pledging eliminates all but fraud risk. (After the failure of Oakwood Deposit Bank of Ohio in February 2002, some municipalities discovered that the bank had pledged the same securities multiple times.) The call report does not indicate the volume of collateralized jumbo CDs, so we could not control directly for pledging.

10 Berger et al. (2000) also note the role of market power and product innovation in bolstering bank profits. In spite of a climate of deregulation, rivals found it difficult to enter bank markets and copy bank innovations. Also, many banks expanded into risky new activities and enjoyed only the "upside" because of robust economic conditions.

11 Another possible explanation for our results is that financial markets punish risky institutions through capital requirements. Flannery and Rangan (2002) argue that stakeholders of large complex banking organizations insisted on a greater cushion in the 1990s because of increasingly sophisticated risk exposures. A market-induced rise in capital would explain the weak jumbo-CD signals—the capital cushion reduced default risk. Even if this conjecture proves correct, it would not undermine the importance of the evidence presented here. Policy discussion to date has implicitly assumed that market signals are equally reliable in all states of the world.

Finally, our evidence carries implications for the current debate over a proposed hike in the deposit-insurance ceiling.[12] Community bankers have argued forcefully for an increase in the ceiling (Independent Community Bankers of America, 2000). In the 1990s, large banks merged at a record pace, producing sizable cost savings and putting intense pressure on small banks to cut expenses. At the same time, small banks lost consumer loans and retail deposits to tax-exempt credit unions. Community bankers contend that a higher coverage ceiling would improve their ability to attract large household deposits, retirement accounts and municipal deposits. And, they note that rising prices have considerably eroded the real value of coverage since 1980. Economists have countered that raising the deposit-insurance ceiling would weaken depositor pressure to contain bank risk. (For example, see Vaughan and Wheelock, 2002). And weaker deposit pressure would imply weaker jumbo-CD signals. The evidence presented here suggests that jumbo-CD signals yield no valuable supervisory information—at least in the current institutional and economic environment. So, raising the deposit-insurance ceiling now would not rob supervisors of valuable supervisory information. Of course, our evidence says nothing about the optimal coverage ceiling from a surveillance perspective—lowering the deposit-insurance ceiling might significantly increase the power of jumbo-CD signals by increasing default risk. A complete assessment of this issue—and the overall value of jumbo-CD signals—may have to await greater heterogeneity in banking conditions.[13]

---

[12] Our results carry mixed implications for the debate over requiring large complex banking organizations to issue subordinated debt. On the one hand, the evidence suggests that available jumbo-CD data would do little to enhance off-site surveillance, thereby clearing the way for experimentation with sub-debt signals. On the other, if the "unique sample period" explanation of our results is true, then it is likely that the surveillance value of sub-debt signals will vary with the business cycle. Other things equal, such time variation would lower the net benefit of mandatory sub-debt.

[13] Expanding our data set to include the 1980s would not offer much insight into the surveillance value of jumbo-CD signals. Call-report conventions for reporting jumbo-CD maturities were much different—and, more importantly, much cruder—in the 1980s than in the 1990s. Also, FDICIA changed the regulatory environment significantly in 1991. Supervisors need information about the surveillance tools that would add the most value in the current economic and institutional environment; evidence from a previous state-of-the world would be of little use.

REFERENCES

Altman, Edward I., and Anthony Saunders. "Credit Risk Measurement: Developments over the Last Twenty Years." *Journal of Banking and Finance* 21 (December 1997), 1721-42.

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.

Barth, Mary E., William H. Beaver, and Wayne R. Landsman. "Value-Relevance of Banks' Fair Value Disclosures under SFAS No. 107." *The Accounting Review* 71 (October 1996), 513-537.

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* (2nd Quarter 1998), 2-20.

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

Berger, Allen N.; Seth D. Bonime; Daniel M. Covitz; and Diana Hancock. "Why are Bank Profits so Persistent? The Roles of Product Market Competition, Informational Opacity, and Regional/Macroeconomic Shocks." *Journal of Banking and Finance* 24 (2000): 1203-1235.

Berger, Allen N., and Sally M. Davies. "The Information Content of Bank Examinations." *Journal of Financial Services Research* 14 (October 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 (June 1998), 333-58.

Bliss, Robert R. "Market Discipline and Subordinated Debt: A Review of Some Salient Issues." Federal Reserve Bank of Chicago *Economic Perspectives* (1st Quarter 2001), 24-45.

Bliss, Robert R. and Mark J. Flannery. "Market Discipline in the Governance of U.S. Bank Holding Companies: Monitoring versus Influence." In *Prudential Supervision: What Works and What Doesn't,* edited by Frederic Mishkin. Chicago: University of Chicago Press, 2001.

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: The Case of Junk Bonds at Savings and Loan Associations." *Journal of Money, Credit, and Banking* 26 (February 1994), 146-64.

Cargill, Thomas F. "CAMEL Ratings and the CD Market." *Journal of Financial Services Research* 3 (December 1989), 347-58.

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

Cole, Rebel A., Barbara G. Cornyn, and Jeffrey W. Gunther. "FIMS: A New Monitoring System for Banking Institutions." *Federal Reserve Bulletin* 81 (January 1995), 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 (August 1994), 439-59.

Covitz, Daniel M., Diana Hancock, and Myron L. Kwast, "Market Discipline in Banking Reconsidered: The Roles of Deposit Insurance Reform, Funding Manager Decisions and Bond Market Liquidity." Board of Governors Working Paper, 2002-46 (October).

Crabbe, Leland, and Mitchell A. Post. "The Effect of a Rating Downgrade on Outstanding Commercial Paper." *Journal of Finance* 49 (March 1994), 39-56.

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

Curry, Timothy J., Peter J. Elmer, and Gary S. Fissel. "Regulator Use of Market Data to Improve the Identification of Bank Financial Distress." Federal Deposit Insurance Corporation Working Paper 2001-01, December, 2001.

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

DeYoung, Robert. "Birth, Growth, and Life or Death of Newly Chartered Banks." Federal Reserve Bank of Chicago *Economic Perspectives* (3rd Quarter 1999), 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 (December 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.

Estrella, Arturo. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." *Journal of Business Economics and Statistics* 16 (April 1998), 198-205.

Estrella, Arturo, and Frederic S. Mishkin. "Predicting U.S. Recessions: Financial Variables as Leading Indicators." *The Review of Economics and Statistics*, 80 (February 1998), 45-61.

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

Flannery, Mark J. "The Faces of 'Market Discipline.'" *Journal of Financial Services Research* 20 (October/December 2001), 107-19.

Flannery, Mark J. "Asymmetric Information and Risky Debt Maturity Choice." *Journal of Finance* 41 (March 1986), 19-37.

Flannery, Mark J. "Retail Bank Deposits as Quasi-Fixed Factors of Production." *American Economic Review* 72 (June 1982), 527-36.

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 (February 1999), 14-34.

Flannery, Mark J. and Kasturi Rangan. "Market Forces at Work in the Banking Industry: Evidence from the Capital Buildup of the 1990s. Working paper, University of Florida, September 2002.

Flannery, Mark J., and Sorin M. Sorescu. "Evidence of Bank Market Discipline in Subordinated Debenture Yields: 1983-1991." *Journal of Finance* (September 1996), 1347-77.

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.

Goldberg, Lawrence G. and Sylvia C. Hudgins. "Depositor Discipline and Changing Strategies for Regulating Thrift Institutions." *Journal of Financial Economics* 63 (February 2002), 263-74.

Goldberg, Michael A. and Peter R. Lloyd-Davies. "Standby Letters of Credit: Are Banks Overextending Themselves?" *Journal of Bank Research* 16 (Spring 1985), 29-39.

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* (3rd Quarter 2000), 30-5.

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* (2nd Quarter 2001), 2-9.

Hall, John R., Thomas B. 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, April 2003.

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*. Federal Deposit Insurance Corporation: Washington DC, 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 (May 1988), 203-211.

Independent Community Bankers of America. *Federal Deposit Insurance: Time for Reform and Increased Coverage*. ICBA: Washington, DC 2000.

James, Christopher. "Heterogeneous Creditors and LDC Lending." *Journal of Monetary Economics* 25 (June 1990), 325-46.

James, Christopher. "The Losses Realized in Bank Failures." *Journal of Finance* 46 (September 1991), 1223-42.

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

Jordan, John S. "Depositor Discipline at Failing Banks." Federal Reserve Bank of Boston *New England Economic Review* (March/April 2000), 15-28.

Kahn, Charles, George Pennacchi, and Ben Sopranzetti. "Bank Deposit Rate Clustering: Theory and Empirical Evidence." *Journal of Finance* 56 (December 1999), 2185-214.

Keeley, Michael C. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80 (1990), 1183-98.

Krainer, John and Jose A. Lopez. "Incorporating Equity Market Information into Supervisory Monitoring Models." *Journal of Money, Credit, and Banking,* forthcoming.

Kroszner, Randall S. and Philip E. Strahan. "Obstacles to Optimal Policy: The Interplay of Politics and Economics in Shaping Bank Supervision and Regulation Reforms." In *Prudential Supervision: What Works and What Doesn't,* edited by Frederic Mishkin. Chicago: University of Chicago Press, 2001.

Lang, William W. and Douglas Robertson. "Analysis of Proposals for a Minimum Subordinated Debt Requirement." *Journal of Economics and Business* 54 (2002), 115-36.

Malkiel, Burton G. *A Random Walk Down Wall Street.* New York: W. W. Norton, 1999.

Marino, James A. and Rosalind Bennett. "The Consequences of National Depositor Preference." *FDIC Banking Review* 12 (1999), 19-38.

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), 1029-51.

Meyer, Laurence H. "Supervising Large Complex Banking Organizations: Adapting to Change." In *Prudential Supervision: What Works and What Doesn't*, edited by Frederic Mishkin. Chicago: University of Chicago Press, 2001.

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

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

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

Park, Sangkyun and Stavros Peristiani. "Market Discipline by Thrift Depositors." *Journal of Money, Credit, and Banking* 30 (August 1998), 347-64.

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.

Rajan, Raghuram G. "Comment on Bliss and Flannery." In *Prudential Supervision: What Works and What Doesn't*, edited by Frederic Mishkin. Chicago: University of Chicago Press, 2001.

Reidhill, Jack and John O'Keefe. "Off-Site Surveillance Systems." In *History of the Eighties: Lessons for the Future, Volume 1*. Federal Deposit Insurance Corporation: Washington DC, 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* 23:2 (1994), 69-75.

Vaughan, Mark D. and David Wheelock. "Deposit Insurance Reform Is It Déjà Vu All Over Again?" Federal Reserve Bank of St. Louis *Regional Economist* (October 2002), 5-9.

Table 1

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. We use the term “risk pricing” to mean a price or quantity response to a change in bank risk.) 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 risk.

Authors	Issuer of Jumbo Certificate of Deposit	Sample Dates	Yield or Runoff?	Was Bank Risk Priced?
Crane (1976)	Bank	1974	Yield	Somewhat
Goldberg and Lloyd-Davies (1985)	Bank	1976-82	Yield	Yes
Baer and Brewer (1986)	Bank	1979-82	Yield	Yes
Hannan and Hanweck (1988)	Bank	1985	Yield	Yes
James (1988)	Bank	1984-86	Yield	Yes
Cargill (1989)	Bank	1984-86	Yield	Yes
James (1990)	Bank	1986-87	Yield	Yes
Keeley (1990)	Bank	1984-86	Yield	Yes
Ellis and Flannery (1992)	Bank	1982-88	Yield	Yes
Cook and Spellman (1994)	Thrift	1987-88	Yield	Yes
Crabbe and Post (1994)	Bank	1986-91	Runoff	No
Brewer and Mondschean (1994)	Thrift	1987-89	Yield	Yes
Park (1995)	Bank	1985-92	Both	Yes
Park and Peristiani (1998)	Thrift	1987-91	Both	Yes
Jordan (2000)	Bank	1989-95	Both	Yes
Goldberg and Hudgins (2002)	Thrift	1984-94	Runoff	Yes
Hall, King, Meyer, and Vaughan (2003)	Bank	1988-90, 1993-95	Both	Yes

Table 2

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

This table demonstrates that supervisors downgraded safe-and-sound banks to unsatisfactory condition frequently 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 safe-and-sound (composite CAMELS of "one" or "two") sample banks at each year-end that were downgraded to unsatisfactory status (composite CAMELS of "three," "four," or "five") in the following 12 months. The data also 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	2,190 5,482	34 665	1.55 12.13	699
1991	1 2	1,959 5,277	22 402	1.12 7.62	424
1992	1 2	2,291 5,980	7 175	0.31 2.93	182
1993	1 2	2,911 5,726	9 153	0.31 2.67	162
1994	1 2	3,091 4,885	8 94	0.26 1.92	102
1995	1 2	3,284 4,522	10 117	0.30 2.59	127
1996	1 2	3,242 3,741	7 118	0.22 3.15	125
1997	1 2	3,022 3,105	19 133	0.63 4.28	152
1998	1 2	3,067 3,047	19 182	0.62 5.97	201
1999	1 2	3,088 3,320	12 178	0.39 5.36	190
2000	1 2	3,226 3,684	3 46	0.09 1.25	49

Table 3a

Are residuals from jumbo-CD regressions potentially good proxies for bank risk?

Yield Regressions

As one measure of default premiums, we use residuals from year-by-year regressions of jumbo-CD yields on non-default-risk control variables. This table displays the results for the 1991 through 1999 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 explanatory variables include controls for interest-rate levels, regional economic conditions, state tax and banking laws, 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 derivative risk rankings. The overall fit of the equations and the significance of the individual coefficients suggest that the residuals should prove good proxies for bank risk.

	Fourth quarter of:								
Independent variables	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
Intercept	2.7833*** (0.1426)	1.9547*** (0.1213)	0.7836** (0.3493)	1.9608*** (0.0720)	9.0065*** (0.2960)	5.4670*** (0.2851)	8.5020*** (0.4149)	3.9230*** (0.2003)	3.5785*** (0.1371)
Maturity-weighted Treasury yield	0.6626*** (0.0227)	0.7265*** (0.0314)	0.9492*** (0.1142)	0.4118*** (0.0131)	-0.6374*** (0.0511)	-0.0110 (0.0538)	-0.5948*** (0.0781)	0.2856*** (0.0394)	0.2933*** (0.0276)
Maturity	0.0037 (0.0537)	-0.1186*** (0.0386)	0.2070*** (0.0757)	0.3344*** (0.0224)	0.3364*** (0.0220)	0.2625*** (0.0303)	0.4475*** (0.0361)	0.1634*** (0.0277)	0.3006*** (0.0157)
Maturity-Treasury interactive	-0.3188*** (0.0415)	-0.6192*** (0.0428)	-0.9487*** (0.0950)	-0.1361*** (0.0236)	-0.0150 (0.0293)	-0.1147** (0.0478)	1.8436*** (0.1961)	-0.4783*** (0.0604)	-0.1901*** (0.0287)
Holding-company dummy	-0.0826*** (0.0315)	-0.1473*** (0.0276)	-0.0943*** (0.0221)	-0.0460** (0.0203)	-0.0441* (0.0234)	-0.0401* (0.0206)	-0.0090 (0.0224)	0.0015 (0.0193)	-0.0206 (0.0149)
Brokered-deposit dummy	-0.0053 (0.0482)	-0.0295 (0.0386)	0.0723*** (0.0278)	0.0948*** (0.0241)	0.1455*** (0.0263)	0.0322 (0.0233)	0.0875*** (0.0245)	0.0976*** (0.0202)	0.1348*** (0.0149)
MSA dummy	-0.1225*** (0.0272)	-0.2408*** (0.0229)	-0.1424*** (0.0181)	-0.0461*** (0.0164)	0.0139 (0.0187)	-0.0397** (0.0165)	-0.0475*** (0.0177)	-0.0820*** (0.0148)	-0.0539*** (0.0111)
Adjusted R^2	0.1185	0.2221	0.2085	0.2165	0.0911	0.0362	0.0346	0.0504	0.0944
F-statistic for significance of the independent variables	35.09***	68.09***	66.33***	74.40***	31.74***	13.66***	13.43***	19.62***	37.36***

Table 3b

Are residuals from jumbo-CD runoff regressions potentially good proxies for bank risk?

Runoff Regression

As one measure of default premiums, we use residuals from year-by-year regressions of jumbo-CD runoffs on non-default-risk control variables. This table displays the results for the 1991 through 1999 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 explanatory variables include controls for interest-rate levels, regional economic conditions, state tax and banking laws, 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 derivative risk rankings. The overall fit of the equations and the significance of the individual coefficients suggest that the residuals should prove good proxies for bank risk.

	Fourth quarter of:								
Independent variables	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**
Intercept	-13.762*** (1.8657)	14.857*** (1.8494)	5.5251 (5.8027)	10.339*** (1.8074)	-6.9792 (5.4923)	36.432*** (4.9228)	10.399 (6.6026)	-19.386*** (3.6018)	-4.7548 (3.3625)
Maturity-weighted Treasury yield	2.5234*** (0.2972)	-4.4935*** (0.4781)	-1.1881 (1.8965)	-1.6023*** (0.3285)	2.4237** (0.9473)	-6.0304*** (0.9292)	-1.0810 (1.2420)	4.7071*** (0.7087)	1.4900** (0.6759)
Maturity	-3.3476*** (0.7031)	-1.7241*** (0.5878)	-0.0419 (1.2579)	-5.9864*** (0.5613)	-3.3929*** (0.4088)	-0.1672 (0.5236)	-1.6437*** (0.5736)	-2.3116*** (0.4976)	-0.7675** (0.3846)
Maturity-Treasury interactive	-3.1371*** (0.5423)	-7.3444*** (0.6523)	1.4821 (1.5777)	6.6188*** (0.5926)	2.1247*** (0.5430)	2.5669*** (0.8251)	-10.896*** (3.1211)	-6.7920*** (1.0854)	0.6501 (0.7037)
Holding-company dummy	-1.2949*** (0.4124)	-1.1549*** (0.4210)	-0.6797* (0.3676)	0.4064 (0.5087)	1.0616** (0.4344)	0.1928 (0.3564)	-0.1177 (0.3567)	-0.1512 (0.3465)	0.5742 (0.3665)
Brokered-deposit dummy	-0.1757 (0.6305)	0.5957 (0.5890)	0.4205 (0.4625)	1.0892* (0.6055)	0.9330* (0.4883)	1.9252*** (0.4021)	1.1022*** (0.3894)	1.4801*** (0.3631)	2.8030*** (0.3652)
MSA dummy	-0.8961** (0.3554)	-1.4245*** (0.3488)	-0.5580* (0.3010)	0.8209** (0.4112)	0.6598* (0.3467)	0.4380 (0.2852)	0.8013*** (0.2810)	0.2242 (0.2666)	0.8891*** (0.2727)
Adjusted R^2	0.0155	0.0188	0.0028	0.0228	0.0219	0.0086	0.0056	0.0049	0.0111
F-statistic for significance of the independent variables	5.00***	5.51***	1.71***	7.20***	7.87***	3.93***	2.94***	2.72***	4.92***

Table 4
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 to unsatisfactory status over the next two years.

	Independent Variables	Symbol	Hypothesized Relationship to Downgrade Risk
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 5

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 years *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 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 with 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)
Credit Risk	PAST-DUE-90	0.376*** (0.039)	0.328*** (0.040)	0.239*** (0.047)	0.304*** (0.060)
Credit Risk	NONACCRUING	0.235*** (0.029)	0.199*** (0.030)	0.291*** (0.036)	0.178*** (0.045)
Credit Risk	OREO	0.220*** (0.030)	0.216*** (0.032)	0.145*** (0.031)	0.167*** (0.043)
Credit Risk	COMMERCIAL-LOANS	0.009*** (0.003)	0.013*** (0.003)	0.009** (0.004)	0.002 (0.005)
Credit Risk	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)
Leverage Risk	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)
Liquidity Risk	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)
Control Variables	CAMELS-2	0.633*** (0.062)	0.517*** (0.068)	0.509*** (0.087)	0.432*** (0.102)
Control Variables	BAD-MANAGE	0.488*** (0.051)	0.401*** (0.054)	0.478*** (0.061)	0.466*** (0.069)
	Number of Observations Pseudo-R^2 -2 log likelihood testing whether all coefficients (except the intercept) = 0	8,494 0.219 7211.785***	8,065 0.226 6053.423***	7,836 0.208 3946.413***	8,058 0.161 2444.832***

Table 5 (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 6

Do measures of jumbo-CD default premiums or runoffs add value in 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 implies a lower rate of both types of errors and, thus, a better surveillance tool. Each cell in columns two through seven 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 bank surveillance. Risk rankings produced by the downgrade-prediction model (column two) perform considerably better than a random ranking, which would produce an area of approximately 50 percent. At the same time, risk rankings based on jumbo-CD default premiums or runoffs (columns three through six) perform little better than random risk rankings. Moreover, adding jumbo-CD default premiums to the downgrade-prediction model (column seven) did not improve its out-of-sample performance.

Downgrade Years (1)	**Downgrade Model (2)**	**Yield-Spread (3)**	**Yield Residuals (4)**	**Percentage Runoff (5)**	**Runoff Residuals (6)**	**Downgrade Model plus Yield Spread and Percentage Runoff (7)**
1992-93	21.58	46.67	47.97	49.62	48.36	21.88
1993-94	21.04	43.92	41.73	45.36	48.97	20.73
1994-95	20.13	49.72	50.11	50.84	51.15	20.20
1995-96	19.17	40.02	39.68	49.49	53.33	21.28
1996-97	15.39	43.74	41.75	40.06	57.41	16.07
1997-98	20.04	44.28	43.85	44.80	54.85	20.44
1998-99	21.43	43.77	43.60	46.14	51.77	21.89
1999-00	19.89	43.34	43.45	40.09	51.70	19.61
Mean All Years	**19.83**	**44.43**	**44.02**	**45.80**	**52.19**	**20.26**

Table 7

Do default-premium or runoff screens contribute to the downgrade-prediction model?

This table provides alternative measures of the contribution of jumbo-CD screens to the downgrade-prediction model. Column 2 of Panel A shows the impact of removing the yield-spread and percentage-runoff screens on the power-curve areas of the enhanced downgrade model (baseline model plus jumbo-CD screens). Column 2 of Panel B notes the impact of removing these screens on the QPS of the enhanced model. Changes in QPS and power-curve areas are expressed in percentage-change terms to permit direct comparisons. Positive percentage changes for QPS or power curve areas imply that removing the variable block weakens model performance. To facilitate interpretation of the percentage changes, columns 3 through 6 in each panel note the impact of other variable blocks—such as the leverage-risk variables (equity-to-asset ratio and return on assets)—on QPS and power-curve areas. The evidence suggests that jumbo-CD screens add little value in bank surveillance. Removing the yield-spread and percentage-runoff screens reduces power-curve areas as well as QPS, that is, removing these screens improves model performance.

Panel A: Percentage Change in Power Curve Area

Downgrade Years (1)	Jumbo CD Variables (2)	Leverage Risk Variables (3)	Credit Risk Variables (4)	Liquidity Risk Variables (5)	Control Variables (6)
1992-93	-1.37	6.54	19.42	6.08	4.57
1993-94	1.50	10.95	23.15	-3.81	14.38
1994-95	-0.30	10.18	9.04	1.28	12.45
1995-96	-3.57	17.73	22.79	1.69	8.34
1996-97	-4.46	9.73	34.70	1.67	18.40
1997-98	-1.95	3.02	1.80	4.15	7.61
1998-99	0.28	-4.31	14.79	2.34	11.94
1999-00	1.38	3.11	21.37	-0.82	18.26
Mean	**-1.06**	**7.12**	**18.38**	**1.57**	**11.99**

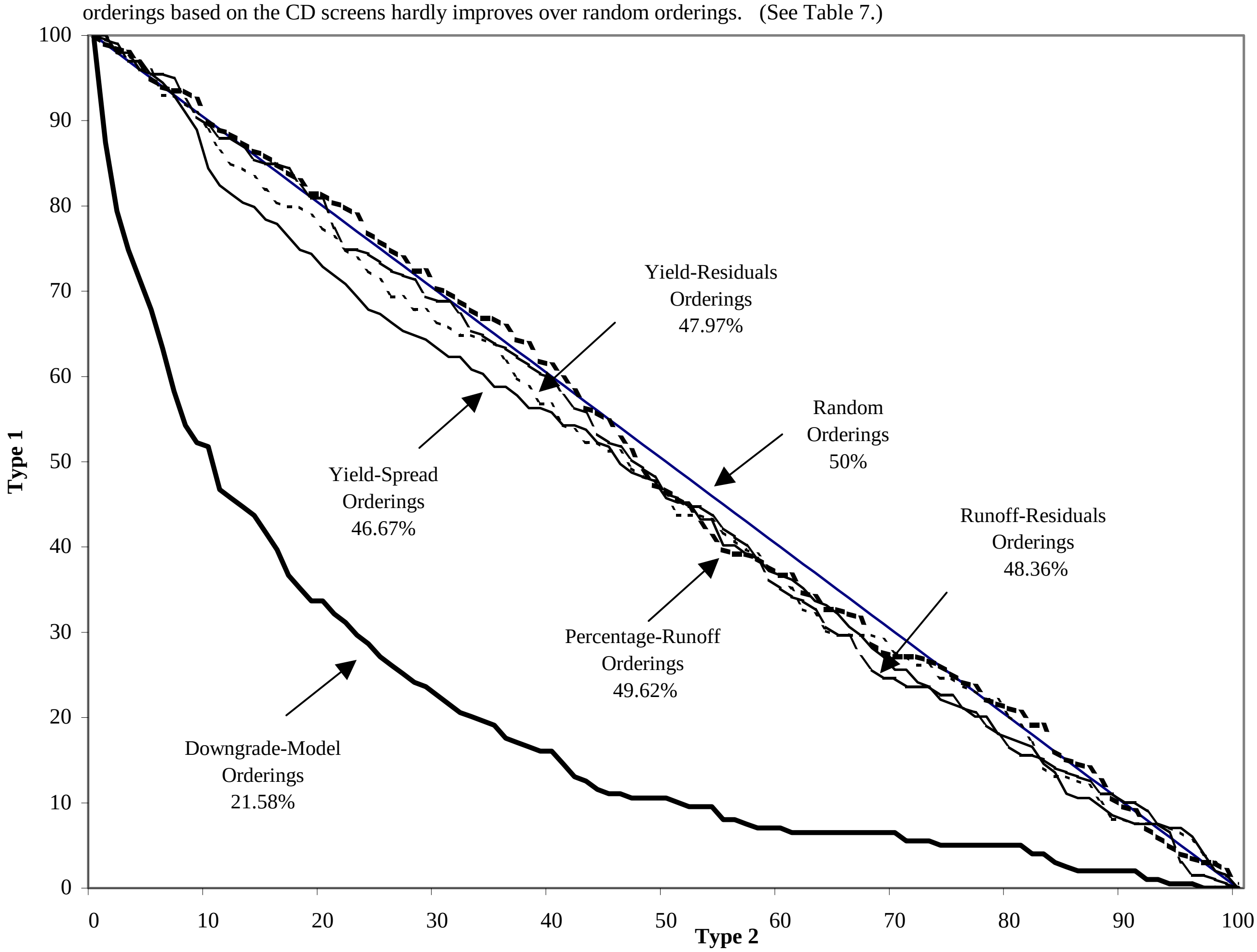
Panel B: Percentage Change in QPS

Downgrade Years (1)	Jumbo CD Variables (2)	Leverage Risk Variables (3)	Credit Risk Variables (4)	Liquidity Risk Variables (5)	Control Variables (6)
1992-93	-0.78	3.33	5.66	4.11	-1.16
1993-94	0.13	6.21	9.31	1.62	-0.94
1994-95	0.00	3.70	5.93	0.74	0.93
1995-96	0.22	0.44	4.39	0.88	2.63
1996-97	-0.19	1.72	3.63	0.76	0.96
1997-98	-0.97	0.81	0.81	0.81	4.21
1998-99	0.00	0.39	2.99	0.26	2.08
1999-00	-0.33	0.77	2.64	0.66	2.42
Mean	**-0.24**	**2.17**	**4.42**	**1.23**	**1.39**

Figure 1: How well do the models predict CAMELS downgrades out-of-sample?

1992-93 Downgrade Predictions Using Year-end 1991 Data

This figure shows that the downgrade-prediction model significantly outperforms all four jumbo-CD screens. Indeed, risk orderings based on the CD screens hardly improves over random orderings. (See Table 7.)



This figure shows the trade-off between type-one and type-two error rates for risk orderings based on the downgrade model and the four univariate jumbo-CD screens (yield spreads, yield residuals, percentage runoff, and runoff residuals). Type-one errors are missed downgrades--that is, one- or two-rated banks not flagged as downgrade risks that subsequently suffer CAMELS downgrades. Type-two errors, in contrast, are over-predicted downgrades--that is, one- or two-rated banks flagged as downgrade risks that do not suffer subsequent downgrades. A desirable early-warning tool minimizes type-one errors for any given level of type-two error. A convenient way to compare orderings is to calculate the area under each list's power curve and express that area as the percentage of the total area in the box. Smaller areas are desired because they imply a simultaneous reduction in both types of errors. The 50 percent line notes the resulting power curve if banks at risk of downgrade are selected randomly.