

Can Feedback from the Jumbo CD Market Improve Bank Surveillance?

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

In recent years, policymakers in the Basel countries have begun exploring strategies for harnessing financial markets to contain bank risk. Indeed, the new Accord counts market discipline, along with supervisory review and capital requirements, as an explicit pillar of bank supervision.¹ A popular proposal for implementing market discipline in the United States would require large banks to issue a standardized form of subordinated debt (Board of Governors 1999; Board of Governors 2000; Meyer 2001). Advocates of

■ Critical feedback from a number of sources greatly improved this work. We would specifically like to thank the examiners and supervisors (Carl Anderson, John Block, Joan Cronin, Ben Jones, Kim Nelson, and Donna Thompson) as well as the economists (Gurdip Bakshi, Rosalind Bennett, Mark Carey, Margarida Duarte, Kathleen McDill, 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, James Thomson, Sherrill Shaffer, Scott Smart, Haluk Unal, Larry Wall, John Walter, and John Weinberg) who provided helpful comments. We also profited from exchanges with seminar participants at Baylor University, the Federal Deposit Insurance Corporation, the Office of the Comptroller of the Currency, and Washington University in St. Louis (Department of Economics and the Olin School of Business), as well as exchanges at the Federal Reserve Surveillance Conference, the Federal Reserve Committee on Financial Structure meetings, and the Financial Management Association meetings. Any remaining errors and omissions are ours alone. The views expressed do not represent official positions of the Federal Reserve Bank of Richmond, the Federal Reserve Bank of St. Louis, or the Federal Reserve System.

¹The cornerstone of supervisory review—the most important of the pillars—is thorough, regularly scheduled, on-site examinations. The Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) requires most U.S. banks to submit to a full-scope examination every 12 months. These examinations focus on six components of safety and soundness—capital protection (C), asset quality (A), management competence (M), earnings strength (E), liquidity risk exposure (L), and market risk sensitivity (S)—CAMELS. At the close of each exam, an integer ranging from 1 (best) through 5 (worst) is awarded for each component. Supervisors then use these component ratings to assign a composite CAMELS rating reflecting overall condition—also on a 1-to-5 scale. In general, banks with composite ratings of 1 or 2 are considered satisfactory while banks with ratings of 3, 4, or 5 are unsatisfactory and subject to supervisory sanctions. (Footnote 10 offers more details about these sanctions.) At year-end 2005, 4.63 percent of U.S. banks held unsatisfactory ratings.

this proposal argue that high-powered performance incentives in the subordinated debt (sub-debt) market will produce accurate risk assessments. And, in turn, these assessments—expressed for risky institutions through rising yields or difficulties rolling over maturing debt—will pressure bank managers to maintain safety and soundness (Calomiris 1999; Lang and Robertson 2002).

Even if financial markets apply little direct pressure to curb risk taking, market data could still enhance supervisory review by improving off-site surveillance.² Off-site surveillance involves the use of accounting data and anecdotal evidence to monitor the condition of supervised institutions between scheduled exams.³ Market assessments could enhance surveillance in three ways: (1) by flagging banks missed by conventional off-site tools, (2) by reducing uncertainty about banks flagged by other tools, or (3) by providing earlier warning about developing problems in banks flagged by these tools (Flannery 2001). Such enhancements would reduce failures over time by enabling supervisors to take action earlier to address safety-and-soundness problems.

One concern about attempts to incorporate market data into surveillance is regulatory burden—current proposals would require large banking organizations to float a standardized issue of sub-debt. That most large banks currently issue sub-debt does not imply the burden is negligible.⁴ Voluntary issuance varies considerably over time with market conditions. For example, the number of sub-debt issues by the top-50 banking organizations rose from 3 in 1988 to 108 in 1995, only to fall to 42 in 1999 (Covitz, Hancock, and Kwast 2002). Moreover, banks currently issuing sub-debt may be choosing maturities unlikely to produce valuable risk signals, so a mandated maturity would still impose a regulatory burden. Before placing additional burden on the banking sector, particularly at a time when other sizable regulatory changes (Basel II) are in the offing, supervisors should first assess the power of risk signals from existing securities.

One potential source of risk assessments that can be mined without increasing regulatory burden is the market for jumbo certificates of deposit (CDs).

² Bliss and Flannery (2001) found that managers of holding companies do not respond to market pressure to contain risk, though Rajan (2001) questioned the ability of their framework to unearth such evidence.

³ Examination is the most effective tool for spotting safety-and-soundness problems, but it is costly and burdensome—costly because of the examiner resources required and burdensome because of the intrusion into bank operations. Surveillance reduces the need for unscheduled visits by prodding bankers to contain risk between scheduled exams. It also helps supervisors plan exams by highlighting risk exposures. For example, if pre-exam surveillance reports indicate a bank has significant exposure to interest rate fluctuations, supervisors will staff the exam team with additional market risk expertise.

⁴ Mandating issuance of a security with specific attributes is tantamount to a tax on capital structure. Although we know of no direct evidence about the burden of this tax, heterogeneity in sub-debt maturities, outstanding volume over time, and the source of issue (bank vs. bank holding company) suggest it is nontrivial.

Jumbo CDs are time deposits with balances exceeding \$100,000. The typical bank relies on a mix of deposits to fund assets—checkable deposits, passbook savings accounts, retail CDs, and jumbo CDs. Both retail and jumbo CDs have fixed maturities (as opposed to checkable deposits which are payable on demand); they differ by Federal Deposit Insurance Corporation (FDIC) coverage. Only the first \$100,000 of deposits is eligible for insurance, so the entire retail CD (which is less than \$100,000) is insured while only the first \$100,000 of a jumbo CD is covered. Checkable deposits, passbook savings, and retail CDs are often collectively referred to as “core deposits” because balances respond little to changes in bank condition and market rates. Full FDIC coverage makes these deposits a stable and cheap source of funding. At year-end 2005, U.S. banks funded on average 67.1 percent of assets with core deposits and 14.4 percent with jumbo CDs. The average jumbo CD balance in the fourth quarter of 2005 was \$330,886; the average balance in 95 percent of the U.S. banks exceeded \$152,115. The average maturity was just over one year. Jumbo CDs are considered a “volatile” liability because relatively large uninsured balances and short maturities force issuing banks to match yields (risk-free rates plus default premiums) available in the money market or lose the funding. This pressure to “price” new conditions quickly makes the jumbo CD market, in theory, an important source of feedback for off-site surveillance.⁵

Potentially valuable jumbo CD data are currently available for most commercial banks. In contrast, only very large banking organizations now issue sub-debt. These organizations may be the most important from a systemic-risk standpoint, but the focus of off-site surveillance—indeed of all U.S. prudential supervision—is on the bank, and most banks do not issue or belong to holding companies that issue sub-debt. Moreover, a negative risk signal from a holding company claim would not, by itself, help supervisors identify the troubled subsidiary. Jumbo CDs constitute a large class of direct claims on both large and small banks. At year-end 2005, U.S. banks with more than

⁵ Since the early 1990s, financial innovation has offered households a growing array of substitutes for traditional bank deposits. As a result, the supply of core deposits has declined secularly, forcing banks to turn to more volatile funding sources such as jumbo CDs. Between 1992 and 2005, for example, the average core deposit-to-total asset ratio for U.S. banks tumbled from 80.1 percent to 67.1 percent, while average jumbo CD dependence jumped from 7.5 percent to 14.4 percent of assets. Increasing reliance on jumbo CDs implies greater exposure to liquidity and market risk—a bad outcome from the perspective of a bank supervisor. At the same time, the \$100,000 ceiling on deposit insurance makes jumbo CD holders savvier about bank risk than other depositors. So the jumbo CD market could exert pressure on bank managers to contain risk—either directly through the impact of higher yields and lower balances on profits or indirectly through supervisory responses to risk signals conveyed by yields and withdrawals. Such pressure would complement supervisory review. Hence, another contribution of this article is to offer insight into the tradeoff by quantifying the potential contribution of jumbo CD data to off-site surveillance. See Feldman and Schmidt (1991) for further discussion of the tradeoff between greater risk exposure and more reliable market data implied by rising jumbo CD dependence. Our results suggest this rising dependence makes supervisors on balance worse off.

\$500 million in assets funded 14.6 percent of assets with jumbo CDs; for banks with less than \$500 million, the average jumbo-CD-to-total-asset ratio was 14.3 percent. Finally, risk signals in the form of yields and withdrawals can be cheaply and easily constructed because banks report jumbo CD interest expense and balances quarterly to their principal supervisor. Also, nearly 30 years of research—much of which relies on these interest-expense and account-balance data—has produced robust evidence of risk pricing in the jumbo CD market.

Data from the jumbo CD market might prove particularly useful in community bank surveillance. Community banks specialize in making loans to and taking deposits from small towns or city suburbs. For regulatory purposes, the Financial Modernization Act of 1999 established an asset threshold of \$500 million—expressed in constant 1999 dollars. At year-end 2005, nearly 90 percent of U.S. banks operated on this scale. Not surprisingly, most failures are community banks. They also frequently operate on extended exam schedules, with up to 18 months elapsing between full-scope, on-site visits. This schedule diminishes the quality of quarterly financial statements, thereby reducing the effectiveness of off-site monitoring.⁶ It is possible that holders of community bank jumbo CDs supplement public financial data with independent “Peter-Lynch-type” research.⁷ Or, inside information about bank condition could leak from boards of directors, which typically include prominent local businesspeople. (Community bank jumbo CDs are often held by such “insiders.”) Thus, sudden changes in yields or withdrawals might signal trouble more quickly or reliably than surveillance tools based on financial statements.

In short, jumbo CDs fund a large portion of bank assets and furnish a cheap source of market data, yet no study has formally tested the surveillance value of yields and withdrawals. We do so with an early warning model and out-of-sample timing conventions designed to mimic current surveillance practices. Specifically, we generate risk rankings using jumbo CD default premiums and quarter-over-quarter withdrawals for banks with satisfactory supervisory ratings. We rank the same banks by CAMELS-downgrade probability as estimated by an econometric surveillance model. Finally, out-of-sample performance for all three rankings is compared over a sequence of two-year windows running from 1992 to 2005, counterfactually as if super-

⁶ Verification of financials is an important source of value created by exams (Berger and Davies 1998; Flannery and Houston 1999). Indeed, recent research has documented large adjustments in asset-quality measures following on-site visits, particularly for banks with emerging problems (Gunther and Moore 2000).

⁷ Peter Lynch ran Fidelity’s Magellan Fund from 1977 to 1990. During this period, fund value rose over 2,700 percent. Lynch was famous for looking past financial statements to the real world, observing consumer and firm behavior in malls, for example. For more details, see Lynch and Rothchild (2000).

visors in the fourth quarter of each year possessed data only up to that point. We find that jumbo CD signals would not have flagged banks missed by the CAMELS-downgrade model or would not have reduced uncertainty about banks flagged by the model. We also find that jumbo CD signals would not have provided earlier warning about developing problems in banks flagged by the CAMELS-downgrade model. These results are broadly consistent with other recent work, so we close by exploring reasons the surveillance value of market data may have been overestimated.

1. PRIOR LITERATURE

Research on the jumbo CD market since the mid-1970s—mostly with 1980s data—has consistently found evidence of risk pricing (see Table 1). Some 20 articles have been published using a mix of time series and panel approaches: 18 articles exploited U.S. data, 11 examined only yields, 4 examined only runoff (i.e., deposit withdrawals), and 5 studied both. Most drew heavily on quarterly financial statements. Only one article—the first contribution to the literature in 1976—found no link between bank risk and yields or runoff. In some ways, the robustness of these results is striking because U.S. samples mostly predate the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA). Before this Act, the majority of failures were resolved through purchases and assumptions, whereby the FDIC offered cash to healthy banks to assume the liabilities of failed ones. So, even though jumbo CD holders faced default risk in theory, many were shielded from losses in practice.⁸

Although evidence from prior literature about our out-of-sample test windows (1992–2005) is thinner, intuition and history make a case for significant risk sensitivity. The handful of articles looking at 1990s data found risk pricing, but no study examined jumbo CD data for the post-2000 period. Nonetheless, economic intuition suggests sensitivity should be strong because of three important institutional changes in the 1990s. First, as noted, the FDICIA directed the FDIC to resolve failures in the least costly way, which implies imposing a greater share of losses on uninsured bank creditors (Benson and Kaufman 1998; Kroszner and Strahan 2001).⁹ This change should

⁸ Before 1991, expected losses had three components: (1) the probability of bank failure, (2) the loss if the failed bank were not purchased by a healthy one, and (3) the probability the failed bank would not be purchased. Even if (1) and (2) were positive, expected losses would still be approximately zero if jumbo CD holders expected all failures to be resolved with purchase and assumptions. The need to model FDIC behavior, therefore, complicates estimation of risk sensitivity for the pre-1991 regime. Suppose, for example, (1) and (2) fall, reducing expected losses, incentives to monitor risk, and jumbo CD risk sensitivity. But the FDIC responds by curtailing implicit coverage—perhaps because of the reduced threat of contagious runs. If large enough, this offsetting effect could induce a rise in measured sensitivity to bank condition.

⁹ As discussed in footnote 8, expected losses equal zero if jumbo CD holders anticipate resolution through purchase and assumptions. But the FDICIA should have changed expectations

have increased expected losses for jumbo CD holders and their incentive to monitor bank condition. Second, the Financial Institutions Reform, Recovery, and Enforcement Act of 1989 required supervisors to disclose serious enforcement actions (Gilbert and Vaughan 2001).¹⁰ Third, in the late 1990s, the FDIC began putting quarterly financial data for individual banks on the Web, along with tools for comparing performance with industry peers. The second and third change should have lowered the cost to jumbo CD holders of monitoring bank condition. Evidence from U.S. banking history also implies our sample should feature strong risk pricing. Gorton (1996), for example, documented a link between discounts on state bank notes and issuer condition during the free-banking era, while Calomiris and Mason (1997) observed sizable differences in yields and runoff for weak and strong Chicago banks prior to the 1932 citywide panic. Friedman and Schwartz (1960) also noted that public identification of banks receiving loans from the Reconstruction Finance Corporation triggered runs in August 1932. More recently, Continental Illinois began hemorrhaging uninsured deposits when the extent of its problems became public in May 1984 (Davison 1997). In all these cases, uninsured claimants monitored and reacted to changes in bank condition, thereby impounding risk assessments into prices or quantities.

Evidence of risk pricing in the jumbo CD market does not imply that yield and runoff data would add value in surveillance. First, stable in-sample estimates of reactions to current bank condition and reliable out-of-sample forecasts of emerging safety-and-soundness problems are not the same thing. Evidence from the market efficiency literature, for example, has demonstrated that trading strategies based on well-documented pricing anomalies, such as calendar effects, size effects, and mean revision, do not offer abnormal returns when tested in real time by fund managers (Roll 1994; Malkiel 2003). Second, just as assessing the profitability of trading rules requires a benchmark, such as the return from an index fund, assessing the surveillance value of market data requires a baseline for current practices. It is not enough to note that jumbo CD signals flag problem banks because supervisors already have systems in place for these purposes. The true litmus test is this: Does integration of

about FDIC behavior. Between 1988 and 1990, jumbo CD holders suffered losses in only 15 percent of bank failures. From 1993 to 1995, they lost money 82 percent of the time.

¹⁰The term “enforcement action” refers to a broad range of powers used to address suspect practices of depository institutions and institution-affiliated parties—the supervisory sanctions mentioned in footnote 1. Typically, these actions are imposed in response to adverse exam findings, but they can also be triggered by deficient capital levels under Prompt Corrective Action or by negative information gathered through off-site surveillance. Usually enforcement actions are implemented in a graduated manner, with informal preceding formal actions. An informal action is the most common; it is simply a private, mutual understanding between a bank and its supervisory agency about the steps needed to correct problems. Formal actions are far more serious. Supervisors resort to them only when violations of law or regulations continue or when unsafe and abusive practices occur. Formal enforcement actions are legally enforceable and, in most cases, publicly disclosed.

yields and runoff into actual surveillance routines consistently and materially improve out-of-sample forecast accuracy?¹¹

Four recent articles have gauged the surveillance value of market data against a current practices benchmark. Evanoff and Wall (2001) compared regulatory capital ratios and sub-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), meanwhile, observed in-sample improvement in model fit when estimated default frequencies (EDFs, as produced by Moody's KMV) were included in an econometric model designed to predict holding company supervisory ratings with accounting data. Krainer and Lopez (2004) also experimented with equity market variables—in this case, cumulative abnormal stock returns as well as EDFs—in a model of holding company ratings. Unlike Gunther, Levonian, and Moore (2001), they assessed value added in one-quarter-ahead forecasts. Like Evanoff and Wall (2001), they noted only a modest improvement in out-of-sample performance. Finally, Curry, Elmer, and Fissel (2003) added various equity signals to an econometric model built to predict four-quarter-ahead supervisory ratings, again witnessing only a slight increase in forecast accuracy.

Recent tests against a surveillance benchmark have advanced the market data literature, to be sure, but the absence of empirical tests modeled on actual practice mutes the potential impact on supervisory policy. Evanoff and Wall (2001), for example, proxied supervisor perceptions of safety and soundness with regulatory capital ratios—a practice that was problematic because capital is the sole criterion only when Prompt Corrective Action (PCA) thresholds are violated. Otherwise, a variety of measures are weighed.¹² In addition, Gunther, Levonian, and Moore (2001) and Krainer and Lopez (2004) conducted performance tests with holding company data—a problematic approach because, as noted, off-site surveillance focuses on individual banks. Indeed, the Federal Reserve, which has responsibility for holding company supervision, does not maintain an econometric model estimated on holding company data.¹³ Finally, Gunther, Levonian, and Moore (2001) and Curry, Elmer, and Fissel (2003) relied on tests unlikely to impress supervisors: the first assessing

¹¹ As discussed in footnote 10, supervisors use enforcement actions to induce banks to address safety-and-soundness problems. Some are quite severe, going as far as permanent removal from the banking industry. The earlier actions are imposed, the more likely problems can be corrected. But enforcement actions impose significant costs on the bank, so supervisors prefer to wait for compelling evidence of serious problems. Hence, jumbo CD signals could add supervisory value by reinforcing conclusions yielded by other surveillance tools, thereby facilitating swifter action.

¹² Footnote 1 discusses the CAMELS framework supervisors use to assess bank condition. In any event, evidence from counterfactual applications of PCA to late 1980s/early 1990s data (Jones and King 1995; Peek and Rosengren 1997) suggests the thresholds are too low to affect supervisor behavior.

¹³ Each article estimated a unique holding company model to benchmark surveillance procedures. Both tested joint hypotheses: (1) the model approximates the one the Fed would use and (2) equity market signals enhance the performance of that model.

in-sample performance only and the second assessing out-of-sample performance with a contemporaneous holdout (rather than a period-ahead sample). Our work improves on this research by employing an econometric model used in surveillance, out-of-sample timing conventions patterned on current practices, and data taken from bank (rather than holding company) financial statements and supervisor assessments. Even more important, we contribute a coherent framework for use in future research on the surveillance value of market data.

2. THE DATA

To test the surveillance value of jumbo CD data, we built a long panel containing financial data and supervisory assessments for all U.S. commercial banks. This data set contained income statement and balance sheet series as well as CAMELS composite and management ratings from 1988:Q1 through 2005:Q4.¹⁴ The accounting data came from the Call Reports—formally the Reports of Condition and Income—which are collected under the auspices of the Federal Financial Institutions Examination Council (FFIEC). The FFIEC requires all U.S. commercial banks to submit such data quarterly to their principal supervisor; most reported items are publicly available. CAMELS ratings were pulled from a nonpublic portion of the National Information Center database; only examiners, analysts, and economists involved in supervision at the state or federal level can access these series. Only one substantive sample restriction was imposed—exclusion of banks with operating histories of under 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 instance, *de novos* often lose money in their early years, so earnings ratios are poor. Extreme values could introduce considerable noise into risk rankings, making it more difficult to assess relative performance. Another reason for dropping *de novos* is that supervisors already monitor these banks closely. The Federal Reserve, for example, examines newly chartered banks every six months until they earn a composite rating of 1 or 2 in consecutive exams.

Although our testing framework improves on prior research, our data still contain measurement error. Only a small number of money center banks issue negotiable instruments that are actively traded, so true market yields are not available for a cross section of the industry. It is possible, however, to construct average yields from the Call Reports for all U.S. banks by dividing quarterly in-

¹⁴Two data notes: (1) Explicit assessment of market risk sensitivity (S) was added in 1997, so pre-1997 composites are CAMEL ratings, and (2) none of our empirical exercises exploits the entire dataset (1988:Q1–2005:Q4); each uses a suitable sub-sample. For example, estimation of the downgrade model ends in 2003 to permit out-of-sample tests on 2004–2005.

terest expense by average balance. Subtracting rates on comparable-maturity Treasuries from these yields produces something that looks like a default premium series. Other researchers have successfully tested hypotheses about bank risk with this approach (for example, James 1988; Keeley 1990; and, more recently, Martinez-Peria and Schmukler 2001). Still, two related types of measurement error must be acknowledged; the proxy is an average rather than marginal measure (and, therefore, somewhat backward looking), and it is a quarterly accounting rather than real-time economic measure.

Measurement error in this series does not imply that jumbo CD data taken from the Call Report lack surveillance value. Jumbo CD holders may react to rising risk by withdrawing funds, and changes in account balances (deposit runoff) can be measured error-free with accounting data.¹⁵ Moreover, distress models based on financial statements have been a cornerstone of public- and private-sector surveillance for decades (Altman and Saunders 1997). Indeed, federal and state supervisors alike give heavy weight to book-value measures of credit risk and capital protection in routine surveillance, yet both contain serious measurement error (Barth, Beaver, and Landsman 1996; Reidhill and O’Keefe 1997). Finally, and most importantly, the supervisory return on jumbo CD signals—or any market signal for that matter—depends not on the value of the signal alone, but rather on that value net of the cost of extraction. Current surveillance routines are built around the Call Reports and, as noted, these reports already contain the data necessary to construct yield and runoff series for jumbo CDs. Even if the marginal surveillance value of jumbo CD signals were low relative to pure market signals because of measurement error, the marginal cost of extracting jumbo CD signals is near zero. The cost of integrating market signals into off-site surveillance is not as low because of the regulatory burden associated with any compulsory security issues and the training burden associated with changes in supervisory practices. It is possible, therefore, that jumbo CD data add more net value than pure market signals. In short, the surveillance value of jumbo CD data is ultimately an empirical issue.

Still, the net contribution of jumbo CD signals to surveillance cannot be positive if measurement error renders the data hopelessly noisy. So, as a check, we performed a simple test on yields and another on runoff—both suggested that bank condition is priced. In the first test, we compared quarterly yields—that is, jumbo CD interest expense divided by average balance—for the 5 percent of banks most and least at risk of failure each year from 1992 to 2005

¹⁵ In the literature, “runoff” is used loosely as a synonym for withdrawals. For this test, we define it as quarter-over-quarter percentage changes in a bank’s total dollar volume of jumbo CDs. Later, we define “simple” deposit runoff similarly.

(the period used in out-of-sample testing).¹⁶ Over this period, yields at high-risk banks topped yields at low-risk banks by an average of 25 basis points. (By way of comparison, the average spread between yields on three-month nonfinancial commercial paper and three-month Treasury bills for 1992 to 2005 was 24 basis points.) Institutional changes in the 1990s appear to have strengthened risk pricing. Despite declining money market rates, the mean spread between “risky” and “safe” banks climbed from 14 basis points for 1992–1997 to 33 for 1998–2005 (difference significant at 1 percent). In the second test, we examined quarterly jumbo CD growth at the 169 U.S. banks that failed between 1992 and 2005 for two distinct periods in the migration to failure: two to four years out and zero to two years out. Mean growth two to four years prior to failing was a healthy 8.4 percent. But in the final two years, quarterly growth turned sharply negative, averaging -4.0 percent—a pattern consistent with jumbo CD holders withdrawing funds to avoid losses.

As a final check, we regressed yields and runoff on failure probability and suitable controls; the results also attested to risk pricing. The sample contained observations for all non-de-novo banks with satisfactory supervisory ratings from 1988:Q1 to 2004:Q4.¹⁷ (Table 2 contains the results.) Both coefficients of interest were “correctly” signed and significant at the 1-percent level, implying a rise in failure risk translated into higher yields and larger runoff: coefficient magnitudes were economically small, but it is important to remember that risk sensitivity is a cardinal concept whereas risk ranking is an ordinal one. Recent back-testing of the Focus Report highlights the difference. The Focus Report is a Call-Report-based, Federal Reserve tool for predicting the impact of a 200-basis-point interest rate shock on bank capital. For the 1999–2002 interest rate cycle, Sierra and Yeager (2004) found that estimates of bank losses were very noisy, but risk rankings based on these losses were quite accurate. Our criterion for assessing jumbo CD data is analogous.

¹⁶ We estimated failure probabilities with the Risk-Rank model—one of two econometric surveillance models used by the Federal Reserve. See footnote 18 for more discussion of this model.

¹⁷ We controlled for factors suggested by academic literature, examiner interviews, and specification tests. These factors included term-to-maturity, the rate on Treasury securities with comparable maturities, economic conditions (dummies for quarters and states in the union), power in local deposit markets (dummy for banks operating in an MSA), access to parent-company support (dummy for banks in holding companies), and demand for funding in excess of local supply (dummy for banks with brokered deposits). The estimation sample included only satisfactory banks to parallel the performance tests of downgrade probability and jumbo CD rankings. Confining the analysis to 1- and 2-rated banks may seem odd, akin to testing risk sensitivities of AA or better corporate debt, but there are theoretical as well as practical justifications. Managers of nonregulated firms operate with considerable latitude up to the point of bankruptcy. Bank managers, in contrast, lose much of their discretion when an unsatisfactory rating is assigned. So market data for 3-, 4-, and 5-rated banks contain assessments of ongoing supervisory intervention as well as inherent risk. Excluding unsatisfactory banks also produces more relevant evidence about the surveillance value of market feedback. Supervisors continuously monitor these institutions, so market data are unlikely to yield new information. But knowledge of deteriorating 1s and 2s would be valued because these banks do not face constant scrutiny between exams.

Surveillance value is measured not by the precision of estimated sensitivities to bank risk, but rather by the improvement in risk rankings traceable to jumbo CD yields and runoff.¹⁸

3. MARKET ASSESSMENTS OF RISK: THE JUMBO CD RANKINGS

The first step in assessing the value of jumbo CD data was obtaining default premiums for all sample banks with satisfactory supervisory ratings. We created two measures—a “simple” and a “complex” default premium—to reduce the likelihood that performance tests would be biased by one, possibly poor, proxy. At the root of each measure was average yield—the ratio of jumbo CD interest expense to average balance, computed with Call Report data for each bank in each quarter. To convert yields into simple default premiums, we adjusted for the average maturity of a bank’s jumbo CD portfolio. To obtain a complex premium series, we used regression analysis to adjust yields for maturity and nonmaturity factors likely to affect jumbo CD rates.¹⁹ Simple

¹⁸ Besides measurement error, there are several idiosyncratic aspects of the jumbo CD market that might weaken risk pricing. Jumbo CD holders often receive other bank services—loan commitments and checking accounts, for example—so the issuer might price the relationship comprehensively. Another potential explanation is that many jumbo CDs are held by state or local governments and are, therefore, practically risk-free. (Most states require banks to “pledge” Treasury or agency securities against uninsured public deposits, thereby eliminating all but fraud risk.) Still another possibility is that many banks no longer fund at the margin with jumbo CDs—these instruments are now essentially core deposits because of the declining cost of commercial paper issuance and the increasing availability of Federal Home Loan Bank advances. A final, related possibility is that posted jumbo CD rates are sticky, “clustering” around integers and even fractions like retail CD rates (Kahn, Pennacchi, and Sopranzetti 1999). These market characteristics may account for modest risk sensitivities in the yield and runoff regressions. Still, evidence presented in this section suggests the data contain information about bank condition, thereby satisfying the necessary condition for jumbo CD risk rankings to add value in surveillance.

¹⁹ Default premiums were obtained with maturity and nonmaturity controls from the Call Report. The reporting convention for maturities changed in the middle of our sample. From 1989 to 1997, the FFIEC required banks to slot jumbo CDs in one of four buckets: “less than 3 months remaining,” “3 months to 1 year remaining,” “1 to 5 years remaining,” and “over 5 years remaining.” In 1997, the two longest maturity buckets became “1 to 3 years remaining” and “over 3 years remaining.” These maturity measures are crude—jumbo CDs in the shortest bucket might have been issued years ago—but they offer the only means of controlling for term structure. We produced simple premiums by first multiplying each bank’s jumbo CD balance for each maturity bucket by that quarter’s yield on Treasuries of comparable maturity. The sum of the resulting values, divided by average jumbo CD balances, approximated that bank’s risk-free yield. Simple default premiums for a quarter were then the difference between a bank’s risk-free yield and its average jumbo CD yield that quarter. Complex premiums controlled for other factors likely to affect jumbo CD demand or supply. Specifically, average yields were regressed on average jumbo CD maturity, maturity-weighted Treasury yield (the portion of a sample bank’s CDs in each maturity bucket, multiplied by that quarter’s yield on a comparable-maturity Treasury), and the same nonmaturity controls used in the data-check equations in Section 3. Regression residuals served as the complex premium series. Carefully controlling in this way for maturity and nonmaturity influences on yields should render the resulting default premium series a cleaner measure of default risk.

and complex default premiums were highly correlated, exhibiting an average year-by-year correlation coefficient of 0.88.

The second step was generating a deposit-runoff series for all banks with satisfactory ratings. When significant transaction or information frictions are present, jumbo CD holders are apt to withdraw funds as failure probability rises (Park and Peristiani 1998). Another reason to examine runoff is that a bank's demand for jumbo CDs could depend on its condition. Billett, Garfinkel, and O'Neal (1998) and Jordan (2000) have documented a tendency for risky banking organizations to substitute insured for uninsured deposits to escape market discipline. If such substitution is important, escalating risk would show up in declining jumbo CD balances rather than rising default premiums. To explore these possibilities, we again computed two measures of runoff: "simple" and "complex." Simple deposit runoff was defined for each sample bank as the quarterly percentage change in jumbo CD balances.²⁰ The complex series was constructed by adjusting simple runoff with the same approach used to identify complex default premiums—that is, regressions of quarterly deposit runoff on maturity and nonmaturity factors likely to affect jumbo CD demand or supply. The correlation coefficient for simple and complex runoff was 35 percent, somewhat less than the correlation between simple and complex default premiums.

4. THE SURVEILLANCE BENCHMARK—DOWNGRADE PROBABILITY RANKINGS

Since the 1980s, econometric models have played an important role in bank surveillance at all three federal supervisory agencies.²¹ We benchmark the performance of these models with the CAMELS-downgrade model developed by Gilbert, Meyer, and Vaughan (2002).²² This model is a probit regres-

²⁰ Technically, a positive number implies growth while a negative number implies runoff. To simplify, we refer to all percentage changes as runoff. By our nomenclature, a bank can experience positive or negative jumbo CD runoff.

²¹ Since the early 1990s, the Federal Reserve has relied on two econometric models, collectively known as SEER—the System for Estimating Examination Ratings. One model, the Risk-Rank model, exploits quarterly Call Report data to estimate the probability of failure over the next two years. The other model, the Ratings model, produces "shadow" CAMELS ratings—that is, the composite that would have been assigned had an examination been performed using the latest Call Report submission. Every quarter, analysts at the Board of Governors feed the data into the SEER models and forward the results to the Reserve Banks. The surveillance unit at each Bank, in turn, follows up on flagged institutions. The FDIC and the OCC use similar approaches in off-site monitoring of the banks they supervise (Reidhill and O'Keefe 1997).

²² The model is discussed in detail here because it is possible in-sample performance has deteriorated since the Gilbert, Meyer, Vaughan (2002) estimation sample ended in 1996. Such deterioration would bias performance tests in this research in favor of the jumbo CD rankings. So, we explain the rationale for the explanatory variables and present evidence of in-sample fit to make the case that the CAMELS-downgrade model is still a good benchmark for current surveillance practices.

sion estimating the likelihood a bank with a satisfactory supervisory rating (a CAMELS 1 or 2 composite) will migrate to an unsatisfactory rating (a 3, 4, or 5 composite) in the coming eight quarters. Explanatory variables were selected in 2000 based on a survey of prior research and interviews with safety-and-soundness examiners. Table 3 describes the independent variables, as well as the expected relationship between each variable and downgrade risk. Table 5 contains summary statistics for these variables. Most variables are financial performance ratios related to leverage risk, credit risk, and liquidity risk—three risks that have consistently produced financial distress in commercial banks (Putnam 1983; Cole and Gunther 1998).

We benchmark current surveillance procedures with a CAMELS-downgrade model. Traditionally, the most popular econometric surveillance tool has been a failure-prediction model. But failures have been rare since the early 1990s, preventing re-estimation of these models. Any resulting “staleness” in coefficients could bias performance tests by compromising the surveillance benchmark used to assess jumbo CD data. Unlike failures, migration to unsatisfactory ratings remains common, so a downgrade model can be updated quarterly. (Table 4 contains 1992–2005 data on downgrade frequency.) Recent research confirms that a CAMELS-downgrade model would have improved slightly over a failure-prediction model in the 1990s (Gilbert, Meyer, and Vaughan 2002). Even more important, a downgrade model is best suited to support current supervisory practice. Institutions with unsatisfactory ratings represent significant failure risks; supervisors watch them closely and constantly to ensure progress toward safety and soundness. Most 1- and 2-rated banks, in contrast, are monitored between exams through quarterly Call Report submissions. As noted, early supervisory intervention improves chances for arresting financial deterioration. So a tool that more accurately flags deteriorating banks with Call Report data would yield the most surveillance value. These considerations have prompted one Federal Reserve Bank to “beta test” a CAMELS-downgrade model in routine surveillance and the Board of Governors to add a downgrade model to the System surveillance framework in 2006.

The CAMELS-downgrade model relies on six measures of credit risk, the risk that borrowers will not render promised interest and principal payments. These measures include the ratio of loans 30 to 89 days past due to total assets, the ratio of loans over 89 days past due to total assets, the ratio of loans in nonaccrual status to total assets, the ratio of other real estate owned to total assets (OREO), the ratio of commercial and industrial loans to total assets, and the ratio of residential real estate loans to total assets. High past-due and nonaccruing loan ratios increase downgrade probability because, historically, large portions of these loans have been charged off. OREO consists primarily of collateral seized after loan defaults, so a high OREO ratio signals poor credit-risk management. Past due loans, nonaccruing loans, and OREO

are backward looking; they register asset quality problems that have already emerged (Morgan and Stiroh 2001). The ratio of commercial and industrial loans to total assets is forward looking because, historically, losses on these loans have been relatively high. The ratio of residential real estate loans to total assets also provides a forward-looking dimension because, historically, the loss rate on mortgages has been relatively low. Other things equal, an increase in dependence on commercial loans or a decrease in dependence on mortgage loans should translate into greater downgrade risk.

The model contains two measures of leverage risk—the risk that losses will exceed capital. Measures of leverage risk include the ratio of total equity (minus goodwill) to total assets and the ratio of net income to average assets (or, return on assets). Return on assets is part of leverage risk because retained earnings are an important source of capital for many banks, and higher earnings provide a larger cushion for withstanding adverse economic shocks (Berger 1995). Increases in capital protection or earnings strength should reduce the probability of migration to an unsatisfactory rating.

Liquidity risk, the risk that loan commitments cannot be funded or withdrawal demands cannot be met at a reasonable cost, also figures in the CAMELS-downgrade model. This risk is captured by two ratios: investment securities as a percentage of total assets and jumbo CD balances as a percentage of total assets. A large stock of liquid assets, such as investment securities, indicates a strong ability to meet unexpected funding needs and, therefore, should reduce downgrade probability. Liquidity risk also depends on a bank's reliance on non-core funding, or "hot money." Non-core funding, which includes jumbo CDs, can be quite sensitive to changes in money market rates. Other things equal, greater reliance on jumbo CDs implies greater likelihood of a funding runoff or an interest expense shock and, hence, a larger risk of receiving a 3, 4, or 5 rating in a future exam.

Finally, the model uses three control variables to capture downgrade risks not strictly associated with current financials. These controls include the natural logarithm of total assets because large banks are better able to reduce risk by diversifying across product lines and geographic regions. As Demsetz and Strahan (1997) have noted, however, such diversification relaxes a constraint, enabling bankers to assume more risk, so the *ex ante* relationship between asset size and downgrade probability is ambiguous. We also add a dummy variable for 2-rated banks because they migrate to unsatisfactory status more often than 1-rated banks. (See Table 4 for supporting evidence.) The list of control variables rounds out with a dummy for banks with management component ratings higher (weaker) than their composite rating. In such banks, examiners have raised questions about managerial competence, even though problems have yet to appear in financial statements.

We estimated the CAMELS-downgrade model for 13 overlapping two-year windows running from 1990–1991 to 2002–2003.²³ Each equation regressed downgrade incidence (1= downgraded, 0= not downgraded) in years $t + 1$ and $t + 2$ on accounting and supervisory data for banks with satisfactory ratings in the fourth quarter of year t . For example, to produce the first equation (1990–1991 in Table 6), downgrade incidence in 1990–1991 was regressed on 1989:Q4 data for all 1- and 2-rated banks that were not de novos. We continued with this timing convention, estimating equations year by year, through a regression of downgrade incidence in 2002–2003 on 2001:Q4 data. Observations ranged from 6,367 (2002–2003 equation) to 8,682 (1995–1996 equation); the count varied because bank mergers and supervisory reassessments altered the number of satisfactory institutions over the estimation period.

The model fit the data relatively well throughout the estimation sample. (Table 6 contains the results.)²⁴ The hypothesis that model coefficients jointly equaled zero could be rejected at the 1 percent level for all 13 equations. The pseudo-R2, the approximate proportion of variance in downgrade/no downgrade status explained by the model, was in line with numbers in prior early warning studies—ranging from 15.0 percent (1994–1995 equation) to 22.6 percent (1991–1992 equation). Estimated coefficients for seven explanatory variables—the jumbo-CD-to-total-asset ratio, the past due and nonaccruing loan ratios, the net-income-to-total-asset ratio, and the two supervisor rating dummies—were statistically significant with expected signs in all eight equations. The coefficient on the logarithm of total assets had a mixed-sign pattern, which is not surprising given *ex ante* ambiguity about the relationship between size and risk. The coefficients on the other six explanatory variables were statistically significant with the expected sign in at least three equations.

Comparing out-of-sample performance of jumbo CD and downgrade probability rankings is not as biased as it may first appear. True, jumbo CD rankings draw on one variable—either default premiums or deposit runoffs—while the downgrade probability rankings draw on 13 variables. But theory suggests

²³ Gilbert, Meyer, and Vaughan (2002) estimated the model for six windows running from 1990–1991 to 1995–1996. We re-estimated the model for these windows because Call Report data have since been revised, which implies slight changes in coefficients. We also wanted to use a consistent approach and consistent data for the entire estimation sample to insure subsequent out-of-sample tests of jumbo CD data were not biased against the surveillance benchmark.

²⁴ 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 at the 5-percent level, and three at the 1-percent level. The pseudo-R2 indicates the approximate proportion of variance in downgrade status explained by the model. Overall, the downgrade-prediction model fit the data well. For all eight regressions, the hypothesis that all model coefficients equal zero could be rejected at the 1-percent level of significance. In addition, eight of the 13 regression variables are significant with the predicted sign in all eight years, and all variables were significant in at least some years.

premiums and runoff should summarize overall bank risk, not just one type of exposure such as leverage or credit risk. Put another way, jumbo CD holders should sift through all available information about the condition of the issuing bank, note any changes in expected losses, and react to heightened exposures by demanding higher yields or withdrawing funds. This process should impound all relevant information—financial as well as anecdotal—into default premiums and deposit runoff just as the econometric model impounds all relevant Call Report data into a CAMELS-downgrade probability.

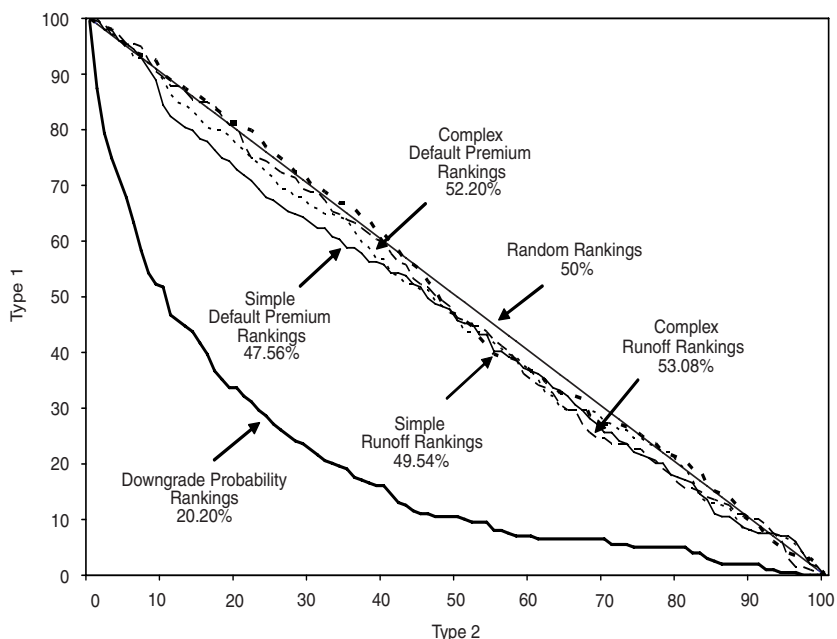
5. ASSESSING OUT-OF-SAMPLE PERFORMANCE: POWER CURVE AREAS

We assessed out-of-sample performance using both Type 1 and Type 2 error rates. Both forecast errors are costly. A missed downgrade to unsatisfactory status—Type 1 error—is costly because accurate downgrade predictions give supervisors more warning about emerging problems. A predicted downgrade that does not materialize—Type 2 error—is costly because unwarranted supervisory intervention wastes scarce examiner resources and disrupts bank operations. A tradeoff exists between the two errors—supervisors could eliminate overprediction of downgrades by assuming no banks are at risk of receiving an unsatisfactory rating in the next two years.

For each risk ranking, it is possible to draw a power curve indicating the minimum achievable Type 1 error rate for any desired Type 2 error rate (Cole, Cornyn, and Gunther 1995). For example, tracing the curve for simple default premium rankings starts by assuming no sample bank is a downgrade risk. This assumption implies all subsequent downgrades are surprises—a 100 percent Type 1 error rate. Because no banks are incorrectly classified as downgrade risks, the Type 2 error rate is zero. The next point on the curve is obtained by selecting the bank with the highest simple default premium (maturity-adjusted spread over Treasury). If that bank suffers a downgrade in the following eight quarters, then the Type 1 error rate decreases slightly. The Type 2 error rate remains zero because, again, no institutions are incorrectly classified as downgrade risks. If the selected bank does not suffer a downgrade, then the Type 1 error rate remains 100 percent, and the Type 2 error rate increases slightly. Selecting banks from highest to lowest default premium and recalculating error rates each time produces a power curve. At the lower right extreme of the curve, all banks are considered downgrade risks—the Type 1 error rate is 0 percent, and the Type 2 error rate is 100 percent. Figure 1 illustrates with the power curves for downgrade probability and jumbo CD rankings for the 1992–1993 test window.

Areas under power curves provide a basis for comparing out-of-sample performance across risk rankings. The area for each ranking is expressed as a percentage of the total area of the box. A smaller percentage implies a lower

Figure 1 How Well Do Jumbo CD and Downgrade Probability Rankings Perform Out-of-Sample? 1992–1993 Test Window



Notes: This Figure depicts the power curves for risk rankings based on jumbo CD default premiums, jumbo CD runoff, and downgrade probabilities (as produced by the CAMELS-downgrade model) for the 1992–1993 out-of-sample test window. These curves reflect the tradeoff between Type 1 and Type 2 errors. (Type 1 errors are missed downgrades, and Type 2 errors are overpredicted downgrades.) A convenient way to compare the performance of risk rankings is to calculate the area under the power curve for each ranking and express that area as a percentage of the total for the box. Smaller areas are desired because they imply a simultaneous reduction in both types of errors. The 50-percent line is the power curve produced when downgrade risks are selected randomly over a large number of trials. The power curves above show that rankings based on downgrade probabilities would have significantly outperformed rankings based on jumbo CD default premiums and runoff. Indeed, jumbo CD rankings would not have improved materially over random rankings.

overall Type 1 and Type 2 error rate and, hence, a more accurate forecast. The area for a “random-ranking” power curve offers an example as well as a yardstick for evaluating the economic significance of differences in forecast accuracy. Random selection of downgrade candidates, over a large number of trials, will produce power curves with an average slope of negative one. Put another way, the area under the random-ranking power curves, on average, equals 50 percent of the total area of the box. Power curve areas can be

compared—jumbo CD ranking against downgrade probability rankings or either ranking against a random ranking—for any error rate. Assessing forecast accuracy this way, though somewhat atheoretical, makes best use of existing data. A more appealing approach would minimize a loss function explicitly weighing the benefits of early warning about financial distress against the costs of wasted examination resources and unnecessary regulatory burden. Then, the relative performance of risk rankings could be assessed for the optimal Type 1 (or Type 2) error rate. The requisite data, however, are not available.

A specific example will clarify the mechanics of the “horse race” we run for risk rankings. To assess the surveillance value of simple default premiums for 1992–1993, we start by assuming it is early 1992, just after fourth quarter 1991 data became available. In accordance with standard surveillance procedures, 1990–1991 downgrade incidences are regressed on 1989:Q4 data for the 13 explanatory variables in the CAMELS-downgrade model. Model coefficients are then applied to 1991:Q4 data to estimate the probability that each 1- and 2-rated bank will migrate to an unsatisfactory condition between 1992:Q1 and 1993:Q4. These banks are then ranked from highest to lowest downgrade probability. At the same time, all banks with satisfactory supervisory ratings are ranked from highest to lowest simple default premium (maturity adjusted spread over Treasury), also using 1991:Q4 data, under the assumption that high spreads map into high downgrade probabilities. After two years, the record of missed and overpredicted downgrades is compiled to generate power curve areas for each ranking. A smaller area for the downgrade probability ranking would imply that simple default premiums added no surveillance value in the 1992–1993 test window.

6. EMPIRICAL EVIDENCE

Downgrade Model Rankings and Jumbo CD Rankings—Full-Sample Results

The evidence suggests jumbo CD default premiums would have contributed nothing to bank surveillance between 1992 and 2005 when used to forecast downgrades two years out. (Columns 2, 3, and 4 of Table 7 contain the relevant power curve areas.) Over the 13 test windows, the average area under the simple default premium power curve (45.63 percent) and the average area under the complex default premium power curve (49.70 percent) did not differ statistically or economically from the random-ranking benchmark (50 percent). In contrast, the average area under the downgrade model power curve (19.66 percent) came to less than half of that benchmark. Power curve areas for individual two-year test windows showed the same patterns. Specifically, downgrade model areas ranged from 15.24 percent (1996–1997) to 22.39 percent (1994–1995); simple default premium areas ran from 41.56 percent (2003–2004) to 50.57 percent (1994–1995); and complex default premium

areas varied from 45.69 percent (1998–1999) to 52.20 percent (1992–1993). The poor performance of jumbo CD rankings relative to downgrade model rankings suggests default premiums would not have flagged banks missed by conventional surveillance. The poor performance relative to the random-selection benchmark suggests default premiums would not have increased supervisor confidence about rankings produced by the CAMELS-downgrade model.

Out-of-sample performance of risk rankings based on jumbo CD runoff was no better. (Columns 5 and 6 of Table 7 contain the relevant power curve areas.) The average area for simple runoff rankings across all test windows was 46.12 percent while the average for complex runoff was 50.47 percent—again, statistically and economically indistinguishable from random selection. And once again, patterns were consistent across individual two-year test windows. Power curve areas for simple runoff rankings varied from 43.56 percent (1999–2000) to 49.54 percent (1992–1993), areas under complex runoff curves from 45.75 percent (1998–1999) to 53.08 percent (1992–1993). This consistently poor performance suggests runoff rankings would not have helped spot downgrade risks two years out between 1992 and 2005.

Changing forecast horizons did not alter the results. Over 13 one-year windows, downgrade model rankings produced an average power curve area of 17.37 percent (standard deviation across test windows of 2.39 percent). In contrast, simple default premium rankings produced an average area of 45.38 percent (standard deviation of 2.95 percent) and complex default premium rankings, an average area of 50.29 percent (standard deviation of 3.26 percent). Areas for runoff rankings were even closer to the random-ranking benchmark—46.79 percent on average for simple runoff (standard deviation across the 13 windows of 2.47 percent) and 50.87 percent for complex runoff (standard deviation of 3.45 percent). Lengthening the forecast horizon to three years yielded similar numbers. This evidence goes to the timeliness of information in jumbo CD rankings. As noted, market data could enhance surveillance by flagging problems before existing tools. But, between 1992 and 2005, jumbo CD data would not have improved over random selection at any forecasting horizon, much less current surveillance procedures. Put another way, feedback from the jumbo CD market would not have provided supervisors with earlier warning about developing problems.

Jumbo CD rankings constructed from both default premiums and runoff did not improve over random selection, either. In theory, price and quantity signals from the jumbo CD market, though weak when used singly, could jointly capture useful information about future bank condition. If so, a model relying only on multiple signals could add value—even if performance relative to the benchmark was poor—by reducing supervisor uncertainty about banks flagged by conventional surveillance. We explored this possibility by estimating a downgrade model with (1) only simple default premiums and

runoff as explanatory variables and (2) only complex default premiums and runoff as explanatory variables. Out-of-sample performance was then tested over a variety of forecasting horizons for 1992–2005. Columns 7 and 8 in Table 7 contain the results for two-year windows; they are representative. Over all 13 tests, the power curve area for the bivariate “simple” model averaged 45.55 percent (standard deviation across individual windows of 2.49 percent) and 50.05 percent for the bivariate “complex” model (standard deviation of 2.99 percent). Further perspective can be gained by comparing these numbers to power curve areas produced by a “pared down” model including only the dummy variables in the baseline CAMELS-downgrade model. The average power curve area for this model across the 13 two-year windows was 30.07 percent. Taken together, this evidence suggests jumbo CD data would not have reduced supervisory uncertainty about banks flagged by conventional surveillance tools.

Downgrade Model Rankings and Jumbo CD Rankings—Sub-Sample Results

Although jumbo CD risk rankings would not have contributed to general surveillance of 1- and 2-rated banks, default premiums and runoff might improve monitoring of specific cohorts such as banks with short jumbo CD portfolios, large asset portfolios, no foreign deposits, low capital ratios, or significant “deposits at risk.”

The marginal-average problem noted earlier could in part account for the weak performance of default premium rankings. As an arithmetic matter, today’s average yield will be more representative of today’s risk levels if jumbo CD maturities are short. To explore this possibility, we replicated all out-of-sample tests described earlier in Section 6 on a sub-sample of banks with weighted-average portfolio maturities under six months. The results did not change. At the two-year horizon, for example, the average area under complex default premium power curves was 42.97 percent. At the one-year horizon, the average area was 42.30 percent. Put simply, long jumbo CD maturities do not account for the poor performance of default premiums.

The jumbo CD market might emit stronger risk signals 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 runoff of community bank jumbos could respond sluggishly to changes in risk no matter how short the maturity of the portfolio. Another reason large bank signals may be more informative is that monitoring costs for their uninsured depositors are lower—these institutions have publicly traded securities and are closely followed by market analysts. To test for an asset-threshold effect, we reproduced all out-of-sample tests from Section 6 on a sub-sample

of banks holding more than the median level of assets. Out-of-sample results for this sub-sample were qualitatively similar to the results from the full sample. Across the 13 two-year test windows, for example, the area under the simple default premium power curves averaged 44.90 percent (compared with 45.63 percent for the full sample). We also tested risk rankings for banks holding more than \$1 billion in assets and for banks with SEC registrations. Each time, we compared results from the large bank sub-sample with results from the remaining sub-sample (i.e., banks holding less than median assets, banks holding less than \$1 billion in assets, and banks with no SEC registration), looking for performance differences across size cohorts. Size-split evidence was consistent: for large as well as community banks, the CAMELS-downgrade model proved to be the far superior surveillance tool, and rankings based on default premiums and runoff barely improved over random rankings.

Jumbo CD default premiums and runoff might improve off-site monitoring of banks with no foreign deposits. The National Depositor Preference Act of 1993 elevated claims of domestic depositors over claims of foreign depositors, reducing expected losses for jumbo CD holders (Marino and Bennett 1999). Domestic holders of jumbo CDs issued by banks with foreign offices may have perceived no default-risk exposure because of the financial cushion provided by foreign deposits. To test for a depositor-preference effect, we screened out banks with foreign deposits and replicated all out-of-sample tests. Again, the results mirrored the full-sample results; for example, for the two-year test windows, the average power curve area under the simple default premium rankings was 45.70 percent, virtually unchanged from the full sample (45.63 percent). Even for banks with no foreign-deposit cushion, jumbo CD rankings contained no useful supervisory information.

Finally, the jumbo CD market might yield clues about emerging problems in banks with high levels of uninsured deposits or low levels of capital. In theory, jumbo CD holders with more exposure—either because their uninsured balances are high or bank capital levels are low—have greater incentive to monitor and discipline risk. So we produced rankings for the quartile of sample banks with the largest volume of “deposits at risk” and the quartile with the lowest ratios of equity-to-assets (adjusted for bank size). Again, default premium and runoff rankings did not improve over random selection, much less conventional surveillance. As a final check, we looked at various intersections of the sub-samples—banks with high deposits at risk *and* low capital, banks with no foreign deposits *and* short jumbo CD maturities, etc. We generated rankings based on default premiums, deposit runoff, and both default premiums and deposit runoff. The results across all tests were consistent—jumbo CD rankings did not improve materially over random rankings at any forecast horizon.

Default Premiums and Runoff as Regressors in the Downgrade Model

Although default premium and runoff perform poorly as independent risk signals, they could add value as regressors in the CAMELS-downgrade model. Indeed, previous research has identified surveillance ratios with this property (Gilbert, Meyer, and Vaughan 1999). To pursue this angle, we estimated an “enhanced” CAMELS-downgrade model, adding both simple and complex measures of premiums and runoff to the 13 baseline explanatory variables. As before, out-of-sample performance was gauged by impact on power curve areas—first when default premiums and runoff were added to the baseline model and then when these variables were dropped from the enhanced model. As a further check, we assessed performance with the quadratic probability score (QPS)—a probit analogue for root mean square error (Estrella 1998; Estrella and Mishkin 1998).²⁵ If default premiums and runoff enhance the CAMELS-downgrade model, removing them from the enhanced model will boost QPS. Columns 9 and 10 in Table 7 contain power curve areas for the simple and complex enhancements of the downgrade model. Column 2 of panel A in Table 8 notes the impact of the two simple jumbo CD series on power curve areas; column 2 of panel B in Table 8 shows the impact of the series on QPS. (Results for complex default premiums and runoff are not reported because they mirror results for the simple series.) To facilitate interpretation, we note the impact on QPS and power curve areas of other variable blocks—such as the leverage-risk variables (equity-to-asset ratio and return on assets) and control variables (log of total assets, dummy for composite rating of 2, and dummy for management component rating weaker than the composite rating)—in columns 3 through 6 of panels A and B in Table 8. In Table 6, changes in QPS and power curve areas are expressed in percentage-change terms to permit direct comparison.

In performance tests for 1992–2005, default premiums and runoff did not enhance the CAMELS-downgrade model. Adding simple versions of the series increased (worsened) average power curve area by 4.17 percent (0.82 percentage points, from 19.66 percent for the baseline model to 20.48 percent). Removing these series from the enhanced downgrade model improved performance slightly by the power curve metric (reduced average power curve area by 0.26 percent) and worsened performance even more slightly by the QPS metric (increased average QPS by 0.06 percent). The leverage-risk variables provide

²⁵ To obtain QPS, we first computed downgrade probability for each sample bank with the CAMELS-downgrade model. Then, we subtracted R_t —a binary variable equal to one if the bank was downgraded in the out-of-sample window and zero if not—from the downgrade probability estimate. Finally, we squared the difference, multiplied the result by two, and averaged across all 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.

some perspective on the economic significance of these changes—dropping both of them increased the power curve area (worsened performance) by an average of 5.20 percent and increased the average QPS (worsened performance) by an average of 2.03 percent. These results held up in tests on the various sample cuts and forecasting horizons described in the previous subsections.

7. DISCUSSION

It is possible that some combination of measurement error and idiosyncrasies in the jumbo CD market accounts for our results. These factors may not be important enough to remove all evidence of risk pricing from jumbo CD data, but they may be important enough to prevent risk rankings based on the data from imparting valuable surveillance information.

But data problems and market frictions are unlikely to explain away the findings. As noted, recent studies using actual debt and equity market data rather than accounting proxies have found only modest surveillance value in market signals. Rather, the economic environment since the early 1990s probably plays an important role. Over this period, bank profitability and capital ratios soared to record highs. Some economists attribute these trends to an unprecedented economic boom that allowed banks to reap the upside of expansions into risky new markets and product lines (Berger et al. 2000). Others argue that stakeholders of large complex banking organizations insisted on greater capital cushions because of increasingly sophisticated risk exposures (Flannery and Rangan 2002). In such a high-profit, high-capital environment, jumbo CD signals—no matter how accurately measured or precisely determined—would convey little information because the benefits of monitoring are so low. Such an explanation would account for the successful use of average yields in bank-risk studies on data from the 1980s—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 data set 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.

8. CONCLUSION

The evidence suggests that feedback from the jumbo CD market would have added no value in bank surveillance between 1992 and 2005. Throughout the decade, risk rankings produced by a CAMELS downgrade—a model chosen to benchmark current surveillance practices—would have significantly outperformed risk rankings based on default premiums and runoff. Moreover, jumbo CD rankings would have improved little over random orderings.

Finally, adding jumbo CD signals to the downgrade model would not have improved its out-of-sample performance. These results hold up for a variety of sample cuts and forecast horizons. Taken together, these results imply that the marginal surveillance value of jumbo CD signals is less than the marginal production cost—even if that cost is very low.

Our results carry mixed implications for proposals to incorporate market data more formally into bank supervision. On the one hand, the evidence suggests available jumbo CD data would do little to enhance surveillance, thereby clearing the way for experimentation with other, “purer” market signals. On the other hand, if the “unique sample period” explanation for our results is true, then it is likely the surveillance value of signals from the market for bank debt and equity will vary over time. Other things equal, such time variation would lower the net benefit of integrating market data into current surveillance routines. Interpreted in this light, our findings imply that future policy and research work on market data should focus on identifying the specific bank claims that yield the most surveillance value in each state of the business cycle. Put another way, our findings—when viewed with other recent research—suggest the supervisory return from reliance on a single market signal through all states of the world may have been overestimated.

Table 1 Do Prior Studies Point to Risk Pricing in the Jumbo CD Market?

Authors	Issuer of Jumbo CD	Country	Sample Dates	Yield or Runoff?	Risk Pricing? Somewhat
Crane (1979)	Bank	United States	1974	Yield	Yes
Goldberg & Lloyd-Davies (1985)	Bank	United States	1976–1982	Yield	Yes
Baer & Brewer (1986)	Bank	United States	1979–1982	Yield	Yes
Hannan & Hanweck (1988)	Bank	United States	1985	Yield	Yes
James (1988)	Bank	United States	1984–1986	Yield	Yes
Cargill (1989)	Bank	United States	1984–1986	Yield	Yes
James (1990)	Bank	United States	1986–1987	Yield	Yes
Keeley (1990)	Bank	United States	1984–1986	Yield	Yes
Ellis & Flannery (1992)	Bank	United States	1982–1988	Yield	Yes
Cook & Spellman (1994)	Thrift	United States	1987–1988	Yield	Yes
Crabbe & Post (1994)	Bank	United States	1986–1991	Runoff	No
Brewer & Mondschean (1994)	Thrift	United States	1987–1989	Yield	Yes
Park (1995)	Bank	United States	1985–1992	Both	Yes
Park & Peristiani (1998)	Thrift	United States	1987–1991	Both	Yes
Jordan (2000)	Bank	United States	1989–1995	Both	Yes
Martinez Peria & Schmuckler (2001)	Bank	Argentina, Chile, Mexico	1981–1997	Both	Yes
Goldberg & Hudgins (2002)	Thrift	United States	1984–1994	Runoff	Yes
Birchler & Maechler (2002)	Bank	Switzerland	1987–2001	Runoff	Yes
Maechler & McDill (2003)	Bank	United States	1987–2000	Runoff	Yes
Hall, King, Meyer, & Vaughan (2005)	Bank	United States	1988–90, 1993–95	Both	Yes

Notes: This table summarizes the literature on risk pricing by jumbo CD holders. (“Bank” refers to commercial banks, bank holding companies, and thrift institutions. “Risk pricing” refers to price or quantity responses to a change in bank condition.) These studies used both cross-section and time-series techniques along with a variety of risk proxies and control variables. The weight of the evidence indicates bank condition is priced, suggesting that jumbo CD data might add value in off-site surveillance.

**Table 2 Do Jumbo CD Data Contain Evidence of Risk Pricing?
Evidence from Regressions of Yields and Runoff on Failure
Probabilities**

Sensitivity of Jumbo CD Yields and Runoff to Failure Probability 1988: Q1–2004:Q4		
Independent Variable	Dependent Variable: Yields Coefficient (Std. Error)	Dependent Variable: Runoff Coefficient (Std. Error)
Failure Probability (Lagged One Quarter)	0.0108*** (0.0038)	-0.3309*** (0.0505)
Maturity-Weighted Treasury Yield	0.6878*** (0.0391)	0.5324 (0.5211)
Average Portfolio Maturity	0.0820*** (0.0209)	-1.8743*** (0.2785)
Maturity-Treasury Interactive	-0.0140*** (0.0147)	-0.3408* (0.1959)
Holding Company Dummy	-0.0595 (0.0136)	-0.9042*** (0.1808)
Brokered Deposit Dummy	0.1403*** (0.0270)	-0.0868 (0.3592)
MSA Dummy	0.1384*** (0.0913)	1.7813 (1.2164)
R ²	0.2288	0.0125
F-statistic Control Variables	394.16***	33.25***
F-statistic Time Dummies	155.69***	21.88***
F-statistic State Dummies	21.53***	15.21***
Observations	229,486	229,486

Notes: This table reports results for regressions of jumbo CD yields and runoff on failure probabilities and various controls. Equations were estimated on a sub-sample of banks with satisfactory CAMELS ratings. Asterisks denote statistical significance at the 10- (*), 5- (**), and 1- (***) percent levels. A positive and significant failure-probability coefficient in the yield equation and/or a negative and significant coefficient in the runoff equation constitute evidence that bank condition is priced. The results indicate that greater risk of failure translates, on average, into higher yields and larger runoff, suggesting that jumbo CD data may have surveillance value in our sample.

Table 3 Which Variables Predict Migration to an Unsatisfactory CAMELS Rating?

	Independent Variables	Symbol	Impact on Downgrade Risk
Credit Risk	Loans past due 30-89 days (% of total assets)	PAST DUE 30	+
	Loans past due over 89 days (% of total assets)	PAST DUE 90	+
	Loans in nonaccrual status (% of total assets)	NONACCRUING	+
Leverage Risk	"Other real estate owned" (% of total assets)	OREO	+
	Commercial & industrial loans (% of total assets)	COMMERCIAL	+
	Residential real estate loans (% of total assets)	RESIDENTIAL	-
	Equity capital minus goodwill (% of total assets)	NET WORTH	-
Liquidity Risk	Net income (% of average assets)	ROA	-
	Book value of investments (% of total assets)	SECURITIES	-
Controls	Time deposits over \$100,000 (% of total assets)	JUMBO CDs	+
	Natural logarithm of total assets (thousands of dollars)	SIZE	Ambiguous
	Dummy for banks with composite CAMELS rating=2	CAMELS-2	+
	Dummy for banks with management rating > composite rating	MANAGEMENT	+

Notes: This table lists the independent variables in the CAMELS-downgrade model. Signs note the hypothesized relationship between each variable and the likelihood of downgrade from a satisfactory (CAMELS 1 or 2 composite) to an unsatisfactory rating (CAMELS 3, 4, or 5 composite). For example, the negative sign on the NET WORTH variable indicates that, other things equal, higher capital levels reduce the likelihood of migration to an unsatisfactory rating over the next two years.

Table 4 How Common Is Migration to Unsatisfactory CAMELS Ratings? Evidence from 1992–2005

Year of Migration to 3, 4, or 5 Rating	Rating at Beginning of Year	Banks with 1 & 2 Rating	Number Migrating to 3, 4, or 5 Rating	Percentage Migrating to 3, 4, or 5 Rating	Total Downgrades to 3, 4, or 5 Rating
1992	1	1,959	22	1.12	425
	2	5,275	403	7.64	
1993	1	2,289	7	0.31	182
	2	5,978	175	2.93	
1994	1	2,919	9	0.31	162
	2	5,742	153	2.66	
1995	1	3,106	8	0.26	102
	2	4,905	94	1.92	
1996	1	3,295	10	0.30	127
	2	4,518	117	2.59	
1997	1	3,250	7	0.22	125
	2	3,744	118	3.15	
1998	1	3,027	19	0.63	154
	2	3,101	135	4.35	
1999	1	3,064	19	0.62	198
	2	3,041	179	5.89	
2000	1	2,843	12	0.42	195
	2	3,084	183	5.93	
2001	1	2,661	12	0.45	231
	2	3,153	219	6.95	
2002	1	2,449	11	0.45	230
	2	3,216	219	6.81	
2003	1	2,283	16	0.70	193
	2	3,101	177	5.71	
2004	1	2,111	10	0.47	124
	2	2,950	114	3.86	
2005	1	2,573	10	0.39	95
	2	3,650	85	2.33	

Notes: This table demonstrates that banks with satisfactory composites ratings (CAMELS 1 or 2) frequently migrate to unsatisfactory ratings (3, 4, or 5), thereby permitting yearly re-estimation of the CAMELS-downgrade model. The data also show that 2-rated banks are much more likely to migrate to unsatisfactory ratings than 1-rated banks.

Table 5 Selected Summary Statistics—Jumbo CD Data and Regressors for the CAMELS-Downgrade Model

	Variable	Median	Mean	Standard Deviation
Credit Risk	PAST DUE 30	0.68	0.90	0.84
	PAST DUE 90	0.07	0.21	0.39
	NONACCRUING	0.19	0.38	0.56
	OREO	0.03	0.20	0.46
	COMMERICAL	7.65	9.28	6.97
Leverage Risk	RESIDENTIAL	14.26	15.91	10.68
	NET WORTH	8.84	9.79	4.74
Liquidity Risk	ROA	1.17	1.23	2.05
	SECURITIES	28.65	30.41	14.87
Controls	JUMBO CDs	8.00	9.33	6.56
	SIZE	11.07	11.21	1.29
	CAMELS-2	1.00	0.61	0.49
	MANAGEMENT	0.00	0.18	0.39
	“Simple” Default Premium	0.47	0.42	2.83
	“Complex” Default Premium	NA	NA	2.18
	“Simple” Deposit Runoff	9.39	19.25	52.02
	“Complex” Default Premium	NA	NA	33.03

Notes: This table contains summary statistics for the independent variables used in the CAMELS-downgrade prediction model, computed over all year-end regression observations from 1989 to 2001. Summary statistics for the default premiums and deposit runoff series used in jumbo CD risk rankings are also provided for comparison. The “complex” measures of premium and runoff are regression residuals, so means and medians are not meaningful, but standard deviations are roughly in line with their “simple” counterparts. The correlation coefficients between the “simple” and “complex” measures are 88 percent for default premiums and 35 percent for the runoff.

**Table 6 How Well Does the CAMELS-Downgrade Model Fit the Data?
Downgrade Years: 1990–2005**

Independent Variable	Period of Downgrade in CAMELS Rating			
	1990–1991	1991–1992	1992–1993	1993–1994
Credit Risk				
Intercept	-2.087*** (0.246)	-0.957*** (0.264)	-0.081 (0.318)	0.048 (0.375)
PAST DUE 30	0.112** (0.021)	0.150*** (0.022)	0.136*** (0.026)	0.174*** (0.033)
PAST DUE 90	0.376*** (0.039)	0.328*** (0.040)	0.239*** (0.047)	0.304*** (0.060)
NONACCRUING	0.235*** (0.029)	0.199*** (0.030)	0.291*** (0.036)	0.178*** (0.045)
OREO	0.220*** (0.030)	0.216*** (0.032)	0.145*** (0.031)	0.167*** (0.043)
COMMERCIAL	0.009*** (0.003)	0.013*** (0.003)	0.009** (0.004)	0.002 (0.005)
RESIDENTIAL	-0.005*** (0.002)	-0.004 (0.002)	-0.004 (0.003)	-0.005 (0.003)
Leverage Risk				
NET WORTH	-0.054*** (0.010)	-0.048*** (0.011)	-0.073*** (0.013)	-0.074*** (0.013)
ROA	-0.241*** (0.035)	-0.318*** (0.039)	-0.200*** (0.043)	-0.263*** (0.051)
Liquidity Risk				
SECURITIES	-0.016*** (0.002)	-0.017*** (0.002)	-0.013*** (0.002)	-0.009*** (0.003)
JUMBO CDs	0.017*** (0.003)	0.019*** (0.003)	0.015*** (0.004)	0.017*** (0.005)
Controls				
SIZE	0.079 (0.017)	-0.029 (0.019)	-0.125*** (0.024)	-0.147*** (0.030)
CAMELS-2	0.633*** (0.062)	0.517*** (0.068)	0.509*** (0.087)	0.432*** (0.102)
MANAGEMENT	0.488*** (0.051)	0.401*** (0.054)	0.478*** (0.061)	0.466*** (0.069)
Number of Observations	8,494	8,065	7,837	8,060
Pseudo-R ²	0.219	0.226	0.209	0.161

Table 6 (Continued) How Well Does the CAMELS-Downgrade Model Fit the Data?

Independent Variable	Period of Downgrade in CAMELS Rating				
	1994–1995	1995–1996	1996–1997	1997–1998	1998–1999
Intercept	-0.780* (0.402)	-0.011 (0.436)	-0.162 (0.415)	-1.371*** (0.388)	-1.603*** (0.352)
Credit Risk					
PAST DUE 30	0.119*** (0.035)	0.164*** (0.035)	0.093*** (0.029)	0.189*** (0.033)	0.186*** (0.030)
PAST DUE 90	0.296*** (0.064)	0.322*** (0.074)	0.347*** (0.057)	0.399*** (0.064)	0.182*** (0.058)
NONACCRUING	0.192*** (0.046)	0.145*** (0.051)	0.187*** (0.044)	0.157*** (0.046)	0.163*** (0.044)
OREO	0.192*** (0.044)	0.153*** (0.052)	0.156** (0.067)	0.091 (0.059)	0.118 (0.087)
COMMERICAL	0.007 (0.005)	0.013*** (0.005)	0.005 (0.005)	0.010** (0.005)	0.015 (0.004)
RESIDENTIAL	-0.002 (0.004)	-0.013*** (0.004)	0.000 (0.003)	-0.009*** (0.003)	-0.004 (0.003)
Leverage Risk					
NET WORTH	-0.032** (0.014)	-0.034*** (0.013)	-0.020* (0.012)	-0.036*** (0.014)	-0.011 (0.010)
ROA	-0.229*** (0.052)	-0.164*** (0.038)	-0.110** (0.044)	-0.393*** (0.063)	-0.133 (0.040)
Liquidity Risk					
SECURITIES	-0.002 (0.003)	-0.010*** (0.003)	-0.011*** (0.003)	-0.015*** (0.003)	-0.007** (0.003)
JUMBO CDs	0.024*** (0.005)	0.020*** (0.005)	0.019*** (0.004)	0.023*** (0.005)	0.008* (0.004)
Controls					
SIZE	-0.150*** (0.033)	-0.202*** (0.035)	-0.101*** (0.030)	-0.150*** (0.032)	-0.071*** (0.027)
CAMELS-2	0.594*** (0.104)	0.589*** (0.103)	0.760*** (0.093)	0.501*** (0.099)	0.716*** (0.078)
MANAGEMENT	0.389*** (0.075)	0.510*** (0.078)	0.535*** (0.081)	0.406*** (0.083)	0.518*** (0.077)
Number of Observations	8,665	8,682	8,585	8,314	7,818
Pseudo-R ²	0.150	0.188	0.223	0.184	0.166

Table 6 (Continued) How Well Does the CAMELS-Downgrade Model Fit the Data?

Independent Variable	Period of Downgrade in CAMELS Rating			
	1999–2000	2000–2001	2001–2002	2002–2003
Credit Risk				
Intercept	-1.118*** (0.360)	-1.061*** (0.358)	-1.788*** (0.342)	-1.528*** (0.342)
PAST DUE 30	0.169*** (0.029)	0.184*** (0.032)	0.170*** (0.026)	0.171*** (0.027)
PAST DUE 90	0.217*** (0.055)	0.417*** (0.059)	0.321*** (0.059)	0.147** (0.061)
NONACCRUING	0.227*** (0.044)	0.165*** (0.045)	0.250*** (0.042)	0.285*** (0.041)
OREO	0.117 (0.076)	0.157* (0.087)	0.175* (0.090)	0.082 (0.073)
COMMERICAL	0.013*** (0.004)	0.013*** (0.004)	0.010*** (0.004)	0.004 (0.004)
RESIDENTIAL	-0.002 (0.003)	0.002 (0.003)	0.001 (0.003)	-0.005 (0.003)
Leverage Risk				
NET WORTH	-0.044*** (0.011)	-0.036*** (0.010)	-0.008 (0.009)	-0.025** (0.010)
ROA	-0.199*** (0.046)	-0.254*** (0.043)	-0.135*** (0.039)	-0.134*** (0.037)
Liquidity Risk				
SECURITIES	-0.002 (0.003)	-0.005** (0.003)	-0.010*** (0.003)	-0.005* (0.002)
JUMBO CDs	0.015*** (0.004)	0.017*** (0.004)	0.023*** (0.004)	0.015*** (0.004)
Controls				
SIZE	-0.099*** (0.029)	-0.106*** (0.028)	-0.069*** (0.026)	-0.065** (0.026)
CAMELS-2	0.780*** (0.079)	0.799*** (0.081)	0.878*** (0.083)	0.797*** (0.083)
MANAGEMENT	0.564*** (0.080)	0.538*** (0.084)	0.338*** (0.092)	0.491*** (0.088)
Number of Observations	7,341	6,968	6,582	6,367
Pseudo-R ²	0.190	0.206	0.210	0.184

Note: See footnote 24.

**Table 7 Do Jumbo CD Default Premiums or Runoff Add Value in Bank Surveillance?
Full-Sample, Two-Year Horizon**

Out-of-Sample Test Window (1)	CAMELS- Downgrade Model (2)	Simple Default Premiums (3)	Complex Default Premiums (4)	Simple Runoff (5)	Complex Runoff (6)	Simple Premium + Model (7)	Complex Premium + Model (8)	Downgrade Model + Simple Premiums/ Runoff (9)	Downgrade Model + Complex Premiums/ Runoff (10)
1992-1993	20.20	47.56	52.20	49.54	53.08	50.51	52.88	21.12	21.67
1993-1994	21.81	47.10	47.48	47.22	48.62	48.31	49.30	23.22	21.07
1994-1995	22.39	50.57	49.57	45.06	50.90	46.18	49.55	22.67	19.94
1995-1996	17.51	43.50	46.81	47.42	48.49	45.52	46.48	18.64	18.87
1996-1997	15.24	46.93	49.01	46.31	51.03	46.39	48.79	18.38	16.16
1997-1998	19.24	46.34	48.12	45.01	49.33	45.22	48.92	19.72	20.59
1998-1999	21.39	47.27	45.69	47.20	45.75	47.24	45.70	21.87	21.45
1999-2000	19.55	45.62	48.83	43.56	48.94	43.41	48.64	19.33	19.19
2000-2001	18.77	44.31	51.78	46.10	52.03	45.38	52.13	20.11	18.47
2001-2002	18.92	43.82	51.78	44.93	51.85	44.40	51.90	19.64	20.23
2002-2003	19.46	45.35	51.61	46.27	51.56	44.34	51.45	19.56	19.36
2003-2004	20.36	41.56	51.30	46.12	52.54	43.02	52.30	21.18	20.51
2004-2005	20.76	43.29	51.89	44.85	52.00	42.18	52.59	20.77	21.14
All Years	19.66	45.63	49.70	46.12	50.47	45.55	50.05	20.48	19.95

Notes: This table summarizes evidence about the surveillance value of jumbo CD data. Each cell in columns 2 through 10 contains the area under the power curve for a specific risk-ranking produced by a specific surveillance tool over a specific test window. Smaller areas imply lower Type 1 and Type 2 error rates and, thus, better performance. Column 2 contains areas for downgrade probability rankings to benchmark current practices. Columns 3 through 10 contain rankings based on various uses of jumbo CD default premiums and runoff. The evidence suggests the data would have added no value in surveillance between 1992 and 2005. Risk rankings produced by the CAMELS-downgrade model (column 2) performed considerably better than random rankings (average power curve area of 50 percent). But, rankings based on default premiums or runoff (columns 3 through 6) as well as rankings based on both series (columns 7 and 8) barely outperformed random rankings. Finally, default premiums and runoff (columns 9 and 10) did not improve out-of-sample performance of the CAMELS-downgrade model.

Table 8 Do Jumbo CD Default Premiums or Runoff Enrich the CAMELS-Downgrade Model?

Out-of-Sample Window (1)	Panel A: Percentage Change in Power Curve Area					
	Default Premiums and Runoff (2)	Leverage Risk Variables (3)	Credit Risk Variables (4)	Liquidity Risk Variables (5)	Control Variables (6)	
1992-1993	-4.36	8.81	13.92	8.00	4.36	
1993-1994	0.64	-1.21	13.24	6.16	12.91	
1994-1995	-0.44	3.93	10.48	0.75	19.36	
1995-1996	-1.18	4.50	14.36	6.98	27.70	
1996-1997	0.13	-0.39	24.65	5.87	16.90	
1997-1998	-1.78	1.58	9.62	4.89	22.24	
1998-1999	0.19	-0.28	9.93	0.05	16.82	
1999-2000	1.98	3.43	16.44	-1.30	20.40	
2000-2001	0.43	4.89	18.28	1.12	19.08	
2001-2002	-0.94	1.15	18.47	2.73	13.59	
2002-2003	0.15	6.52	16.22	1.28	14.12	
2003-2004	1.15	5.64	22.18	1.90	12.64	
2004-2005	0.63	5.79	17.33	1.36	19.62	
Mean	-0.26	5.20	15.78	3.06	16.90	

Table 8 (Continued) Do Jumbo CD Default Premiums or Runoff Enrich the CAMELS-Downgrade Model?

Out-of-Sample Window	Panel B: Percentage Change in QPS					
	Default Premiums and Runoff (2)	Leverage Risk Variables (3)	Credit Risk Variables (4)	Liquidity Risk Variables (5)	Control Variables (6)	
1992-1993	-1.06	3.65	6.25	4.55	-1.70	
1993-1994	0.43	6.45	9.31	1.43	-0.72	
1994-1995	0.20	3.59	5.39	0.60	1.20	
1995-1996	0.00	0.23	3.85	0.68	2.49	
1996-1997	-0.20	1.43	3.46	0.61	1.02	
1997-1998	-0.17	0.86	2.06	1.03	3.26	
1998-1999	0.00	0.40	2.53	0.40	2.39	
1999-2000	0.00	0.91	2.62	0.57	2.50	
2000-2001	-0.21	0.84	3.45	0.31	2.72	
2001-2002	0.66	1.98	5.84	0.85	2.64	
2002-2003	-0.38	2.67	2.96	0.29	2.29	
2003-2004	0.72	0.96	5.97	-1.55	0.72	
2004-2005	0.82	2.45	2.45	-0.49	1.96	
Mean	0.06	2.03	4.32	0.71	1.60	

Notes: This table provides alternative measures of the contribution of simple default premiums and runoff to the CAMELS-downgrade model. Column 2 of Panel A shows the impact on power curve areas of removing the two jumbo CD series from the enhanced downgrade model (baseline model plus premiums and runoff). Column 2 of Panel B notes the impact of removing these series on quadratic probability score (QPS). Changes in the 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 changes, columns 3 through 6 show the impact of removing other variable blocks from the CAMELS-downgrade model, such as the control variables (asset size, dummy for CAMELS ratings, and dummy for management rating of 2). The evidence suggests default premiums and runoff add nothing to the CAMELS-downgrade model.

REFERENCES

- Altman, Edward I., and Anthony Saunders. 1997. "Credit Risk Measurement: Developments Over the Last Twenty Years." *Journal of Banking and Finance* 21: 1721–42.
- Baer, Herbert, and Elijah Brewer III. 1986. "Uninsured Deposits as a Source of Market Discipline: Some New Evidence." Federal Reserve Bank of Chicago *Economic Perspectives* (September/October): 23–31.
- Barth, Mary E., William H. Beaver, and Wayne R. Landsman. 1996. "Value-Relevance of Banks' Fair Value Disclosures Under SFAS No. 107." *The Accounting Review* 71: 513–37.
- Benston, George J., and George G. Kaufman. 1998. "Deposit Insurance Reform in the FDIC Improvement Act: The Experience to Date." The Federal Reserve Bank of Chicago *Economic Perspectives* (2): 2–20.
- Berger, Allen N. 1995. "The Relationship Between Capital and Earnings in Banking." *Journal of Money, Credit, and Banking* 27: 432–56.
- _____, and Sally M. Davies. 1998. "The Information Content of Bank Examinations." *Journal of Financial Services Research* 14: 117–44.
- Berger, Allen N., Seth D. Bonime, Daniel M. Covitz, and Diana Hancock. 2000. "Why are Bank Profits So Persistent? The Roles of Product Market Competition, Informational Opacity, and Regional/Macroeconomic Shocks." *Journal of Banking and Finance* 24: 1203–35.
- Billett, Matthew T., Jon A. Garfinkel, and Edward S. O'Neal. 1998. "The Cost of Market Versus Regulatory Discipline in Banking." *Journal of Financial Economics* 48: 333–58.
- Birchler, Urs W., and Andrea M. Maechler. 2002. "Do Depositors Discipline Swiss Banks?" In *Research in Financial Services: Private and Public Policy* 14, ed. George G. Kaufman. Oxford, U.K.: Elsevier Science.
- Bliss, Robert R., and Mark J. Flannery. 2001. "Market Discipline in the Governance of U.S. Bank Holding Companies: Monitoring Versus Influence." In *Prudential Supervision: What Works and What Doesn't*, ed. Frederic Mishkin. Chicago: University of Chicago Press.
- Board of Governors of the Federal Reserve System, and the United States Department of the Treasury. 2000. *The Feasibility and Desirability of Mandatory Subordinated Debt* (December).

- Board of Governors of the Federal Reserve System. 1999. "Using Subordinated Debt as an Instrument of Market Discipline." Staff Study 172 (December).
- Brewer, Elijah III, and Thomas H. Mondschean. 1994. "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: 146–64.
- Calomiris, Charles W. 1999. "Building an Incentive-Compatible Safety Net." *Journal of Banking and Finance* 23: 1499–1519.
- _____, and Joseph R. Mason. 1997. "Contagion and Bank Failures during the Great Depression: The 1932 Chicago Bank Panic." *American Economic Review* 87: 863–83.
- Cargill, Thomas F. 1989. "CAMEL Ratings and the CD Market." *Journal of Financial Services Research* 3: 347–58.
- Cole, Rebel A., and Jeffrey W. Gunther. 1998. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems." *Journal of Financial Services Research* 13: 103–17.
- Cole, Rebel A., Barbara G. Cornyn, and Jeffrey W. Gunther. 1995. "FIMS: A New Monitoring System for Banking Institutions." *Federal Reserve Bulletin* 81: 1–15.
- Cook, Douglas O., and Lewis J. Spellman. 1994. "Repudiation Risk and Restitution Costs: Toward an Understanding of Premiums on Insured Deposits." *Journal of Money, Credit, and Banking* 26 (August): 439–59.
- Covitz, Daniel M., Diana Hancock, and Myron L. Kwast. 2002. "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. 1994. "The Effect of a Rating Downgrade on Outstanding Commercial Paper." *Journal of Finance* 49: 39–56.
- Crane, Dwight B. 1976. "A Study of Interest Rate Spreads in the 1974 CD Market." *Journal of Bank Research* 7: 213–24.
- Curry, Timothy J., Peter J. Elmer, and Gary S. Fissel. 2003. "Using Market Information to Help Identify Distressed Institutions: A Regulatory Perspective." *Federal Deposit Insurance Corporation Banking Review* 15 (3): 1–16.
- Davison, Lee. 1997. "Continental Illinois and 'Too Big to Fail.'" In *History of the Eighties: Lessons for the Future*, vol. 1. Federal Deposit Insurance Corporation: Washington DC.

- Demsetz, Rebecca S., and Philip E Strahan. 1997. "Diversification, Size, and Risk at Bank Holding Companies." *Journal of Money, Credit, and Banking* 29: 300–13.
- DeYoung, Robert. 1999. "Birth, Growth, and Life or Death of Newly Chartered Banks." Federal Reserve Bank of Chicago *Economic Perspectives* (3): 18–35.
- Ellis, David M., and Mark J. Flannery. 1992. "Does the Debt Market Assess Large Banks' Risk? Time Series Evidence from Money Center CDs." *Journal of Monetary Economics* 30: 481–502.
- Estrella, Arturo. 1998. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." *Journal of Business Economics and Statistics* 16: 198–205.
- _____, and Frederic S. Mishkin. 1998. "Predicting U.S. Recessions: Financial Variables as Leading Indicators." *Review of Economics and Statistics* 80: 45–61.
- Evanoff, Douglas D., and Larry D. Wall. 2001. "Sub-debt Yield Spreads as Bank Risk Measures." *Journal of Financial Services Research* 20: 121–45.
- Feldman, Ron, and Jason Schmidt. 2001. "Increased Use of Uninsured Deposits." Federal Reserve Bank of Minneapolis *Fedgazette* (March): 18–19.
- Flannery, Mark J. 1982. "Retail Bank Deposits as Quasi-Fixed Factors of Production." *American Economic Review* 72: 527–36.
- _____. 2001. "The Faces of 'Market Discipline.'" *Journal of Financial Services Research* 20: 107–19.
- _____, and Joel F. Houston. 1999. "The Value of a Government Monitor for U.S. Banking Firms." *Journal of Money, Credit, and Banking* 31: 14–34.
- Flannery, Mark J., and Kasturi Rangan. 2002. "Market Forces at Work in the Banking Industry: Evidence from the Capital Buildup of the 1990s." University of Florida Working Paper (September).
- Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press.
- Gilbert, R. Alton, and Mark D. Vaughan. 2001. "Do Depositors Care About Enforcement Actions?" *Journal of Economics and Business* 53: 283–311.
- Gilbert, R. Alton, Andrew P. Meyer, and Mark D. Vaughan. 1999. "The Role of Supervisory Screens and Econometric Models in Off-Site

Surveillance.” Federal Reserve Bank of St. Louis *Review* (November/December): 31–56.

_____. 2002. “Could a CAMELS-Downgrade Model Improve Off-Site Surveillance?” Federal Reserve Bank of St. Louis *Review* (January/February): 47–63.

Goldberg, Lawrence G., and Sylvia C. Hudgins. 2002. “Depositor Discipline and Changing Strategies for Regulating Thrift Institutions.” *Journal of Financial Economics* 63: 263–74.

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

Gorton, Gary. 1996. “Reputation Formation in Early Bank Note Markets.” *Journal of Political Economy* 104: 346–97.

Gunther, Jeffery W., and Robert R. Moore. 2000. “Financial Statements and Reality: Do Troubled Banks Tell All?” Federal Reserve Bank of Dallas *Economic and Financial Review* (3): 30–5.

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

Hall, John R., Thomas B. King, Andrew P. Meyer, and Mark D. Vaughan. 2005. “Did FDICIA Improve Market Discipline? A Look at Evidence from the Jumbo CD Market.” Federal Reserve Bank of St. Louis Supervisory Policy Analysis Working Paper (April).

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

James, Christopher. 1988. “The Use of Loan Sales and Standby Letters of Credit by Commercial Banks.” *Journal of Monetary Economics* 22: 395–422.

_____. 1990. “Heterogeneous Creditors and LDC Lending.” *Journal of Monetary Economics* 25: 325–46.

Jones, David S., and Kathleen K. King. 1995. “The Implementation of Prompt Corrective Action.” *Journal of Banking and Finance* 19: 491–510.

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

- Kahn, Charles, George Pennacchi, and Ben Sopranzetti. 1999. "Bank Deposit Rate Clustering: Theory and Empirical Evidence." *Journal of Finance* 56: 2185–214.
- Keeley, Michael C. 1990. "Deposit Insurance, Risk, and Market Power in Banking." *American Economic Review* 80: 1183–98.
- Krainer, John, and Jose A. Lopez. Forthcoming. "Incorporating Equity Market Information into Supervisory Monitoring Models." *Journal of Money, Credit, and Banking*.
- Kroszner, Randall S., and Philip E. Strahan. 2001. "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*, ed. Frederic Mishkin. Chicago: University of Chicago Press.
- Lang, William W., and Douglas Robertson. 2002. "Analysis of Proposals for a Minimum Subordinated Debt Requirement." *Journal of Economics and Business* 54: 115–36.
- Lynch, Peter, and John Rothchild. 2000. *One Up on Wall Street: How to Use What You Already Know to Make Money in the Market*. New York: Simon & Schuster.
- Maechler, Andrea M., and Kathleen M. McDill. 2003. "Dynamic Depositor Discipline in U.S. Banks." Federal Deposit Insurance Corporation Working Paper, 2003.
- Malkiel, Burton G. 2003. "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17 (1): 59–82.
- Marino, James A., and Rosalind Bennett. 1999. "The Consequences of National Depositor Preference." Federal Deposit Insurance Corporation *Banking Review* 12: 19–38.
- Martinez-Peria, Maria Soledad, and Sergio L. Schmukler. 2001. "Do Depositors Punish Banks for Bad Behavior? Market Discipline, Deposit Insurance, and Banking Crises." *Journal of Finance* 56: 1029–51.
- Meyer, Laurence H. 2001. "Supervising Large Complex Banking Organizations: Adapting to Change." In *Prudential Supervision: What Works and What Doesn't*, ed. Frederic Mishkin. Chicago: University of Chicago Press.
- Morgan, Donald P., and Kevin J. Stiroh. 2001. "Market Discipline of Banks: The Asset Test." *Journal of Financial Services Research* 20: 195–208.
- Park, Sangkyun. 1995. "Market Discipline by Depositors: Evidence from Reduced-Form Equations." *Quarterly Review of Economics and Finance* 35: 497–514.

- _____, and Stavros Peristiani. 1998. "Market Discipline by Thrift Depositors." *Journal of Money, Credit, and Banking* 30: 347–64.
- Peek, Joe, and Eric S. Rosengren. 1997. "Will Legislated Early Intervention Prevent the Next Banking Crisis?" *Southern Economic Journal* 64: 268–280.
- Putnam, Barron H. 1983. "Early Warning Systems and Financial Analysis in Bank Monitoring: Concepts of Financial Monitoring." Federal Reserve Bank of Atlanta *Economic Review* (November): 6–13.
- Rajan, Raghuram G. 2001. "Comment on Bliss and Flannery." In *Prudential Supervision: What Works and What Doesn't*, ed. Frederic Mishkin. Chicago: University of Chicago Press.
- Reidhill, Jack, and John O'Keefe. 1997. "Off-Site Surveillance Systems." In *History of the Eighties: Lessons for the Future*, vol. 1. Federal Deposit Insurance Corporation: Washington D.C.
- Roll, Richard. 1994. "What Every CFO Should Know about Scientific Progress in Financial Economics: What is Known, and What Remains to be Resolved." *Financial Management* 23 (2): 69–75.
- Sierra, Gregory E., and Timothy J. Yeager. 2004. "What Does the Federal Reserve's Economic Value Model Tell Us about Interest Rate Risk at U.S. Community Banks?" Federal Reserve Bank of St. Louis *Review* (November/December): 45–60.