Supervisory Information and the Frequency of Bank Examinations

Beverly J. Hirtle and Jose A. Lopez

Bank supervisors need timely and reliable information about the financial condition and risk profile of banks in order to conduct effective supervision. Although such information can be obtained in part from regulatory reports and public disclosures, a key source is the on-site bank examination. Bank examinations enable supervisors to confirm the accuracy of information in regulatory reports. More important, perhaps, the examinations allow supervisors to gather confidential information about banks' financial conditions and to assess qualitative attributes, such as internal controls and risk management procedures, that affect bank risk profiles.

Such confidential information may be valuable to supervisors, yet it is costly to obtain: bank examinations absorb considerable resources on the part of supervisors as well as banks. Thus, supervisors face a trade-off between the timeliness of the information gathered from bank examinations and the costs of obtaining it. In particular, the longer the time since a bank's most recent examination, the higher the likelihood that conditions at the bank will have changed in a way that diminishes the current value of that information. This concern must be balanced against the costs of conducting more frequent examinations.

The potential "time decay" of bank examination information has been a concern for both supervisors and policymakers. Supervisors have developed a number of approaches for allocating scarce examination resources, including off-site monitoring systems to help identify banks whose financial conditions may have deteriorated. Concern about the timeliness of examination information also motivated provisions in the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA), which mandates annual on-site examinations for most commercial banks. In this case, legislators felt that frequent examinations would be useful in limiting losses to the deposit insurance system.

Beverly J. Hirtle is a vice president at the Federal Reserve Bank of New York. Jose A. Lopez, formerly an economist at the Federal Reserve Bank of New York, is now an economist at the Federal Reserve Bank of San Francisco.

In this article, we provide insight into the policy aspects of this informational time decay by assessing how the length of time between bank examinations affects the quality of information available to supervisors. For these purposes, we define the quality of information in terms of how accurately the information from a prior examination reflects the current condition of a bank. Our analysis assumes that examination information incorporates two types of data: information available from public sources and updated regulatory reports, and private supervisory information gathered from on-site examinations. That is, examination findings contain information that is readily available from public financial and regulatory reports as well as private information that can be obtained only through on-site examinations, such as confidential information about a bank's troubled assets or the examiner's assessment of a bank's internal controls. Changes over time in the quality of the examination information will be affected by changes in both of these components.

Because the information in regulatory reports is readily available to supervisors, an on-site examination is not required to update this component.[1] However, supervisory information can be acquired only through an examination. Thus, the rate of decay in the accuracy of this private supervisory information should be the key determinant in the timing of bank examinations. The faster this information decays over time, the more frequent these examinations must be to ensure that the quality of information does not drop below some critical level.

Our analysis suggests that the private supervisory component of examination information ceases to provide useful information about the current condition of a bank after six to twelve quarters, or one and a half to three years. This rate of information decay seems to be cyclical, in that the quality of this private supervisory information appears to decay faster during years in which the U.S. banking industry experiences financial difficulties. Consistent with this finding, our analysis further suggests that the decay rate depends on the initial financial condition of the bank: the rate of decay in the quality of private supervisory information appears to be significantly greater for troubled banks than for healthy ones. This latter result is consistent with the idea that conditions change more rapidly at institutions experiencing financial difficulty and that more frequent examinations of these institutions may be warranted.

Our results provide insight into how often a bank should be examined. The range of six to twelve quarters indicated by our results is really an upper bound, since it reflects the point at which the supervisory information from the previous examination contains *no* useful information about the current condition of the bank.[2] As a matter of practice, it is probably desirable to examine banks somewhat more frequently—that is, when the supervisory information from the previous examination still contains *some* useful information about the bank's current condition. Our results also suggest that more frequent examinations may be warranted during periods of banking industry stress and for banks that are financially troubled. Taken together, these results imply that the annual examination frequency mandated in FDICIA is reasonable, particularly during times of financial difficulties for the banking industry.

The Timing and Frequency of Bank Examinations

The timing and frequency of bank examinations have received increased public scrutiny in recent years, especially in light of the significant loan losses and number of bank failures suffered by the U.S. banking industry during the 1980s and early 1990s.[3] Debate has centered around the idea that supervisors, banks, and the tax-paying public face a trade-off between the costs and benefits of more frequent examinations. On the one hand, more frequent examinations would generate more timely information about the current condition of banks and could allow supervisors to address emerging problems more quickly. This quicker response could reduce the exposure of the deposit insurance system—and ultimately the taxpayer—to loss. On the other hand, examinations are resource-intensive for both banks and supervisors, and maintaining large supervisory and examination staffs can be costly.[4]

The balance of this trade-off has shifted over the years in response to conditions in the banking industry. For instance, the Federal Deposit Insurance Corporation recently reported on the efforts of the federal bank supervisory agencies to extend the examination cycle as a means of reducing the size of their examination staffs, especially during the early-to-mid-1980s (Federal Deposit Insurance Corporation 1997). According to the report, these agencies focused their resources on the institutions most likely to present systemic risk concerns. In many cases, targeted or limited-scope examinations—that is, examinations that assess only selected bank activities or that involve less detailed evaluations of a bank's overall activities, respectively—were used in place of more resource-intensive, full-scope examinations. As a result, the frequency of full-scope examinations fell considerably during this period, especially for smaller banks and banks believed to be in sound financial condition. Taken together, such measures in particular allowed the FDIC and the Office of the Comptroller of the Currency to reduce their examination staffs.

The FDIC report concluded that such cutbacks in examination staffs "reduced the ability of supervisors to detect problems early enough to take corrective actions."[5] As a result, state and federal banking supervisors increased their examination staffs and conducted bank examinations more frequently, on average, as problems in the banking industry increased through the latter part of the 1980s. As these problems intensified, the issue of the frequency and scope of bank examinations increasingly became part of the public policy debate. This process culminated in the passage of FDICIA, which mandates full-scope, on-site examinations each year for U.S. commercial banks.[6]

In general, bank examinations are scheduled at least several months in advance, both to give banks time to prepare and to allow supervisors to develop an overall schedule and individual examination plans that make efficient use of available resources. Given this advanced scheduling, changing the timing of one examination typically also entails rescheduling others to free up the needed resources. Thus, several factors work to reinforce the timing implicit in the original examination schedule. However, even after the passage of FDICIA, supervisors continue to have some discretion over the timing of examinations for individual banks. To some extent, the size and perceived condition of a bank can influence the planned time between full-scope examinations, but there is now less scope for supervisors to lengthen this time period without violating FDICIA's annual examination requirement. Supervisors can and do accelerate full-scope examinations and undertake targeted examinations if other factors indicate that problems are developing at a particular bank. In fact, supervisors employ fairly extensive off-site monitoring procedures—including the use of statistical models—to help identify those banks where problems might be emerging.[7]

THE CAMEL RATING SYSTEM

A key outcome of an examination is a supervisory rating of the bank's overall financial condition. For the purposes of this article, we focus on these supervisory ratings—known as CAMEL ratings—as a proxy for the information resulting from a bank examination. CAMEL ratings, which are assigned by examiners at the conclusion of an examination, are numerical ratings of the quality of a bank's financial condition, risk profile, and overall performance. This rating system is used by the three federal banking supervisors—the OCC, the FDIC, and the Federal Reserve—as well as by the Office of Thrift Supervision, the National Credit Union Administration, and state banking supervisors to provide a convenient summary rating of each bank's condition at the time of the examination. In addition, CAMEL ratings are increasingly being used for other supervisory purposes, such as setting deposit insurance rates and expediting bank applications for various regulatory purposes.

CAMEL ratings, which are assigned by examiners at the conclusion of an examination, are numerical ratings of the quality of a bank's financial condition, risk profile, and overall performance.

The acronym CAMEL refers to the five components of a bank's condition assessed by examiners: *C*apital adequacy, *A*sset quality, *M*anagement, *E*arnings, and *L*iquidity.[8] Ratings are assigned for each component, and a composite rating is assigned for the overall condition and performance of the bank. These component and composite ratings are assigned on a scale of 1 to 5, with 1 representing the highest rating (strongest performance) and 5 representing the lowest (weakest performance). Banks with composite CAMEL ratings of 1 or 2 are generally considered to present few, if any, significant supervisory concerns, while those with ratings of 3, 4, or 5 are considered to present moderate to extreme degrees of supervisory concern.[9]

The CAMEL rating is only one aspect of the information produced during a bank examination. Examiners also put together a detailed report that describes the bank's activities and management structure; assesses the bank's performance along the dimensions reflected in the CAMEL rating as well as in other areas; and, where appropriate, contains recommendations for changes and improvements in the bank's policies and procedures. This report is backed by extensive examination work papers detailing the process leading to the examiners' conclusions. The CAMEL rating, while not a comprehensive indicator of all this information, nonetheless provides a convenient summary measure of the examination findings.

A bank's CAMEL rating is known only by the bank's senior management and appropriate supervisory staff at the relevant supervisory agencies. CAMEL ratings are never made publicly available, even on a lagged basis.

All of this examination material, including the CAMEL rating, is highly confidential. A bank's CAMEL rating is known only by the bank's senior management and appropriate supervisory staff at the relevant supervisory agencies. CAMEL ratings are never made publicly available, even on a lagged basis. Thus, to a considerable degree, the CAMEL rating reflects the private supervisory information gathered during a bank examination as well as whatever public and regulatory information is available about the bank's condition. For this reason, we use the ratings as our indicator of the private supervisory information arising from bank examinations.

Previous Research on the Information in CAMEL Ratings

Other researchers have examined the role of CAMEL ratings in providing information about the financial condition of banks. For instance, Berger and Davies (1994) examine the information content of CAMEL ratings by testing for stock price reactions when new ratings are assigned. Despite the fact that CAMEL ratings are confidential, the authors find that rating downgrades seem to lead to negative excess stock returns. They interpret this result as evidence that examinations generate valuable private information and that rating downgrades reveal unfavorable private information about bank conditions. Similarly, DeYoung, Flannery, Lang, and Sorescu (1998) find that CAMEL ratings contain information useful to the market for subordinated, bank holding company debt.

Berger, Davies, and Flannery (1998) find that BOPEC ratings—the supervisory ratings given to bank holding companies—contain information about bank conditions that goes beyond the information in market data, such as bond-rating downgrades.[10] In particular, they find that supervisory data and market information Granger-cause (or are useful in forecasting) one another, suggesting that both supervisors and the financial markets have some unique information. Finally, Barker and Holdsworth (1993) find evidence that CAMEL ratings are significant predictors of bank failure, even after controlling for a wide range of publicly available information about the condition and performance of banks. Taken together, these papers suggest that supervisory ratings contain information about the condition and performance of banks that is not available to the public.

These papers suggest that newly assigned CAMEL ratings contain relevant information. Some researchers have also studied how that relevance changes over time. For example, Gilbert (1993) addresses the issue to some extent by finding that more frequent examinations reduced losses to the Bank Insurance Fund, which covers government losses when a bank fails. Cole and Gunther (1995, 1998) find that the information contained in CAMEL ratings decays quickly with respect to predicting bank failure from 1986 to 1992. In particular, they find that a model using publicly available financial data is a better indicator of the likelihood of bank failure than the previous CAMEL rating is once the rating is more than one or two quarters old. These two studies address the issue of information decay directly; however, the primary purpose of CAMEL ratings is not to identify future bank failures, but to provide an assessment of banks' overall conditions at the time of the examinations.

Focusing on this aspect of supervisory ratings, Berger, Davies, and Flannery (1998) examine BOPEC ratings in relation to market-based data and find that only very recent examinations provide useful information. The information appears to become much less useful, or "stale," over time. In our analysis, we focus directly on the time decay of the supervisory information in CAMEL ratings and the decay's impact on assessing the current condition of a bank. Thus, we view our article as complementary to, and an extension of, this general line of research.

Structure of the Data Set

The basic data set used in our analysis consists of the composite CAMEL ratings assigned to banks from 1989 to 1995 by the Federal Reserve, the FDIC, the OCC, and state banking supervisors. Each CAMEL rating was given following a full-scope, on-site examination. We eliminated from our sample any ratings associated with targeted or limited-scope examinations. CAMEL ratings are not always assigned during such examinations and, if they are, may not reflect the most up-to-date information about the full scope of a bank's activities.

For each CAMEL rating in the sample, we know the as-of date of the examination (the date as of which the condition of the bank is assessed), the supervisory entity that conducted the examination, and the identity of the bank. We matched each observation to the corresponding bank's income and balance sheet data, as reported in the *Report of Condition and Income* (the *Call Report*) maintained by the bank supervisory agencies, for the quarter prior to the as-of date of the examination. These *Call Report* data will serve as our proxy for the information available from regulatory reports and other public sources about the bank's condition at the time of the examination.

To assess how quickly the private supervisory information from a bank examination decays, we linked each observation to the CAMEL rating from the previous full-scope examination for that bank. That is, for each CAMEL rating in the sample, we know the lagged composite CAMEL rating as well as the date of the previous full-scope examination. With this information, we can calculate the time since the last examination, a key variable in our analysis.[11]

An overview of this element of the data set appears in Table 1. The table presents the number of full-scope bank examinations in our sample from 1989 to 1995, sorted by the time since the last examination. The number of examinations per year varies significantly. From about 7,000 examinations in 1989, the sample size drops to just under 4,000 in 1991 before rising again to over 8,000 starting in 1992. This variation is due to several factors. During 1990 and 1991, the number of full-scope examinations performed by the FDIC dropped significantly, while the number of limited-scope examinations rose. Given that our sample is based on full-scope examinations, this shift resulted in a sharp decline in the size of the data set.

Table 1

Number of Full-Scope Examinations in a Year, Sorted by Quarters since Last Full-Scope Examination

Quarters since Last Examination	1989	1990	1991	1992	1993	1994	1995
1	102	57	20	102	87	79	54
2	911	427	198	625	722	494	321
3	1,347	820	324	805	1,131	952	718
4	1,581	1,057	580	1,037	2,780	3,273	2,704
5	1,191	795	557	601	1,523	1,784	1,932
6	717	463	598	564	623	561	1,034
7	380	251	389	324	342	369	417
8	298	184	324	389	270	397	317
9	188	99	209	371	126	269	244
10	97	41	154	369	114	132	97
11	76	34	129	354	82	57	51
12	58	36	121	364	72	32	23
13-14	38	28	212	750	161	53	18
15-16	8	7	83	640	208	52	8
17 or more	5	7	82	1,029	757	333	74
Total	6,997	4,306	3,980	8,324	8,998	8,837	8,012

Source: Authors' calculations, based on data from the Board of Governors of the Federal Reserve System.

However, following the passage of FDICIA in 1991 with its requirement for annual full-scope examinations, the number of examinations in the data set rose significantly.

Looking down the columns for each year, we see that about three-quarters of the examinations took place within six quarters of the prior examination. There is a clear pattern of clustering of lagged examinations at three to five quarters, particularly in the latter part of the sample period. This clustering is consistent with the supervisory goal of ensuring that each bank has an annual full-scope examination. Finally, there is significant variation across the years in the share of the sample for which the time since the last examination was more than twelve quarters. The early years of the sample contain relatively few such observations, but their numbers increase sharply during 1992 and 1993 before declining significantly in later years. This sharp increase most likely reflects the impact of FDICIA, as the various supervisory agencies made efforts to examine more banks in response to the requirement for annual full-scope examinations.[12]

Empirical Approach: The Off-Site and Examination Models of CAMEL Ratings

To explore the question of how quickly private supervisory information generated during an examination decays, we develop two empirical models to predict banks' composite CAMEL ratings. The first is based on the FIMS model used by the Federal Reserve for off-site monitoring purposes (see Cole, Cornyn, and Gunther [1995] for details). Like the FIMS model, ours uses income and balance sheet data from banks' *Call Reports* to predict composite CAMEL ratings.[13] The specific variables included in the model are listed in Box A. These variables are intended to capture the five CAMEL rating components as well as other influences—such as regional factors and the time since the last full-scope exam—that could help determine the CAMEL rating. Because the variables used in the model do not incorporate the information gathered by supervisors through on-site exams, we call this model the "off-site model."

We estimated this model for each year in the sample period.[14] The overall fit is quite good with the R^2 goodness-of-fit statistics ranging from 0.50 to 0.70.[15] Although the specific variables that enter the model with statistically significant coefficients differ from year to year, a core set of variables have consistent signs and are significant in nearly every year. These variables include the log of total assets, the equity-to-capital ratio, the current and lagged ratios of net income to total assets, the ratio of residential mortgages to total loans, and the ratio of consumer loans to total loans. The coefficients on these variables suggest that, all else equal, larger banks, banks with higher capital and net income ratios, and banks with higher proportions of comparatively less risky residential mortgages and consumer loans tend to receive

Box A: Explanatory Variables Used in the Empirical Models

Capital Adequacy
- equity-to-capital ratio
- four-quarter change in equity-to-capital ratio

Asset Quality
- log of total assets
- four-quarter change in log of total assets
- loan-to-asset ratio
- commercial and industrial loans as share of total loans
- one-to-four-family mortgages as share of total loans
- real estate loans as share of total loans
- consumer loans as share of total loans
- loans past due thirty to eighty-nine days as share of total assets
- loans past due ninety or more days as share of total assets
- nonperforming loans as share of loan loss reserves
- loan loss reserves as share of total loans
- net charge-offs in year before examination as share of total assets
- year-over-year change in net charge-offs as share of total assets
- provisions in year before examination as share of total assets
- year-over-year change in provisions as share of total assets

Management
- interest rate risk exposure (assets minus liabilities that mature or reprice in more than five years)
- insider loans as share of total assets

Earnings
- ratio of net income to total assets in year before examination
- net-income-to-assets ratio lagged one year

Liquidity
- cash as share of total assets

Other Variables
- dummy variables for quarter in which examination took place (Q1, Q2, Q3, Q4)
- dummy variables for bank's Federal Reserve District
- dummy variables for agency conducting examination (Federal Reserve Bank, Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, or state regulator)
- dummy variables for number of quarters since previous examination

better CAMEL ratings. In contrast, banks with higher loan-to-asset ratios, higher amounts of past due and nonaccrual loans, higher ratios of nonperforming loans to loan loss reserves, and higher interest rate risk exposures consistently receive worse CAMEL ratings.

In addition to estimating the off-site model, we estimated a second model that includes the previous composite CAMEL rating for each bank. Because this model includes the private supervisory information contained in these lagged CAMEL ratings, we call it the "examination model." The model already contains variables that control for information from updated regulatory reports, so any additional explanatory power from the lagged CAMEL rating is assumed to reflect private supervisory information.[16] By comparing the ability of the two models to explain current CAMEL ratings as the age of the lagged CAMEL rating increases, we can assess how long this supervisory information provides additional useful information on the current condition of the bank.

To conduct this comparison, we allow the coefficients on the lagged CAMEL rating to differ according to the length of time since the previous examination. In particular, we divide the observations in each year of the sample into fifteen distinct categories according to the time since the previous examination. We then let the lagged CAMEL rating enter the model with a different coefficient for each category.[17] In this way, we can test how the explanatory power of lagged CAMEL ratings varies as the ratings age.[18] This approach provides a *within-sample diagnostic*, meaning that we can test the explanatory power of the lagged CAMEL ratings on the same sample of examinations used to estimate the model.

Before presenting our empirical results, it is worth discussing the role of the fifteen dummy variables reflecting the time since the previous examination. They are included to capture the effects of any independent factors that might cause a relationship between the value of the current CAMEL rating and the time since the last examination. In that way, we can be assured that the coefficients on the interacted, lagged CAMEL ratings are capturing just the influence of the private information from the previous examination rather than these other factors. In fact, the hypothesis that these time-related factors are not meaningful (that is, the coefficients on the dummy variables are jointly equal to zero) is strongly rejected for each year in the sample, indicating that there is some independent influence of the time since the last examination on the value of the current CAMEL rating.[19]

The results of the within-sample diagnostic are presented in Table 2, which contains the coefficient estimates on the interacted, lagged CAMEL ratings in the examination model for each year in the sample. The end

Table 2
COEFFICIENTS ON INTERACTED, LAGGED CAMEL RATINGS IN THE EXAMINATION MODEL

Quarters since Last Examination	1989	1990	1991	1992	1993	1994	1995
1	2.158*	2.663*	0.803	1.893*	2.680*	3.209*	3.237*
	(0.236)	(0.437)	(0.581)	(0.259)	(0.306)	(0.367)	(0.399)
2	2.309*	2.343*	1.631*	1.347*	2.556*	2.607*	3.058*
	(0.093)	(0.133)	(0.179)	(0.099)	(0.105)	(0.134)	(0.169)
3	2.221*	2.500*	1.811*	1.728*	2.531*	2.785*	2.736*
	(0.078)	(0.109)	(0.161)	(0.094)	(0.092)	(0.105)	(0.131)
4	2.241*	2.345*	1.931*	1.728*	2.370*	2.572*	2.842*
	(0.078)	(0.105)	(0.132)	(0.091)	(0.072)	(0.077)	(0.091)
5	2.306*	2.580*	2.203*	1.624*	2.274*	2.578*	2.569*
	(0.089)	(0.130)	(0.143)	(0.117)	(0.086)	(0.093)	(0.096)
6	2.209*	2.767*	2.164*	1.596*	2.525*	2.650*	2.747*
	(0.117)	(0.158)	(0.150)	(0.126)	(0.152)	(0.167)	(0.151)
7	2.163*	2.113*	1.772*	1.324*	2.154*	2.292*	2.207*
	(0.161)	(0.200)	(0.148)	(0.142)	(0.219)	(0.212)	(0.224)
8	1.557*	2.0448	1.872*	1.524*	2.386*	2.221*	2.513*
	(0.198)	(0.237)	(0.160)	(0.140)	(0.285)	(0.223)	(0.303)
9	1.661*	2.138*	1.429*	1.245*	1.988*	2.497*	2.914*
	(0.251)	(0.344)	(0.212)	(0.128)	(0.373)	(0.277)	(0.399)
10	1.786*	1.332*	1.892*	1.300*	2.498*	1.958*	1.615*
	(0.340)	(0.587)	(0.222)	(0.134)	(0.417)	(0.378)	(0.477)
11	1.579*	1.623*	1.202*	1.187*	1.731*	1.628*	1.992*
	(0.373)	(0.497)	(0.261)	(0.137)	(0.405)	(0.485)	(0.734)
12	1.924*	1.805	1.752*	1.205*	2.157*	2.556*	1.229
	(0.606)	(1.265)	(0.296)	(0.123)	(0.568)	(1.087)	(0.696)
13-14	0.253	1.014	1.736*	1.158*	1.990*	0.988	0.779
	(0.527)	(0.700)	(0.258)	(0.100)	(0.280)	(0.638)	(0.553)
15-16	0.135	-0.553	1.486*	1.025*	1.816*	2.035*	0.697
	(0.828)	(2.874)	(0.333)	(0.120)	(0.204)	(0.523)	(0.661)
17 or more	2.075*	-0.072	1.250*	0.742*	0.597*	0.945*	0.669
	(0.772)	(1.284)	(0.467)	(0.107)	(0.131)	(0.189)	(0.414)
Memo:							
R^2	0.824	0.811	0.768	0.741	0.737	0.694	0.692

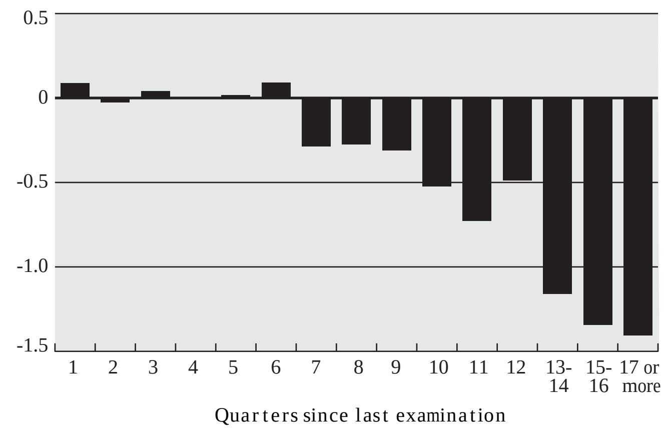
Source: Authors' calculations.

Notes: The coefficients are for the independent variables produced by interacting the lagged CAMEL ratings with dummy variables reflecting the amount of time since the last examination. Standard errors are presented in parentheses. An asterisk indicates that the coefficient is significantly different from zero at the 5 percent level.

of the shading indicates the point at which the lagged CAMEL rating generally no longer enters the model with statistical significance and thus ceases to provide useful information in modeling current CAMEL ratings. Clearly, this result varies across the sample. For 1989, 1990, and 1995, the lagged CAMEL rating is not significant beyond eleven to twelve quarters.[20] However, for the other years, CAMEL ratings older than three years provide some information regarding the current condition of the bank.

Although these results indicate that relatively old CAMEL ratings have explanatory power, further analysis shows that the value of the private supervisory information contained in the ratings decays as it ages. This evidence arises from the size of the coefficients on the lagged CAMEL ratings in addition to their statistical significance. Overall, the hypothesis that the coefficients on the lagged CAMEL ratings are stable across the age categories is strongly rejected. Furthermore, the size of the coefficients declines as the age of the lagged CAMEL rating increases, even while remaining significant. In the context of our model, smaller coefficients imply that changes in the value of the lagged CAMEL ratings have less of an impact on the value of current CAMEL ratings, even though they continue to provide some explanatory power. As shown in the chart, although the decline in the size of the coefficients is not monotonic, there is a general pattern consistent with the idea that the relationship between lagged and current CAMEL ratings decays as the age of the lagged CAMEL rating increases.[21] As the chart illustrates, there is a sharp drop-off in the size of the coefficients once the lagged CAMEL rating is more than six quarters old, suggesting that lagged CAMEL ratings have their greatest impact before they reach this age.

Average Coefficients on Lagged CAMEL Ratings, Sorted by Time since Last Examination, 1989-95
Relative to Coefficient for Four-Quarter-Old Lagged CAMEL Ratings



Source: Authors’ calculations.

Out-of-Sample Analysis

The results discussed thus far all represent a within-sample analysis of the information content of lagged CAMEL ratings, where the significance tests are carried out on the same set of data used to estimate the models. To enhance our understanding of how the value of private supervisory information changes over time, we also conduct several *out-of-sample tests*; that is, tests of the predictive power of the lagged CAMEL ratings using data other than those used to estimate the models. Out-of-sample tests are of interest for two related reasons.

First, the tests provide a more robust assessment of a model’s ability to explain current CAMEL ratings. By using data outside of the estimation sample, we can assess whether the estimated model is stable over time and across different sets of observations. In our analysis, this distinction amounts to asking whether the decay rate of private supervisory information indicated by the examination model reflects the particular observations in a given year or whether the relationship is more general.

Second, out-of-sample tests more closely mirror the situation facing bank supervisors. Supervisors have information about recent bank examinations and therefore can analyze the relationship between lagged and current CAMEL ratings for those banks. Based partly on this analysis, supervisors need to infer how quickly the private supervisory information from other banks is likely to deteriorate and therefore how quickly these banks need to be examined. This situation is essentially an out-of-sample forecasting problem.

To conduct this out-of-sample analysis, we estimate our two models using data from one year and then use the estimated coefficients to forecast the CAMEL ratings to be assigned during the following year. For example, we estimate the off-site and examination models using the 1989 sample and use the coefficient estimates to forecast the CAMEL ratings for the examinations in the 1990 sample. This procedure gives us two separate forecasts of CAMEL ratings for 1990, one based on each model.

To evaluate the quality of these CAMEL rating forecasts, we need statistical tools that differ from those used in the within-sample tests of the statistical significance of the regression coefficients. The forecasts from the off-site and examination models are actually probability forecasts that a bank will receive a CAMEL rating of either 1, 2, 3, 4, or 5. For example, such a forecast might be that the bank has a 30 percent chance of being rated 2; a 50 percent chance of being rated 3; a 20 percent chance of being rated 4; and a 0 percent chance of being rated 1 or 5. We use a standard measure of forecast accuracy, known as the logarithmic scoring rule (LSR), to evaluate such multistate probability forecasts. The LSR measure examines how much weight a model's forecast places on the outcome that actually occurred. That is, if the CAMEL rating for a particular examination was 2, the LSR would assess the quality of the forecast by looking only at the probability assigned to that outcome. Under the mathematical assumptions used in computing the LSR measure (Box B), smaller LSR values imply more accurate forecasts.[22]

Box B: Model Comparisons Using the Logarithmic Scoring Rule

The off-site and examination models used in this article are ordered logit models, which provide probability forecasts for each of the five possible CAMEL ratings. In mathematical form, such an out-of-sample forecast, denoted P_n, is a (5x1) vector in which the i^{th} element represents the forecasted probability of being in state *i*. For example, the out-of-sample forecast might be P_n = [0; 0.30; 0.50; 0.20; 0]. Accuracy measures for such forecasts relate the performance of the forecasts to actually observed outcomes. Let R_n be an indicator vector such that if the CAMEL rating is *i* (where *i* = 1,...,5), then the i^{th} element equals one and zero otherwise. For example, if bank *n* receives a CAMEL rating of 4, then R_n = [0; 0; 0; 1; 0].

The accuracy measure used here, known as the logarithmic scoring rule (LSR), examines how much weight the probability forecast places on the actual outcome. That is, if the CAMEL rating for a particular examination were 2, the LSR would assess the accuracy of the forecast only by looking at the probability it assigned to that outcome. The mathematical formula for the LSR is

$$LSR = -\frac{1}{N}\sum_{n=1}^{N} \log\left(\sum_{i=1}^{5} P_{in}{*}R_{in}\right),$$

where N is the number of banks for which forecasts are generated. Since R_{in} equals one only for the CAMEL rating actually observed, the LSR is simply the average of the negative, logged value of the probability forecast for the rating actually observed. LSR can take on values in the interval $[0,\infty]$ with smaller values implying greater accuracy.

The LSR measure permits model comparison by creating performance rankings. For example, if the LSR value for the probability forecasts from model A (denoted $P_{A,n}$) is smaller than that for the forecasts from model B (denoted $P_{B,n}$), then model A is said to be more accurate than model B. However, an important question is whether this observed difference in LSR values is statistically significant or just an artifact of the data sample. To examine this point, Diebold and Mariano (1995) propose several tests for determining whether the difference is statistically different from zero.

Generally, the null hypothesis under scoring rule g is $E[g(P_{An}, R_n)] = E[g(P_{Bn}, R_n)]$, or equivalently, $E[d_n] = E[g(P_{An}, R_n) - g(P_{Bn}, R_n)] = 0$. For LSR, d_n for a single observation is

$$d_n = -\log\left(\sum_{i=1}^{5} P_{Ani}{*}R_{ni}\right) + \log\left(\sum_{i=1}^{5} P_{Bni}{*}R_{ni}\right).$$

To examine this null hypothesis, we simply calculate the difference between the scores from our two models for each observation in the sample and regress it against a constant. If this coefficient is statistically different from zero, then the aggregate scores for the sample as a whole differ significantly; that is, the observed performance ranking is statistically significant.

This measure of forecast accuracy permits model comparison by creating performance rankings. For example, if the LSR value for the probability forecasts from model A is smaller than that for the forecasts from model B, then model A can be said to be more accurate than model B. However, an important question is whether this observed difference in LSR values is statistically significant or just an artifact of the data sample. To examine this point, Diebold and Mariano (1995) propose several tests for determining whether the difference is statistically different from zero. We use a simple variant of one of their suggestions (Box B). Specifically, we calculate the difference between the scores from our two models for each observation in the sample. We then regress this difference against a constant term and test to see whether the constant is statistically different from zero. This procedure is equivalent to testing whether the aggregate scores for the sample as a whole differ significantly.

The basic results of the out-of-sample analysis are presented in Table 3, which contains the comparison of LSR values for each year of the data set.[23] In the early years of the sample, the LSR values for the examination model are significantly smaller than those for the off-site model for examinations up to six to seven quarters old; that is, the difference between the two values is positive and significant. After 1991, this cutoff point increases to ten to twelve quarters. In other words, the results suggest that the private supervisory information contained in CAMEL ratings continues to provide useful information in predicting the current condition of a bank for six to twelve quarters after the previous examination. After this point, there appears to be little value in the information contained in the prior CAMEL rating.

Overall, the examination model generates more accurate forecasts than the off-site model up to a certain point. An alternative way to express this result is to examine the models' integer forecasts of the CAMEL ratings; that is, the expected CAMEL rating, rounded to the nearest integer, based on the models' probability forecasts. These forecasted CAMEL ratings can then be compared with the observed CAMEL ratings. For the 1990 data, the off-site model correctly predicted about 67 percent of the realized CAMEL ratings for banks that had lagged ratings up to six quarters old. The examination model improved this performance by correctly predicting roughly 75 percent of the realized ratings. However, for banks with older lagged CAMEL ratings, both models perform equally, with about 40 percent accuracy. For all the years in our sample, the off-site and examination models exhibit this difference in forecast performance before the cutoff point, but not after. Again, this result indicates that the private supervisory information in lagged CAMEL ratings from full-scope examinations decays over time and is not useful in predicting the current CAMEL ratings after a certain point.[24]

In fact, some of the results suggest that after a certain point, using lagged CAMEL ratings to predict current ones may actually be detrimental to producing accurate forecasts. In some instances, the score for the *off-site* model is significantly smaller than for the *examination* model, indicating that the former produces more accurate forecasts than the latter. For example, in Table 3, for observations in

Table 3
DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS
Full Sample

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.167*	0.173	0.840	0.189*	0.297*	0.429*
2	0.223*	0.250*	0.119*	0.137*	0.184*	0.307*
3	0.218*	0.215*	0.138*	0.137*	0.202*	0.165*
4	0.160*	0.145*	0.162*	0.103*	0.146*	0.179*
5	0.152*	0.137*	0.101*	0.095*	0.153*	0.151*
6	0.179*	0.130*	0.090*	0.145*	0.278*	0.146*
7	0.029	0.069*	0.076*	0.077*	0.121*	0.129*
8	-0.045	0.046	0.025	0.082*	0.096*	0.097*
9-10	-0.104	-0.061	0.071*	0.077*	0.076*	0.109*
11-12	-0.177	-0.095*	0.037	0.083*	0.098*	0.053
13 or more	-0.212*	-0.157*	-0.030*	-0.027*	-0.073*	-0.073

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

which the lagged CAMEL rating is thirteen or more quarters old, the off-site model has a significantly smaller LSR in all but one year in the sample, and thus it provides more accurate predictions of current CAMEL ratings in these years. These results imply that these aged CAMEL ratings add no value in assessing a bank's current condition.

Finally, the results suggest that there is significant variation over the sample period in the useful life of supervisory information from prior examinations. This variation may reflect changes in the condition of the U.S. banking industry over the sample period. In particular, the private supervisory information contained in CAMEL ratings appears to decay more rapidly during the early part of the sample period, when the U.S. banking industry was experiencing financial stress, than during the latter part of the sample period, when the industry experienced more robust performance. Because we would expect the condition of banks to change more rapidly during periods of financial stress, we would also expect a faster rate of information decay.

Our findings suggest that the rate of decay in private supervisory information is considerably faster for banks experiencing some degree of financial difficulty, regardless of the overall condition of the banking industry.

To explore our results further, we divided the data into subsets according to the initial financial condition of the bank. Specifically, for each year, we divided the data sample into observations with lagged CAMEL ratings of 1 or 2 (indicating little reason for supervisory concern at the time of the previous examination) and observations with lagged CAMEL ratings of 3, 4, or 5 (indicating moderate to severe degrees of supervisory concern). We then compared the LSR measures for our CAMEL forecasting models for each of these subsamples. These results are reported in Tables 4 and 5.

As Table 4 indicates, the results for the subsample with lagged CAMEL ratings of 1 or 2 are very similar to those for the overall sample. The results indicate that the lagged CAMEL ratings cease to provide useful information about the current condition of a bank after six to twelve quarters have elapsed and that this information decays faster in the early part of the sample, when the U.S. banking industry was experiencing financial distress. The similarity between these results and the overall results for the sample is not surprising, since the majority of observations (between 70 and 90 percent) have lagged CAMEL ratings of 1 or 2.

As indicated in Table 5, the findings for observations with lagged CAMEL ratings of 3, 4, or 5 are considerably different.[25] The point at which the lagged CAMEL rating ceases to provide useful information regarding the current CAMEL ratings is significantly earlier than it is for the overall sample.[26] The information in prior CAMEL ratings seems to be no longer useful after just three to six quarters. Further, the cyclical pattern that

Table 4
DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS
Subsample with Lagged CAMEL Ratings of 1 or 2

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.150*	0.252*	0.082*	0.210*	0.158*	0.331*
2	0.147*	0.114*	0.165*	0.131*	0.164*	0.231*
3	0.145*	0.169*	0.112*	0.175*	0.137*	0.150*
4	0.119*	0.108*	0.168*	0.139*	0.136*	0.164*
5	0.133*	0.111*	0.134*	0.118*	0.125*	0.146*
6	0.126*	0.103*	0.111*	0.138*	0.283*	0.141*
7	0.043	0.048	0.073*	0.091*	0.160*	0.127*
8	-0.045	0.028	0.078*	0.096*	0.105*	0.091*
9-10	-0.058	-0.051	0.036*	0.085*	0.082*	0.109*
11-12	-0.207	-0.042	0.008	0.078*	0.111*	0.086
13 or more	-0.182	-0.109*	0.001	-0.013	-0.049*	-0.056

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

was evident in both the overall sample and in the subsample with lagged CAMEL ratings of 1 or 2 does not emerge in these results. Taken together, these findings suggest that the rate of decay in private supervisory information is considerably faster for banks experiencing some degree of financial difficulty, regardless of the overall condition of the banking industry.

What do these results imply for the basic question motivating this article, namely, how often should a bank be examined? To answer this question, it is important to understand that the tests described above provide an upper-bound for the length of time that prior CAMEL ratings provide useful information about current bank conditions. That is, beyond the six-to-twelve-quarter range is where the lagged CAMEL rating contains *no* useful information about the current condition of a bank. In practice, supervisors would probably wish to examine a bank before this point, when the private information from the prior examination continues to have *some*, though diminished, value.

Table 5

DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS

Subsample with Lagged CAMEL Ratings of 3 to 5

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.220	-0.143	0.089	0.142	0.725*	0.771*
2	0.345*	0.377*	0.071*	0.143*	0.215*	0.433*
3	0.365*	0.303*	0.173*	0.071*	0.374*	0.221*
4	0.294*	0.278*	0.142*	0.002	0.208*	0.305*
5	0.240*	0.275*	-0.005	0.020	0.321*	0.200*
6	0.568*	0.309*	0.000	0.112	0.193	0.359
7	-0.091	0.141	0.096	-0.072	-0.684	0.186
8	-0.049	0.136	-0.139	-0.060	-0.062	0.502
9-10	-0.383	-0.102	0.152*	-0.024	-0.042	0.133
11-12	0.223	-0.340	0.100	0.154	-0.106	NA
13 or more	-0.980	-0.450	-0.160*	-0.115*	-0.163	-0.391

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

Finally, in thinking about the optimal time between examinations, the results suggest that this horizon may vary. During periods of financial stress in the banking industry, the quality of private supervisory information appears to decay faster than it does in more stable periods, suggesting that the optimal time between examinations may be shorter in times of stress. Further, the rate of information decay is markedly greater for banks that are themselves financially troubled, regardless of the state of the overall industry. This finding implies, rather sensibly, that it is desirable to examine troubled institutions more often than healthy ones, although the optimal examination interval for any particular bank will vary from the averages discussed here.[27]

ROBUSTNESS CHECKS

To examine the robustness of our results to the choices we made in setting up the analysis, we conducted two additional sets of tests. We examined the performance of our two models on out-of-sample observations, both to test the robustness of the results and to mirror more closely the actual situation faced by bank examiners. The approach we chose—year-ahead forecasts—is only one way of setting up such an out-of-sample test. As discussed in Granger and Huang (1997), out-of-sample analysis for models of this type can also be conducted by holding out a random part of the sample for a given year and using that holdout sample for the out-of-sample analysis.[28] We use this approach to test whether the results discussed above are due solely to the year-ahead forecast analysis.

Table 6 contains the results of our holdout sample prediction analysis.[29] For each year in the sample, we estimated the two models over a randomly selected 75 percent of the total sample. These estimated models for each year were then used to predict the CAMEL ratings on the remaining 25 percent of the sample. We again compared the accuracy of these predictions using the LSR measure.

The holdout sample prediction results are broadly similar to those for the year-ahead forecast analysis. The examination model exhibits better performance than the off-site model for observations with CAMEL ratings that are six to twelve quarters old; that is, the differences in

LSR values between the examination and off-site models are positive and significant for this portion of the holdout sample. The previously observed cyclical pattern is less evident, but the results nonetheless provide some indication that the information in lagged CAMEL ratings decays less rapidly in the latter years of the sample. The weaker cyclical pattern may be due to the considerably smaller number of out-of-sample observations available using this type of analysis. The smaller sample size reduces the power of the statistical tests to determine whether the accuracy measures for the two models differ significantly. Overall, however, the holdout sample results support the findings of the year-ahead forecasts, suggesting that our analysis is not overly sensitive to the structure of the out-of-sample analysis.

For the second set of robustness tests, we focus directly on the question whether the time between full-scope examinations can be treated as an exogenous variable in our two models. We have assumed that the models capture the relevant explanatory variables used by examiners in determining CAMEL ratings. However, it might be the case that in scheduling examinations, supervisors have additional information—not present in our empirical specifications—about the extent to which conditions at a bank have changed since the last examination. Using such information, supervisors might schedule more frequent examinations for banks whose financial conditions are less stable and less frequent examinations for those with more stable conditions. In that case, the time since the last examination would be an endogenous variable, rather than an exogenous one as we have assumed. That is, the time since the last examination may be a function of the current CAMEL rating that we are trying to model.

FDICIA's requirement for annual full-scope examinations seems reasonable, particularly for banks whose initial financial condition is troubled or when the banking system as a whole is experiencing financial stress.

Some preliminary evidence in favor of such endogeneity is presented in Table 7, which shows the cumulative distribution of the time since the last full-scope examination in percentage terms. The first column of the table presents the cumulative distribution for all observations aggregated across the seven years in the sample, while the other columns report the results for subsamples divided by current CAMEL ratings. Clearly, the time between examinations for banks with CAMEL ratings of 3 to 5 is shorter than it is for banks with ratings of 1 or 2. About 45 percent of banks with ratings of 1 or 2 had a lagged full-scope examination within four quarters, compared with almost 60 percent for banks with ratings of 3 to 5. Although this difference diminishes as the time between examinations increases (by eight quarters, the percentages are nearly equal), it may be the case that the time since the last examination is a function of the current CAMEL rating. The existence of such endogeneity might lead our

Table 6

DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES FOR THE HOLDOUT SAMPLE, SORTED BY TIME SINCE LAST EXAMINATION

Full Sample

	(LSR1-LSR2)						
Quarters since Last Examination	1989	1990	1991	1992	1993	1994	1995
1	0.177	-0.040	0.226	0.175*	0.241*	0.229	0.238
2	0.224*	0.266*	0.228*	0.179*	0.280*	0.279*	0.150
3	0.237*	0.195*	0.123	0.177*	0.178*	0.242*	0.077
4	0.136*	0.168*	0.100*	0.142*	0.150*	0.113*	0.128*
5	0.104*	0.184*	0.134*	0.215*	0.137*	0.222*	0.141*
6	0.154*	0.237*	0.121*	0.072*	0.145*	0.142*	0.144*
7	0.057	0.121*	0.099*	0.059	0.093	0.106*	0.106*
8	0.070	0.135	0.005	0.017	0.049	0.069	0.108
9-10	-0.007	-0.175	0.053	0.071*	-0.008	0.058	0.028
11-12	0.077	-0.060	-0.038	0.017	0.045	0.144*	-0.085
13 or more	-0.057	-0.610	0.035	-0.008	-0.042	-0.083	-0.069

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

empirical tests to overstate the length of time that a lagged CAMEL rating continues to provide useful information about the current condition of a bank.

To test for this possible endogeneity, we use two distinct methods. First, we use a logistic regression relating the probability that the CAMEL rating changes (either upward or downward) to the time between examinations. If the time between examinations were strictly endogenous, we would expect to find no significant relationship between these two variables: supervisors would schedule examinations at the point when conditions at the bank had changed sufficiently to warrant a change in the CAMEL rating. In contrast, if the time between examinations were exogenous, we would expect to see a positive relationship between the time since the last examination and the probability of a change in the CAMEL rating.

The results of this regression are reported in Table 8. Clearly, the coefficient on the time since the last examination is positive and significant for each year of the sample. Although the overall fit of the regressions is poor (the R^2 statistics are quite low), these results support the idea that the time since the last examination is not significantly endogenous.

Table 7
CUMULATIVE DISTRIBUTION OF TIME SINCE LAST FULL-SCOPE EXAMINATION
Complete Sample and Divided by Current CAMEL Rating

Quarters since Last Examination	Complete Sample (Percent)	Current CAMEL Rating of 1 or 2 (Percent)	Current CAMEL Rating of 3 to 5 (Percent)
1	1.0	0.9	1.4
2	8.5	6.7	16.2
3	20.8	17.5	35.5
4	47.1	44.7	58.0
5	64.1	62.3	71.8
6	73.3	72.5	76.8
7	78.3	77.8	80.3
8	82.7	82.6	83.3
9	85.8	85.8	85.8
10	87.8	87.8	87.7
11	89.4	89.4	89.3
12	90.8	90.8	90.9
13-14	93.3	93.1	94.2
15-16	95.4	95.1	96.7
17 or more	100.0	100.0	100.0
Memo: Number of observations	49,455	40,252	9,203

Source: Authors' calculations, based on data from the Board of Governors of the Federal Reserve System.

To explore this question further, we conducted a second test that explicitly attempts to control for the endogeneity of the time between examinations. We begin this test by modeling the time since the last examination as a function of variables that are correlated with it, but not with the current CAMEL ratings. The fitted values from this model should therefore be free of this possible endogeneity. By substituting these fitted values for the dummy variables for the actual time since the last examination in our two earlier models, we expect that the generated CAMEL rating forecasts will not be affected by any endogeneity between the time since the last examination and the current CAMEL rating. If the LSR results based on these modified models are found to be similar to those for the versions that do not control for potential endogeneity, then this finding would provide additional evidence that such endogeneity is not influencing our results.[30]

In technical terms, we model the time between examinations using an econometric technique known as hazard modeling.[31] The explanatory variables used in estimating the hazard models were the changes in the core balance sheet and income statement variables that form the basis of the off-site and examination models. Although the levels of these variables are significant determinants of current CAMEL ratings, it seems reasonable to assume that their lagged values, and thus the changes in their values, are exogenous. After the hazard models have been estimated, they can be used to generate predicted probabilities

Table 8
LOGISTIC REGRESSION RESULTS: PROBABILITY OF CAMEL RATING CHANGE AS A FUNCTION OF TIME SINCE LAST EXAMINATION

	1989	1990	1991	1992	1993	1994	1995
Constant	-1.096 *	-1.522 *	-1.456 *	-0.996 *	-1.032 *	-1.283 *	-1.470 *
	(0.057)	(0.078)	(0.073)	(0.042)	(0.035)	(0.039)	(0.052)
Time since last examination (months)	0.016 *	0.033 *	0.031 *	0.017 *	0.011 *	0.013 *	0.018 *
	(0.004)	(0.005)	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)
R^2	0.002	0.010	0.026	0.013	0.008	0.005	0.004
Number of observations	6,998	4,306	3,980	8,324	8,998	8,837	8,012

Source: Authors' calculations.

Notes: The dependent variable equals 1 when the CAMEL rating changes (increases or decreases) and is zero otherwise. R^2 statistics are those derived for limited dependent variable models in Estrella (1998). An asterisk indicates that the coefficient is significantly different from zero at the 1 percent level.

Table 9

DIFFERENCES IN LOGARITHMIC SCORING RULE VALUES BASED ON FITTED HAZARD FUNCTIONS, SORTED BY TIME SINCE LAST EXAMINATION FOR THE ONE-YEAR-AHEAD FORECASTS OF CAMEL RATINGS

Full Sample

	(LSR1-LSR2)					
Quarters since Last Examination	1989/ 1990	1990/ 1991	1991/ 1992	1992/ 1993	1993/ 1994	1994/ 1995
1	0.159*	0.143	0.199*	0.141*	0.279*	0.391*
2	0.236*	0.242*	0.168*	0.059*	0.170*	0.295*
3	0.226*	0.199*	0.172*	0.073*	0.195*	0.179*
4	0.165*	0.146*	0.163*	0.114*	0.150*	0.187*
5	0.155*	0.118*	0.136*	0.100*	0.152*	0.155*
6	0.179*	0.125*	0.138*	0.134*	0.283*	0.189*
7	0.021	0.064*	0.100*	0.075*	0.144*	0.149*
8	-0.052	0.013	0.036	0.085*	0.095*	0.125*
9-10	-0.124*	-0.011	0.065*	0.075*	0.110*	0.136*
11-12	-0.185	-0.116*	0.031	0.096*	0.104	0.141*
13 or more	-0.199	-0.070	-0.017	0.009	-0.088*	-0.024

Source: Authors' calculations.

Notes: Each figure gives the difference in values for banks examined in that year (column) whose lagged CAMEL ratings were of the corresponding age (row). The figures represent the difference between the LSR value for the off-site model (LSR1) and the LSR value for the examination model (LSR2). A positive (negative) value indicates that the examination (off-site) model produces a more accurate forecast than the off-site (examination) model. An asterisk indicates that the difference is significantly different from zero at the 5 percent level using the specified Diebold-Mariano test. The end of the shading indicates the point at which the difference between the LSR values is no longer statistically positive.

that the time between examinations falls into specified ranges. We substituted these predicted probabilities for the dummy variables representing the actual time since the last examination.[32]

Table 9 presents the LSR comparison results for the off-site and examination models using the estimated survivor function for the examinations in each year. The results are quite similar to those reported in Table 3. In the early years of the sample, the LSR value for the examination model is less than the value for the off-site model for examinations up to six or seven quarters old. After 1991, this cutoff point increases to roughly nine to twelve quarters after the examination. The results for the two subsamples of CAMEL ratings (not reported in the tables) are similar to those in Tables 4 and 5. Thus, the out-of-sample forecast results do not appear to be sensitive to our attempts to control for the potential endogeneity of the time since the last examination. Based on these results, as well as on the logit results reported above, it does not appear that our conclusions are being driven by an endogenous relationship between the current CAMEL rating and the time since the previous examination.

CONCLUSION

This article examines the frequency with which supervisors should examine banks by assessing the decay rate of the private supervisory information gathered during full-scope examinations. Such information is costly to obtain since it can be gathered only during on-site examinations. Thus, the question of how quickly the information's value erodes has important implications for both supervisors and banks. The more quickly this information decays, the more frequently examinations need to take place in order for supervisors to have access to accurate information about the current condition of banks.

Our results suggest that CAMEL ratings cease to provide any useful information about the current condition of a bank after about six to twelve quarters. Thus, examinations should take place at least at this frequency, since supervisors would probably want to examine a bank while the information from the previous examination continues to have some value. Our results indicate that supervisory information tends to decay more rapidly for banks with weaker CAMEL ratings (3, 4, or 5). Thus, for these institutions, a somewhat shorter examination cycle may be justified. In this light, FDICIA's requirement for annual full-scope examinations seems reasonable, particularly for banks whose initial financial condition is troubled or when the banking system as a whole is experiencing financial stress. Of course, the optimal examination frequency for any particular bank can and will deviate from the average results presented here.

ENDNOTES

The authors thank Bob DeYoung, Fred Herriman, Larry Radecki, Marc Saidenberg, Phil Strahan, Joe Tracy, two anonymous referees, and participants at the 1998 meeting of the Federal Reserve System Committee on Financial Structure and Regulation and the Federal Reserve Bank of New York Banking Studies seminar for helpful comments and suggestions. Leslie DuPuy and Oba McMillan provided excellent research assistance in preparing the data set for this article.

1. An important qualification to this statement is that the verification of the accuracy of regulatory reports is one aspect of on-site examinations.

2. Note that our results reflect the average pattern of information decay across the examinations in the sample; the optimal examination timing for individual banks will differ from these averages.

3. Profits in the banking industry fell sharply through the mid-to-late 1980s, reflecting large loan losses in several lending sectors, including agriculture, energy, developing countries, and real estate. Profits, as measured by return on equity, did not return to pre-downturn levels until 1992. Failures also rose sharply during this period, reaching a high of more than 250 per year in the late 1980s (see Federal Deposit Insurance Corporation [1997]).

4. According to estimates by the Federal Deposit Insurance Corporation (FDIC), the field examination staffs of the three federal bank supervisory agencies—the FDIC, the Federal Reserve, and the Office of the Comptroller of the Currency (OCC)—and the fifty state banking supervisors totaled about 9,000 in 1994. For more information, see Federal Deposit Insurance Corporation (1997).

5. See Federal Deposit Insurance Corporation (1997, p. 428).

6. The exception is very small banks with supervisory ratings that indicate few, if any, significant supervisory concerns; these banks can be examined once every eighteen months.

7. For example, the Federal Reserve uses the Financial Institutions Monitoring System (FIMS) for this purpose (see Cole, Cornyn, and Gunther [1995] for details).

8. The formal name of the rating system is the Uniform Financial Institutions Rating System, although it is commonly known as the CAMEL rating system. In 1997, a sixth component was added, reflecting a bank's *S*ensitivity to market risk. The expanded rating system is known as the CAMELS rating system. Because our data sample extends only through 1995, none of the examinations in our sample includes this new component.

9. See Federal Financial Institutions Examination Council (1996) for a detailed description of the CAMEL rating system and an interpretation of the component and composite ratings.

10. Bank holding companies are examined separately from their bank subsidiaries. The BOPEC rating assigned at the conclusion of such an examination reflects the conditions of the holding company's *B*ank subsidiaries, *O*ther nonbank subsidiaries, *P*arent company, *E*arnings, and *C*apital adequacy.

11. We focus on the time between full-scope bank examinations, so banks in our sample may have had either a targeted or limited-scope examination between full-scope examinations. In such cases, supervisors will have had the opportunity to update some of their private information about the bank's condition. As discussed above, however, such examinations generally do not result in comprehensive assessments of a bank's condition. Therefore, by examining the time interval between full-scope examinations, we likely obtain the best indication of the time decay of the private supervisory information.

12. Note that there are a significant number of observations with prior examinations more than six quarters old, even in 1994 and 1995, well after the passage of FDICIA, which set an outside limit of eighteen months between examinations. About 75 percent of the observations have intervening, limited-scope examinations that occurred within six quarters of the current examination, suggesting the efforts made by supervisors to make a full transition to FDICIA's requirements. Furthermore, the relatively small number of observations during 1989 and 1990 for which the time between examinations is fairly long may partly reflect the source data used in constructing the data set. Because the source data contained increasingly sparse information on examinations before 1989, our data set for 1989 and 1990 excludes examinations of banks whose previous examinations were not recorded in the source data.

13. Technically, the statistical approach used is an ordered logit model. CAMEL ratings have discrete values, so a standard linear regression model—which assumes that the dependent variable is continuous—would be inappropriate. The ordered logit model is specifically designed to handle discrete dependent variables, such as CAMEL ratings, whose values are ordinally related (that is, 1 implies "strongest performance," while 5 implies "weakest performance"). See Maddala (1983) for a detailed discussion of ordered logit models.

14. We conduct our analysis on annual, cross-sectional data sets, as opposed to a panel data set, for two reasons. First, a simple likelihood ratio test rejects the null hypothesis that the model coefficients are constant across the years. Second, because examiners must allocate their

Note 14 continued
scarce resources over future examinations, we felt that focusing on year-ahead CAMEL rating forecasts generated from annual data sets would more closely mirror examiner behavior.

15. The R^2 statistic is the goodness-of-fit measure developed by Estrella (1998) specifically for limited dependent variable models. The statistic is roughly analogous to the R^2 statistic used in linear regressions because its value ranges between zero (for a model with no explanatory power) and one (for a model with complete explanatory power).

16. The information contained in the lagged CAMEL rating reflects both private supervisory information and past values of the public information on bank condition. To isolate the effects of the supervisory information, we also estimated a version of the examination model that controlled for the publicly available information component. In particular, we estimated an ordered logit model that regressed lagged CAMEL ratings on lagged values of the publicly available independent variables listed in Box A. From this model, we calculated a fitted value of the lagged CAMEL rating using the predicted probabilities that the rating was equal to a 1, 2, 3, 4, or 5. We then subtracted this fitted value from the actual lagged CAMEL rating. We interpret this residual as reflecting the information in the lagged CAMEL rating stemming just from the private supervisory information. We then substituted this residual for the actual lagged CAMEL rating in the ordered logit equation for the current CAMEL rating. The results of the subsequent out-of-sample forecast analysis were nearly identical to those for the examination model using the actual lagged CAMEL rating, suggesting that this variable primarily reflects private supervisory information.

17. Using mathematical notation, we can summarize the off-site model for a given year as $y_i = f\left(\gamma x_i + \sum_{j=2}^{15} \beta_j I(\text{lag})_{ij} + \varepsilon_i\right)$, where y_i is the current CAMEL rating for bank i; γ is the vector of coefficients on the independent variables x_i listed in Box A (except for the indicator variables for the time since last examination); the $I(lag)_{ij}$'s are the indicator variables corresponding to the time since the last examination for bank i; the β_j's are the corresponding coefficients; and ε_i is the error term. The examination model for a given year is

$$y_i = f\left(\gamma x_i + \sum_{j=2}^{15} \beta_j I(lag)_{ij} + \sum_{j=1}^{15} \theta_j {*} I(lag)_{ij} {*}\, lag\text{CAMEL}_i + \varepsilon_i\right),$$

where $lag\text{CAMEL}_i$ is the value of the lagged CAMEL rating for bank i from the previous examination and the θ_j's are the corresponding coefficients. The difference between the models is simply the inclusion of interacted, lagged CAMEL variables. The coefficients on the x_i variables, particularly on the core set of variables, do not significantly change when the interacted, lagged CAMEL variables are included in the specification.

18. We also estimated a constrained version of the examination model in which the coefficient on the lagged CAMEL rating does not vary with the time since the last examination. Based on a likelihood ratio test, these constraints are clearly rejected for every year in the sample. This finding indicates that there is meaningful variation in the coefficients on the lagged CAMEL ratings as the age of the rating increases. However, as the out-of-sample forecast results (discussed in the next section) were not significantly affected by these constraints, our primary results are not overly sensitive to the way in which the lagged CAMEL ratings enter the examination model.

19. Using this model specification, we assume that the time since the last full-scope examination is an exogenous variable; that is, it does not depend on the current CAMEL rating. A plausible argument can be made that the variable is endogenously determined, especially with respect to lower rated banks. Although we cannot conclusively prove that the variable is exogenous, we provide indirect evidence later on. We maintain the assumption throughout the analysis that follows.

20. The coefficient on lagged CAMEL ratings that are seventeen or more quarters old in the 1989 regression is an exception, since it is statistically significant.

21. To try to formalize this observation, we test the hypothesis that the coefficients on lagged CAMEL ratings that were twelve or more quarters old were smaller than the coefficients on lagged CAMEL ratings that were four quarters old. (We selected four quarters as being representative of relatively new CAMEL ratings, but the results are not sensitive to this choice.) For all cases, the coefficients on the older CAMEL ratings were smaller than those on the four-quarter-old CAMEL ratings, and these differences were statistically significant about half the time. In particular, in all but two of the sample years, at least half the coefficients on the older CAMEL ratings were significantly smaller. These results lend support to the more informal observation that the size of the coefficients tends to decrease as the age of the CAMEL ratings increases.

22. Estrella and Mishkin (1996) recommend using the logarithmic scoring rule to evaluate probability forecasts derived from models estimated using maximum likelihood estimation techniques (such as the ordered logit models used here). The reason for this is that the LSR mimics the weights used in maximizing the likelihood function to obtain the parameter estimates.

23. To maximize the power of the Diebold-Mariano test used in the analysis, we use a higher level of aggregation for the lagged CAMEL ratings. Specifically, we group into three categories all observations for which the prior examination occurred nine or ten quarters ago, eleven or twelve quarters ago, and thirteen or more quarters ago, rather than into the seven categories used in the ordered logit estimation.

24. We also estimated versions of our model in which we attempted to assess the impact of limited-scope examinations on these results. Overall, about 20 percent of the observations in the sample have limited-scope examinations between the full-scope examinations. The distribution of these observations is uneven across years and concentrated in 1991 to 1994, the period during which the supervisory agencies were in the midst of the transition to FDICIA. For this analysis, we substituted the time since the last limited-scope examination and its associated CAMEL rating for the time since the last full-scope examination and its CAMEL rating. The empirical results differ somewhat from the results presented in Table 3. The adjusted results suggest that the information contained in lagged CAMEL ratings decays within six to eight quarters, and the strong cyclical pattern in Table 3 is not evident. The difference in results may be attributable to the fact that the adjustment for limited-scope examinations reduces the number of observations with "old" lagged CAMEL ratings to the point where the statistical tests on this part of the sample have greatly diminished power. Alternatively, the results could reflect the fact that limited-scope examinations are not as in-depth as full-scope examinations and may not produce information of as high a quality. The difference in our results could reflect the fact that this lower quality information simply decays faster than the information derived from full-scope examinations. This interpretation is supported by the results presented in Table 3, which suggest that the information from lagged full-scope CAMEL ratings persists *even when there has been an intermediate limited-scope examination.* Based on this interpretation, we do not view the limited-scope results as undercutting our findings about the persistence of information from full-scope examinations.

25. The subsample of banks with lagged CAMEL ratings of 3, 4, or 5 makes up between 10 and 30 percent of the yearly samples. This smaller sample size reduces the power of the Diebold-Mariano tests upon which our results are based, especially for the reduced number of banks with older CAMEL ratings. However, the sample size for the figures just beyond the cut-over points (that is, the figures after which our inference is most relevant) remains large enough to permit valid inference.

26. Note that, for the 1992 results in Table 5, the LSR values for the examination model are lower than those for the off-site model up to four quarters since the last examination. There is no clear intuition why this is also the case at nine to ten quarters. This result is probably due to random noise in the data set.

27. In fact, the conclusion that financially troubled institutions should be examined more frequently is directly incorporated into FDICIA. For example, the previously noted exception for small banks is disallowed for banks with CAMEL ratings of 3, 4, or 5. In fact, the eighteen-month examination window is disallowed, regardless of CAMEL rating, for a number of reasons related to bank soundness and important changes in bank control. These additional criteria help ensure that potentially troubled banks are examined more frequently than healthy ones. Moreover, the FDICIA-specified intervals between examinations are meant to be outer limits; bank supervisors have the discretion to conduct more frequent examinations. The figures in Table 1 suggest that this discretion is often used.

28. Granger and Huang (1997) distinguish between forecasting, which involves estimating a model on a set of observations and then applying these estimates to observations from a future period, and prediction, which involves estimating a model on a subset of the observations from a given period and then applying these estimates to other observations from the same time period. In the discussion in the text, we use these terms in a manner consistent with these definitions.

29. Because we are no longer using the year-ahead forecast analysis, we have out-of-sample results for 1989.

30. Note that the use of these fitted values for the time since the last examination can be viewed, at least in spirit, as an instrumental variable estimation technique.

31. See Kiefer (1988) for a survey of hazard modeling. For our particular hazard model, we specify the baseline hazard function as a Weibull function, which allows the hazard rate (the probability that an examination occurs in a given quarter after the last examination) to increase or decrease as the time since the last examination increases. In our estimates, we found evidence that the hazard rate was increasing, suggesting that banks were more likely to be examined as the time since the last examination increased.

32. The hazard model results were used to create variables representing the probability of an examination occurring one, two, three, four, and five or more years after the lagged examination. Note that this is a higher level of aggregation than the one used in the results presented in Tables 3-5, where fifteen dummy variables were used. The reduction in the number of time variables was performed to facilitate the estimation of the ordered logit models used in the analysis. The results are not sensitive to this reduction.

REFERENCES

Barker, D., and D. Holdsworth. 1993. "The Causes of Bank Failures in the 1980s." Federal Reserve Bank of New York Research Paper no. 9325.

Berger, A. N., and S. M. Davies. 1994. "The Information Content of Bank Examinations." In PROCEEDINGS OF THE 1994 CONFERENCE ON BANK STRUCTURE AND COMPETITION, 615-28. Federal Reserve Bank of Chicago.

Berger, A. N., S. M. Davies, and M. J. Flannery. 1998. "Comparing Market and Supervisory Assessment of Bank Performance: Who Knows What When?" Board of Governors of the Federal Reserve System Finance and Economics Discussion Series Paper no. 1998-32.

Cole, R. A., B. G. Cornyn, and J. W. Gunther. 1995. "FIMS: A New Monitoring System for Banking Institutions." FEDERAL RESERVE BULLETIN, January: 1-15.

Cole, R. A., and J. W. Gunther. 1995. "A CAMEL Rating's Shelf Life." Federal Reserve Bank of Dallas FINANCIAL INDUSTRY STUDIES, December: 13-20.

———. 1998. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems." JOURNAL OF FINANCIAL SERVICES RESEARCH 13: 103-17.

DeYoung, R., M. J. Flannery, W. Lang, and S. Sorescu. 1998. "The Informational Advantage of Specialized Monitors: The Case of Bank Examiners." Federal Reserve Bank of Chicago Research Department Working Paper no. 98-4.

Diebold, F. X., and R. Mariano. 1995. "Comparing Predictive Accuracy." JOURNAL OF BUSINESS AND ECONOMIC STATISTICS 13: 253-64.

Estrella, A. 1998. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." JOURNAL OF BUSINESS AND ECONOMIC STATISTICS 16: 198-205.

Estrella, A., and F. S. Mishkin. 1996. "The Evaluation of Out-of-Sample Probability Forecasts Based on Maximum Likelihood Estimates." Unpublished paper, Federal Reserve Bank of New York.

Federal Deposit Insurance Corporation. 1997. "An Examination of the Banking Crises of the 1980s and Early 1990s." In HISTORY OF THE EIGHTIES: LESSONS FOR THE FUTURE. Vol. 1, 1-86 and 428. Washington, D.C.: Federal Deposit Insurance Corporation.

Federal Financial Institutions Examination Council. 1996. "Uniform Financial Institutions Rating System." FEDERAL REGISTER 61, no. 245: 67021-9.

Gilbert, R. A. 1993. "Implications of Annual Examinations for the Bank Insurance Fund." Federal Reserve Bank of St. Louis REVIEW, January-February: 35-52.

Granger, C. W. J., and L.-L. Huang. 1997. "Evaluation of Panel Data Models: Some Suggestions from Time Series." University of California, San Diego, Department of Economics Discussion Paper no. 97-10.

Kiefer, N. M. 1988. "Economic Duration Data and Hazard Functions." JOURNAL OF ECONOMIC LITERATURE 26: 646-79.

Maddala, G. S. 1983. LIMITED-DEPENDENT AND QUALITATIVE VARIABLES IN ECONOMETRICS. Cambridge: Cambridge University Press.