Journal of Financial Services Research 13:2 103–117 (1998)

Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems

REBEL A. COLE
Employment Policies Institute, Washington, DC

JEFFERY W. GUNTHER
Federal Reserve Bank of Dallas, TX

Abstract

On-site examinations are regulators' primary tool for monitoring the financial condition of federally insured depository institutions. In this paper, we assess the speed with which the information content of the supervisory rating assigned during bank exams—the CAMEL rating—decays. This is an important issue because cost and regulatory burden considerations often cause CAMEL ratings to be assigned relatively infrequently. As a benchmark for information content, we use econometric forecasts of bank failures generated by applying a probit model to publicly available accounting data. When compared with all CAMEL ratings available at a given point in time, the econometric forecasts provide a more accurate indication of failure. Further analysis reveals that this overall finding reflects the tendency for a CAMEL rating's information content to deteriorate noticeably beginning in the second or third quarter after the rating initially was assigned.

1. Introduction

In evaluating the financial performance and condition of banks, regulators use a combination of on-site examinations and off-site surveillance systems. During an on-site exam, regulators visit a bank's offices to evaluate its financial soundness and compliance with laws and regulatory policies, to assess the quality of its management team, and to evaluate its systems of internal control. Based on the findings of the exam, regulators assign the bank a composite rating, known by the acronym CAMEL, which refers to the five components of the regulatory rating system: capital adequacy, asset quality, management, earnings, and liquidity.[1] Because financial conditions often change rapidly, the supervisory ratings assigned to banks on the basis of on-site exams would not be expected to remain applicable for long periods of time. For this reason, the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) mandated annual exams.[2] However, the speed with which a bank's financial condition can change may at times reduce the applicability of even annual exam ratings, thereby creating a potentially important role for more frequent off-site monitoring. In this regard, regulators long have relied on off-site monitoring systems, or early warning models, to supplement the CAMEL ratings derived from periodic on-site exams and to provide up-to-date assessments of the financial status of individual banks. A recent example of this type of off-site monitoring

system is the Federal Reserve's Financial Institutions Monitoring System (FIMS), which was implemented in 1993.[3]

Even though several empirical studies have assessed the information content of regulatory ratings, few have attempted to measure the speed with which changing financial conditions may reduce a CAMEL rating's applicability. Based on regressions explaining the rates paid on certificates of deposit, Cargill (1989) concludes that CAMEL ratings are primarily proxies for market information. In contrast, Hirschhorn (1987), Flannery and Houston (1994), and Berger and Davies (1994) provide evidence that on-site bank examinations lead to the dissemination of private (i.e. confidential regulatory) information about the financial condition of individual banks. Berger and Davies argue that changes in confidential examination ratings are reflected in subsequent changes in the information publicly available from the quarterly *Reports of Condition and Income* (call report) that banks file with the regulatory agencies. Changes in call report data, such as adjustments to nonperforming assets, reserves, capital, net income, and other financial measures, act as a conduit transmitting a portion of exam results to the public. However, Berger, King, and O'Brien (1991) and Jones and King (1995) find evidence that financial reports using book values often overstate a bank's true market-value financial condition, pointing to the need for frequent exams. Similarly, the Federal Deposit Insurance Corporation (1997) argues that reductions in examiner resources and associated increases in the average interval between on-site exams made it difficult for regulators to identify emerging banking difficulties during the 1980s and most likely raised the final cost of resolving the banking problems of that decade.

In this paper, we empirically assess the degree to which the ability of off-site monitoring systems to predict bank failures provides a valuable supplement to on-site exams. Our assessment is based on the speed with which changing financial conditions reduce the accuracy of CAMEL ratings in anticipating bank failure. As a benchmark for information content, we use forecasts of bank failures produced by applying a probit model to publicly available accounting data. A previous study by Gilbert and Park (1994) concluded that off-site monitoring systems frequently provide an earlier indication of financial deterioration at banks that ultimately fail than on-site exams.

The on-site examination process and the CAMEL ratings it generates have numerous important uses, the full treatment of which is beyond the scope of this paper.[4] Reflecting the scope of on-site exams, CAMEL ratings incorporate a bank's financial condition, its compliance with laws and regulatory policies, and the quality of its management and systems of internal control. Regulators do not expect all poorly rated banks to fail but rather focus attention on early intervention and take action designed to return troubled banks to financial health. Given the multiple dimensions of CAMEL ratings, their primary purpose is *not* to predict bank failures.

Nevertheless, success in discriminating between failing and surviving banks provides a convenient metric for assessing the speed with which changing financial conditions can reduce the applicable information content of previously assigned CAMEL ratings. And relative accuracy in anticipating bank failure is a particularly suitable measure of the importance of early warning models as a supplement to exam ratings. Our analysis takes into account the length of time between on-site exams and subsequent failures, defined here as regulatory closures.[5] If changing financial conditions reduce the information

content of previously assigned CAMEL ratings, then any tendency for our off-site econometric model to identify failures more accurately than assigned CAMEL ratings should be greater for banks with relatively distant exam dates.

Our findings for the 1988–1992 period suggest that the information content of exam ratings can decay fairly quickly. For the two-year period from the second quarter of 1988 through the first quarter of 1990, the ability of exam ratings to discriminate between failing and surviving banks matches or exceeds that of our off-site monitoring system only when the ratings are no more than two quarters old. For the period beginning with second-quarter 1990 and ending in first-quarter 1992, the ability of exam ratings to identify failures matches the accuracy of our econometric forecasts only when the ratings are no more than one quarter old. These findings point to a large role for off-site early warning models in the pivotal supervisory task of identifying emerging bank failures.

2. An early warning model

Various off-site monitoring models have been developed by the regulatory agencies to complement the CAMEL rating system. Most have relied on call report data. The success of these systems in identifying emerging problems obviously hinges on the degree to which the banks report their financial results truthfully and accurately during the periods between on-site exams.

For this study, we develop a parsimonious econometric model of bank failure to serve as our early warning model. Reflecting the binary nature of our dependent variable, we use the probit model to estimate the relationship between a set of financial ratios measured at year-end 1985 for all 13,966 U.S. insured commercial banks and the likelihood of bank failure during the two-year period from the second quarter of 1986 through the first quarter of 1988.[6] We repeat this exercise using the financial ratios measured at year-end 1987 and bank failures during the period from second-quarter 1988 through first-quarter 1990.

In the probit model, we specify an unobservable index variable as a linear function of bank-specific characteristics and a disturbance term. Following standard practice, we assume positive values of the index variable are associated with failure, while nonpositive values are associated with survival. The resulting likelihood function is based on the standard normal cumulative distribution function. We maximize the likelihood function using the iteratively reweighted least squares algorithm.

We use seven financial variables, each measured as a percentage of gross assets (net assets plus reserves), to characterize the financial position of individual banks. Measures of capital adequacy, asset quality, earnings, and liquidity—four of the five components of the CAMEL rating system—are utilized.[7] These types of variables have been found significant in previous studies of bank failure (see, e.g., Sinkey, 1975; Bovenzi, Marino, and McFadden, 1983; Korobow and Stuhr, 1983; Gajewski, 1989; Demirguc-Kunt, 1989; Thomson, 1992; Cole and Gunther, 1995).

The expected effects on bank failure of capital adequacy, asset quality, and earnings are as follows. Total equity capital acts as a buffer protecting a bank's solvency against financial losses and is expected to reduce the probability of failure. Asset quality difficulties are measured by loans 90 days or more past due, nonaccrual loans, and other

real estate owned, which consists primarily of foreclosed real estate assets. A positive relationship is expected between asset quality problems and the probability of failure. High net income generally reflects a lack of financial difficulties and so is expected to reduce the likelihood of failure.

We use two measures of bank liquidity: investment securities and large certificates of deposit ($100,000 or more). On average, investment securities tend to be more liquid than loans and enable a bank to minimize fire-sale losses when responding quickly to unexpected demands for cash. In contrast, a reliance on purchased funds, such as large certificates of deposit, rather than core deposits, often reflects aggressive financial strategies and typically entails higher and more variable levels of interest expense. As a result, the probability of failure should depend negatively on investment securities and positively on large certificates of deposit.

As shown in table 1, the estimation results indicate that the variables included in the model are important indicators of bank survivability and that each affects the probability of failure in the expected fashion for both periods examined. Each variable is significant at the 1% level for both of the time periods we examine, and each consistently possesses the expected sign.

The coefficients in column 1 of table 1 were used in conjunction with year-end 1987 financial data to generate out-of-sample forecasts of the probability of failure for individual banks over the two-year period from the second quarter of 1988 through the first quarter of 1990. Similarly, the coefficients in column 2 of table 1 were used in conjunction

Table 1. Estimation results for the probit model of bank failure

Variable	1985 Data	1987 Data
Intercept	− 1.83	− 1.59
	(0.13)	(0.13)
Equity capital	− 9.55	− 11.49
	(1.32)	(1.43)
Past due loans	19.43	18.29
	(1.98)	(2.64)
Nonaccrual loans	12.15	10.78
	(1.38)	(1.61)
Other real estate owned	10.31	11.81
	(2.01)	(1.48)
Net income	− 7.84	− 6.96
	(1.37)	(1.60)
Investment securities	− 1.36	− 2.48
	(0.29)	(0.31)
Large certificates of deposit	2.34	2.50
	(0.25)	(0.27)
R^2	0.38	0.49

Notes: Standard errors are in parenthesis. $R^2 = 1 - L_\Omega/L_\omega$, where L_Ω is the value of the log-likelihood function maximized with respect to both the intercept and the explanatory variables, and L_ω is the value of the log-likelihood function maximized with respect to the intercept parameter alone. All variables are measured as a percentage of gross assets.

with year-end 1989 financial data to generate a second set of out-of-sample forecasts for the period from second quarter 1990 through first quarter 1992. A comparison of the predicted probabilities of failure for each of the two forecast periods with the failures that actually occurred establishes a sense of the early warning model's predictive accuracy.

3. The information content of CAMEL ratings

Regulators use the information obtained through on-site exams to rate the five components of bank performance included in the CAMEL rating system—capital adequacy, asset quality, management, earnings, and liquidity—on a scale of 1 to 5 as follows: 1, strong performance; 2, satisfactory performance; 3, performance that is flawed to some degree; 4, marginal performance that is significantly below average; and 5, unsatisfactory performance that is critically deficient and in need of immediate remedial action.

Once the five component ratings have been assigned, a composite, or overall rating is derived, again on a scale from 1 to 5. The five composite rating levels are described as follows in the *Commercial Bank Examination Manual* prepared by the Board of Governors of the Federal Reserve System: 1, an institution that is basically sound in every respect; 2, an institution that is fundamentally sound but has modest weaknesses; 3, an institution with financial, operational, or compliance weaknesses that give cause for supervisory concern; 4, an institution with serious financial weaknesses that could impair future viability; and 5, an institution with critical financial weaknesses that render the probability of failure extremely high in the near term.

The ability to predict bank failures provides a measure of the information content of CAMEL ratings and establishes a basis for evaluating the value of off-site monitoring systems as a supplement to the CAMEL rating system. Exams usually are conducted on an annual or less frequent basis, and the resulting CAMEL ratings are assigned only after the exams have been completed. As a result, the most recent ratings available on a given date may be based on exams completed a few days, a few months, or even a few years earlier. Given that financial conditions are subject to frequent and sometimes severe change, we expect the relative accuracy of a CAMEL rating in identifying failure to be a decreasing function of the length of time since the rating was assigned.

To test this hypothesis, we assess the relative accuracy of the CAMEL ratings on individual banks at year-end 1987 in predicting failures during the two-year period from the second quarter of 1988 through the first quarter of 1990.[8] Whereas 2,254 of the 9,880 examination ratings for year-end 1987 were based on financial data from the fourth quarter of that year, the majority of ratings available at year-end 1987 were based on financial data from the previous quarters. The CAMEL ratings based on exams conducted near the end of 1987 may be expected to provide a better indication of survivability than those based on earlier exams.

This exercise then is repeated for the CAMEL ratings at year-end 1989. That is, we assess the relative accuracy of the year-end 1989 ratings in predicting failures during the two-year period from the second quarter of 1990 through the first quarter of 1992. The ratings based on exams conducted near the end of 1989 may be expected to provide a better indication of survivability than those based on earlier exams.

3.1. Out-of-sample results using 1987 data

Although it is a simple matter to rank the banks based on the predicted probabilities of failure generated by the probit model, establishing a comparable ranking based on the CAMEL rating data is more difficult. To produce such a ranking, we first sort the 9,880 banks, for which we are able to obtain CAMEL rating data at year-end 1987, from worst to best based on their composite CAMEL ratings.[9] This sorting procedure produces five groups of banks with a declining likelihood of failure but does not identify differences in the likelihood of failure *within* each of the five composite rating groups. In an attempt to identify such differences for banks possessing the same composite rating, we use the separate component ratings assigned by examiners for each of the five performance categories included in the CAMEL rating system. More specifically, after we sort the banks from worst to best *composite* CAMEL ratings, we then sort the banks *within* each of the five composite ratings from worst to best, based on the equally weighted arithmetic average of their *component* CAMEL ratings.[10] Examiners typically do not view the average component rating as a basis for ranking banks within each composite rating class, but some such ranking procedure is necessary to form a comparison with the results of our off-site monitoring system. However, even this ranking procedure generates ties for banks with the same composite rating and the same average component rating. As a result, the banks are sorted randomly within groups possessing the same composite and component ratings. Using the resulting overall ranking as our guide, we expect the banks with the worst CAMEL ratings to be the most likely to fail.

Figure 1 shows the accuracy of both the probit model and the year-end 1987 CAMEL ratings in identifying failures during the subsequent two-year period (April 1988–March 1990). The vertical axis of figure 1 measures the proportion of failed banks identified as surviving, which we will refer to as the type-1 error rate, and the horizontal axis measures the proportion of surviving banks identified as failed, or the type-2 error rate. The bank failure data are collected from FDIC press releases. Of the 9,880 banks in our sample, 244 failed during the two-year period (April 1988–March 1990) examined.

Both the probit model and the CAMEL ratings are fairly accurate in identifying failures, as indicated by the high degree of curvature in the trade-off between the type-1 and type-2 error rates. However, for any given level of type-2 error, the probit model identifies failures more accurately. At a type-2 error rate of 10%, the probit model's estimates misclassify 9.8% of the bank failures, whereas the CAMEL ratings misclassify 22.5%. Based on the standard chi-square test for pair-matched samples, this difference is highly significant, with a test statistic of 20.4.

To provide an indication of how much better more recently assigned CAMEL ratings identify failing banks, we limit the sample to CAMEL ratings based on financial data from the fourth quarter of 1987. Of the 9,880 U.S. insured commercial banks used in this analysis, 2,254 had CAMEL ratings based on financial data from the fourth quarter. In figure 2, we plot the difference in type-1 error rates, for each level of type-2 error, between the ranking based on CAMEL ratings assigned as of the fourth quarter and the ranking based on the probit model forecasts. The difference in type-1 error rates for the banks with CAMEL ratings assigned as of the fourth quarter is represented by the thinnest solid line in figure 2. Positive values for the difference in type-1 error rates correspond to relatively

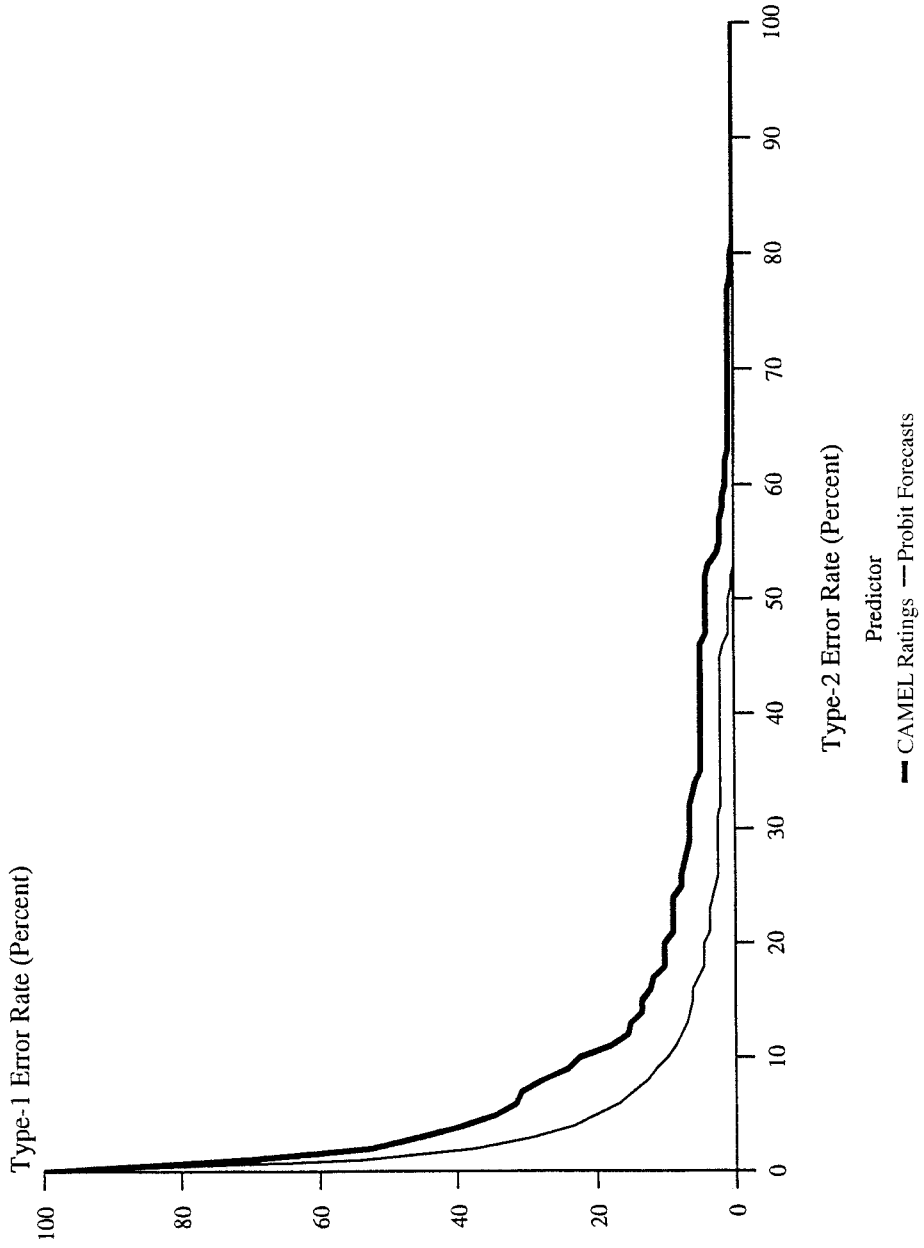

Figure 1. Errors in identifying bank failures using 1987 data.

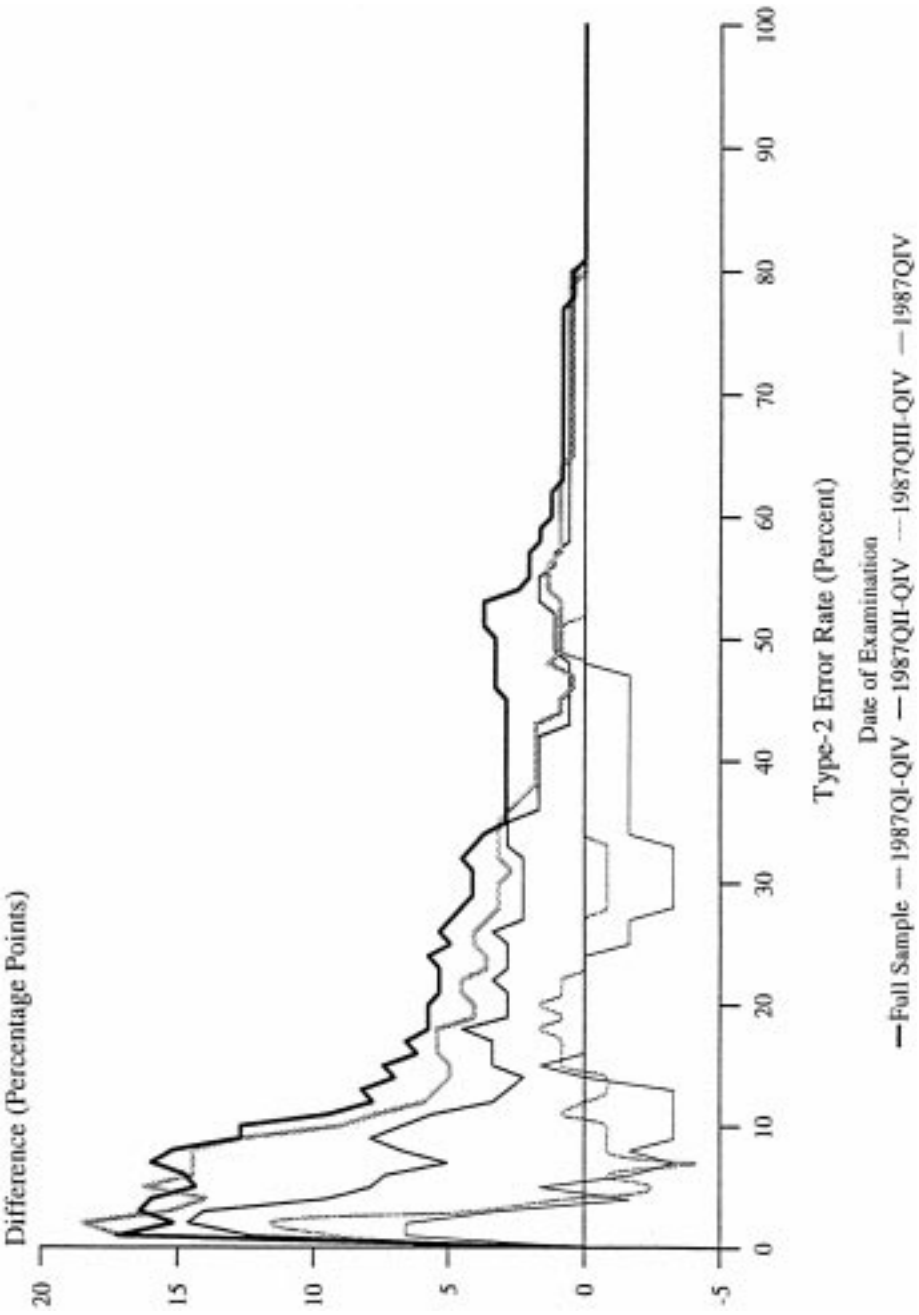

Figure 2. Difference between CAMEL rating and probit forecast type-1 error rates using 1987 data.

high levels of accuracy for the probit model, while the reverse is true for negative values. The probit model appears to be slightly more accurate at low levels of type-2 error. However, overall the two rankings identify failures equally well, as indicated by the tendency for the difference in type-1 error rates to fall in both positive and negative territory, while never straying far from zero.

This finding for banks examined in the fourth quarter contrasts with the results based on the full sample of banks. As shown by the thickest solid line in figure 2, the difference in type-1 error rates for the full sample of banks is consistently nonnegative and reaches a maximum of about 17 percentage points, reflecting the probit model's relatively high level of accuracy when applied to all banks, irrespective of exam date.

However, the results for recently examined banks at times may attribute a higher degree of accuracy to the off-site monitoring system than could be achieved in actual practice. Banks that are discovered to have underreported credit difficulties often are required to refile their most recent call report. As a result, the call report data for some of the banks we analyze may have been revised, so they may not reflect the information actually available to analysts during the historical period we study. These considerations suggest that, in practice, the accuracy of recently assigned CAMEL ratings in identifying failures may exceed the accuracy of off-site monitoring systems.

We now augment the sample of banks rated as of fourth quarter 1987 with banks rated as of the third quarter of that year. Of the 9,880 insured commercial banks used in this analysis, 4,529 had CAMEL ratings based on financial data from the third or fourth quarter. Once again, we sort these individual banks from worst to best, based on their composite CAMEL ratings and average component CAMEL ratings, with the expectation that those with the worst ratings would be the most likely to fail during the two-year period from second quarter 1988 through first quarter 1990. The on-site and off-site systems' levels of accuracy again are comparable, with little difference overall between the type-1 error rates, as shown by the thin dotted line in figure 2.

Substantial differences in accuracy occur only when we enlarge the sample of banks rated as of the third or fourth quarter of 1987 with banks rated as of the second quarter of that year. Of the 9,880 insured commercial banks used in this analysis, 6,358 had CAMEL ratings based on financial data from the second, third, or fourth quarter of 1987. When the banks with three-quarter-old CAMEL ratings are included in the analysis, the accuracy of the CAMEL ratings in identifying bank failures is appreciably less than that of the forecasts generated by the off-site monitoring system, as shown in figure 2. Hence, there is a substantial reduction in the relative ability of CAMEL ratings to identify bank failures when the sample of banks with exams conducted one and two quarters earlier is augmented with banks examined three quarters earlier.

The deterioration in the accuracy of CAMEL ratings continues when banks with four-quarter-old CAMEL ratings are included in the analysis. Of the 9,880 banks analyzed, 7,872 had CAMEL ratings based on financial data from the first through fourth quarters of 1987. For this broader sample of banks, the ratings from the off-site monitoring system are substantially more accurate in identifying bank failures than the CAMEL ratings. The reduction in accuracy attributable to relatively old CAMEL ratings causes the overall accuracy of CAMEL ratings to fall substantially below that of the off-site monitoring system.[11]

These results indicate that the applicable information content of CAMEL ratings can deteriorate rather quickly, pointing to the conclusion that off-site monitoring systems provide regulators with valuable information on bank survivability over and above the information generated by the on-site examination process.

3.2. Out-of-sample results using 1989 data

We now sort the 12,198 banks for which we are able to obtain year-end 1989 CAMEL rating data from worst to best, based on their composite CAMEL ratings.[12] We again sort these banks within each of the five possible composite ratings from worst to best, based on the arithmetic average of their five component CAMEL ratings.

Figure 3 shows the accuracy of the probit model and year-end 1989 CAMEL ratings in identifying failures during the subsequent two-year period of interest (April 1990–March 1992). Consistent with our results based on the 1987 data, both the probit forecasts and the CAMEL ratings again are fairly accurate in identifying failures, with the probit model producing more accurate results. For a type-2 error rate of 10%, the probit model's estimates misclassify 7.9% of the bank failures, whereas the CAMEL ratings misclassify 20.3%. Based on the standard chi-square test for pair-matched samples, this difference is highly significant, with a test statistic of 23.4.

Of the 12,198 insured commercial banks used in this analysis, 2,794 had CAMEL ratings based on financial data from fourth quarter 1989. In figure 4, we plot the difference in type-1 error rates, for each level of type-2 error, between the ranking based on CAMEL ratings assigned as of the fourth quarter and the ranking based on the probit model forecasts. Consistent with our results for the 1987 data, neither ranking dominates for all levels of type-2 error.

Next, we enlarge the sample of banks rated as of fourth quarter 1989 with banks rated as of the third quarter of that year. Of the 12,198 banks analyzed, 5,808 had CAMEL ratings based on financial data from the third or fourth quarter. In contrast to our results based on the 1987 data, the probit model forecasts now are noticeably more accurate for most levels of type-2 error than the ranking based on CAMEL ratings. That is, for the 1989 data, a substantial reduction in the relative ability of CAMEL ratings to identify bank failures occurs when the sample including only exams conducted one quarter earlier is enlarged to include exams conducted two quarters earlier.

Consistent with our previous results, this difference in accuracy becomes greater when banks with a most recent exam dated three quarters before the forecast period also are included in the analysis. Of the 12,198 banks analyzed, 8,536 had CAMEL ratings based on financial data from the second, third, or fourth quarter of 1989. When the banks with three-quarter-old CAMEL ratings are included in the analysis, the accuracy of the CAMEL ratings in identifying bank failures is appreciably less than that of the forecasts generated by the off-site monitoring system, as shown in figure 4. And the same is true when banks with four-quarter-old CAMEL ratings are included. Of the 12,198 insured commercial banks analyzed, 10,404 had CAMEL ratings based on financial data from the first through fourth quarters of 1989.

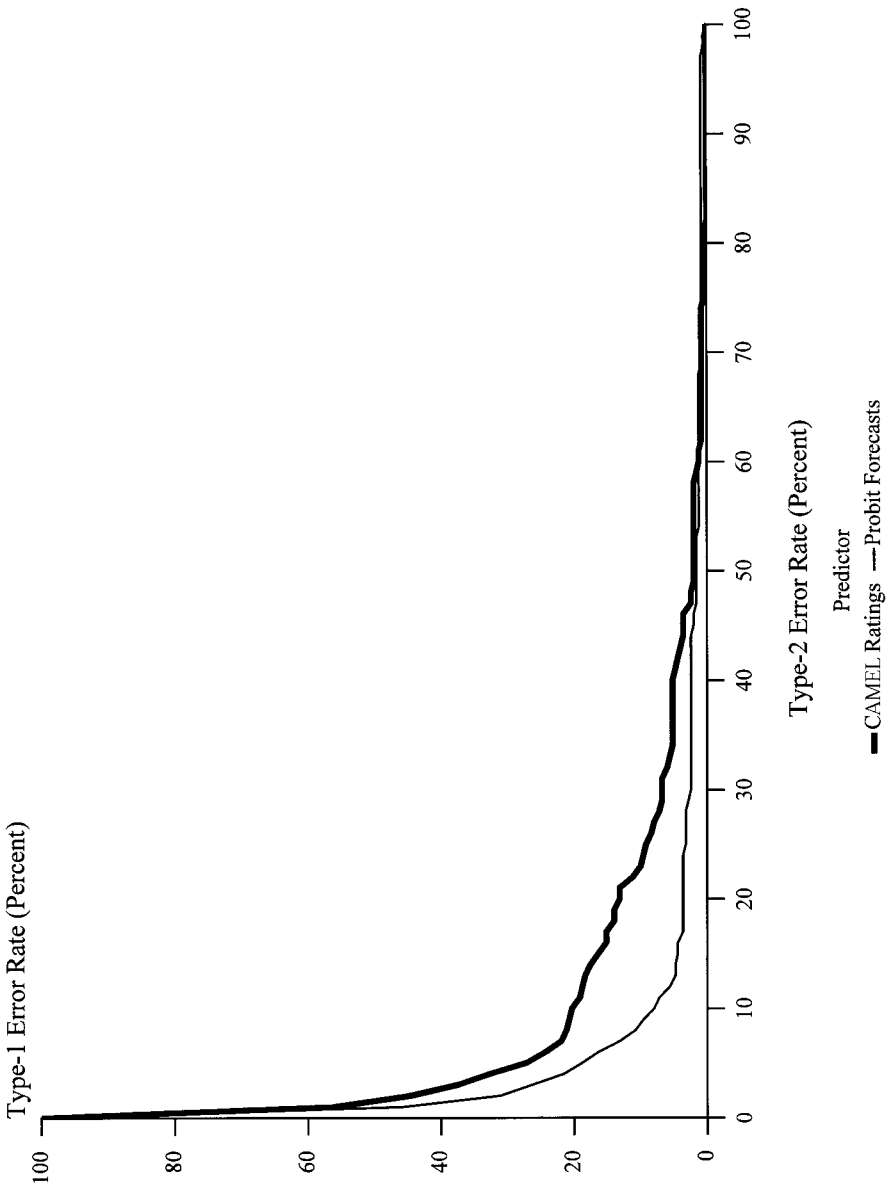

Figure 3. Errors in identifying bank failures using 1989 data.

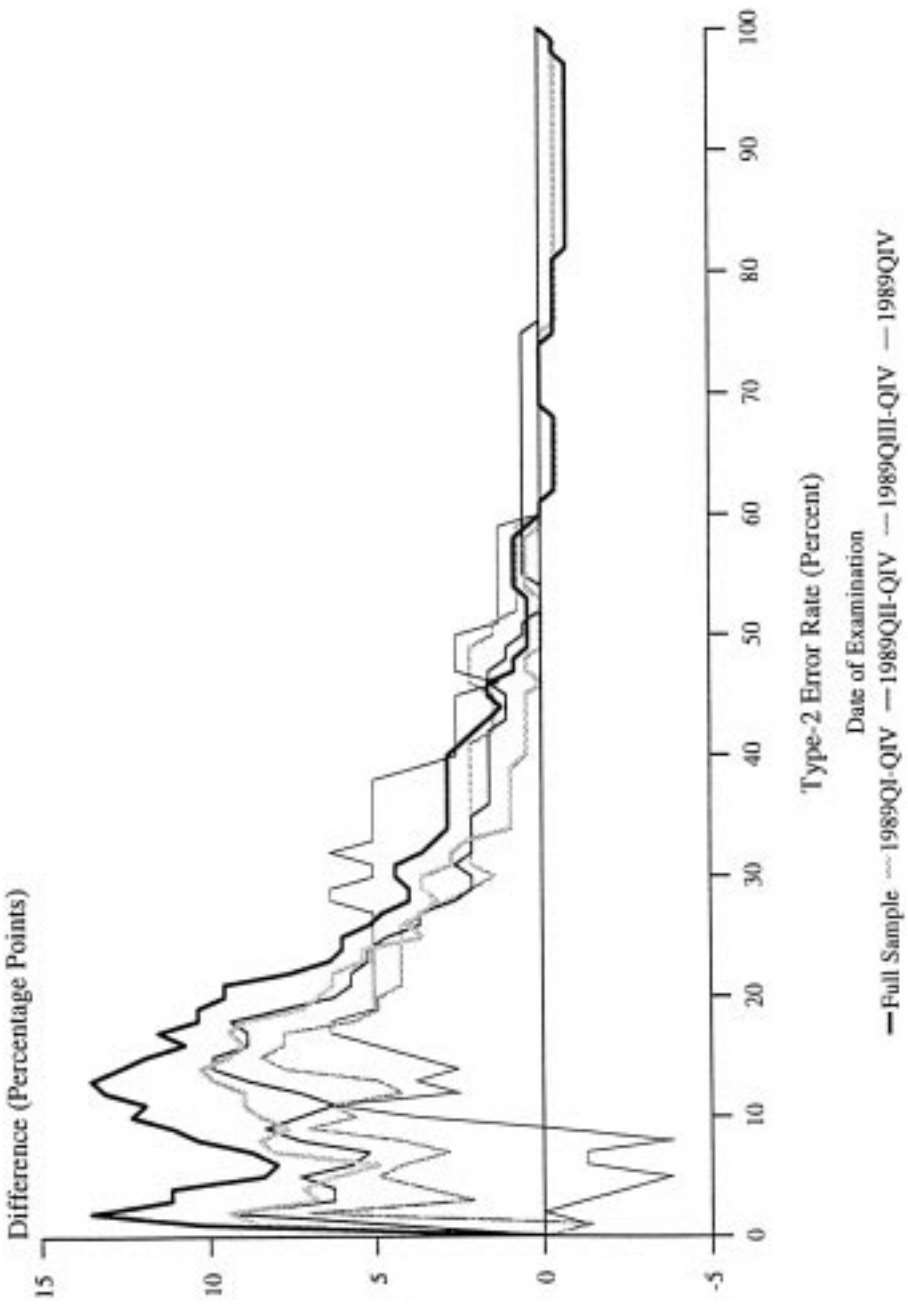

Figure 4. Difference between CAMEL rating and probit forecast type-1 error rates using 1989 data.

4. Conclusion

The findings reported here suggest that the information content of the CAMEL ratings derived from on-site bank examinations can decay fairly rapidly. For the two periods we examine, second quarter 1988 through first quarter 1990 and second quarter 1990 through first quarter 1992, the ability of CAMEL ratings to identify bank failures matches or exceeds that of our off-site econometric model only when the ratings are based on exams conducted no more than one or two quarters prior to the forecast period. For the banks with ratings more than one or two quarters old, the probit model provides a more accurate indication of survivability.

Even though these results highlight the usefulness of even relatively simple off-site monitoring systems as a complement to on-site exams, the findings of this study should not be construed as detracting from the critical dependence of a successful banking supervision program on the examination process. The examination process and the CAMEL ratings it generates have numerous important uses, many of which are quite distinct from the relatively narrow task of identifying bank failures. Moreover, off-site monitoring systems, such as the one used here, depend on the integrity of accounting data, which is enhanced through periodic exams.

It should also be noted that, by limiting our focus to 1988–1992, a relatively compact period of rapid changes and financial difficulties in the banking sector, the analysis here may not be representative of more ''normal'' periods. The early warning model we use is based on estimates of the relationship between financial variables and bank failures. However, from the mid-1940s until the early 1980s, no more than 20 bank failures occurred in any one year. Similarly, in 1993 and 1994, there were only 42 and 13 bank failures, respectively. The paucity of recent failures effectively precludes efforts to update the model. Whether off-site monitoring systems, based on the failure experience of the 1980s and early 1990s, are capable of accurately identifying any relatively far removed failures that might emerge in the future is an open question.

Acknowledgments

We are especially grateful to the editor, George Kaufman, for numerous comments and suggestions. Our work also has benefited from comments by Mark Flannery, Alton Gilbert, Michael Gordy, and an anonymous referee. The views expressed are those of the authors and should not be attributed to the Federal Reserve Bank of Dallas, the Board of Governors of the Federal Reserve System, or the Employment Policies Institute.

Notes

1. The CAMEL rating system was revised on January 1 1997, to include a sixth component. The new *S* component focuses on the sensitivity of individual banks to market risk, such as the risk arising from changes in interest rates. This study focuses on the original CAMEL rating system because that was the rating system used during the period investigated. Because the original CAMEL rating system implicitly

considered factors related to market risk, the introduction of the *S* component in the revised rating system, CAMELS, is not expected to result in significant changes to composite ratings.

2. Although generally requiring annual on-site exams, FDICIA permits banks that are small, well-capitalized, and highly rated to be examined only once every 18 months. Depositories classified by regulators as ''problem banks''—those with composite CAMEL ratings of 4 or 5—typically are examined twice per year.
3. The Fed uses FIMS not only to track the financial condition of individual banks and banking organizations between on-site exams but also to direct examination resources. An overview of FIMS is provided by Cole, Cornyn, and Gunther (1995).
4. A detailed overview of the rating system, as revised on January 1 1997, is found on pages 67021–29 of the *Federal Register*, Vol. 61, No. 245, December 19 1996.
5. It is important to note that regulatory closure typically occurs at some point after a depository institution becomes insolvent on a market-value, or even book-value, basis. For example, Cole (1993) reports that the average thrift closed during the 1980s had been insolvent for 18 months at the time of closure. Even though bank regulators did not grant capital forbearance to such an extreme degree as thrift regulators, they often did not close banks at the time of insolvency, as evidenced by the billions of dollars in losses incurred by the FDIC bank insurance fund. For the purposes of this study, the lag between insolvency and regulatory closure may impart a bias against the accuracy of CAMEL ratings, if examiners rate the likelihood of insolvency rather than the likelihood of closure.
6. Failures are identified starting in the second quarter of 1986, rather than the first quarter, to impose a one-quarter lag in the estimated relationship. Edited call report data generally are not available until 45–70 days after the end of each quarter. Failures occurring in the first quarter are excluded from the analysis.
7. It is not difficult to envision numerous additional variables as potential candidates for inclusion in our bank failure model. However, we feel that the relatively parsimonious model we employ is well suited for this study. To the extent that the inclusion of additional variables could improve the model's accuracy, our comparison of on- and off-site monitoring systems understates the value of early warning models.
8. Three primary dates typically are associated with an exam: the start date, the end date, and the ''as of'' date. The ''as of'' date derives its name from the fact that it is the date for the financial data on which the CAMEL rating is based. We use the ''as of'' date to match CAMEL ratings with the forecasts from our off-site monitoring system, which also are dated based on the date of the financial data used. As a result, the date of a CAMEL rating does not imply that the rating actually was available at that time, because a lag naturally would occur between the date of the financial data used and the date on which the CAMEL rating actually was assigned. The dates given in the text for the CAMEL ratings should be interpreted with this processing lag in mind. The one-quarter lag between the date of the CAMEL ratings and the start of our forecast period accommodates the expected processing lag.
9. The number of banks included in our analysis is limited by our access to historical CAMEL rating data. Of the 13,365 U.S. insured commercial banks possessing a call report for year-end 1987 and also meeting the other requirements of our study, we were able to obtain year-end 1987 CAMEL ratings for 9,880, or 74%. Of the 9,880 banks, 9,740 were rated based on a ''full scope'' exam, another 134 had ratings associated with ''limited scope'' exams, and the remaining 6 were the subject of ''targeted'' exams, which focus exclusively on specific areas of supervisory concern. The results reported here are qualitatively identical when the analysis is limited to ''full scope'' exams.
10. Although the equal treatment of the five component ratings is somewhat arbitrary, we also used several alternative schemes to weight the five component ratings. The results are not qualitatively different when alternative schemes are used.
11. For example, looking separately at the 2,008 banks with CAMEL ratings based on financial data from 1986 or earlier, the 10% with the worst CAMEL ratings includes only 59% of the subsequent failures, whereas the 10% with the highest probabilities of failure, as generated by the probit model, includes 95% of the subsequent failures. Similarly large differences in accuracy occur for banks examined in the first and second quarters of 1987. The CAMEL ratings' level of accuracy in identifying failures tends to decline as the time since the last exam increases, whereas, interestingly, the probit model's level of accuracy is fairly uniform across the groupings based on exam date.
12. Of the 12,442 U.S. insured commercial banks possessing a call report for year-end 1989 and also meeting the other requirements of our study, we were able to obtain year-end 1989 CAMEL ratings for 12,198, or

98%. Of these 12,198 banks, 251 failed during the two-year period examined. Also, of the 12,198 banks, 7,912 were rated based on a ''full scope'' exam. The results reported here are qualitatively identical when the analysis is limited to ''full scope'' exams.

References

Berger, Allen N., and Sally M. Davies, ''The Information Content of Bank Examinations,'' Finance and Economics Discussion Series No. 94–20, Board of Governors of the Federal Reserve System, 1994.

Berger, Allen N., Kathleen Kuester King, and James M. O'Brien, ''The Limitations of Market Value Accounting and a More Realistic Alternative,'' *Journal of Banking and Finance* 15 (September 1991), 753–783.

Bovenzi, John F., James A. Marino, and Frank E. McFadden, ''Commercial Bank Failure Prediction Models,'' *Economic Review*, Federal Reserve Bank of Atlanta (November 1983), 14–26.

Cargill, Thomas F., ''CAMEL Ratings and the CD Market,'' *Journal of Financial Services Research* 3 (December 1989), 347–358.

Cole, Rebel A., ''When Are Thrift Institutions Closed? An Agency-Theoretic Model,'' *Journal of Financial Services Research* 7 (December 1993), 283–307.

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

Cole, Rebel A., and Jeffery W. Gunther, ''Separating the Likelihood and Timing of Bank Failure,'' *Journal of Banking and Finance* 19 (September 1995), 1073–1089.

Demirguc-Kunt, Asli, ''Modeling Large Commercial-Bank Failures: A Simultaneous-Equations Approach,'' Working Paper 8905, Federal Reserve Bank of Cleveland, 1989.

Federal Deposit Insurance Corporation, *History of the Eighties—Lessons for the Future* (December 1997).

Flannery, Mark J., and Joel F. Houston, ''Market Responses to Federal Examinations of U.S. Bank Holding Companies,'' mimeo, University of Florida, 1994.

Gajewski, Gregory R., ''Assessing the Risk of Bank Failure,'' *Proceedings of a Conference on Bank Structure and Competition*, Federal Reserve Bank of Chicago (May 1989), pp. 432–456.

Gilbert, R. Alton, and Sangkyun Park, ''The Value of Early Warning Models in Bank Supervision,'' mimeo, Federal Reserve Bank of St. Louis, 1994.

Hirschhorn, Eric, ''The Information Content of Bank Examination Ratings,'' Federal Deposit Insurance Corporation *Banking and Economic Review* (July–August 1987), 6–11.

Jones, David S., and Kathleen Kuester King, ''The Implementation of Prompt Corrective Action: An Assessment,'' *Journal of Banking and Finance* 19 (June 1995), 491–510.

Korobow, Leon, and David P. Stuhr, ''The Relevance of Peer Groups in Early Warning Analysis,'' *Economic Review*, Federal Reserve Bank of Atlanta (November 1983), 27–34.

Sinkey, Joseph F., Jr., ''A Multivariate Statistical Analysis of the Characteristics of Problem Banks, *Journal of Finance* 30 (March 1975), 21–36.

Thomson, James B., ''Modeling the Regulator's Closure Option: A Two-Step Logit Regression Approach,'' *Journal of Financial Services Research* 6 (May 1992), 5–23.