WORKING PAPER SERIES

The Contribution of On-site Examination Ratings to an Empirical Model of Bank Failures

David C. Wheelock
Paul Wilson

Working Paper 1999-023A
http://research.stlouisfed.org/wp/1999/99-023.pdf

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

The Contribution of On-site Examination Ratings to an Empirical Model of Bank Failures

David C. Wheelock

Research Department
Federal Reserve Bank of St. Louis
P. O. Box 442
St. Louis, MO 63166

David.C.Wheelock@stls.frb.org

and

Paul W. Wilson*

Department of Economics
University of Texas
Austin, TX 78712

wilson@eco.utexas.edu

November 1999

*The views expressed in this paper do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. We thank Joan Cronin, Alton Gilbert, Andrew Meyer, and Mark Vaughan for comments and assistance with acquisition of the data used in this paper, and Heidi Beyer for research assistance. Research for this paper was conducted while Wilson was a visiting scholar at the Federal Reserve Bank of St. Louis.

ABSTRACT

This paper investigates how well regulator examinations predict bank failures, and how best to incorporate examination information into an econometric model of time-to-failure. We estimate proportional hazard models with time-varying covariates and find that examiner ratings help explain the failure hazard. Both the overall rating of a bank's condition and management, *i.e.,* the composite CAMELS rating, and ratings of specific components contain information. In addition, we find that the marginal "effect" of ratings is non-linear, in that the impact of a rating downgrade on the probability of failure is larger, the weaker a bank's initial rating.

1. INTRODUCTION

In the wake of post-war record numbers of bank and thrift failures, Congress enacted the Federal Deposit Insurance Corporation Improvement Act of 1991, which mandated annual safety and soundness examinations of nearly all U.S. commercial banks. Such examinations yield ratings assigned on a scale from 1 (strong) to 5 (critically deficient) for six categories—capital adequacy (C), asset quality (A), management quality (M), earnings (E), liquidity (L), and sensitivity to market risk (S).[1] A composite rating, reflecting the examiner's overall assessment of the bank's condition, also is assigned. Gilbert (1993) and Cole and Gunther (1998) find that supervisory examinations of banks can identify accurately "problem" banks, *i.e.,* those banks at greatest risk of failing, if done frequently.

Examinations are, however, costly to perform. Regulators have thus long relied on statistical models or financial ratio "screens" to identify banks at risk of failure so as to better devote scarce supervisory resources to problem banks (see Gilbert *et al.,* 1999). This approach relies largely on information banks file quarterly about their financial condition (*i.e.,* Reports of Condition and Income, or "call reports"). Such information is public, and numerous studies have found that call report information is useful for identifying the characteristics that make banks more likely to fail (*e.g.,* Cole and Gunther, 1998; Gilbert *et al.,* 1999; Wheelock and Wilson, 1999). Indeed, Cole and Gunther (1998) find that a simple econometric model using call report data predicts failures better than do examiner ratings that are more than six months old. With the important caveat that on-site examinations might enhance the accuracy of bank-reported information, Cole and Gunther's results suggest that on-site examinations might yield little information about the likelihood of failure beyond that contained in a bank's call report.[2]

Although call reports are updated frequently, the information they contain might not

[1]Sensitivity to market risk has been evaluated only since 1997.

[2]The purpose of on-site examinations is to identify problem banks, not to predict failures per se. Problem banks receive extra scrutiny and may be subject to prompt corrective action designed to prevent their failure. Thus, a finding that examinations do not predict failures in a statistical sense would not necessarily indicate that they fail to identify problem banks.

fully or accurately reflect a bank's true condition. Balance sheet information is reported at book, rather than market, values, and reported information about loans and other aspects of a bank's operations is not fully comprehensive. On-site examinations, by contrast, have the potential to provide a more detailed picture of a bank's condition. Their relative infrequency, however, implies that a bank's condition might change markedly since its most recent examination.

Differences in timeliness and possible information content suggest that statistical models of the probability of bank failure could be enhanced by incorporating both call report and examination information in them. This paper investigates that possibility. Combining examination and call report information in a statistical model is, however, not straightforward. First, for most banks, the interval between exams exceeds one quarter. Thus, observations on exam variables are of a different vintage than those derived from call reports. Second, call and examination reports contain a myriad of information that might be incorporated into the statistical model.

To address the first issue, we estimate a proportional hazards model of time-to-failure that allows for time-varying covariates. Observations on each variable are updated as they occur, which need not be at the same frequency. Thus, the model can accommodate quarterly updating of variables derived from the call reports with the less frequent updating of exam ratings. Moreover, the model explicitly accounts for the possibility that the age of information about a bank's condition can affect how well that information explains the probability of failure.

To address the second issue, we begin by specifying a standard, reasonably comprehensive model of bank failure based on call report information. We then investigate whether adding examination information enhances the model's ability to explain the probability of failure. We investigate whether the various exam component ratings add information not reflected in the composite rating, and we examine whether the marginal effect of an exam rating differs by the specific rating assigned. And, finally, we control for possible differences in the assignment (and reporting) of ratings by the various federal and state regulatory agencies.

We find that examiner ratings contribute statistically significant explanatory power to an empirical model of time-to-failure, even though most banks were examined no more than once a year during the period covered by our study. Our research finds that at least two component ratings contain information about the likelihood of failure beyond that reflected in the composite examination rating. Moreover, we find that the "effect" of CAMEL ratings on the probability of failure is non-linear, with some evidence that the likelihood of failure is more than proportionately larger for a lower-rated bank than for a higher-rated bank with otherwise identical attributes. This result could reflect non-linearity in the assignment of ratings by examiners or the higher frequency of examinations for low-rated banks. Finally, we find that exam ratings alone do not explain failure as well as an empirical model that includes both exam ratings and call report information. Statistical models best explain bank failure when they incorporate both call report and examination information.

The next section discusses the possible information content of regulatory examinations of banks. It also describes our data, econometric model of time-to-failure, and how we incorporate examination ratings into the model. The following sections describe our estimation methodology and present results and conclusions.

2. BANK EXAMINATIONS

Most U.S. commercial banks are examined on-site at least annually by the relevant federal regulatory agency. Safety and soundness composite exam ratings are assigned on a scale from 1 to 5, where a rating of "1" reflects a bank "that is basically sound in every respect," a "2 " indicates a bank "that is fundamentally sound, but with modest weaknesses," a "3" indicates a bank "with financial, operational, or compliance weaknesses that give cause for supervisory concern," a "4" indicates "an institution with serious financial weaknesses that could impair future viability," and a "5" indicates that a bank has "critical financial weaknesses that render the probability of failure extremely high in the near term." In addition, ratings on a scale from 1 to 5 are also assigned for the six CAMELS components, where a rating of "1" indicates "strong performance" in that category, a "2" indicates "satisfactory performance," a "3" indicates "performance that is flawed to some degree," a "4" indicates "marginal performance that is significantly below average," and a "5" indicates "unsatisfactory performance that is critically deficient and in need of immediate remedial action."[3]

Numerous studies have found evidence that regulatory examinations obtain unique information about a bank's condition. Hirschhorn (1987), for example, finds that bank stock prices reflect examiner ratings almost immediately, suggesting alternatively that the market and examiners discover information about banks at the same time, or that examination results are inferred by market participants through predictable actions, such as changes in loan loss reserves, of examined banks.[4]

Berger and Davies (1994) find that examination rating downgrades are followed by declines in market values (a symmetric effect for upgrades was not detected, suggesting that market participants learn of good news about a bank's condition prior to an examination). Flannery and Houston (1999) find that bank holding company valuations can be affected by the occurrence of examination, although the specific effect can vary with general economic conditions and other factors. For a sample of holding companies in 1988, they found that the occurrence of examination increased market values, suggesting that the market interpreted an examination as certifying the quality of a bank's accounting information. In 1990, however, they found a negative (though statistically insignificant) effect. Because 1990 was a poor year for banks generally, and because regulators tend to examine problem banks more frequently than healthy banks, the occurrence of an examination in 1990 might

[3] See Cole *et al.* (1995) for additional discussion.

[4] Examination reports, including CAMELS ratings, are reported to senior bank management, but are otherwise confidential.

have been interpreted by the market as evidence that a bank's condition was weaker than revealed by public information.

By contrast, Berger *et al.* (1998) find only weak evidence that bank holding company examination ratings Granger-cause market valuations. They find, however, that changes in examination ratings Granger-cause changes in private market ratings of bank holding company debentures (bond ratings also Granger-cause examination ratings, suggesting that regulators and bond rating agencies uncover different information that turns out to be useful to the other).[5]

That safety and soundness examinations appear to obtain information that affects market assessments of bank condition suggests that exam ratings might help explain the probability of failure. We thus obtained CAMEL composite and component ratings for safety and soundness examinations conducted on all U.S. commercial banks since 1984 for use in a time-to-failure model.[6]

3. MODEL SPECIFICATION

We model the failure hazard for banks using the Cox proportional hazards model which we estimate by the method of partial likelihood (Cox, 1972, 1975). This approach offers several advantages: (1) it is easy to incorporate time-varying covariates; (2) the model is semi-parametric in the sense that we do not need to specify the distribution function; and (3) the partial likelihood approach avoids specification of the baseline hazard function, which may be viewed as capturing unobserved heterogeneity among banks failing at different times. Details are given in the Appendix.

[5] Several other studies have found that examiner ratings reflect aspects of a bank's condition. Swindle (1995) finds that the assignment of a low (3–5) rating by examiners is followed by an increase in bank capital, suggesting that regulators impose discipline on problem banks. Jones and King (1995), on the other hand, find that examination ratings reflect examiner assessments of banks' risk-weighted capital. DeYoung (1998) and Siems and Barr (1998) find that examination ratings are also correlated with a bank's productive efficiency—more efficient banks tend to have higher management and overall examination ratings. Finally, Whalen and Thomson (1988), Cole *et al.* (1995), and Gilbert *et al.* (1999) examine the extent to which examination ratings can be explained by a bank's financial information.

[6] Because ratings for sensitivity to market risk (S) were not given before 1997, they are excluded from our analysis.

Numerous studies have used call report information to estimate bank failure probability models. Most select independent variables to match the capital adequacy, asset quality, management, earnings, liquidity and market risk characteristics of banks that on-site examinations seek to discern (recent examples include Cole *et al.* (1995), Cole and Gunther (1998), and Wheelock and Wilson (1999)). Here we use data from the quarterly call reports to specify variables similar to those used in Wheelock and Wilson (1999):

- Capital adequacy:
  CAPAD = total equity/total assets.
- Asset quality:
  L2A = total loans/total assets.
  RELOAN = real estate loans/total loans.
  CILOAN = commercial and industrial loans/total loans.
  OTHRE = other real estate owned/total assets.
  UNCOL = income earned, but not collected on loans/total assets.
- Earnings:
  EARN = net income after taxes/total assets.
- Liquidity:
  LIQ = (fed funds sold − federal funds purchased)/total assets.
- Miscellaneous factors:
  SIZE = log of total assets.
  HOLD = 1 if 25% or more of equity is held by a multi-bank holding company; 0 otherwise.
  BR1 = 1 if bank is located in a state allowing limited branching; 0 otherwise.
  BR2 = 1 if bank is located in a state allowing unlimited branching; 0 otherwise.[7]

We expect that the lower a bank's capital/assets ratio is, the higher the bank's probability of failure. An observed capital/assets ratio of zero or less does not necessarily mean that a bank has "failed," however, because capital and assets are reported as book values in the call reports. In addition, failure date is defined here as the date of closure by regulatory authorities. Nevertheless, we expect to find a strong association between book value capital and the likelihood of failure.

[7] For a summary of state branching laws, see the Annual Reports of the FDIC.

The ratios of other real estate owned to total assets, and of earned, but uncollected income to total assets, are measures of asset quality. Other real estate owned reflects foreclosed property. High values of either ratio would indicate problem loans; hence we expect positive coefficients on each variable in the failure model. The ratios of commercial and industrial loans to total loans, and of real estate loans to total loans, are measures of portfolio concentration. In the 1980s and early 1990s, numerous banks suffered loan losses in declining real estate markets. If this was general, we would expect a higher concentration in real estate loans would have increased the likelihood of failure. Finally, high earnings and liquidity are expected to reduce the likelihood of failure.

We include bank size, a dummy variable indicating whether or not a bank was a member of a holding company, and a dummy variable reflecting the branch banking laws of the state in which the bank operated, as additional control variables. Size and the ability to branch might enable banks to diversify, and thereby lower risk, while a holding company might inject capital into weak subsidiaries to keep them from failing.

Our primary objective is to investigate whether examination ratings contribute information to the empirical model not otherwise reflected in the financial and environmental variables above. Further, we investigate whether the exam component ratings contain information not otherwise captured by the composite CAMEL rating. In addition, we test whether the marginal impact of exam ratings differs by rating, *e.g.,* whether the probability of failure increases more when a bank's rating is reduced from a "4" to a "5" than it is when the rating is reduced from a "1" to a "2." Thus, we estimate several specifications of the model and use likelihood ratio tests to compare the explanatory power of alternative specifications.

To allow for nonlinear effects with respect to the five possible values of the composite CAMEL rating, we define:

CAMEL2 — equals 1 if CAMEL = 2, 0 otherwise;
CAMEL3 — equals 1 if CAMEL = 3, 0 otherwise;
CAMEL4 — equals 1 if CAMEL = 4, 0 otherwise;

CAMEL5 — equals 1 if CAMEL = 5, 0 otherwise.

We define similar dummy variables corresponding to the possible values of the component ratings for capital adequacy (C), asset quality (A), management (M), earnings (E), and liquidity (L):

C2 — equals 1 if C = 2, 0 otherwise;
C3 — equals 1 if C = 3, 0 otherwise;
C4 — equals 1 if C = 4, 0 otherwise;
C5 — equals 1 if C = 5, 0 otherwise;

A2 — equals 1 if A = 2, 0 otherwise;
A3 — equals 1 if A = 3, 0 otherwise;
A4 — equals 1 if A = 4, 0 otherwise;
A5 — equals 1 if A = 5, 0 otherwise;

M2 — equals 1 if M = 2, 0 otherwise;
M3 — equals 1 if M = 3, 0 otherwise;
M4 — equals 1 if M = 4, 0 otherwise;
M5 — equals 1 if M = 5, 0 otherwise;

E2 — equals 1 if E = 2, 0 otherwise;
E3 — equals 1 if E = 3, 0 otherwise;
E4 — equals 1 if E = 4, 0 otherwise;
E5 — equals 1 if E = 5, 0 otherwise;

L2 — equals 1 if L = 2, 0 otherwise;
L3 — equals 1 if L = 3, 0 otherwise;
L4 — equals 1 if L = 4, 0 otherwise;
L5 — equals 1 if L = 5, 0 otherwise.

Finally, we control for the examining agency. Federally-chartered (national) banks are examined by the Office of the Comptroller of the Currency (OCC). State-chartered banks are examined by state authorities and by the Federal Reserve bank (FRB) of the district in which they are located if they are members of the Federal Reserve System, or by the Federal Deposit Insurance Corporation (FDIC) if they are not Federal Reserve members. Although the various examining agencies follow the same general guidelines for examining banks, by including dummy variables for the examining agency, we can control for idiosyncrasies across agencies in either the conduct or reporting of examination information.[8] We define the following additional dummy variables:

$$\text{FRB} = \begin{cases} 1 & \text{if exam is conducted by the Federal Reserve System,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\text{OCC} = \begin{cases} 1 & \text{if exam is conducted by the Office of the Comptroller of the Currency,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\text{STATE} = \begin{cases} 1 & \text{if exam is conducted by state bank regulators;} \\ 0 & \text{otherwise;} \end{cases}$$

$$\text{JOINT} = \begin{cases} 1 & \text{if exam is conducted by two or more agencies,} \\ 0 & \text{otherwise.} \end{cases}$$

For purposes of interpreting coefficients on these dummy variables, the reference group consists of exams conducted by the Federal Deposit Insurance Corporation.

4. ESTIMATION RESULTS

Our sample includes all banks that filed a call report for the first quarter 1987 and were chartered before June 30, 1984. Omitting newly-chartered banks avoids idiosyncracies in the operations and examinations of new banks that could cause the relationship between their financial statements, exam ratings, and likelihood of failure to differ from those of established banks. For estimation, we define time beginning at the date of the first quarter 1987 call report, which corresponds to the calendar date March 31, 1987; the time scale is the same for all banks. Each variable is assumed to remain constant until a new value is observed. Thus, variables derived from call reports are typically updated at regular quarterly intervals.[9] Other variables, such as examination ratings, are updated at

[8] Our examination information come from an internal Federal Reserve database. Records for the other agencies, particularly in the early years of our sample, are known to be incomplete.

[9] Call report variables are occasionally missing. We replace missing values with the most recently observed value. However, we drop from the sample banks with three consecutive missing call reports, and treat them as censored on the date of the third missing report. Banks that are acquired or merge are censored at the date of those events.

irregular and less frequent intervals when they occur. A bank's exam rating, for example, is assumed to remain in effect and unchanged until the bank's next examination.[10]

Specification Tests:

Table 1 summarizes the 17 models we estimated. Models #1–14 each include the financial ratios derived from the call reports, bank size, holding company membership and the branch banking dummy variables. Model #1 includes no additional variables. For models #2–14, Table 1 gives the number of additional parameters, relative to model #1, as well as the optimized value of the log-(partial) likelihood function for each specification. Model #1 is nested within each of models #2–14, and thus provides a baseline comparison for examining the information content of the bank exam variables.

Among models #1–14, there are 55 possible pairwise comparisons where likelihood-ratio tests can be used to test restrictions in one model nested within a more general model. These various pairwise tests and their outcomes are summarized in Table 2. For example, model #1 is nested within model #2—model #2 expands the basic model (#1) by adding the CAMEL composite rating. Table 2 reveals that the null hypothesis that the coefficient on CAMEL in model #2 equals zero is rejected at better than .01. Similarly, model #3 expands the basic model (#1) by adding the component variables C, A, M, E, and L. If the coefficients on C, A, M, E, and L in model #3 are zero, then model #3 reduces to model #1. Table 2 indicates that a likelihood ratio test rejects the null hypothesis that the coefficients on the component rankings are simultaneously equal to zero. Model #4 includes both the CAMEL composite and the component variables in a single model, and thus nests models #2–3 as well as model #1. Likelihood ratio tests reject the restrictions implied by models #1–3. Inclusion of examination ratings in the hazard model enhances

[10] Our treatment of the time-varying covariates means that each variable is defined as the level of a step function in continuous time. Although in principle the variables may actually vary continuously over time, they are observed only at discrete intervals. An alternative approach would be to interpolate values for each variable between their observation times, but this would substantially increase the cost of computation. Moreover, interpolation involves additional assumptions that may not be warranted. Our treatment of the data is consistent with the literature on time-varying covariates in hazard models.

the model's explanatory power. And, the component ratings appear to contain information not otherwise reflected by the composite rating.

Model #5 is similar to model #2, except that in model #5, the dummy variables CAMEL2, ..., CAMEL5 replace the single variable CAMEL. Letting $\beta_2, \ldots, \beta_5$ denote the coefficients on CAMEL2, ..., CAMEL5, respectively, if $\beta_2 = \frac{2}{3}\beta_3 = \frac{1}{2}\beta_4 = \frac{2}{5}\beta_5$, then model #5 reduces to model #2.[11] Table 2 indicates that this null hypothesis is rejected at better than .01. Model #6 is similar to model #3, except that in model #6 the single variable C is replaced by the dummy variables C2, ..., C5; A is replaced by A2, ..., A5; *etc.* Imposing restrictions on the coefficients of these variables similar to those on CAMEL2, ..., CAMEL5 discussed above reduces model #6 to model #3. A likelihood ratio test however rejects these restrictions at better than .01. Model #7 combines CAMEL2, ..., CAMEL5, C2, ..., C5, ..., L2, ..., L5 in a single model and thus nests models #1–6. Table 2 reveals that likelihood ratio tests reject restrictions implied by each model #1–6 at better than .01. These results indicate that the ability of the empirical model to explain the failure hazard rate is enhanced by allowing the marginal impact of exam ratings to vary by specific rating. The evidence thus suggests that the effect of a rating change is nonlinear.

Models #8–13 are identical to models #2–7 respectively, except that models #8–13 include the four dummy variables FRB, OCC, STATE, and JOINT to capture possible effects of different examining agencies. Table 2 reveals that likelihood ratio tests reject the null hypotheses that coefficients on these four agency dummies are simulataneously equal to zero in each model #8–13. Model #13 nests models #1–12, and likelihood ratio tests reject the restrictions implied by each of these models.

Model #14 generalizes model #13 by adding interaction terms involving the products of the four agency dummy variables and the 24 dummy variables describing the bank exam ratings that were used in model #13; thus, model #14 has 124 parameters more than the

[11] Letting β_0 represent the coefficient on CAMEL in model #2, then models #2 and #5 are equivalent if $\beta_2 = 2\beta_0$, $\beta_3 = 3\beta_0$, $\beta_4 = 4\beta_0$, and $\beta_5 = 5\beta_0$. These conditions imply $\beta_0 = \frac{1}{2}\beta_2 = \frac{1}{3}\beta_3 = \frac{1}{4}\beta_4 = \frac{1}{5}\beta_5$, or $\beta_2 = \frac{2}{3}\beta_3 = \frac{1}{2}\beta_4 = \frac{2}{5}\beta_5$ in model #5.

basic specification in model #1, and 96 more parameters than in model #13. Although model #14 is the most general of our models, the null hypothesis that coefficients on all the interaction terms are equal to zero cannot be rejected by a likelihood-ratio test; indeed, the likelihood-ratio statistic (minus twice the difference of the log-partial-likelihoods) is 82.855, and with 96 degrees of freedom, this results in a p-value for the test of 0.8234. Thus, the 96 interaction terms add very little information to the hazard function. Similar likelihood-ratio tests using model #14 also fail to reject the restrictions implied by models #7 and #12, which were rejected by model #13. Moreover, while the restrictions implied by models #4, #6, #9, and #10 are rejected by model #14, they are rejected with a larger p-value than was the case with model #13. This is suggestive of overfitting in model #14.

The bottom panel in Table 1 shows the log-partial likelihoods obtained for three additional models, labeled #15–17. Model #15 includes only the four environmental variables: SIZE, HOLD, BR1, and BR2. Likelihood-ratio tests indicate that this specification is easily rejected by every other specification that we estimated. This model's strong rejection by model #1 gives an indication of the importance of the financial ratios from the call report data in explaining bank failure hazards. Model #16 resembles model #10, in that each includes the CAMEL composite and component ratings, and four agency dummies. In model #16, however, the eight financial ratios have been omitted. The likelihood-ratio statistic for a test of model #16 versus model #10 is 516.08, which easily rejects the restriction in model #16. This again indicates the importance of the financial ratios for explaining bank failures. Similarly, model #17 resembles model #13, except that once again the financial ratios have been omitted in model #17. The likelihood-ratio statistic for a test of model #17 versus model #13 is 567.4476, and so the restriction in model #17 is strongly rejected, again illustrating the importance of the financial ratios in explaining bank failure hazards.

Parameter Estimates:

The likelihood-ratio tests indicate that model #13, which includes the financial ratios derived from call reports, the environmental variables, the CAMEL composite and component rankings entered as dummy variables, and the agency dummy variables is the superior specification. Estimation results for model #13 are shown in Table 3. The coefficient on the ratio of equity to assets (CAPAD) is negative and highly significant, which should be no surprise since banks with low net worth have less cushion against losses. In addition, two asset quality variables—OTHRE and UNCOL—have positive, significant coefficients. Both variables reflect the effects of banks' having made risky loans in the past, and the postive signs on these coefficients indicate that bad loan portfolios increase the hazard of failure.

The ratios of real estate loans to total assets (RELOAN), commercial and industrial loans to total assets (CILOAN), and total loans to assets (L2A) are all insignificant. We thus find no evidence that loan portfolio concentration affects the likelihood of failure.

Higher earnings reduce the risk of bank failure, as indicated by the negative sign on EARN in Table 3. High liquidity (LIQ) appears to increase the likelihood of failure, though this may reflect the narrow definition of liquidity that missing data forced us to use. Data on cash and short-term Treasury security holdings are not available for large numbers of banks in some quarters. We elected to use net federal funds sold as a measure of liquidity, rather than limiting our sample to just those banks with no missing data.

The significant, negative coefficient on SIZE suggests that larger banks are less prone to failure, perhaps because they are generally better diversified or, possibly, because regulators devote more resources to avoiding the failure of larger banks. The positive estimate for the coefficient on the multi-bank holding company dummy (HOLD) suggests that members of holdings companies are more likely to fail. Although holding companies might inject capital into weak subsidiaries, in the 1980s the insolvency of holding companies sometimes resulted in the closure (and, hence, declaration of failure) of all subsidiary banks, insolvent or not. This effect could be driving our result.

Of the two branching variables, only BR2—reflecting unlimited statewide branching—

has a significant coefficient. Its positive sign indicates that banks located in states with unlimited branching are more likely to fail than similar banks in unit banking states. Wheelock and Wilson (1999), by contrast, obtain a negative coefficient for this variable—indicating that the ability to branch reduces the failure hazard. Wheelock and Wilson (1999), however, use a sample that excluded banks with less than $50 million of assets (almost half of all banks during the period studied). While branching enables geographic diversification, the positive sign on BR2 obtained here alongside the result in Wheelock and Wilson (1999) suggests that the liberalization of branching laws in the 1980s and early 1990s increased competitive pressures on small banks with few branches.

Of the disaggregated exam variables in model #13, the only statistically significant coefficients are those on three of four liquidity variables (L3, L4, and L5). These coefficients are positive, and their magnitudes increase moving from L3 to L4 to L5. This indicates that banks receiving increasingly critical (higher numeric) scores on the liquidity component of bank exams are more likely to fail. Although most of the coefficients on the examination variables in model #13 are statistically not significant, the likelihood ratio tests reported in Table 2 show that these variables do add information about bank failures. The available data, however, do not contain enough information to obtain precise estimates of the 24 individual coefficients on these variables.

For purposes of comparison, Table 4 reports estimation results for model #10. This model resembles model #13, except that model #10 implicitly assumes linear effects for the five possible composite and component ratings. In model #10, the coefficients on the composite rating (CAMEL), the management (M) component rating, and the liquidity (L) component rating are positive and statistically significant. These indicate that more critical (higher numeric) ratings produce an increase in the failure hazard. In addition, the significant coefficients on the management and liquidity ratings suggest that these exam components are not proxied well by the financial condition variables derived from the call reports.

Tables 5 and 6, which report coefficient estimates for models #16 and #17, illustrate further how well examination ratings explain bank failure. These models omit all financial variables derived from the call reports. In model #16, exam composite and component ratings enter the model linearly. Here the CAMEL composite rating, and the capital and liquidity component ratings have positive and statistically significant coefficients. Model #17 includes dummy variables for each exam composite and component rating, thereby allowing specific rating scores to have distinct marginal impacts on the likelihood of failure. Here again, the coefficients on the CAMEL composite and the capital and liquidity component ratings are most frequently statistically significant. Of course, in terms of overall explanatory power, these models are inferior to ones that include both financial data and examination ratings.

5. CONCLUSIONS

Commercial bank regulators use both on-site examinations and statistical models based on firm-reported financial data to identify banks at risk of failure. Although expensive to conduct, on-site examinations conceivably can probe deeply into a bank's condition to reveal information not contained in the quarterly call reports. Previous research, however, suggests that the information content of on-site examinations decays fairly rapidly.

In this paper we investigate whether the ratings of banks generated by on-site examinations contribute information to an empirical model of bank failure based otherwise on self-reported financial information. We do so using a proportional-hazards model with time-varying covariates. This model accounts explicitly for the time between the observation of an independent variable and the event of failure, and permits the use of independent variables that are observed at different times or frequencies. We find that examination ratings can discriminate between banks that fail or survive. When included as covariates in an econometric model, examination ratings contribute information about the probability of failure not otherwise captured by variables derived from the call reports. By themselves, however, examination ratings explain the probability of failure less well than when financial variables derived from the call reports are also included in the model. We find, moreover, that the explanatory power of the model is enhanced by allowing examination ratings to have non-linear effects on the probability of failure. Our empirical estimates indicate that the weaker a bank's initial rating, the more the probability of failing increases when the bank's rating is lowered.

We have not conducted an exhaustive search for call report variables that might duplicate fully the information contained in examination ratings; the sheer volume of information in the call reports makes such a search impossible. Our results suggest, however, that information from on-site examinations can enhance the explanatory power of a parsimoniously specified empirical model of bank failure, and thus could prove useful to regulators and other researchers interested in predicting failures.

APPENDIX: THE PROPORTIONAL HAZARDS MODEL

We write the hazard function for failure by the ith bank as $\lambda_i(t \mid \boldsymbol{x}_i(t), \boldsymbol{\beta})$; the hazard varies with time t, and is conditioned on covariates contained in the vector $\boldsymbol{x}_i(t)$ which vary over time and a corresponding vector of parameters, $\boldsymbol{\beta}$.

For each bank i in the sample, we observe a vector of covariates $\boldsymbol{x}_i(t)$ at $(J_i - 1)$ distinct times $t_{i1} < t_{i2} < \ldots < t_{i,J_i-1}$. In addition, at time $t_{iJ_i} \geq t_{i,J_i-1}$ we either observe failure, or the observation on the ith bank is censored at time t_{iJ_i}. Time is measured by calendar time since the first observation date, which is identical for all banks in the sample, so that $t_{i1} = 0 \ \forall \ i$. The data used in $\boldsymbol{x}_i(t)$ corresponding to time t_{ij}, $j = 1, \ldots, (J_i - 1)$, are assumed to reflect measurable characteristics of bank i over the interval $[t_{ij}, t_{i,j+1})$ for $j = 1, \ldots, J_i - 1$. At time t_{iJ_i} we assume that these characteristics are represented by $\boldsymbol{x}_i(t_{J_i-1})$. Hence $\boldsymbol{x}_i(t_{ij} + \Delta t) = \boldsymbol{x}_i(t_{ij}) \ \forall \ 0 \leq \Delta t < t_{i,j+1}$, $j = 1, \ldots, J_i - 1$. Our model is time-varying in the sense that covariates may vary across intervals, although they are assumed constant within intervals of time $[t_{ij}, t_{i,j+1})$.

We estimate the hazard of failure using the partial likelihood method suggested by Cox (1972, 1975). Define the indicator variable

$$d_i = \begin{cases} 1 & \text{if bank } i \text{ fails at time } t_{J_i}; \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

Then the contribution by the ith bank to the partial likelihood is

$$L_i(\boldsymbol{\beta}) = \frac{\lambda_i(t_{J_i} \mid \boldsymbol{x}_i(t_{J_i}), \boldsymbol{\beta})^{d_i}}{\sum\limits_{m \in R_i} \lambda_m(t_{J_i} \mid \boldsymbol{x}(t_{J_i}), \boldsymbol{\beta})}, \tag{A.2}$$

where $R_i = \{m | t_{J_m} \geq t_{J_i}, \ m = 1, \ldots, N\}$ is the risk set associated with bank i, *i.e.,* the set of banks that did not fail before bank i failed. Censored banks do not enter the numerator of any contribution to the partial likelihood, although they enter the denominator for contributions by banks that fail before the date of censoring.

We specify the proportional hazards as

$$\lambda_i(t \mid \boldsymbol{x}_i(t), \boldsymbol{\beta}) = \overline{\lambda}(t) \exp(\boldsymbol{x}_i(t)\boldsymbol{\beta}), \tag{A.3}$$

where $\overline{\lambda}(t)$ is the baseline hazard. Substituting this into (A.2), taking logs, and summing across individuals yields the partial log-likelihood for the entire sample:

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left\{ d_i \boldsymbol{x}_i(t_{J_i})\boldsymbol{\beta} - \log \left[\sum_{m \in R_i} \exp(\boldsymbol{x}_m(t_{J_i})\boldsymbol{\beta}) \right] \right\}. \qquad (A.4)$$

The baseline hazard drops out when (A.3) is substituted into (A.2); the model is semi-parametric in the sense that the parameter vector $\boldsymbol{\beta}$ is estimated without specifying the baseline hazard. In principal, although the baseline hazard $\overline{\lambda}(t)$ varies only over t and not over i, it is evaluated at different times t for different banks i and therefore may be viewed as capturing individual heterogeneity among banks disappearing at different times.

The advantage of the partial likelihood estimation approach in our case lies in the fact that only part of the hazard function need be specified; we do not have to specify the baseline hazard, nor do we need to define the density or survivor functions. For purposes of hypothesis testing, the partial likelihood (A.4) may be regarded as an ordinary log-likelihood concentrated with respect tot the baseline hazard; see Andersen and Gill (1982) and Johansen (1983). Efron (1977) discusses the efficiency loss from using the partial likelihood as opposed to the full information likelihood.

REFERENCES

Andersen, P.K., and R.D. Gill, "Cox's Regression Model for Counting Processes," *Annals of Statistics* 10 (1982), 1100–1120.

Berger, Allen N. and Sally M. Davies. "The Information Content of Bank Examinations," working paper, May 1994.

Berger, Allan N., Davies, Sally M., and Mark J. Flannery. "Comparing Market and Supervisory Assessments of Bank Performance: Who Knows What When?" working paper, March 1998.

Cole, Rebal A., and Gunther, Jeffery, W. "FIMS: A New Monitoring System for Banking Institutions," *Federal Reserve Bulletin,* January 1995, pp. 1-15.

Cole, Rebal A., and Gunther, Jeffery, W. "Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems," *Journal of Financial Services Research* 13 (no. 2), 1998, pp. 103-17.

Cox, D.R., "Regression Models and Life-Tables (with discussion)," *Journal of the Royal Statistical Society* 34B (1972), 187–220.

Cox, D.R., "Partial Likelihood," *Biometrika* 62 (1975), 269–276.

DeYoung, Robert. "Management Quality and X-Inefficiency in National Banks," *Journal of Financial Services Research* 13 (February 1998), pp. 5-22.

Efron, B., "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association* 72 (1977), 557–565.

Flannery, Mark J. and Joel F. Houston. "The Value of a Government Monitor for U.S. Banking Firms," *Journal of Money, Credit, and Banking* 31 (no. 1) , February 1999, pp. 14-34.

Hirschhorn, Eric. "The Informational Content of Bank Examination Ratings," *FDIC Banking and Economic Review* 5 (no. 4), July/August 1987, pp. 6-11.

Johansen, S., "An Extension of Cox's Regression Model," *International Statistical Review* 51 (1983), 165–174.

Gilbert, R. Alton. "Implications of Annual Examinations for the Bank Insurance Fund," *Federal Reserve Bank of St. Louis Review* 75 (no. 1), January/February 1993, pp. 35-52.

Gilbert, R. Alton, Meyer, Andrew P. and Mark D. Vaughan. "Alternative Approaches to Off-Site Surveillance: A Comparison of Supervisory Screens and Econometric Models as Tools for Allocating Examination Resources," working paper, Federal Reserve Bank of St. Louis, 1999.

Jones, David S. and Kathleen Kuester King. "The Implementation of Prompt Corrective Action: An Assessment," *Journal of Banking and Finance* 19 (no. 3-4), June 1995, pp. 491-510.

Siems, Thomas F. and Richard S. Barr. "Benchmarking the Productive Efficiency of U.S. Banks," *Federal Reserve Bank of Dallas Financial Industry Studies,* December 1998, pp. 11-24.

Swindle, C. Sloan. "Using CAMEL Ratings to Evaluate Regulator Effectiveness at Commercial Banks," *Journal of Financial Services Research* 9, 1995, pp. 123-41.

Whalen, Gary and James B. Thomson. "Using Financial Data to Identify Changes in Bank Condition," *Federal Reserve Bank of Cleveland Economic Review* 24 (no. 2), Second Quarter 1988, pp. 17-26.

Wheelock, David C. and Paul W. Wilson. "Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions," *Review of Economics and Statistics,* forthcoming, 1999.

TABLE 1

Model Specifications

Model	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	Additional Parameters	LLF
1	X	X							0	−3388.2139
2	X	X	X						1	−3303.5866
3	X	X		X					5	−3249.0527
4	X	X	X	X					6	−3243.4481
5	X	X			X				4	−3291.0709
6	X	X				X			20	−3230.9167
7	X	X			X	X			24	−3218.6902
8	X	X	X				X		5	−3297.4191
9	X	X		X			X		9	−3241.8107
10	X	X	X	X			X		10	−3235.8311
11	X	X			X		X		8	−3284.4527
12	X	X				X	X		24	−3222.6509
13	X	X			X	X	X		28	−3210.1499
14	X	X			X	X	X	X	124	−3168.7224
15	X								—	−5322.0670
16	X		X	X			X		—	−3528.9352
17	X				X	X	X		—	−3493.8735

NOTE: Xs in columns (a)–(h) indicate the variables included in each specification:

(a) environmental variables (SIZE, HOLD, BR1, BR2);
(b) financial ratios (CAPAD, L2A, RELOAN, CILOAN, OTHRE, UNCOL, EARN, LIQ);
(c) CAMEL;
(d) C, A, M, E, L;
(e) CAMEL2, ..., CAMEL5;
(f) C2, ..., C5; A2, ..., A5; ...; L2, ..., L5;
(g) four agency dummies;
(h) four agency dummies interacted with CAMEL2, ..., CAMEL5; C2, ..., C5; A2, ..., A5; ...; L2, ..., L5.

In addition, for models #1–#14, the column labeled “Parameters” gives the number of additional parameters contained in each model, relative to the basic specification in Model #1, while the column labeled “LLF” gives the optimized value of the log-likelihood function for each model.

TABLE 2

Likelihood Ratio Tests, Models #1–#14

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	.	.	.	.	.	.	.	.	.	.	.	.	.
2	x	.	.	.	.	.	.	.	.	.	.	.	.
3	x	.	.	.	.	.	.	.	.	.	.	.	.
4	x	x	x	.	.	.	.	.	.	.	.	.	.
5	x	x	.	.	.	.	.	.	.	.	.	.	.
6	x	.	x	.	.	.	.	.	.	.	.	.	.
7	x	x	x	x	x	x	.	.	.	.	.	.	.
8	x	a	.	.	.	.	.	.	.	.	.	.	.
9	x	.	x	.	.	.	.	.	.	.	.	.	.
10	x	x	x	.	.	.	.	x	x	.	.	.	.
11	x	x	.	.	a	.	.	x	.	.	.	.	.
12	x	.	x	.	.	x	.	.	x	.	.	.	.
13	x	x	x	x	x	x	x	x	x	x	x	x	.
14	x	x	x	a	x	b	o	x	a	b	x	o	o

NOTE: A character other than "." in row j, column k indicates that model k is nested within model j. An "x" indicates that a likelihood ratio test rejects model k in favor of model j at .01 or better; an "a" indicates that a likelhood ratio test rejects model k in favor of model j at .05 but not at .01. A "b" indicates model k is rejected at .1 but not at .05, and an "o" indicates that model k cannot be rejected in favor of model j at the .1 level.

TABLE 3

Coefficient Estimates for Model #13

partial log-likelihood: -3210.1499

Variable	Coefficient	Std. Err.	z	p − value	95%Conf. Int.	
CAPAD	−48.83	4.2319	−11.539	0.000	−57.124	−40.536
L2A	0.42861	0.43325	0.989	0.323	−0.42054	1.2778
RELOAN	−0.45333	0.32567	−1.392	0.164	−1.0916	0.18498
OTHRE	6.3820	1.0336	6.174	0.000	−1.0916	0.18498
UNCOL	25.703	6.6682	3.855	0.000	12.633	38.772
CILOAN	0.28488	0.37571	0.758	0.448	−0.45150	1.0213
EARN	−3.9112	0.43951	−8.899	0.000	−4.7726	−3.0498
LIQ	2.2534	0.61723	3.651	0.000	1.0437	3.4632
SIZE	−0.18275	0.05326	−3.431	0.001	−0.28713	−0.07836
HOLD	0.20725	0.10315	2.009	0.045	0.0050864	0.40941
BR1	0.015083	0.10781	0.140	0.889	−0.19621	0.22638
BR2	0.23812	0.14416	1.652	0.099	−0.044424	0.52066
CAMEL2	−1.8697	1.1694	−1.599	0.110	−4.1617	0.42229
CAMEL3	0.34844	1.2399	0.281	0.779	−2.0816	2.77852
CAMEL4	0.93798	1.2955	0.724	0.469	−1.6012	3.47716
CAMEL5	1.6076	1.3217	1.216	0.224	−0.98281	4.19805
C2	0.39715	0.99742	0.398	0.690	−1.5578	2.35207
C3	1.3856	1.0562	1.312	0.190	−0.68445	3.45568
C4	0.87801	1.0810	0.812	0.417	−1.2408	2.99678
C5	1.0007	1.0967	0.912	0.362	−1.1487	3.15017
A2	0.69829	0.73692	0.948	0.343	−0.74605	2.14263
A3	0.44031	0.73855	0.596	0.551	−1.0072	1.88785
A4	0.10857	0.75029	0.145	0.885	−1.3620	1.57911
A5	0.14641	0.75784	0.193	0.847	−1.3389	1.63175
M2	−0.10758	0.72787	−0.148	0.883	−1.5342	1.31903
M3	−0.36948	0.78526	−0.471	0.638	−1.9086	1.16960
M4	−0.47863	0.78572	−0.609	0.542	−2.0186	1.06136
M5	−0.07453	0.78617	−0.095	0.924	−1.6154	1.46633
E2	0.038873	0.75020	0.052	0.959	−1.4315	1.50925
E3	−0.82040	0.80120	−1.024	0.306	−2.3907	0.74993
E4	−0.80820	0.79111	−1.022	0.307	−2.3587	0.74236
E5	−0.83203	0.79505	−1.047	0.295	−2.3903	0.72623
L2	0.18000	0.35920	0.501	0.616	−0.52402	0.88401
L3	0.96761	0.38897	2.488	0.013	0.20525	1.7300
L4	1.3539	0.40374	3.353	0.001	0.56257	2.1452
L5	1.9710	0.41280	4.775	0.000	1.1619	2.7801
FRB	−0.23489	0.15996	−1.468	0.142	−0.54840	0.07863
OCC	−0.39489	0.09628	−4.101	0.000	−0.58359	−0.20618
STATE	−0.41410	0.14461	−2.864	0.004	−0.69753	−0.13067
JOINT	−0.39632	0.17806	−2.226	0.026	−0.74531	−0.04734

TABLE 4

Coefficient Estimates for Model #10

partial log-likelihood: −3235.8311

Variable	Coefficient	Std. Err.	z	p − value	95%Conf. Int.	
CAPAD	−48.009	4.1319	−11.619	0.000	−56.107	−39.910
L2A	0.31680	0.42675	0.742	0.458	−0.51961	1.1532
RELOAN	−0.50057	0.33226	−1.507	0.132	−1.1518	0.15066
OTHRE	6.4528	1.0650	6.059	0.000	4.3655	8.5402
UNCOL	25.226	6.7844	3.718	0.000	11.929	38.523
CILOAN	0.30872	0.37456	0.824	0.410	−0.42540	1.0429
EARN	−3.9798	0.43594	−9.129	0.000	−4.8342	−3.1253
LIQ	2.3961	0.58775	4.077	0.000	1.2441	3.5481
SIZE	−0.18012	0.05285	−3.408	0.001	−0.28371	−0.076546
HOLD	0.21192	0.10256	2.066	0.039	0.010902	0.41293
BR1	0.026822	0.10728	0.250	0.803	−0.18345	0.23709
BR2	0.26088	0.14365	1.816	0.069	−0.020676	0.54243
CAMEL	0.62791	0.18080	3.473	0.001	0.27356	0.98226
C	0.11652	0.14135	0.824	0.410	−0.16051	0.39356
A	−0.08669	0.10380	−0.835	0.404	−0.29012	0.11675
M	0.14757	0.06511	2.266	0.023	0.019956	0.27517
E	−0.15408	0.10041	−1.535	0.125	−0.35088	0.04272
L	0.56544	0.06687	8.456	0.000	0.43438	0.69649
FRB	−0.19478	0.16029	−1.215	0.224	−0.50895	0.11939
OCC	−0.35768	0.09419	−3.797	0.000	−0.54230	−0.17307
STATE	−0.39571	0.14337	−2.760	0.006	−0.67670	−0.11472
JOINT	−0.41895	0.18262	−2.294	0.022	−0.77687	−0.06103

TABLE 5
Coefficient Estimates for Model #16
partial log-likelihood: -3528.9352

Variable	Coefficient	Std. Err.	z	p − value	95%Conf. Int.	
SIZE	−0.19257	0.04412	−4.364	0.000	−0.27905	−0.10609
HOLD	0.23420	0.09795	2.391	0.017	0.04221	0.42619
BR1	0.09365	0.10460	0.895	0.371	−0.11137	0.29867
BR2	0.10810	0.14154	0.764	0.445	−0.16931	0.38550
CAMEL	0.96663	0.22715	4.256	0.000	0.52143	1.4118
C	0.64571	0.16222	3.980	0.000	0.32776	0.96366
A	0.07796	0.10905	0.715	0.475	−0.13578	0.29170
M	0.02778	0.06167	0.451	0.652	−0.09308	0.14865
E	−0.03537	0.09122	−0.388	0.698	−0.21415	0.14341
L	0.81688	0.06648	12.287	0.000	0.68658	0.94719
FRB	−0.15784	0.17048	−0.926	0.355	−0.49197	0.17628
OCC	0.00415	0.08938	0.046	0.963	−0.17104	0.17934
STATE	−0.28591	0.15214	−1.879	0.060	−0.58410	0.01228
JOINT	−0.24188	0.17948	−1.348	0.178	−0.59365	0.10988

TABLE 6

Coefficient Estimates for Model #17

partial log-likelihood: −3493.8735

Variable	Coefficient	Std. Err.	z	p − value	95%Conf. Int.	
SIZE	−0.18382	0.04442	−4.138	0.000	−0.27088	−0.09676
HOLD	0.20046	0.09856	2.034	0.042	0.00729	0.39364
BR1	0.07038	0.10525	0.669	0.504	−0.13590	0.27667
BR2	0.08666	0.14454	0.600	0.549	−0.19663	0.36996
CAMEL2	−1.76590	1.1518	−1.533	0.125	−4.02342	0.49162
CAMEL3	0.70142	1.2312	0.570	0.569	−1.7116	3.1144
CAMEL4	1.42145	1.3219	1.075	0.282	−1.1695	4.0124
CAMEL5	2.76330	1.3717	2.015	0.044	0.07485	5.4517
C2	1.40751	0.99743	1.411	0.158	−0.54741	3.3624
C3	2.74635	1.0768	2.551	0.011	0.63590	4.8568
C4	2.94427	1.1132	2.645	0.008	0.76245	5.1269
C5	3.34625	1.1385	2.939	0.003	1.1149	5.5776
A2	0.70048	0.76054	0.921	0.357	−0.79015	2.1911
A3	0.35710	0.77696	0.460	0.646	−1.1657	1.8799
A4	0.18803	0.80350	0.234	0.815	−1.3868	1.7629
A5	0.59163	0.81206	0.729	0.466	−0.99998	2.1832
M2	−0.32204	0.70749	−0.455	0.649	−1.7087	1.0646
M3	−0.78257	0.75774	−1.033	0.302	−2.2677	0.70257
M4	−1.09873	0.75677	−1.452	0.147	−2.5820	0.38451
M5	−0.74669	0.75614	−0.988	0.323	−2.2287	0.73532
E2	0.09842	0.76313	0.129	0.897	−1.3973	1.5941
E3	−0.71083	0.81632	−0.871	0.384	−2.3108	0.88913
E4	−0.62190	0.80504	−0.773	0.440	−2.1998	0.95593
E5	−0.32423	0.80304	−0.404	0.686	−1.8982	1.2497
L2	0.27714	0.38095	0.727	0.467	−0.46952	1.0238
L3	1.24307	0.40888	3.040	0.002	0.44168	2.0445
L4	1.77659	0.42401	4.190	0.000	0.94555	2.6076
L5	2.67743	0.42916	6.239	0.000	1.8363	3.5186
FRB	−0.21544	0.16801	−1.282	0.200	−0.54474	0.11386
OCC	−0.07239	0.09220	−0.785	0.432	−0.25309	0.10831
STATE	−0.34280	0.15328	−2.236	0.025	−0.64321	−0.04238
JOINT	−0.21785	0.17624	−1.236	0.216	−0.56327	0.12758