Early warning models for bank supervision: Simpler could be better

Julapa Jagtiani, James Kolari, Catharine Lemieux, and Hwan Shin

Introduction and summary

Capital adequacy has long been central to the regulatory oversight of banking systems around the world. Capital is crucial to bank safety and soundness, because it represents the cushion available to financial institutions to withstand unanticipated losses. By monitoring capital levels, supervisors seek to predict which financial institutions are most likely to be at risk if subjected to an earnings or asset quality shock. In 1991, the passage of the Federal Deposit Insurance Corporation Improvement Act (FDICIA) emphasized capital levels as a key benchmark to use in determining appropriate supervisory interventions to aid ailing financial institutions. Intervention when an institution is beginning to experience problems may allow it to avoid failure.

Over the past two decades, various off-site monitoring systems have been created to identify developing financial problems at banking institutions between on-site examinations. Supervisors use the output from these monitoring or early warning system (EWS) models to determine which organizations need increased supervisory scrutiny, identify specific areas of concern, accelerate on-site examinations of institutions showing financial deterioration, and allocate more experienced or more specialized examiners to institutions with financial problems.

The current models used by bank regulators focus on predicting either CAMELS[1] downgrade or bank failure.[2] In this article, we develop EWS models that focus on identifying banks that will have inadequate capital in the following year. Specifically, our models predict banks with an early stage of capital distress, with a primary capital to assets ratio falling below the 5.5 percent minimum capital adequacy standard (the relevant capital standard for this period). Earlier identification of capital inadequacy would enable supervisors to identify firms at risk and manage timely supervisory interventions.

Based on samples of banks in the late 1980s and early 1990s, we test our EWS models empirically using financial and economic data for individual banks. We chose 1988–90 as the sample period, rather than a more recent period, in order to have a sufficient number of problem banks in the sample. Also, since most troubled banks are those with assets of less than $1 billion, we focus on these banks in our study. We also exclude banks with less than $300 million in assets.

Our objective is to develop a model that predicts one of two states—capital adequate versus capital inadequate—where the latter state represents capital levels that fall below a minimum threshold employed by bank supervisors during this period. Although we use a well-recognized regulatory threshold for adequate capital, our main objective is to examine an early stage of financial distress, rather than regulatory compliance with capital standards.[3]

Our choice of the capital to asset ratio as a plausible proxy for the early onset of financial distress is supported by previous research. Estrella, Park, and Peristiani (2000) examined the relationship between different capital ratios and bank failure and found that the simple capital to assets ratio (leverage ratio) predicts bank failure as well as more complex risk-weighted capital ratios over one-year or two-year horizons. In addition, they recommended using the

Julapa Jagtiani is a senior financial economist at the Federal Reserve Bank of Kansas City; James Kolari is the Chase Professor of Finance at Texas A&M University; Catharine Lemieux is a vice president at the Federal Reserve Bank of Chicago; and Hwan Shin is an assistant professor of Finance at the University of Texas at Tyler. *The authors thank Kevin Birtch, Robert DeYoung, David Marshall, and Bill Tracy for their comments and suggestions, James Harvey for his assistance on the regulatory information, and Carlos Gutierrez and Loretta Kujawa for their dedicated research assistance.*

simple capital ratio as a tool to provide a timely signal of the need for supervisory action.

Our empirical results reveal that banks with impending capital deficiency are much different from other banks in terms of their financial condition. Also, our EWS models are able to detect the early onset of financial distress in commercial banks one year in advance with a reasonable degree of accuracy. Importantly, we compare three EWS models—a simple logit model that includes only the lagged capital ratio and lagged change in capital ratio, a more complete logit model, and a non-parametric trait recognition analysis (TRA) model. Our results suggest that simple linear models perform better than more complex EWS models such as the TRA. This is interesting given the substantial disagreement in the statistical forecasting literature on the merits of non-parametric approaches versus statistical models such as logit. We conclude that EWS models could be useful to bank regulators as both an off-site surveillance tool and a supplement to on-site examinations by supervisory personnel. In addition, simple linear logit EWS approaches could perform just as well as or better than highly sophisticated models in flagging capital-inadequate banks one year ahead.[4]

In the next section, we review the types of EWS models currently being used by the major bank regulatory agencies and survey the literature on the performance of these models. Then, we introduce our three models and compare their results.

EWS models in use and literature

This section describes the various EWS models used by federal bank regulators (the Federal Deposit Insurance Corporation [FDIC], the Federal Reserve, and the Office of the Comptroller of the Currency [OCC]) and reviews the literature on early warning models. It is important to note that all predictions from EWS models involve a set of trade-offs. That is, a model's accuracy may be measured by two broad types of misclassification—Type-1 error and Type-2 error. For the purposes of this study, Type-1 errors or false positives are more serious than Type-2 errors or false negatives. For example, a misclassification of unsatisfactory banks as satisfactory (Type-1 error) can be costly and result in a bank failure that might have been prevented by early supervisory intervention. On the other hand, the cost of misclassifying satisfactory banks as unsatisfactory (Type-2 error) is limited to the cost of focusing unnecessary supervisory resources on a healthy bank. We begin with an overview of the FDIC's model.

The FDIC model

The FDIC began developing EWS models in the mid-1980s. Its current model, called the SCOR (Statistical CAMELS Off-Site Ratings) model, uses Call Report data to identify banks likely to experience a downgrade at the next on-site examination, using stepwise ordered logit analysis. According to Collier et al. (2003), the SCOR model has played an important role in the FDIC supervisory process—that is, in off-site monitoring, in resource allocation for examinations, and in tracking industry trends. The SCOR model predicts the probability that a bank will be assigned a specific rating (1, 2, 3, 4, or 5, where 1 is best and 5 is worst) and the probability that a bank's rating will be downgraded. The following financial variables are included: total equity capital, loan loss reserve, past due loans 30–89 days, past due loans 90+ days, nonaccrual loans, other real estate owned, net charge-offs, provision for loan losses, net income, cash dividends declared, volatile liabilities, liquid assets, and loans and long-term securities.[5] All variables are in ratios as a percentage of total assets. The model is currently set to flag banks with at least a 35 percent probability of being downgraded. Tests of the accuracy of the FDIC model, as reported in Collier et al. (2003), indicate that approximately two-thirds of the institutions that were actually downgraded were not flagged by the model, and approximately two-thirds of the institutions that the model did flag for downgrades were not downgraded. The model's accuracy relies on the accuracy of its financial information inputs.[6]

The Federal Reserve model

In the 1990s, the Federal Reserve System moved from a surveillance process that relied on screening key financial ratios to one that incorporated econometric models to predict financial conditions. In 1995, the Federal Reserve developed the current EWS model called SEER (System to Estimate Examination Ratings). There are two parts of the SEER model—the SEER Rating model and the SEER Risk Rank model.

The SEER Rating model:

Like the FDIC's SCOR model, the Federal Reserve's SEER model predicts the probability that a bank will be assigned each of the five possible ratings, using a stepwise multinomial logit analysis. The SEER rating is then calculated as the sum of the five rating levels multiplied by their respective probabilities. In addition, the SEER Rating model is often used to classify banks into satisfactory (1- or 2-rated) and unsatisfactory (3-, 4-, or 5-rated) categories.

The model utilizes two previous quarters of Call Report and examination data. The model includes approximately 45 financial and non-financial variables in the analysis. Through a stepwise selection process, different variables are selected to be included in the final model in each quarter. The variables that have consistently remained statistically significant in each quarter are proxies for, among other things, capital adequacy, asset quality, earnings performance, and liquidity. These are variables typically used by examiners in evaluating the condition of a bank. In addition, the prior composite CAMELS rating is also included in the SEER Rating model.[7] The economic variables, such as unemployment, income per capita, and permits per capita, are usually not statistically significant.

The SEER Risk Rank model:

The objective of the model is to estimate the likelihood that a bank will fail or become critically undercapitalized (according to prompt corrective action standards) during the subsequent two years. The model was originally re-estimated each quarter using a stepwise probit regression based on data on failed and undercapitalized banks from the prior years. However, as the number of failed or critically undercapitalized banks decreased throughout the 1990s, a pooled cross-section and time-series model based on data from 1985 to 1991 was developed to capture enough failed institutions to estimate the model. The Risk Rank model is no longer re-estimated.

Both of the SEER models seem to perform well. According to the original validation tests in Cole, Cornyn, and Gunther (1995), based on ten separate quarterly estimates using data from December 1989 to March 1992, the SEER Rating model was highly accurate in predicting banks with 1 and 2 ratings. That is, 77.5 percent of 1-rated banks and 79.9 percent of 2-rated banks were correctly predicted. In addition, out of their 262 sampled failing banks, 97.7 percent received a SEER rating of 5; 1.9 percent received a 4 rating; and the remaining 0.4 percent received a 3 rating. None of the banks received a SEER rating of 1 or 2.[8] In addition, the models have been revalidated on an annual cycle, and, accordingly, a significant number of enhancements have been made to the model since 1995.

The OCC model

The OCC uses a monitoring system called Canary that consists of a diverse array of supervisory and economic predictive models and tools, organized into four components—Benchmarks, Credit Scope, Market Barometers, and Predictive Models. The OCC also uses the FDIC's SCOR model (described earlier) as one component of the off-site monitoring system.[9]

Peer Group Risk Models (PGRMs) are one type of predictive model used by the OCC. PGRMs are a series of econometric models designed to predict the potential impact of different economic scenarios on a bank's return on assets (ROA) over the next three years and how a bank will perform relative to similar asset-based bank peer groups.[10] Banks are classified into 11 different groups based on the specialization in their loan portfolios and whether they are de novo banks.[11] The model uses various economic indicators, including interest rates, prices, wages, unemployment rates, and bankruptcy. Bank-specific financial variables include nonperforming loans, provisions for loan and lease losses, and capital–asset ratios. In addition, specific financial and economic factors that are likely to impact banks in certain peer groups are included, such as farm incomes and farm output prices for those in the agricultural peer group. Information about performance and accuracy of the OCC model's results is currently not publicly available.

Review of EWS literature

Several research studies examine the factors that are important in determining bank failure or important for off-site surveillance.

Whalen (1991) examines a particular type of EWS model called a Cox proportional hazards model, which produces estimates of the probability that a bank with a given set of characteristics will survive longer than some specified length of time into the future. Using a relatively small set of publicly available explanatory variables, the model identifies both failed and healthy banks with a high degree of accuracy. A large proportion of banks that subsequently failed are flagged as potential failures in periods prior to their actual demise. The author concludes that reasonably accurate EWS models can be built and maintained at relatively low cost.

Thomson (1991) models bank failures of all sizes based on Call Report data using a logit regression analysis. The probability that a bank will fail is a function of capital adequacy, asset quality, management quality, earnings performance, and the relative liquidity of the portfolio. These are CAMELS-motivated proxy variables. Thomson finds that the majority of these factors are significantly related to the probability of failure as much as four years before a bank fails.

Cole and Gunther (1998) develop a model of bank failure, using a probit analysis to estimate the relationship between a set of financial ratios and the likelihood of bank failure during the subsequent two-year period. They find that the information content of the CAMELS ratings derived from on-site examinations can decay fairly rapidly. The ability of CAMELS ratings to identify bank failures matches or exceeds that of their off-site EWS model only when the ratings are based on exams conducted no more than one or two quarters prior to the forecast period. However, they also note that the effectiveness of off-site EWS models depends on the integrity of accounting data, which is enhanced through periodic exams. Thus, EWS is not meant to substitute for on-site examinations.

Gunther and Moore (2000) note that Call Report data are often subject to revisions. They find evidence of a strong relationship between on-site exams and Call Report revisions, pointing to a substantial auditing role for on-site exams. This suggests that the accuracy of EWS models may be overstated when using Call Report real-time data. However, they find no empirical evidence of any substantial impact of the data revisions on the model's ability to predict downgrades.[12] Thus, real-time data are still useful in the EWS model for off-site monitoring.

Gilbert, Meyer, and Vaughan (1999) compare univariate and multivariate models' ability to predict bank failures. They find that the "best" single variable varied from year to year, and that only multivariate models could provide consistently accurate predictions. One of the most significant variables in all their tests was the equity ratio. Our work builds upon their results by expanding the variables and interactions included in the analysis.

Finally, Krainer and Lopez (2003) suggest that market information may be included in the EWS model. They find that changes in stock prices tend to precede changes in supervisory BHC ratings by at least nine months. They conclude that equity market information can be useful to supervisors.

Comparison of three EWS models

We test our three EWS models using financial and economic data for sample banks. The models are 1) a simple logit model comprising only the lagged capital ratio and lagged change in capital ratio, 2) a more complete stepwise logit (parametric or statistical) model, and 3) the non-parametric method of trait recognition analysis (TRA).

We developed our set of explanatory variables for predicting banks that will become capital inadequate based on a wide variety of on- and off-balance-sheet measures currently used by regulators to gauge bank risks—including those used in the Fed's SEER model and the FDIC's SCOR model (for example, profitability, loan risk, operational risk, liquidity risk, interest rate gap, bank size, derivatives exposure, loan commitments, years in the banking business, and changes in loan compositions). We also incorporate a number of control variables that reflect economic conditions, including information on business bankruptcy filings in the state in the past year, rural versus urban location of the bank, and income per capita and permits per capita in the state where the banks are located. Table 1 lists the 42 explanatory variables we use in our analysis.

We use data from the Call Reports for year-end 1987, 1988, 1989, and 1990. The empirical methodologies are described in box 1 (logit) and box 2 (TRA). Our research methodology is implemented in two steps.

In step one, for the original sample using year-end 1989 data, we assigned each sample bank a dummy value of 1 (adequate capital) if the ratio of primary capital to assets was equal to or greater than 5.5 percent, and 0 (inadequate capital) otherwise.[13] We assembled financial and economic data for the original sample one year prior to the capital inadequacy event using year-end 1988 data. We also used the year-end 1987 data in the calculation of the lagged change variables. We then developed our in-sample models (logit and TRA models). The models seek to classify banks—adequately capitalized versus inadequately capitalized—correctly in 1989. Banks that were already inadequately capitalized in 1988 are excluded from the in-sample analysis, since supervisors are more interested in those banks that were adequately capitalized in 1988 but might become capital inadequate in 1989.

In step two, we coded the data for the 1990 holdout sample as 0 or 1 based on the 5.5 percent primary capital ratio as of year-end 1990. One-year prior data for the independent variables were passed through the logit and TRA models for holdout sample banks. Banks that were already inadequately capitalized in 1989 are not included in this holdout sample test for the reason described earlier. This is to test the predictive accuracy of our EWS models (developed during the original sample period) in predicting capital inadequacy in the out-of-sample period (1990). As such, the holdout sample test allows us to observe and compare the predictive ability of the three EWS models with data that were not employed in their development.

The final sample, after dropping institutions with missing data, consists of 477 banks in the original sample and 499 banks in the holdout sample. Of the 477 banks in the original sample, 26 banks that were capital adequate in 1988 became capital inadequate in 1989. Of the 499 banks in the holdout sample, 38 banks that were capital adequate in 1989 became capital inadequate in 1990.

For supervisory purposes, the most meaningful measure of accuracy is the ability to classify banks correctly in a future period rather than in the previous periods. Therefore, we focus on the models' predictive accuracy in the holdout period. Below, we discuss the results, which are presented in table 2 (p. 56).

TABLE 1

List of explanatory variables

Variable descriptions	Xi
Special characteristics	
Dummy for de novo banks (1 if less than 10 years)	X3
Dummy for MSA (1 if in metropolitan statistical area)	X4
Number of full-time employees to total assets	X18
Financial variables (Call Report data)	
Lagged capital ratio	X1
Lagged change in capital ratio	X2
Total assets	X5
Net income after tax to total assets	X6
Net interest income plus noninterest income to noninterest expenses	X17
Provision of loan and lease losses to total assets	X13
Other borrowed funds to total assets	X14
Core deposits to total deposits	X15
Total cash dividends to total assets	X16
Short-term interest rate gap to total assets	X19
Loans made to insiders to total assets	X21
Insured deposits to total liabilities	X22
Jumbo CDs to total assets	X23
Short-term assets to short-term liabilities	X24
Total loans to core deposits	X25
Total loans to total deposits	X26
Average maturity of assets	X27
Nonperforming loans (more than 90 days accruing) to assets	X28
Nonperforming loans (more than 90 days nonaccruing) to assets	X29
Agricultural nonperforming loans to total agricultural loans	X30
C&I nonperforming loans to total C&I loans	X31
Consumer nonperforming loans to total consumer loans	X32
Real estate nonperforming loans to total real estate loans	X33
Foreign exchange transactions to total assets	X34
Off-balance-sheet interest rate risk (all FX derivatives) to assets	X35
Off-balance-sheet loan commitments to total assets	X36
Other real estate loans to total assets	X37
Noninterest expenses to total assets	X7
Investment securities (book value) to total assets	X40
Growth and volatility variables	
Average quarterly loan growth over the year, agricultural loans	X8
Average quarterly loan growth over the year, C&I loans	X9
Average quarterly loan growth over the year, CRE loans	X10
Average quarterly loan growth over the year, consumer loans	X11
Average quarterly loan growth over the year, mortgage loans	X12
Volatility of consumer loan volume	X41
Economic factors	
Number of bankruptcy filings	X39
Business bankruptcy filings information	X42
Income per capita (personal income to labor force at state level)	X20
Number of permits per capita at state level	X38

As we can see in table 2, panel A, the simple logit model, with only lagged capital ratio and lagged change in capital ratio included as explanatory variables, performed quite well in the holdout period. The overall accuracy of prediction is 79.96 percent—with 21.05 percent Type-1 error and 19.96 percent Type-2 error. The model accurately predicted 30 capital inadequate banks, but failed to accurately predict eight of the 38 capital inadequate banks. Of the 461 capital adequate banks, the model mistakenly predicted 92 of them (19.96 percent) as being capital inadequate.

Table 2, panel B shows the results from the more complete logit model, which includes all 42 explanatory variables in the analysis. Through a stepwise selection process, only 16 variables are included in the final logit model.[14] Interestingly, the model did not perform as well as the simple logit model in predicting capital inadequate banks in the holdout period. While

BOX 1

Logit models

Logit is one of the most commonly employed parametric EWS models in business academic studies as well as bank regulatory practice—see Amemiya (1973, 1981) for a detailed discussion of this technique. The logit model has the statistical property of not assuming multivariate normality among the independent variables.

The dependent variable is the log of the bank's odds of capital inadequacy versus capital adequacy, as shown in equation 1, where P_i is the probability that bank i is a member of the capital inadequate group of banks as opposed to the capital adequate group of banks.

1) $\text{Log}\,[P_i/(1–P_i)] = a_{1i}X_{1i} + a_{2i}X_{2i} + \ldots\, a_{ni}X_{ni}.$

One advantage of logit models is that statistical software is readily available. However, as in any other parametric model, logit is not well suited to exploring interactions between large numbers of variables due to losses in degrees of freedom. Relatedly, interaction variables are typically computed by multiplying two variables, which tends to lose information.

For example, consider two banks, one with a high return on assets and low capital ratio and the other with a low return on assets and high capital ratio. The product of the return on assets and capital ratio would be moderate in level in both cases; as such, their interaction would lose information about each of the component ratios.

Model 1: Simple logit

The predicted probability of becoming capital inadequate one year from now is estimated from two simple factors X_{1i} and X_{2i}, as shown in equation 2, where X_{1i} is bank i's lagged capital ratio and X_{2i} is bank i's lagged change in capital ratio.

2) $\text{Log}\,[P_i/(1–P_i)] = a_{1i}X_{1i} + a_{2i}X_{2i}.$

Model 2: Stepwise logit

The predicted probability of becoming capital inadequate one year from now is estimated from 42 different financial and economic variables (listed in table 1), rather than just two capital ratio variables as in model 1. Only 16 of these 42 factors entered the model through a stepwise logit procedure.

Panels A and B of table 2 present the logistic results of model 1 and model 2, respectively.

the overall accuracy of prediction is 83.17 percent (higher than the overall accuracy of prediction of the simple model), the complete logit model failed to predict 76 percent of the capital inadequate banks. Specifically, the complete model's Type-1 error is 76.32 percent and Type-2 error is 11.93 percent. The model failed to accurately predict 29 of the 38 capital inadequate banks, and mistakenly predicted 55 of the 461 good banks to be capital inadequate.

Table 2, panel C presents the results of the TRA model, with all 42 explanatory variables and the interactions included in the analysis. This is the most complex of the three models. Its overall accuracy of prediction was exceptional at 99.23 percent in the original sample period, but dropped to 77.35 percent in the holdout period. The TRA model's Type-1 error is 47.37 percent and Type-2 error is 20.61 percent. In terms of accurately predicting capital inadequate banks, the TRA model performed better than the complete logit model, but not as well as the simplest logit model, which has the best prediction record with the smallest number of Type-1 errors. The TRA model failed to accurately predict 18 of the 38 capital inadequate banks, and mistakenly predicted 95 of the 461 good banks to be capital inadequate.

However, the TRA model may be viewed as having an advantage over logit models. In the modeling process, the TRA identified 89 safe features and 80 unsafe features—banks with safe features are less likely to encounter financial distress in the near future, whereas banks with unsafe features are those whose capital ratios are likely to fall below an adequate level within a one-year time frame. These identified safe and unsafe traits can serve as the starting point for on-site reviews.[15]

An important aspect of the TRA model is highlighted by this collection of traits. It takes into consideration that financial variables are related; thus, the same variable can be "good" or "bad" depending on the value of other variables. This non-parametric approach identifies collections of factors that together can be red flags for supervisors, providing a more detailed picture of the potential issues the institution may be facing. Unlike parametric estimations, TRA does not assume independence among the explanatory variables. In addition, TRA is a flexible enough approach to identify separate features for "good' and "bad" states, so that "good" and "bad" traits are not mirror images of each other.

Overall, both the logit and TRA models developed in this article would enable bank supervisors to identify most capital inadequate banks one year ahead of time. It is interesting, and surprising, to see that the simplest model (the logit model with only lagged capital

BOX 2

TRA models

Unlike logit models, the TRA approach is a non-parametric technique that identifies systematic patterns in the data. TRA was originally developed in the hard sciences and used to predict the risk of earthquakes and the location of oil and uranium fields. In recent years, it has been applied to predicting bank failure. Unlike traditional econometric models, TRA avoids assumptions about the underlying distribution or independence of the variables.

TRA is most closely associated with neural network models in that it seeks to exploit information contained in complex interactions of the independent variable set. A unique aspect of TRA is that variable interactions could be formed to be consistent with the logic of a financial analyst, rather simple cross products of variables. For example, an interaction variable could be defined as high return on assets and low capital ratio, or vice versa. As such, information about components of interaction variables is not lost.

Performing a TRA analysis requires a number of steps. For a detailed discussion of the TRA technique, see Kolari, Caputo, and Wagner (1996), Kolari, Glennon, Shin, and Caputo (2002), and Jagtiani, Kolari, Lemieux, and Shin (2002). We provide a very brief summary of the process below.

Our TRA analysis utilizes all 42 explanatory variables listed in table 1. In building the TRA model, we follow the following steps:

We select cutoffs for each of the 42 independent variables to divide banks into three categories (low, medium, and high) for each variable. We then recode the variables into binary strings. For example, bank *i*'s variable vector may be [$X1$,$X2$,$X3$] = [010011]. This implies that bank *i* is in the middle (01), low (00), and upper (11) segments of the distributions of $X1$, $X2$, and $X3$, respectively. The six-digit code [010011] is known as a "trait" of the *i*th bank. Using variables three at a time, we can produce a long list of traits for each bank.

The trait matrix is a list of all the traits for each bank using the binary strings from the previous step. The pattern of these binary strings, or traits, is useful in discriminating between good and problem banks. For example, traits commonly found in capital adequate banks but seldom found in capital inadequate banks are retained as "good traits" or "safe features." Traits commonly found in capital inadequate banks but seldom found in capital adequate banks are retained as "bad traits" or "unsafe features."

Now we can construct a matrix of the number of good traits on the X axis and the number of bad traits on the Y axis. Each bank is put into a cell in this matrix based on the number of its good and bad traits. This step is called voting.

If a bank has far more good votes (good traits) than bad votes (bad traits), then it is identified as capital adequate, and vice versa for capital inadequate banks.

Finally, we use the voting matrix to classify out-of-sample observations to verify the prediction accuracy of the model in the post-sample period. These banks were not used in making the traits and voting matrix. This allows us to see whether the trait recognition method is able to discriminate between capital inadequate and capital adequate banks in the post-sample period.

Like any other EWS models, TRA models are subject to some drawbacks. The most serious difficulty is the considerable hands-on manipulation required by the researchers, such as creating and inputting cut-points as well as selecting the minimum and maximum percentage definitions of features.

In addition, no statistical measures of variables' importance are produced in the TRA analysis, since it is not a statistical approach. This shortfall, however, may not be an issue if the research goal is the model's prediction accuracy rather than the significance of individual variables. Also, TRA has an advantage over other techniques in generating a list of good and bad traits that may well be useful to bank supervisors in better understanding a bank's strengths and weaknesses.

ratio and lagged change in capital ratio) turned out to be just as useful, if not more so, than the sophisticated models in the holdout sample tests. Its Type-1 error is smaller than that of the other models. Type-1 errors are the most serious, as they mean that the model failed to warn regulators that the bank was going to fall below capital standards in the near future.

The accuracy of the simplest logit model may be partly driven by the inclusion of the lagged capital ratio, since banks' capital ratios do not usually change drastically from one year to the next. However, it is important to point out that our analysis excludes all banks that were already inadequately capitalized. Our results suggest that it would be beneficial for supervisors to allocate more examination resources to banks whose capital levels are approaching the minimum adequate level and to closely supervise and monitor these banks.

Figure 1, panel A demonstrates that the simple logit model (with only lagged capital ratio and lagged change in capital ratio) outperforms the TRA model. Its Type-1 errors are typically smaller than the TRA's

TABLE 2

Results of the three models

A. Simplest logit model (only lagged capital ratio and lagged change in capital ratio as explanatory variables) at *PL* = 0.08

Model: Log[$Pi/1-Pi$] = 3.56 – 94.91 (X1) – 8.82 (X2).

Original sample			Holdout sample		
	Predicted			Predicted	
Actual	Inadequate	Adequate	Actual	Inadequate	Adequate
Inadequate	14	12 (46.15%)	Inadequate	30	8 (21.05%)
Adequate	111 (24.61%)	340	Adequate	92 (19.96%)	369
Overall accuracy prediction: 74.21%			Overall accuracy prediction: 79.96%		

B. Stepwise logit model (with 16 explanatory variables included) at *PL* = 0.10

Model: Log[$Pi/1-Pi$] = 4.38 – 74.88 (X1) – 7.44 (X2) – 163.8 (X6) – 0.12 (X9) + 0.002 (X10) + 0.03 (X11) +88.36 (X13) + 62.98 (X28) – 16.60 (X30) – 54.02 (X31) + 36.28 (X32) + 15.78 (X33) + 31.53 (X37) – 15.83 (X39) – 6.13 (X40)

Original sample			Holdout sample		
Actual	Predicted		Actual	Predicted	
	Inadequate	Adequate		Inadequate	Adequate
Inadequate	13	13 (50.0%)	Inadequate	9	29 (76.32%)
Adequate	44 (9.76%)	407	Adequate	55 (11.93%)	406
Overall accuracy prediction: 88.05%			Overall accuracy prediction: 83.17%		

C. Trait recognition analysis (TRA) model

Original sample			Holdout sample		
Actual	Predicted		Actual	Predicted	
	Inadequate	Adequate		Inadequate	Adequate
Inadequate	24	2 (2.60%)	Inadequate	20	18 (47.37%)
Adequate	0 (0.00%)	451	Adequate	95 (20.61%)	366
Overall accuracy prediction: 99.23%			Overall accuracy prediction: 77.35%		

for any given Type-2 errors. Figure 1, panel B compares the complete logit model (with all 42 variables included in the analysis and 16 variables kept in the model) and the TRA model (most complicated and non-parametric). The plot suggests that TRA outperforms the complete stepwise logit model.

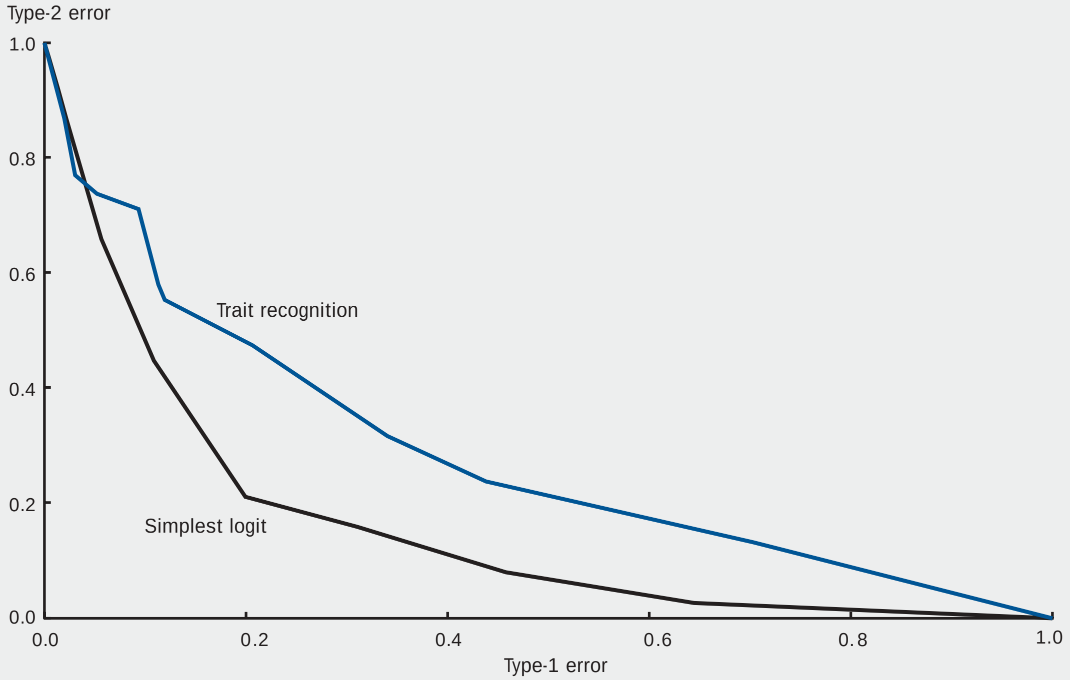
Conclusion

This article focuses on the continuing body of work that attempts to develop early warning systems (EWS) for detecting deterioration in a bank's financial condition. While mandatory on-site examinations provide supervisors with the opportunity to review all information and to develop a periodic view of the institution's financial condition, supervisors need a means of identifying at-risk institutions so that supervisory actions can be taken during the period between exams. Regulators currently have EWS models, which attempt to flag troubled institutions that are likely to fail using logit analysis.

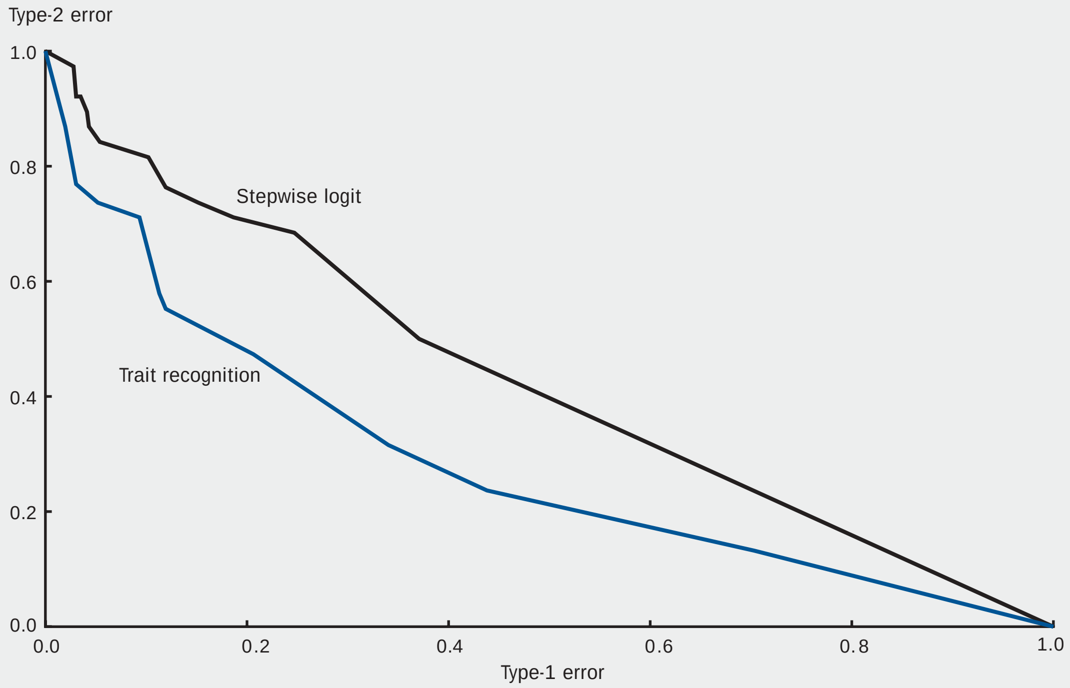
FIGURE 1

Comparing Type-1 and Type-2 error

A. Simplest (naïve) logit model versus TRA model



B. Stepwise logit model versus TRA model



Our study departs from the existing EWS models currently used by bank regulators in two respects. First, our models attempt to identify institutions that are likely to become capital inadequate (not failures) in the next year. FDICIA, passed in 1991, contains a series of mandatory and discretionary actions supervisors can impose as a bank's capital deteriorates, up to and including closure if capital falls below 2 percent of assets. Our EWS models could provide bank supervisors with this early warning of the onset of distress. Second, in addition to just flagging problem banks, we also explore alternative statistical and non-parametric models ranging from a simple logit model to a complicated trait recognition analysis non-parametric model. This process has allowed us to gain insights into what information is most important in terms of predicting pending capital problems at commercial banks.

Using 1987, 1988, and 1989 data for a sample of medium-sized banks, we developed three different EWS models in this article. The first and simplest model estimates banks with inadequate capital based on last period's level of capital and the trend in capital levels using a logit regression analysis. The second model is the complete logit stepwise analysis, which employs a wide variety of financial and economic variables similar to those in the supervisory models, with the exception that it predicts the likelihood of capital adequate or inadequate rather than failure or CAMELS downgrades. The third model is the non-parametric TRA model, which relies on patterns on the data to identify combinations of traits that are most often found in banks with adequate capital levels and combinations of traits found in banks that have inadequate capital levels.

The in-sample results indicate that all three models are able to predict banks with inadequate capital levels. These results show that it is possible to predict early stages of financial distress using off-site data. Comparing performance of the three models in the original sample period, the TRA model's record was superior. It accurately predicted all adequately capitalized banks and only incorrectly classified two banks with inadequate capital as having adequate capital. However, the true value to supervisors is performance in out-of-sample tests.

Using the developed models, we conducted out-of-sample tests to see which approach provides the "best" estimate of adequate and inadequate capital levels based on 1988, 1989, and 1990 data. We define "best" as missing the smallest number of inadequately capitalized banks—in other words, having the smallest number of Type-1 errors. This is the measure that is most relevant to supervisors because it is more costly to miss detecting deterioration in a bank's financial condition than it is to more closely supervise an adequately capitalized bank.

Results presented for each model for the holdout sample indicate that while the complete logit model has the best prediction record overall, the simple logit model has the best prediction record for missing the smallest number of capital inadequate banks (smallest number of Type-1 errors). The simplest logit model has 79.95 percent accuracy in predicting inadequately capitalized banks, compared with only 23.68 percent for the complete stepwise logit model and 52.63 percent for the TRA model.

With the caveat that our empirical results are based on a recessionary sample period and should be tested under different economic conditions, our results imply that the probability of tripping a capital tripwire, which is an effective early indicator of financial deterioration, could be predicted using off-site data. Existing supervisory models generally focus on the probability of failure and CAMELS downgrades.

In addition, we found that a simple logit model could do quite well in predicting the likelihood of deteriorating capital levels in both in-sample and out-of-sample tests. However, these models provide little insight into the root cause of the problem. Unlike the simple logit model, the TRA technique provides more detailed information—that is, information on the combined impact of the various groups of explanatory variables. This additional information could provide supervisors with a head start in identifying the root cause of changes in a bank's financial condition. Therefore, using a TRA technique in conjunction with the current logit EWS models could potentially enhance off-site monitoring effectiveness.

Although a complete cost/benefit analysis is beyond the scope of this article, we feel that the marginal cost of developing and maintaining this additional feature is relatively small compared with the potential cost of a bank failure or more frequent on-site examinations. Understanding the root cause of the bank's problems will likely enhance the effectiveness of supervisory intervention and help guide examination resources toward the institutions with the greatest potential for problems.

To give some perspective on the importance of identifying emerging problem banks, it is important to point out that unless banks have identified problems, they are generally examined every 12 to 18 months depending on their asset size. However, the value of supervisory ratings tends to deteriorate as soon as six months after on-site examinations (see Cole and Gunther, 1998 and 1995). Thus, this additional modeling effort could be viewed as a way of augmenting existing supervisory ratings as more current information becomes available.

Ideally, the current logit EWS models may be used to identify those institutions where additional questions should be asked, and the TRA model may be used to provide supervisors with a better idea about which questions to ask. Knowledge of safe and unsafe features could guide examiners as they develop an integrated evaluation of the financial condition of the entire organization instead of focusing on isolated individual ratios. For example, a high return on assets (ROA) is not always an indication of sound financial condition. High ROA in a de novo institution, in the presence of rising delinquencies, or combined with a rising assets per employee ratio, may be an indication of problems.

Finally, there is one important caveat. No model can accurately predict financial stress if the financial data are not accurate. One of the benefits of on-site examinations is the potential to identify inaccurate or misleading financial reporting, particularly regarding reserves for problem loans and loan write-offs. Previous research suggests that any off-site monitoring system that relies completely on Call Report data is subject to errors due to misreporting of financial data. Off-site EWS models should be regarded as an aid to, but not as a substitute for, the critical roles of examiner experience and judgement.

NOTES

[1]CAMELS stands for capital adequacy, asset quality, management quality, earnings, liquidity, and sensitivity to market risk.

[2]The regulators' failure prediction model predicts banks that are likely to declare equity insolvency or become critically undercapitalized, with a ratio of tangible equity to average assets of less than 2 percent. Our model's orientation to the 5.5 percent primary capital to assets ratio differs substantively from the more traditional surveillance approaches.

[3]Our model predicts banks that would have a primary capital to asset ratio falling below 5.5 percent in the next 12 months. Our results are expected to hold regardless of the regulatory capital standards in place. The change in regulatory capital standards from "leverage ratio" to "risk-based" on December 12, 1990, had no impact on our results.

[4]It is important to point out that while the high accuracy of the simple logit model may be partially driven by the inclusion of lagged capital ratios, which do not seem to change drastically from one year to the next, our analysis does not include those banks that were previously capital inadequate.

[5]These ratios incorporate four quarters of data. Since these ratios can be significantly affected by mergers, SCOR uses merger-adjusted data, where financial statements of the merged banks are combined for the quarters preceding the merger.

[6]Collier, Forbush, Nuxoll, and O'Keefe (2003) note the failure of Best Bank and First National Bank of Keystone as examples where the FDIC model predicted good ratings just prior to the failures because of the misstatement of Call Report data.

[7]According to Cole, Cornyn, and Gunther (1995), inclusion of additional variables would not significantly improve the accuracy of the out-of-sample estimation of the SEER model. In most cases, their inclusion would degrade such accuracy. It is also interesting to note that the SEER Rating model does use current CAMELS ratings as one of the explanatory variables, while the FDIC's SCOR model does not.

[8]The model performed much better than using CAMELS ratings alone. The relatively poor performance of the CAMELS ratings in predicting bank failure during the subsequent two-year period is attributable to the fact that CAMELS ratings available at any given date are based upon information that is more dated than that for the off-site monitoring systems.

[9]See the OCC's Examiner's Guide to Problem Bank Identification, Rehabilitation, and Resolution (2001) for a detailed description of monitoring tools used by the OCC.

[10]The importance of economic scenarios in bank failure has been documented in a study conducted by the OCC staff (see Wentzler, Hiemstra, and Jacques, 2001).

[11]The separation of de novo banks in the OCC model is supported by previous research (see DeYoung, 2003a and 2003b and Hunter, Verbrugge, and Whidbee, 1996). De novo banks are less likely to fail during the first few years and become more likely to fail than small, established banks over time. In addition, de novo banks are more affected by adverse economic conditions than other banks.

[12]They did not test the impact of data revisions on the model's ability to predict bank failure.

[13]The capital benchmark of 5.5 percent was the measure in place during the sample period.

[14]The explanatory variables selected to be included in the stepwise logit model are lagged capital ratio, lagged change in capital ratio, net income after tax to total assets, average quarterly loan (agricultural) growth over the year to total loans, average quarterly loan (C&I) growth over the year to total loans, average quarterly loan (commercial real estate) growth over the year to total loans, average quarterly loan (consumer) growth over the year to total loans, provisions for loan and lease losses to total assets, nonperforming loans (more than 90 days accruing) to total assets, agricultural nonperforming loans to total agricultural loans, C&I nonperforming loans to total C&I loans, consumer nonperforming loans to total consumer loans, real estate nonperforming loans to total real estate loans, other real estate loans to total assets, and investment securities to total assets.

[15]In addition, less capitalized banks are also more likely to become targets of a takeover (see Wheelock and Wilson, 2000 and Brewer, Jackson, Jagtiani, and Nguyen, 2000).

REFERENCES

Brewer, E., W. Jackson, J. Jagtiani, and T. Nguyen, 2000, "The price of bank mergers in the 1990s," *Economic Perspectives*, Federal Reserve Bank of Chicago, First Quarter, pp. 2–23.

Cole, R., B. Cornyn, and J. Gunther, 1995, "FIMS: A new monitoring system for banking institutions," *Federal Reserve Bulletin*, Board of Governors of the Federal Reserve System, January, pp. 1–15.

Cole, R., and J. Gunther, 1998, "Predicting bank failures: A comparison of on- and off-site monitoring systems," *Journal of Financial Services Research*, Vol. 13, No. 2, pp. 102–117.

Collier, C., S. Forbush, D. Nuxoll, and J. O'Keefe, 2003, "The SCOR system of off-site monitoring," *FDIC Banking Review*, forthcoming.

De Young, R., 2003a, "De novo bank exit," *Journal of Money, Credit, and Banking*, forthcoming.

__________, 2003b, "The failure of new entrants in commercial banking markets: A split-population duration analysis," *Review of Financial Economics*, Vol. 12, No. 1, pp. 7–33.

Estrella, A., S. Park, and S. Peristiani, 2000, "Capital ratios as predictors of bank failure," *Economic Policy Review*, Federal Reserve Bank of New York, Vol. 6, No. 2, July, pp. 33–52.

Gilbert, A., A. Meyer, and M. Vaughan, 1999, "The role of supervisory screens and econometric models in off-site surveillance," *Review*, Federal Reserve Bank of St. Louis, November/December, pp. 31–56.

Gunther, J., and R. Moore, 2000, "Early warning models in real time," Federal Reserve Bank of Dallas, Financial Industry Studies, working paper, No. 1-00, October.

Hunter, W. C., J. Verbrugge, and D. Whidbee, 1996, "Risk taking and failure in de novo savings and loans in the 1980s," *Journal of Financial Services Research*, Vol. 10, pp. 235–272.

Jagtiani, J., J. Kolari, C. Lemieux, and H. Shin, 2002, "Looking for trouble: Early detection of inadequate capitalization of U.S. commercial banks," *International Company and Commercial Law Review*, Vol. 13, No. 7, July, pp. 269–279.

Kolari, J., M. Caputo, and D. Wagner, 1996, "Trait recognition: An alternative approach to early warning systems in commercial banking," *Journal of Business, Finance, and Accounting*, Vol. 23, pp. 1415–1434.

Kolari, J., D. Glennon, H. Shin, and M. Caputo, 2002, "Predicting large U.S. commercial bank failures," *Journal of Economics and Business*, Vol. 54, No. 4, July/August, pp. 361–387.

Krainer, J., and J. Lopez, 2003, "Using equity market information to monitor banking institutions," *Economic Letter*, Federal Reserve Bank of San Francisco, January 24.

Office of the Comptroller of the Currency, 2001, *Examiner's Guide to Problem Bank Identification, Rehabilitation, and Resolution*, Washington, DC.

Thomson, J., 1991, "Predicting bank failures in the 1980s," *Economic Review*, Federal Reserve Bank of Cleveland, Vol. 27, First Quarter, pp. 9–20.

Wentzler, N., S. Hiemstra, and K. Jacques, 2001, "Contagion effects, macroeconomic conditions and the probability of bank failure," Office of the Comptroller of the Currency, working paper.

Whalen, G., 1991, "A proportional hazards model of bank failure: An examination of its usefulness as an early warning tool," *Economic Review*, Federal Reserve Bank of Cleveland, First Quarter, pp. 21–31.

Wheelock, D., and P. Wilson, 2000, "Why do banks disappear? The determinants of U.S. bank failures and acquisitions," *Review of Economics and Statistics*, Vol. 82, pp. 127–138.