A Proportional Hazards Model of Bank Failure: An Examination of Its Usefulness as an Early Warning Tool

by Gary Whalen

Gary Whalen is an economic advisor at the Federal Reserve Bank of Cleveland. The author would like to thank Paul Bauer and William Curt Hunter for helpful comments.

Introduction

The number of U.S. bank failures jumped sharply in the mid-1980s and has remained disturbingly high, averaging roughly 170 banks a year over the 1985–1990 period. Furthermore, large-bank failures have become increasingly common. For a variety of reasons, the timing of closures and the resolution techniques used have severely strained the resources of the Federal Deposit Insurance Corporation (FDIC). These developments have stimulated a great deal of debate about the causes of costly bank closures and about alternative ways to prevent them. One focus of this debate has been on the appropriate roles of market versus regulatory discipline. A necessary condition for effective discipline by either force is the ability to identify high-risk banks accurately at a reasonable length of time prior to failure without the use of expensive and time-consuming on-site examinations. This requires the use of some sort of statistical model, conventionally labeled an "early warning model," to translate bank characteristics into estimates of risk. There is considerable debate about whether models of sufficient accuracy can be built using only currently available accounting data.[1]

This study examines a particular type of early warning model called a Cox proportional hazards model, which basically 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. The sample consists of all banks that failed between January 1, 1987 and October 31, 1990 and a randomly selected group of roughly 1,500 nonfailed banks. Using a relatively small set of publicly available explanatory variables, the model identifies both failed and healthy banks with a high degree of accuracy. Furthermore, a large proportion of banks that subsequently failed are flagged as potential failures in periods prior to their actual demise. The classification accuracy of the model over time is impressive, since the coefficients are based on 1986 data and are not updated over time. In short, the results demonstrate that reasonably accurate early warning models can be built and maintained at relatively low cost.

The following section describes the proportional hazards model (PHM) in general terms and compares it to alternative statistical early

1 This is the opinion reflected in Randall (1989), for example.

warning models. A short discussion of sampling issues follows. Section III contains a more detailed discussion about the specification of the model estimated in this paper, and section IV presents the model's estimation results and classification accuracy. The final section contains a brief summary and conclusions.

I. The Proportional Hazards Model

Of the large number of early warning/failure prediction studies that have been done, most have employed discriminant analysis or probit/logit techniques to construct the models. These models are designed to generate the probability that a bank with a given set of characteristics will fall into one of two or more classes, most often failure/nonfailure.[2] Further, the predicted probabilities are of failure/nonfailure at some unspecified point in time over an interval implied by the study design.

Like these statistical techniques, a PHM can be used to generate estimates of the probability of bank failure or, alternatively, of survival. However, a PHM has several advantages relative to these other types of models, including the ability to produce estimates of probable time to failure. In fact, it can be used to generate a survival profile for any commercial bank (the estimated probability of survival longer than specified times as a function of time). The other types of models yield only the probability that a bank will fail at some point in time over some specified period, but provide no insight on when the failure will occur over this period. Additionally, a PHM does not require the user to make assumptions about the distributional properties of the data (for example, multivariate normality) that may be violated. In the one somewhat dated study of bank failures in which a PHM is estimated and used, the model is also found to be slightly more accurate than alternative models (see Lane, Looney, and Wansley [1986, p. 525]).

The dependent variable in a PHM is time until failure, T. The survivor function, which represents the probability of surviving longer than t periods, has the following general form:

$$(1) \quad S(t) = \text{Prob}\,(T > t) = 1 - F(t)\,,$$

where $F(t)$ is the cumulative distribution function for the random variable, time to failure. The probability density function of t is equal to $f(t) = -S'(t)$. Given these definitions, the general form of the so-called hazard function is then

$$(2) \quad h(t) = \lim_{dt \to 0} \frac{P(t < T < t + dt \mid T > t)}{dt} = \frac{-S'(t)}{S(t)}.$$

The hazard function specifies the instantaneous probability of failure given survival up to time t.

A number of different types of hazard models can be specified, depending on the assumptions made about the nature of the failure time distribution.[3] In the PHM, the hazard function is assumed to have the following rather simple form:

$$(3) \quad h(t \mid X, B) = h_0(t)\, g(X, B)\,,$$

where X represents a collection of characteristic variables assumed to affect the probability of failure (or, alternatively, of survival) and B stands for the model coefficients to be estimated that describe how each characteristic variable affects the likelihood of failure. The first part of this expression, $h_0(t)$, is a nonparametric term labeled the baseline hazard probability. This probability depends only on time. To obtain the failure probability in a particular case, the baseline hazard probability is shifted proportionally by the parametric function that is the second part of the expression. In the Cox variant used in this paper, the second function is assumed to have an exponential form. That is, the Cox PHM has the following form:

$$(4) \quad h(t \mid X, B) = h_0(t)\, e^{X'B}.$$

The related survivor function for the Cox PHM, which is used to calculate the probability that a commercial bank with a given set of characteristics will survive longer than some given amount of time into the future, is as follows:

$$(5) \quad S(t \mid X, B) = S_0(t)^{q}\,,$$

where $q = e^{X'B}$ and

$$S_0(t) = \exp\left[-\int_0^t h_0(u)\; du\right].$$

2 In some studies, the categorization of banks into risk classes is made on the basis of confidential CAMEL ratings.

3 See, for example, the excellent review of a number of hazard models in Kiefer (1988) or Kalbfleisch and Prentice (1980).

As in the hazard function, the first part of this expression, $S_0(t)$, is called the baseline survival probability and depends only on time. It is the same for every bank. To calculate survival probabilities for any bank, it is necessary to choose the relevant time horizon that determines the relevant baseline probability and then plug the values of its characteristic variables into the formula.

The PHM does have several disadvantages, although some of these are shared by competing failure prediction models. Perhaps the most important drawback is that estimation of the PHM requires data on the time to failure. As many others have noted, there is a distinction in banking between insolvency (an economic event) and failure (a regulatory event). That is, bank failure represents a regulatory decision. So whether one uses a PHM or a logit model, it is actually the regulatory closure rule that is being modeled. This can be problematic when one is analyzing bank failures over the late 1980s. During this time, regulators had to resolve a number of large distressed holding companies in Texas, where financial problems were concentrated in some but not all of a holding company's banks (generally, the lead or large subsidiary banks). Typically, closure of the insolvent units was delayed while attempts were made to dispose of the entire organization. Thus, in some cases, the reported financial condition of the larger subsidiaries of these holding companies suggests that they were probably insolvent prior to resolution, while smaller, sometimes numerous coaffiliate banks exhibited relatively healthy financials even shortly before closure. Failure to control for these circumstances in some way could significantly affect the coefficients and classification accuracy of any type of estimated early warning model, but the nature of the adjustment is critically important for PHMs given the nature of the dependent variable.

Empirically, this problem can be dealt with in a number of ways. Some researchers have added a consolidated holding-company size variable to estimated bank failure equations (see Gajewski [1989]). Others have estimated two-equation systems: a solvency equation and a failure equation, adding holding company variables to the latter (see Thomson [1989]). Alternatively, one could take the view that smaller bank affiliates in unit banking states are the functional equivalent of branches and so should be consolidated into one or more of the larger subsidiary banks in failure prediction studies.[4] Another, somewhat cruder, solution that is generally equivalent to consolidation is simply to exclude some or all of the smaller bank subsidiaries of the holding companies in question. This is the approach taken here. I include in the estimation sample only the larger subsidiaries (more than $500 million in total assets) of the large Texas holding companies that failed.[5]

One is still left with the problem of somewhat ambiguous dates of failure for some of the large Texas holding companies. For example, in several cases, resolution transactions that were announced (indicating that the company was judged to be failing as of a specific date) ultimately collapsed, and the institutions were not closed until some later date. Here, following standard practice, I use the failure date designated by the FDIC (typically the date that FDIC funds are disbursed).

Another possible disadvantage of the simple PHM is the assumption that the values of the explanatory variables remain constant over the time horizon implicit in the specification. Obviously, this may not be the case, and if this assumption is violated, classification accuracy of estimated PHMs could suffer. It is possible to estimate PHMs that relax this assumption (with so-called time-varying covariates).[6] However, this complicates the analysis and is not undertaken here.

II. Sampling

Using the entire population of banks to generate early warning models is typically not done, since this method is costly and requires substantial computer time and suitable hardware and software. Practically, models of comparable accuracy can be built and maintained much more easily and cheaply using a sample of banks. This is the approach taken here.

In bank failure studies, sampling is an important issue, since it can significantly affect the reported results. One common approach — the one used in the only PHM study done to date — is the use of a matched sample. In this type of approach, the sample initially consists of some collection of failed banks. Then, for each failed

■ 4 In fact, limited consolidation was authorized under a change in Texas branching laws in 1987, and was done in varying degrees by several of the state's multibank holding companies.

■ 5 However, I include *all* bank affiliates of failed holding companies in the holdout sample.

■ 6 For a discussion of time-varying covariates, see Kalbfleisch and Prentice (1980).

bank included in the sample, the researcher adds one or more nonfailed banks determined to be peers. This method is tedious and costly and requires numerous subjective judgments on the part of the researcher. It also is infeasible to use when analyzing relatively recent failures, since close matching is simply not possible. Furthermore, it is not clear that models developed using matched samples could be easily updated/reestimated, and updating may be necessary to preserve model accuracy.

I rejected relying solely on random sampling because of the danger of too few failed banks and because of cost considerations. Instead, I employed a choice-based sampling approach similar to that used in numerous other failure prediction studies. Specifically, the data set includes all banks that failed between January 1, 1987 and October 31, 1990 for which complete data could be obtained and that were in operation for at least three full years prior to failure. The nonfailed portion of the sample consists of roughly 1,500 randomly selected banks. The estimation sample is comprised of the 1987 and 1988 failures and approximately 1,000 of the nonfailed banks. The remainder of the failed and nonfailed banks comprise the holdout sample.

III. The Specific Form of the Model

The Approach of Lane, Looney, and Wansley

Lane, Looney, and Wansley (1986), hereafter referred to as LLW, estimate two different versions of PHMs using a relatively small sample of banks that failed over the 1979–1983 period and a matched sample of nonfailed banks. One version, labeled a one-year model, is designed to generate a survivor function that permits the user to predict the probability that a bank with a given set of characteristics will survive longer than times ranging from roughly zero to 12 months into the future. Another version, the two-year model, allows the user to predict survival probabilities ranging from roughly 12 to 24 months into the future. In their sample, LLW pool failures from all the years examined and use stepwise methods to select a relatively small subset of 21 financial condition variables for use as explanatory variables. They do not employ any local economic condition variables.

In the one-year model, LLW find the following ratios to be significant and include them in the final form of the estimated equation: the log of the commercial loans to total loans ratio, the total loans to total deposits ratio, the log of the total capital to total assets ratio, and the log of the operating expense to operating income ratio. In the two-year model, the ratios included are the total loans to total assets ratio, the log of the commercial loans to total loans ratio, the log of the total capital to total assets ratio, the log of the operating expense to operating income ratio, the log of the municipal securities to total assets ratio, and the rate of return on equity.[7] It is interesting that none of the loan quality variables that LLW examine is found to be significant in either model. However, their set of loan quality variables does not include a measure of nonperforming loans, since such data were not reported by banks over the period examined. This may have lowered the classification accuracy of LLW's models, because nonperforming loan data are probably a better leading indicator of incipient asset quality problems than variables such as loan loss provisions or net chargeoffs, and asset quality problems are a primary cause of bank failure. The out-of-sample classification accuracy of these relatively simple models is good, although the holdout sample is relatively small and the time period examined is quite short.

The Current Model

I designed the model used here to produce estimates of the probability that a bank with some given set of characteristics will survive longer than times ranging from roughly zero to 24 months into the future. To accomplish this, I measure the dependent variable, the time to failure, as the time in months from the end of 1986 to the failure date for each failed bank in the estimation sample. For all nonfailed banks in the estimation sample, I censor the time to failure at 24 months, since these banks are known to have survived at least this amount of time into the future.[8] I measure all of the explanatory variables for both failed and nonfailed banks as of year-end 1986. This approach

7 LLW use log values in some cases to transform explanatory variables that appear to be non-normally distributed. This is done because the authors estimate competing discriminant models that require the explanatory variables to be multivariate normal.

8 An additional advantage of the PHM is that it can accommodate censored failure times.

is feasible given the relatively large number of failed banks over the 1987–1988 period and the sampling method used. So, unlike LLW, I do not pool failures from different years. Furthermore, by estimating a single model with a 24-month time horizon, I incorporate the implicit assumption that these survival probabilities depend only on a single set of explanatory variables.

The Explanatory Variables

In general, I employ subsets of a relatively small number of "typical" financial ratios used in previous bank failure prediction studies as explanatory variables in this study. All of these are publicly available numbers drawn from the year-end reports of the Federal Financial Institutions Examination Council's Reports on Condition and Income, known as call reports. The variable names and definitions, along with the 1986 mean values for banks in the estimation sample, appear in the appendix. I do not use loan classification data drawn from examination reports for a variety of reasons, the most important of which is that such data are available only at irregular intervals.[9]

The only other type of explanatory variable used in this study is a single indicator of "local" economic conditions. Recently, a consensus has emerged that such variables have a significant impact on the probability of bank failure and should somehow be incorporated into the analysis. However, an examination of previous research reveals that this has not typically been done in the past. In those studies that use local economic variables, the standard approach is to add one or more as explanatory variables in the estimated failure equation. The identity of the variables and the precise forms of these relationships differ considerably. Some researchers have found that such variables are significant and aid classification accuracy.

More recent studies have used state-level economic variables such as the change in personal income, unemployment, or real estate construction. Some employ a form of state economic diversification variable, while others simply add variables designed to capture the importance of the energy or farm sector in a given state. In a few studies, economic data from the county level or the metropolitan statistical area are employed.

It seems inappropriate to simply add farm- or energy-sector variables to failure prediction equations. Although it is true that downturns in these industries appear to be highly correlated with bank failures in the recent past, there is no reason to believe that this pattern will repeat itself in the future (in the Northeast or the Southeast in the early 1990s, for example). If one deems it desirable to add local economic variables to a bank failure model (and this may not be the best way to proceed), a preferable approach would be to use local variables such as unemployment, employment, or some construction series that reflect local economic shocks regardless of their source.

I employ a state-level variable rather than a more local variable for several reasons. Incorporating more local variables into the analysis is much more tedious and costly. It would also be more difficult to update such variables over time. Furthermore, it is not clear that using more local variables would produce more accurate failure probabilities than state-level data. Previous research indicates that two of the most useful leading indicators of economic conditions at the state level are movements in building permits and initial unemployment claims.[10] Here, only one state variable is used: the percentage change in state residential housing permits issued over the three-year period ending in the year in which the other explanatory variables are measured.

Realistically, the response of the financial condition of any individual bank to local economic conditions varies across banks and changes over time as managers react to anticipated movements in relevant local and national economic variables. This view suggests that perhaps a more correct approach (and a much more ambitious one) would be to use only forecasted bank financial condition variables in the failure prediction equation. The values of these variables would be based on forecasts of local or regional economic conditions generated using separate models (see Goudie [1987], for example). Alternatively, one might develop state-level leading economic index series and sequential probability models, which can be used to generate the probability of a local recession, and then use these probabilities in a failure prediction model (see Phillips [1990]). Neither of these approaches is attempted here.

■ 9 It would be interesting to add the currently confidential data on 30- to 89-day nonperforming loans to the model to see if this resulted in a substantial increase in explanatory power. Such data are likely to be highly correlated with classified loans and are available at regular intervals.

■ 10 See Whalen (1990). The leading-indicator variables could also reflect the divergence between actual and anticipated local economic conditions, which should be an important determinant of bank asset quality and therefore of the probability of failure.

TABLE 1

PHM Estimation Results

Covariate	Coefficient	Standard error	t statistic
LAR	0.0242	0.0055	4.39
OHR	0.1766	0.0339	5.21
ROA	–0.0499	0.0193	–2.58
CD100R	0.0105	0.0050	2.07
NPCR	–0.1419	0.0086	–16.56
PCHP64	–0.0120	0.0019	–6.26

NOTE: Model chi square with six degrees of freedom: 1490.75. See appendix for variable definitions.
SOURCE: Author's calculations.

IV. Model Estimation and Results

I derive the survivor function from the underlying hazard function that is actually estimated.[11] Although the focus in this study is on the former, it should be noted that the coefficients from the hazard function appear in the survivor function unchanged. As a result, in a survivor function, coefficients can be expected to exhibit counterintuitive signs. Variables that are expected to be positively associated with the probability of survival, like return on assets (*ROA*), will exhibit negative coefficients. Similarly, variables that are expected to be negatively associated with the probability of survival, such as the overhead expense ratio (*OHR*), will have positive coefficients.

The survivor function consists of estimated baseline survival probabilities ($S_0[t]$ for various t's) and a vector of estimated model coefficients (the B vector), which I use to generate survival probabilities for banks, given their particular set of characteristics. I estimate a number of alternative models with differing sets of explanatory variables. The estimation results for one of these model specifications appear in table 1. I focus only on a single model because this allows the classification results to be examined in detail. However, I obtain similar classification results using the other specifications.

All of the estimated coefficients exhibit the correct sign and are highly significant. However, it should be noted that, as in multiple regression, collinearity among explanatory variables can be and is a problem. Therefore, this specification, like the others examined, is necessarily parsimonious. The variables that consistently exhibit the strongest statistical relationships to the probability of bank survival are *OHR*, the large certificate of deposit dependence ratio, the loan to asset ratio, the primary capital ratio, the nonperforming loan ratio, the net primary capital ratio, and the change in housing permits variable. It is interesting to note that the commercial real estate loan variable is never found to be significant in any version of the equation estimated, possibly reflecting the somewhat aggregated form of the variable used. A construction loan variable was not employed, and this type of activity is generally viewed as the riskiest form of commercial real estate lending.

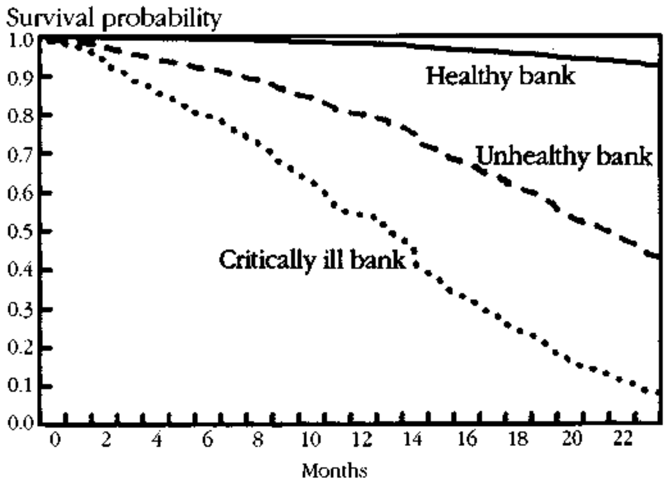
As noted above, the models estimated here can be used to generate the probability that a bank will survive longer than t units, where t can take on any value from roughly zero to 24 months. This is done by substituting the relevant X, B, and baseline survival probabilities into equation (5). Allowing t to vary over the entire permissible range for a bank with some given set of characteristics results in the survival profile for that bank. Thus, this profile shows the probability that some particular bank will survive longer than each possible t value, and vividly portrays the model's estimate of the health of a particular institution. Three illustrative profiles are presented in figure 1.

The top curve depicts the survival profile for a typical "healthy" bank. This profile is derived by inserting the 1986 mean values of the explanatory variables for the nonfailed banks in the estimation sample into the estimated survivor function. Thus, the curve shows that the estimated probability of a healthy bank surviving longer than any number of months ranging from roughly zero to 24 is high — above 0.9. The intermediate profile is for a hypothetical "unhealthy" bank. In this case, the explanatory variable values are set at the 1986 mean value for the banks in the estimation sample that failed in 1988 (that is, those that survived roughly 12 to 24 months into the future). The vertical distance between the two curves represents the estimated reduction in survival probability for the unhealthy bank relative to the healthy bank at every time horizon. The estimated probability

■ 11 I used the Survival Module of SYSTAT to estimate the model, a routine that employs the partial likelihood approach to estimate the B coefficients. This approach does not require that the form of the baseline hazard be specified. For tied failure times, I use Breslow's generalization of the Cox likelihood function. For details, see Steinberg and Colla (1988), appendix C.

FIGURE 1

Survivor Profiles for Three Hypothetical Banks



SOURCE: Author's calculations.

TABLE 2

In-Sample Classification Accuracy

Time horizon (months)	Type I	Type II
12	23 (14.0)	139 (11.8)[a]
18	34 (13.5)	115 (10.6)[b]
24	36 (10.8)	94 (9.3)

a. Of the 139 banks, 107 subsequently failed within 12 to 24 months. Thus, only 32 type II errors occurred for nonfailed banks (3.2 percent).
b. Of the 115 banks, 60 subsequently failed after 18 months. Thus, only 55 type II errors occurred for nonfailed banks (5.5 percent).
NOTE: In the estimation sample, 164 banks failed within 12 months, 252 failed within 18 months, and 333 failed within 24 months. The number of nonfailed banks used in the estimation sample is 1,008. Thus, the type II error rates at the 12-month, 18-month, and 24-month horizons are based on 1,177 banks, 1,089 banks, and 1,008 banks, respectively. The percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

of the unhealthy bank surviving longer than 24 months is roughly 0.46. The bottom curve is the survival profile for a hypothetical "critically ill" bank: The values of all the explanatory variables are set at the 1986 mean values for those banks in the estimation sample that failed within 12 months (that is, 1987 failures). Because the values of the explanatory variables for this group of banks are indicative of very high risk and greater likelihood of failure, the survival profile lies well below that of both of the other groups. The estimated 24-month survival probability for the critically ill bank is just 0.11.

Tables 2 through 8 present the classification results produced using the estimated model. The analysis of classification accuracy and the types of classification errors made using an estimated model are the acid tests of the worth of a potential early warning model.

In the analysis presented here, I focus only on predicted 12-, 18-, and 24-month survival probabilities. In order to use the estimated models to classify banks as failures or nonfailures at each of these time horizons in and out of sample, the generated survival probabilities must be compared to some critical probability cutoff value. Typically, the proportions of failed and nonfailed banks in the estimation sample are used to determine the cutoff values. This is the approach taken here. In the estimation sample used in this study, the probabilities of a bank surviving beyond 12, 18, and 24 months are roughly 0.88, 0.81, and 0.75, respectively. These are the cutoff values used in the analysis. Thus, if a bank's estimated 24-month survival probability is less than 0.75, it is predicted to fail within two years. If its estimated survival probability is greater than 0.75, it is predicted to survive longer than 24 months.

Type I and type II errors are defined in the typical fashion: The former is a bank that failed over some specified time horizon during which it was predicted to survive, and the latter is a bank that survived beyond some specified time horizon during which it was predicted to fail. Both types of errors are important in evaluating the potential usefulness of an early warning model. Obviously, a good model should exhibit low type I error rates. Missing failures typically implies delayed resolution, higher resolution costs, or both. However, if an early warning model is to be useful in allocating scarce examination resources, type II error rates should also be relatively low. One exception to this general rule is illustrated below. In particular, the categorization of a prediction as a type II error depends on the time period and the time horizon examined. Some type II errors could actually represent banks that ultimately fail in some future period. In evaluating the accuracy of any early warning model, it is useful to identify how many banks fall into this category of type II error, since they actually represent a success.

The estimated models are quite accurate in-sample (see table 2). The type I and type II error rates are typically in the 10 to 15 percent range, and the overall classification accuracy is

TABLE 3

Out-of-Sample Classification Accuracy: 1988 Failed Banks

Time horizon (months)	Type I	Type II
	1987 Data	
12	21 (12.4)	—
18	18 (10.7)	—
24	11 (6.5)	—

NOTE: Total number of failed banks in the sample is 169. Percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

TABLE 4

Out-of-Sample Classification Accuracy: 1989 Failed Banks

Time horizon (months)	Type I	Type II
	1987 Data	
12	—	121 (73.3)
18	13 (15.5)	67 (40.6)
24	12 (7.3)	—
	1988 Data	
12	20 (12.1)	—
18	14 (8.5)	—
24	7 (4.2)	—

NOTE: Total number of failed banks is 165. Of these, 84 failed in the first six months of 1989. Percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

above 85 percent for the 12- and 18-month time horizons. The results for the 24-month time horizon are slightly better. Furthermore, a relatively large proportion of the type II errors at the 12- and 18-month time horizons are banks that ultimately failed before 24 months elapsed. Thus, the model was signaling that these banks were potential failures prior to their actual closure.

However, the important yardstick of success for a failure prediction or early warning model is its out-of-sample forecasting accuracy. To obtain insight on this issue, I use the estimated model to generate survival probabilities for all banks in the estimation and holdout samples using data for 1987, 1988, and 1989. Obviously, data are not available for all banks for all years. For example, only 1987 data exist for the 1988 failures. I never reestimate the model coefficients, and use the same cutoff values detailed above. The results for every year are presented for each of the various subsamples in tables 3 to 8.

Turning first to table 3, it is apparent that the model does a relatively good job of identifying the 1988 failed banks. The type I error rate declines from 12.4 percent at the 12- month horizon to 6.5 percent at the 24-month horizon. No type II errors are possible for this subsample.

Table 4 shows the results for the 1989 failures using 1987 and 1988 data. Note that the type I error rates remain relatively low. A look at the type II errors again demonstrates that the model does a reasonably good job of providing an early warning of high-risk banks. For example, using 1987 data, 73.3 percent of the 1989 failures were predicted to fail within 12 months (that is, by year-end 1988).

Results for the 1990 failures (table 5) are similar. The type I error rates are virtually the same as those for the banks that failed in previous years. And again, relatively high proportions of the banks that ultimately failed in 1990 are identified as potential problems in 1987 and 1988.

Table 6 contains the 1987–1989 results for the nonfailed banks used in the estimation sample. Because none of these banks failed, no type I errors are possible. The number and rate of type II errors for this nonfailed subsample are quite low. Table 7 contains virtually identical results for a holdout sample of nonfailed banks.

Finally, table 8 presents results for the largest possible sample. The total number of banks and the numbers classified as failures and nonfailures necessarily change through time. For the 1987 data, for example, the total number of failed banks at the 12-month time horizon consists of all the 1988 failures. The total number of nonfailed banks consists of the 1989 and 1990 failures and the roughly 1,500 nonfailed banks in the estimation and holdout samples. At the 18-month time horizon, those banks that failed in the first six months of 1989 are removed from the nonfailed subsample and considered to be failures. At the 24-month time horizon, all of the 1989 failures are removed from the nonfailed subsample and counted as failures. I use the same procedure to define the subsamples in subsequent years. This exercise perhaps gives the best idea of the potential usefulness of a PHM as an early warning model.

TABLE 5

Out-of-Sample Classification Accuracy: 1990 Failed Banks

Time horizon (months)	Type I	Type II
	1987 Data	
12	—	69 (56.7)
18	—	87 (71.3)
24	—	97 (79.5)
	1988 Data	
12	—	92 (75.4)
18	9 (10.1)	25 (20.5)
24	9 (7.4)	—
	1989 Data	
12	15 (12.3)	—
18	8 (6.6)	—
24	2 (1.6)	—

NOTE: 1990 failed-bank data through October 31. Total number of failed banks is 122. Of these, 89 failed in the first six months of 1989. Percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

TABLE 6

Out-of-Sample Classification Accuracy: Nonfailed Estimation Sample

Time horizon (months)	Type I	Type II
	1987 Data	
12	—	29 (2.9)
18	—	58 (5.8)
24	—	101 (10.1)
	1988 Data	
12	—	33 (3.3)
18	—	68 (6.7)
24	—	116 (11.5)
	1989 Data	
12	—	20 (2.0)
18	—	41 (4.1)
24	—	78 (7.7)

NOTE: Total number of nonfailed banks in the estimation sample is 1,008. Percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

The model appears to perform quite well. In each year, type I error rates are relatively low for all three time horizons. Similarly, type II error rates are also quite low, particularly if the impact of misclassification of subsequent failures is considered. For example, when 1987 data are used and subsequent failures are excluded, the type II error rates for the 12-, 18-, and 24-month horizons fall to 2.7 percent, 5.6 percent, and 9.4 percent, respectively. As noted above, type II errors attributable to misclassification of banks that ultimately fail are not undesirable but rather indicate the ability of the model to identify subsequent failures early. The model appears to perform this task quite well.

The fact that the classification accuracy does not decline over time even though the model coefficients are not reestimated is encouraging. It indicates that the relationship between the explanatory variables and bank survival probabilities represented by the estimated model is relatively stable. This is a desirable characteristic of an early warning model, since it obviates the need to update the model coefficients or to change the specification frequently.[12]

V. Summary and Conclusions

The results strongly suggest that a PHM with a relatively small number of explanatory variables constructed only from publicly available data could be an effective early warning tool. The overall classification accuracy of the estimated model is high, while both type I and type II error rates are relatively low. Furthermore, the model flags a considerable proportion of failures early.

Many further refinements (in variables or in specification, for example) are possible. In particular, it would be interesting to determine if the currently confidential data on 30- to 89-day nonperforming loans would have a significant impact on the explanatory power of this type of equation. It would also be interesting to investigate the relationship between the model's predictions and CAMEL ratings, which reflect additional nonpublic information generated at considerable cost.

■ 12 Although the errors are not examined in detail, a cursory review reveals that a considerable number of the type II out-of-sample errors involved Texas banks. This is an important consideration given the earlier discussion of insolvency versus failure, and may imply that the accuracy of the model is even slightly higher than indicated by the classification results reported in the tables.

TABLE 7

Out-of-Sample Classification Accuracy: Nonfailed Holdout Sample

Time horizon (months)	Type I	Type II
	1987 Data	
12	—	12 (2.4)
18	—	26 (5.1)
24	—	40 (7.8)
	1988 Data	
12	—	8 (1.6)
18	—	26 (5.1)
24	—	43 (8.4)
	1989 Data	
12	—	8 (1.6)
18	—	17 (3.3)
24	—	36 (7.1)

NOTE: Total number of nonfailed banks in the holdout sample is 510. Percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

TABLE 8

Out-of-Sample Classification Accuracy: Maximum Sample

Time horizon (months)	Type I	Type II
	1987 Data	
12	21 (12.4)	231 (12.8)[a]
18	31 (12.3)	238 (13.8)[b]
24	23 (6.9)	238 (14.5)[c]
	1988 Data	
12	20 (12.1)	133 (8.1)[d]
18	23 (9.1)	119 (7.7)[e]
24	22 (7.7)	159 (10.5)
	1989 Data	
12	15 (12.3)	28 (1.8)
18	8 (6.6)	58 (3.8)
24	2 (1.6)	114 (7.5)

a. 190 of these subsequently failed.
b. 154 of these subsequently failed.
c. 97 of these subsequently failed.
d. 92 of these subsequently failed.
e. 25 of these subsequently failed.
NOTE: When year-end 1987 data are used, the sample consists of the 1988, 1989, and 1990 failures and the nonfailed estimation and holdout samples. The number of failed and nonfailed banks at each time horizon depends on the year and time horizon examined. The percentage of banks misclassified is in parentheses.
SOURCE: Author's calculations.

Finally, it will be interesting to see how accurately the model forecasts failures in 1991 and beyond. Some believe that the reasons why banks are encountering financial difficulties at present are somehow different than those faced during the 1980s by southwestern banks, which make up a large part of the sample used to estimate this model. Many argue that effective monitoring of bank financial conditions requires disclosure of additional detailed information on the market value of assets and liabilities. If the estimated PHM exhibits the same degree of accuracy reported here over the next several years, it suggests that neither of these views is correct.

APPENDIX

Estimation Sample: 1986 Mean Values[a]

	1987 failures	1988 failures	Nonfailures
LAR	63.28	64.41	50.82
COMLR	19.33	21.03	11.25
CRELR	9.39	13.14	8.92
CD100R	18.19	22.59	8.44
ROA	–4.57	–2.03	0.59
OHR	5.04	4.68	3.34
PCR	4.31	7.26	9.20
NPCR	–5.62	1.52	7.64
NCOR	6.71	3.13	1.53
NPLR	16.01	9.26	3.10
PCHPxy	–32.53	–39.06	6.09
ASSETS	37.36	271.72	238.38

a. Assets measured in millions of dollars. All other variables are measured in percentages.
SOURCE: Author's calculations.

Variable Definitions

LAR: Total loans/total assets
COMLR: Commercial and industrial loans/total assets
CRELR: Commercial real estate loans/total assets
CD100R: Total domestic time deposits in denominations of $100,000 or more/total assets
ROA: Consolidated net income/average total assets
OHR: Operating expenses/average total assets
PCR: Primary capital/average total assets
NPCR: *PCR* less (total nonperforming loans/average total assets)
NCOR: Total net chargeoffs/average net loans plus leases
NPLR: Total nonperforming loans/total loans plus leases
PCHPxy: Percent change in state's residential housing permits measured over the 198*x* to 198*y* period

References

Gajewski, Gregory R. "Assessing the Risk of Bank Failure," in *Proceedings from a Conference on Bank Structure and Competition,* Federal Reserve Bank of Chicago, May 1989, pp. 432–56.

Goudie, A.W. "Forecasting Corporate Failure: The Use of Discriminant Analysis within a Disaggregated Model of the Corporate Sector," *Journal of the Royal Statistical Society,* vol. 150 (1987), pp. 69–81.

Kalbfleisch, J.D., and R.L Prentice. *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons, Inc., 1980.

Kiefer, Nicholas M. "Economic Duration Data and Hazard Functions," *Journal of Economic Literature,* vol. 26 (June 1988), pp. 646–79.

Lane, William R., Stephen W. Looney, and James W. Wansley. "An Application of the Cox Proportional Hazards Model to Bank Failure," *Journal of Banking and Finance,* vol. 10, no. 4 (December 1986), pp. 511–31.

Phillips, Keith R. "The Texas Index of Leading Economic Indicators: A Revision and Further Evaluation," Federal Reserve Bank of Dallas, *Economic Review,* July 1990, pp. 17–25.

Randall, Richard E. "Can the Market Evaluate Asset Quality Exposure in Banks?" Federal Reserve Bank of Boston, *New England Economic Review,* July/August 1989, pp. 3–24.

Steinberg, Dan, and P. Colla. *Survival: A Supplementary Module for SYSTAT.* Evanston, Ill.: Systat, Inc., 1988.

Thomson, James B. "An Analysis of Bank Failures: 1984–1989," Federal Reserve Bank of Cleveland, Working Paper 8916, December 1989.

Whalen, Gary. "Time Series Forecasts of Regional Economic Activity: Recent Evidence from Ohio," Federal Reserve Bank of Cleveland, unpublished manuscript, 1990.