Predicting failure in the commercial banking industry

Tatom, John

Networks Financial Institute at Indiana State University

5 August 2011

Predicting Failure in the Commercial Banking Industry*

August 5, 2011

Reza Houston, MBA
Scott College of Business
Indiana State University
800 Sycamore St.
Terre Haute, IN 47809
Reza.houston@indstate.edu

John A. Tatom
Director of Research
Networks Financial Institute
Scott College of Business
800 Sycamore St.
Terre Haute, IN 47809
John.tatom@indstate.edu

*This article is preliminary and confidential and may not be quoted without permission. We are indebted to David VanHoose for comments; no responsibility for errors or shortcomings is implied.

Predicting Failure in the Commercial Banking Industry

Abstract

The ability to predict bank failure has become much more important since the mortgage foreclosure crisis began in 2007. The model proposed in this study uses proxies for the regulatory standards embodied in the so-called CAMELS rating system,[1] as well as several local or national economic variables to produce a model that is robust enough to forecast bank failure for the entire commercial bank industry in the United States. This model is able to predict failure (survival) accurately for commercial banks during both the Savings and Loan crisis and the mortgage foreclosure crisis. Other important results include the insignificance of several factors proposed in the literature, including total assets, real price of energy, currency ratio and the interest rate spread.

Classification codes: G01, G17, G21, G33

Keywords: bank failure; banking crises; CAMELS ratings

I. Introduction

Since 2007, the ability to forecast bank failure has become quite relevant. Early in the mortgage foreclosure crisis (and since 1995), there were very few bank failures. For example, from 2003 through 2007, an average of two commercial banks failed each year. However, the last several years have seen more bank failures than in any period since the early 1990's. As a result of the Savings and Loan crisis (S&L crisis), which occurred from the 1980s to

[1] Prior to January 1, 1997, the Federal Reserve used the CAMEL system, which does not include a variable for sensitivity to risk (S). CAMEL refers to scores for capital adequacy , quality of assets, management, equity and liquidity, respectively. . http://www.federalreserve.gov/boarddocs/press/general/1996/19961224/

approximately 1994, Congress passed the Financial Institutions Reform, Recovery, and Enforcement Act of 1989 (FIRREA) and the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA). These tightened restrictions on savings and loan institutions and helped the FDIC to avoid failures and costly bailouts. It was generally believed that these laws, especially the latter, had solved the "too big to fail" problem.

The current state of the economy, including historically low interest rates, and the ability of Government Sponsored entities, investment banks and other real estate financial companies to continue to employ many of the practices that led to the mortgage foreclosure crisis suggest that the relatively high rate of bank failures will continue into the immediate future. The usefulness of a model that can predict failure accurately remains high and is likely to remain so for future financial crises, recessions or other periods of relatively large numbers of failures.

Over the years, researchers have used a wide variety of methods to predict bank failure. Failure prediction models are normally constructed using either internal or external bank data. External data include a variety of economic measures that affect a bank's income and cost and asset values, thereby affecting its solvency. Internal data collected by regulators is often broken down into six main categories- capital adequacy, asset quality, management, earnings, liquidity, and sensitivity to risk in order to compute CAMELS ratings. These ratings give a bank a score from one to five with one being the highest rank and five being the lowest. Regulators normally review each bank annually since CAMELS ratings have been shown become stale more quickly than off-site models using call report data. (Cole and Gunther 1995). One primary source of internal data for researchers in the United States, and that is heavily relied upon here, is the quarterly Report of Condition and Income (Call Report), which is available through the FDIC.[2]

This paper demonstrates the effectiveness of binary models in forecasting failure for the entire commercial banking industry. Section II provides a comprehensive review of previous techniques used to predict bank failure and a comparison between the various models. Section III details the data and methods used here. Section IV provides initial results using logit and probit models and examines whether a model appropriate for the S & L crisis remains useful for the mortgage foreclosure crisis. Because the selection of a critical value for assessing the failure probability can have a relatively large effect on the prediction results, the effect a change in the critical value on these results is examined. Section V demonstrates the ability of the model to predict failure in the entire banking industry for crisis periods over the past twenty years. Section VI summarizes key findings and provides conclusions.

II. A Variety of Techniques

The most prominent statistical methods used to predict failure of banks are logistic regression, neural networks, and multivariate discriminant analysis. The most popular intelligence techniques involving neural networks (NN) include multi-layer perception (MLP), back-propagation neural networks (BPNN), and learning vector quantization (LVQ). More recently developed techniques used to determine probability of failure include Support Vector Machine (SVM) and trait recognition. The variety of techniques speaks to the demand for an accurate model. Previous researchers have found prediction accuracy from several of these

[2] Both individual and bulk data are available. Individual quarterly data going back to 2001 are on the FDIC website: www.fdic.gov. Bulk data are available either through the FFIEC Central Data Repository's Public Data Distribution web site: https://cdr.ffiec.gov/public/ and also through the Chicago Federal Reserve in SAS format.

models to be roughly similar to that of logistic regression. A brief summary of these techniques is provided below.

Artificial neural networks are used to model various problems based on multiple layers: generally an input layer, a hidden layer, and an output layer. Use of artificial neural networks is popular in a variety of fields including agriculture, business, and manufacturing. Perhaps the most popular technique, multi-layer perceptron (MLP), utilizes a feed-forward method of mapping sets of inputs onto outputs in which information never moves backward through the model and one or more of the layers are hidden (Boyacioglu, Kara et al. 2009). Since neural networks adapt to given inputs, in order to employ them in the real-world a large variety of data must be provided. This is the primary drawback and one reason why these approaches are not used here.

Discriminant analysis (DA) is similar to logistic regression in that the dependent "discriminant score" variable can be thought of as a function of the inputs. This technique differs from logistic regression because discriminant analysis is modeled using a normal distribution. The primary goal of this technique is to determine in which group a particular subject should be placed. Unlike logistic regression, discriminant analysis can be used to place subjects into more than two categories. This method is broken down into several specific categories including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and multivariate discriminant analysis (Demyanyk and Hasan 2009).

Multivariate discriminant analysis (MDA) is popular in several areas, including corporate finance and marketing. This technique has several drawbacks. Aside from the possibly inaccurate assumption that the data follows a normal distribution, MDA also requires equal variance-covariance matrices between groups. However Ohlson (1980) notes that this may not be important if MDA is only used for discriminating between groups.

One of the most recent modeling techniques is support vector machine (SVM) which has often been used for machine learning. Several researchers have found that it is able to predict accurately (Vapnik 1995; Boyacioglu, Kara et al. 2009). Recently, other techniques have emerged for analyzing failure, including case-based reasoning, decision trees, and soft computing. However, these methods have been used infrequently for one reason or another.

The Cox proportional hazard model (PHM) has been used occasionally to determine the likelihood of bank failure over a specific period of time. Unlike several other hazard models, the Cox model is considered a semi-parametric model with both parametric and non-parametric inputs. Research using this model shows that it is able to predict bank failure well in advance (Whalen 1991; Shumway 1999; Wheelock and Wilson 2000).

II.1 Binary Models

Logistic regression has long been applied to predict bank failure. This method uses a non-normal distribution in order to determine the probability of an event occurring. The dependent variable, Z_i, can be calculated based on the odds ratio $P_i/(1-P_i)$.

To determine whether an observation will survive or fail, a critical value, c, is used. In failure literature when $P_i \geq c$, this is defined as a failure. The critical value is generally set at 0.5 which equalizes the area for both success and failure. As Barr and Siems (1999) demonstrate, altering this critical value can significantly change the number of Type I and Type II failures. They provide their results using different critical values and recommend leaving the critical value at 0.5. Ohlson(1980) uses a logit model to analyze corporate bankruptcy and finds that size of a firm and measures of its liquidity, financial structure, and performance are all significant factors in determining bankruptcy probability.

The probit model is another popular binary model used in the failure literature and the principal model used here. Unlike the logit model, probit analysis assumes a standardized normal distribution. Like the logit model, the probit relies on linear regression:

$$I_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \tag{1}$$

where I_i is the unobservable utility index. The probit model also uses a critical value. Typically, if the probability of failure calculated by the model is greater than 0.5, then the bank is expected to fail. Research into these models has found that their accuracy is similar to that of logit models (Hanweck 1977; Barr and Siems 1999).[3]

Kolari, Glennon, et. al. (2002) use a sample of 100 large banks and employ both a logit model and non-parametric trait recognition. They divide their data into an in-sample and out-sample in order to test their ability to predict failure. While their study finds that trait recognition has superior predictive accuracy, the authors note that both models are at least 90% accurate when predicting failure.

Halling and Hayden (2007) employ a two-step survival model by combining survival time analysis and a multi-period logit model. They use data on 1,250 Austrian banks from 1995-2002. Their results indicate that bank age is significant and that their two step model outperforms other logit models. Several further attempts to combine logit and probit models

[3] Gerken and Dimmock (2011) use a probit model to predict failure among investment managers. Their model is quite successful, but does not predict as well as various bank faiure models. The latter benefit from more readily quantifiable performance factors and also may be subject to somewhat more discretion in determining fraud.

with other techniques, such as discriminant analysis, in order to increase accuracy have also been successful (Canbas, Canbas et al. 2005).

Barr and Siems (1999) use proxies for the CAMELS ratings with data envelopment analysis (DEA) efficiency as the proxy for the management quality variable in order to predict failure. They use 445 surviving and 294 failed banks during the S & L crisis. In order to keep control for the wide variance between small, medium, and large banks and also de novo and older banks, the authors use only banks that are at least ten years old and had between $20 million and $200 million in total assets. Barr and Siems find that their model out-performs many previous logistic models. First, prediction accuracy in their one year and two year models are 92.4% and 94.8% respectively. Their results are better than those of Hanweck, whose prediction accuracy was found to be 80.8% and 91.1% respectively. This improvement in accuracy is worth a second look to determine if the model can be applied as well to a broader array of commercial banks and in other time periods.

Data envelopment analysis (DEA) is a method of examining production efficiency. This is done by transforming a given number of inputs into a given number of outputs. DEA is a non-parametric approach which makes it less complex than other productivity methods such as stochastic frontier analysis. It is important to note that there is no equation for the relationship between the inputs and the outputs. DEA has been used in several studies to evaluate bank failure (Barr, Seiford et al. 1993; Barr and Siems 1999; Kao and Liu 2004). Barr and Siems (1996) make use of the CAMELS variables and include one variable indicating local economic conditions. They find that failed banks exhibit significantly less efficiency given one year and two years of quarterly data predicting failure. Kao and Liu (2004) make use of DEA to predict the performances of two dozen banks in Taiwan. They also find that this method is able to accurately predict bank failure.

II.2 A Comparison of Prediction Models

There have been several attempts to find the most accurate model for bank failure. Boyacioglu et. al. (2009) use a data set of 21 failed banks and 44 non-failed banks to test the ability of MLP, CLNN, self-organizing map (SOM), LVQ, MDA, and several other models to predict failure. They use 43 of these banks as a training set and forecast failure likelihood for the remaining 22 banks. Data were collected from 1996-2003, when Turkey faced a rise in bank failures. Twenty financial ratios were collected, with many of the ratios fulfilling the role of the CAMELS variables. The researchers find that while MLP is the most accurate, several other neural networks perform well. Logistic regression performs adequately.

Lee, Booth, and Alam (2005) use four sets of banks to compare the accuracy of BPNN, SOM, logistic regression, and descriptive analysis when predicting bankruptcy of Korean firms. They use models with two year and three year data and determine that BPNN outperforms the other models. The logistic model in their study is able to predict bankruptcy about 65% of the time (Lee, Booth et al. 2005). Min and Lee (2005) examine firm bankruptcy and find that SVM out-performs MDA, BPNN, and a logit model(Min and Lee 2005). De Ceca and Mora (2005) examine 30 varieties of winter barley crops in Spain in order to determine whether varieties could be placed into three categories: accept, reject, and test (unsure). They determine that LVQ outperforms QDA when classifying seeds into categories.

Wu et. al. (2007) examine a new model, referred to as the genetic-based SVM (GA-SVM) model. They use a sample of 88 firms, including 22 failed firms. Their results indicate that the GA-SVM predicts failure of firms better than the BPNN, DA, SVM, logit, and probit models. The real-valued genetic algorithm used in this paper is the first of its kind to predict bankruptcy(Wu, Tzeng et al. 2007). Nonetheless, in this study the most prominent approach, logistic probit and logit models, are employed.

III. Data

Many researchers have developed successful proxies for the CAMELS variables over the last two decades. In several of these, the model accuracy is above 90% on average (Barr and Siems 1999; Kolari, Glennon, et. al., 2002). However, the majority of these studies have focused upon finding model accuracy for a very small and restricted sample of the overall commercial banking industry. This study uses a much larger sample than previous work. Incorporating the most significant elements from previous research should provide a starting point for the development of such an appropriate model.

Data were collected from individual quarterly commercial bank call reports[4] and government data from 1988-1994 and 2006-2010 when bank failures were most common in order to assess the ability of the model. Table 1 provides the variables used.

The final column in Table 1 (RSSDID) identifies the bank-specific measure used to calculate each variable. The FDIC assigns RSSD ID's which correspond to individual items on the financial statements of commercial banks.

The model combines several of the most significant variables from previous studies. The most popular proxy for capturing capital adequacy in previous literature is total equity divided by either total assets or total loans. To remain consistent with Barr and Siems (1999), total loans is used here (C01). As banks trend toward failure, their equity position is likely to decrease, thus a negative relationship is expected between C01 and failure.

[4] Call report data was downloaded from the Chicago Fed using the website: http://www.chicagofed.org/webpages/banking/financial_institution_reports/commercial_bank_data.cfm

Table 1: Description of Variables		
Variable	**Description**	**RSSDID**
Capital adequacy	Total Equity Capital/ Total Loans & Leases	RCFD3210/RCFD1400
Asset Quality	Total Loans Not Accruing + Loans 90+ Days Late/Total Assets	(RCFD1403+RCFD1407)/RCFD2170
Management Quality	See Below	
Earnings	Net Income / Total Assets	RCFD4340/RCFD2170
Liquidity	Total Loans & Leases, Gross/ Total Deposits	RCFD1400/RCFD2200
SPREADLAG	1-qtr lagged spread between 10 yr. Treasury notes and 3-mo. Treasury bills	
RPELAG	1-qtr lagged change in the natural log of real price of energy	
FOREC_INV	Percentage of total mortgages in establishment state in foreclosure inventory	
AGE	Current year- Date of Establishment	
TA	Ln(TA)	Ln(RCFD2170)
DEA Inputs		
Input 1	Cash	RCFD0010
Input 2	Salaries and Benefits	RCFD4135
Input 3	Other real estate owned + Premises and Fixed Assets	RCFD2150 + RCFD2145
Input 4	Other noninterest expense	RCFD4092
Input 5	Deposits + Fed Funds Purchased + Mortgage obligations + Bank Liabilities on Acceptances + Subordinated Debt + Other Borrowed Money	RCFD6636+RCFD2800+(RCFD2950-RCFD2200-RCFD2800-RCFD2930)+RCFD3200+RCFD2850
DEA Outputs		
Output 1	Total loans and leases, net of unearned income	RCFD2122
Output 2	Securities Held to Maturity + Securities available for sale + Federal funds Sold and securities purchased + Investments in unconsolidated subsidiaries and associated companies	RCFD0390+RCFD2146+RCFD1350+RCFD2130
Output 3	Deposits—noninterest bearing	RCFD6631

The asset quality variable used is loans not accruing plus loans over 90 days late/ total assets. This ratio (A) increases as banks holds more uncollectible debt on their books, so a positive relation is expected. DEA efficiency is used here to capture the quality of management (M).[5] Like Barr and Siems (1996), this model uses proxies for the CAMELS ratings, with DEA efficiency as a proxy for the management variable. Since better management is associated with an improvement in efficiency, the management variable (M) is expected to be negatively related

[5] DEA efficiency is calculated using FEAR software written by Dr. Paul Wilson for use in the R statistical program. Free downloads of this software are available on Dr. Wilson's website at: http://www.clemson.edu/economics/faculty/wilson/Software/FEAR/agree-to-license.html (Accessed on 6/17/2011)

to failure. The model developed by Haslem et al (1999) is used to construct the DEA efficiency measure.

For earnings, the most popular ratio in the literature (net income/total assets) is used (E) and is expected to have a negative influence on the probability of failure. Finally, the ratio total loans/total deposits is used as a proxy for bank liquidity (L). As banks lend more, they become less liquid and less able to quickly repay their current liabilities. Liquidity has been shown to be a serious cause of concern for failed banks. Therefore, a negative relationship is expected between liquidity and failure.[6]

In order to capture adverse economic conditions, several economic variables were examined. The first is the lagged spread between the 10-Year Treasury yield and the 3-Month Treasury yield, both measured on a constant maturity basis, (SPREADLAG). A larger spread indicates that investors have positive expectations concerning the economy, which should improve future earnings, but the spread also indicates capital losses may not have been fully reflected in the CAMELS measure. Therefore the spread should have a positive effect on the .probability of failure. The second factor (RPELAG) is the change in the natural log of the real price of energy (lagged one quarter). This is added in order to account for energy price changes over the period of the study. Large increases in energy prices reduce economic activity and market income, raising loan defaults, lowering bank income and promoting the rush to failure. A third measure, the natural log of the currency ratio, the ratio of currency held by the public to checkable deposits and sweeps, balances, lagged one quarter, indicates the level of uncertainty and lack of confidence in banks, as well as a broader measure of a lack of liquidity in the economy, according to some monetary economists. During crisis periods, people tend to hold

[6] Bologna (2011) finds that liquidity measures are significant in predicting bank failures in 2008-09, one of the first studies to examine the recent experience, ex post.

more cash relative to bank deposits, straining bank liquidity, increasing fire sales of assets and raising the probability of failure. Another potential influence on the probability of failure is the percentage of mortgages within a bank's incorporation state that are in the foreclosure inventory (FOREC_INV). Larger numbers of foreclosures are expected to result in lost income, deplete capital, and increase the failure rate. The final two variables are the age of each institution (AGE, in years), and the natural log of total assets (TA). Because of the wide range of assets, it was important to identify an effect that size or age might have on an institution's likelihood of failure.

Previous research has often focused only on specific segments of commercial institutions such as non-de novo banks or banks with certain amounts of total assets. For instance, Barr and Siems (1999) use only commercial banks that are non-de novo and have total assets between $20 million and $200 million. The goal here is a model that can be applied to the entire commercial banking industry, without restrictions. To achieve this goal, both the in-sample data and out-sample data were drawn from the commercial banking sector as a whole. In order to eliminate as much bias as possible, several rules were followed when selecting banks for this study. First, all institutions in the initial in-sample are domestic institutions.[7] The ratio of randomly selected surviving banks to failed banks in each quarter was set at 2:1 to provide wide representation of surviving banks. When selecting surviving institutions, first, failed institutions were broken into three total asset sizes: less than $100 million, between $100 million and $1 billion, and greater than $1 billion; twice as many surviving banks were selected from each group using a random number generator. In the end, an in-sample of 1470 banks was obtained, with 963 observations coming from the S & L Crisis period (1990Q1-1993Q4) and 507 observations from the recent crisis period (2008Q3-2009Q4). There are no existing studies of bank failure forecasts or

[7] Institutions in which foreign assets accounted for 100% of total assets were eliminated.

determinants during the foreclosure crisis period. The original forecast sample includes banks' observations during all four quarters of 2010. During that period, there is information on 92 failures and 184 random surviving institutions.

Two years of quarterly data were collected for each institution prior to failure (survival) in order to examine the relative performance of the model using one quarter, one year, and two years of information prior to failure (survival) of institutions. When using the one-year and two-year based prediction models, quarterly data for each variable were averaged to provide a single measure for each institution.

IV. Results

IV.1. Initial Sample

The initial in-sample dataset includes 1470 institutions (for both the S & L and mortgage foreclosure crises), one third of which failed. Both a probit and logit model are used to analyze the initial data. It was found that the capital adequacy, asset quality, and earnings variables are by far the most significant predictors of failure in the model.

Table 2: Probit Model

	2 Year Model			1 Year Model			1 Quarter Model		
Variable	Coefficient	Prob.		Coefficient	Prob.		Coefficient	Prob.	
C	-1.9464	0.0000	**	-1.1940	0.0065	**	-0.6170	0.0903	
C01	-6.4304	0.0000	**	-8.3303	0.0000	**	-11.1219	0.0000	**
A	19.1623	0.0000	**	15.6855	0.0000	**	14.9228	0.0000	**
M	1.9202	0.0000	**	1.8740	0.0000	**	0.8342	0.0215	*
E	-114.5240	0.0000	**	-89.0657	0.0000	**	-36.2035	0.0000	**
L	-0.2445	0.1725		-1.4141	0.0000	**	-0.9921	0.0000	**
SPREADLAG	0.0659	0.4239		0.1620	0.0072	**	0.2277	0.0000	**
McFadden R-squared	0.7139			0.7601			0.7468		

* Significant at the 0.05 level
** Significant at the 0.01 level

Table 2 gives the coefficients and significance of the variables used in the probit regressions. The RPELAG, currency ratio and TA variables are generally not found to be significant in this study and with the exception of SPREADLAG in Tables 2 and 3 are dropped (results are available upon request). Their omission has no detectable influence on the reported results. The foreclosure inventory variable and age are also not found to be significant in this sample and also are dropped.

All of the significant variables have the expected sign with the exception of the efficiency/management measure, which is significant but has the "wrong" sign with both the two-year and one-year information sets. Previous research, such as Barr and Siems (1999), has found a significant negative relation between efficiency and failure in both the two and one year models. However, their study confined banks to a minimum and maximum total asset level and was for a period just before the worst of the S&L crisis. The data here indicate that as the failure date approaches, banks become less efficient. The mean efficiency measure for the failed banks is below that for the surviving banks, for example. To find out what causes this effect, the correlations between the CAMEL variables are examined.

Table 3: CAMEL Correlations in the 2 Year Probit Model

	C01	A	M	E	L
C01	1.0000	-0.0135	0.0216	0.0372	-0.0003
A		1.0000	-0.0119	**-0.4349**	0.0068
M			1.0000	**0.0742**	**0.0619**
E				1.0000	0.0148
L					1.0000

Table 3 shows the correlations between the CAMEL variables in the two-year in-sample information set. The correlations between management quality and earnings, asset quality and earnings, and management quality and liquidity are relatively high and statistically significant at the five percent level. Excluding the earnings variable from the model changes the sign of the efficiency variable to the expected relationship to failure, but the fit of the estimate deteriorates. The correlation between management quality and earnings in the one year and one quarter models are similarly large, 0.1264 and 0.0868, respectively. To anticipate later results, increasing the sample size actually does result in a negative relationship between management quality and failure (see section V). To identify how well the model fits the data, the McFadden pseudo-R^2 measure is used. All three estimates exhibit Pseudo-R^2 values between 0.7139 and 0.7468.

Figure 1: In-sample Probability of Failure
Two Year Probit Model

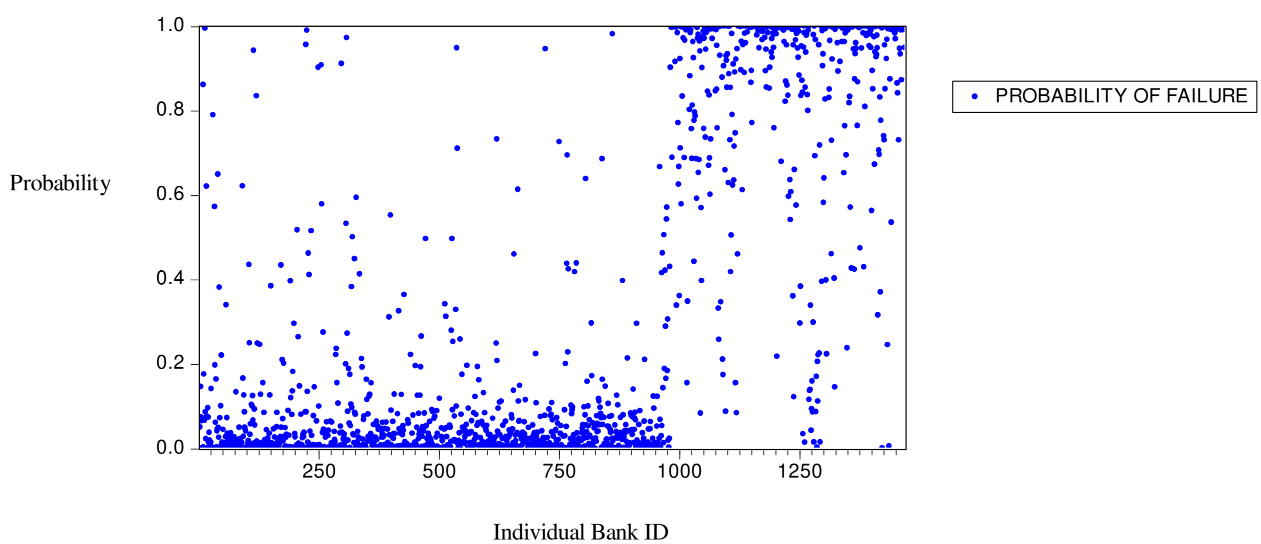


Figure 1 shows a scatterplot of the probability of failure for each bank included during the quarter in which they eventually failed (survived)**.** The probability of failure is between zero and one and banks numbered 1 through 980 on the horizontal axis survived while banks numbered 981 through 1470 failed. There is a clear visual distinction between the two groups of banks.

Table 4: Logit Model									
	2 Year Model			1 Year Model			1 Quarter Model		
Variable	Coefficient	Prob.		Coefficient	Prob.		Coefficient	Prob.	
C	-3.1919	0.0003	**	-1.1459	0.1951		0.6990	0.3831	
C01	-14.7636	0.0000	**	-18.9428	0.0000	**	-28.8225	0.0000	**
A	30.8250	0.0000	**	25.2097	0.0000	**	25.5293	0.0000	**
M	3.4609	0.0001	**	3.2388	0.0002	**	0.9146	0.2209	
E	-216.1389	0.0000	**	-168.5839	0.0000	**	-57.6631	0.0000	**
L	-0.3231	0.4484		-3.2072	0.0000	**	-3.0406	0.0000	**
SPREADLAG	0.1514	0.3534		0.3262	0.0076		0.5802	0.0000	**
McFadden R-squared	0.7201			0.7673			0.7740		

* Significant at the 0.05 level
** Significant at the 0.01 level

The logit model estimates in Table 4 use the same variables as the probit model in Table 2. The results are similar to those in Table 2. The signs of the significant variables are the same as in Table 2. It is important to note the larger pool of data in this study compared with that of Barr and Siems (1999) and others. Because of fewer restrictions on the sample data, one might expect accuracy to be lower, but the sample size is much larger than in previous studies, which presumably compensates with improved accuracy. It is not possible to infer whether this model would be more accurate than previous studies.

Table 5: Percentage of Correct In-Sample Predictions in Probit Model						
	Barr and Siems			**Current Model**		
	Survivals	Failures	Combined	Survivals	Failures	Combined
Two Year	92.1%	82.9%	88.9%	97.55%	86.12%	93.74%
One Year	94.4%	89.5%	92.4%	97.96%	86.73%	94.22%
One Quarter	-	-	-	98.16%	87.35%	94.56%

Table 5 shows the comparison of in-sample accuracy between the Barr and Siems (1999) probit model and the probit model in Table 2 for two years, one year, and one quarter of information. The Barr and Siems (1999) study does not include estimates for a one-quarter window of information. They examine commercial bank failures (survivors) from 1986-1988. In Barr and Siems (1999) and the current model here, accuracy is quite high, though the current model performs slightly better overall. In both cases, overall accuracy of the one-year information variables is better than using the two-year window of information. However, the estimate based on one-quarter data here has slightly better accuracy than that based on one-year information. The superior performance here is all the more impressive given the inclusion of a broader set of banks and the higher failure rates and longer sample here.

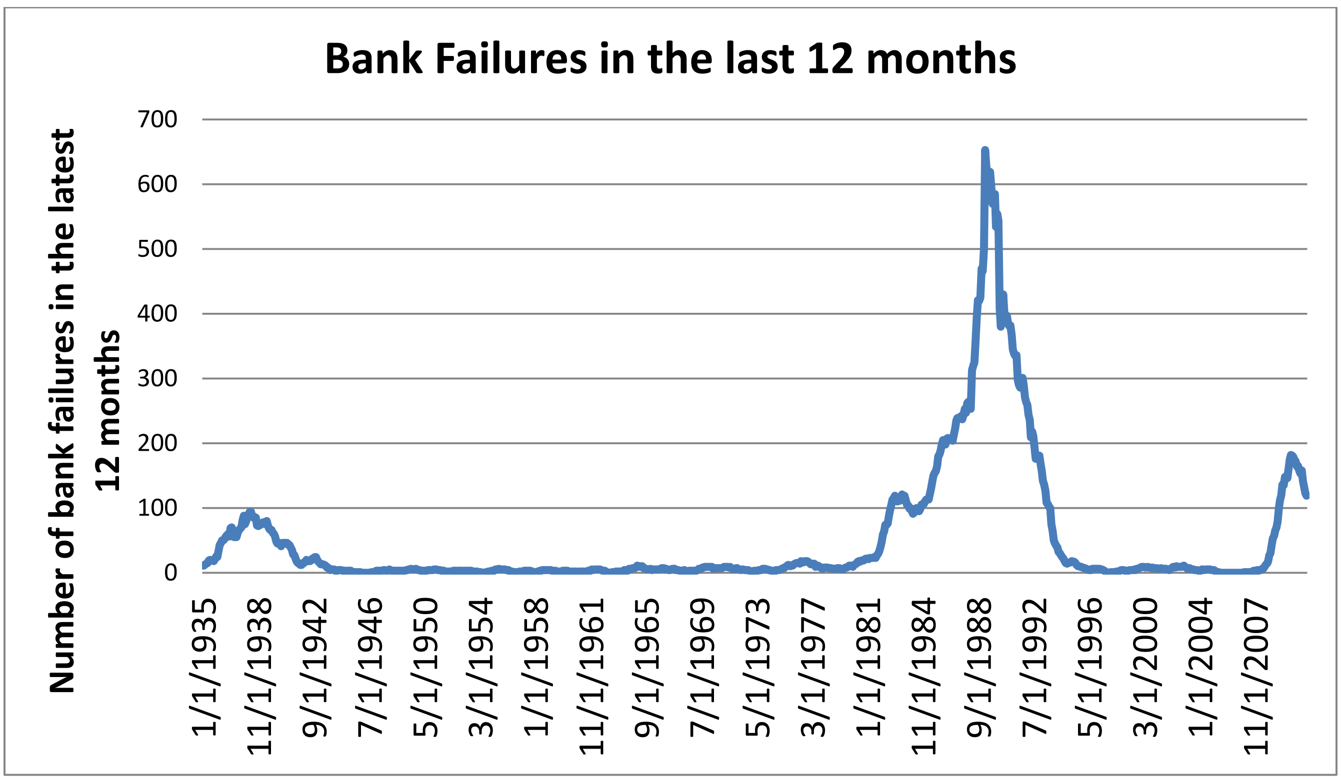
IV.2. Are the S&L and Foreclosure Crises Failure Models Different?

Figure 2 below shows the annual number of failures for the latest 12 month period from 1935 through March 2011. In terms of the number of failures, the S & L Crisis and the mortgage foreclosure crisis were similar. From January 1, 1984 to December 31, 1994 a total of 131 commercial banks failed annually compared with 122 annually from 2008Q3 to 2010Q3. It appears that regulations such as FIRREA and FDICIA have not been as successful as previously thought, although this could be due to the deregulation of the banking industry during the mid and late 1990's or to differences in the types or extent of shocks.

It should be emphasized, however, that 2,655 banks failed from 1984 through 1994, while only 306 failed in the nearly three years of the mortgage foreclosure crisis covered here. The annual pace of failure has declined from a peak of 182 in the twelve months ending in May, 2010 to 119 in the twelve months ending in June 2011.

Other differences between the S & L Crisis and the mortgage foreclosure crisis include the proportion of commercial bank to savings bank failures and the presence of "zombie" banks. Between 1984 and 1994, 1448 commercial banks and 1207 savings institutions failed. During the first two years of the mortgage foreclosure crisis, 250 commercial banks and 51 thrifts failed. The difference in the proportion of commercial banks to savings banks illustrates that the two crises might affected banks differently. During the S & L crisis, the presence of zombie banks and S&Ls (institutions whose net worth is less than zero but who continued to operate) helped drive up CD rates and drive down loan rates. These institutions were willing to take greater risks in order to remain solvent which led to increased cost pressures for other banks and at least partially contributed to the failure of other, more solvent banks.

Figure 2
Bank Failures Show Two Surges since the 1930s



Source: FDIC

Several changes have taken place since the S & L crisis including the rapid merging of banks, which accounts for the larger size of banks on average today, though size does not appear to affect the probability of failure here. Certainly among the top 10 banks or so, recent experience indicates that too-big-to-fail remains in force. Some might argue that the recent period of failures is fundamentally different from the previous high failure period because it was associated with the foreclosure and financial crises and the Great Recession. However, both periods were associated with excessive real estate loan problems. The key difference may simply be that there were far fewer thrift institutions after 1994, even relative to the decline in the number of banks.

IV.3. Forecasting

To test the flexibility and forecasting ability of the model in more detail, the estimate in Table 2 is used to forecast failures during an out-of-sample period, the four quarters of 2010. The critical value for failure probability (c) is set equal to 0.50, following previous research. If the probability of failure calculated from the probit model was greater than this value, it is classified as a failure. Raising the critical value for the criterion for failure raises type II error for failure, but reduces the extent of type I error, incorrectly identifying a failing bank as a survivor. A policy maker or regulator is interested in a highly accurate forecast of failure, but also, if not more so, in minimizing type I error. The results indicate that the model is able to accurately predict bank failure for randomly selected surviving banks and all actual failing banks. As expected (not shown), the results from both the logit and probit models are quite similar with negligible differences in prediction ability. The coefficients in Table 2 are used to forecast survival and failure in Table 6.

It should be noted that the current model is able to predict failure accuracy even given only one-quarter data (panel 6c). Accuracy is generally better in the out-of-sample period than in-sample, except for survivors with two-year and one-year data. The combined and failure forecasts outperform the in-sample forecasts and both exceed 95 percent with all three information sets.

Table 6
Model Accuracy In-Sample and 2010 Out-of-Sample Forecasts
Using Various Information Windows for Data

Panel 6a: 2-Year Probit Model						
	In-sample			Out-sample		
	Survive	Fail	Combined	Survive	Fail	Combined
% Correct	97.55%	86.12%	93.74%	93.48%	98.91%	95.29%
% Incorrect	2.45%	13.88%	6.26%	6.52%	1.09%	4.71%

Panel 6b: 1-Year Probit Model						
	In-sample			Out-sample		
	Survive	Fail	Combined	Survive	Fail	Combined
% Correct	97.96%	86.73%	94.22%	92.93%	98.91%	94.93%
% Incorrect	2.04%	13.27%	5.78%	7.07%	1.09%	5.07%

Panel 6c: 1-Quarter Probit Model						
	In-sample			Out-sample		
	Survive	Fail	Combined	Survive	Fail	Combined
% Correct	98.16%	87.35%	94.56%	95.65%	96.74%	96.01%
% Incorrect	1.84%	12.65%	5.44%	4.35%	3.26%	3.99%

IV.4. Forecasting Failure in a Larger Sample

A second experiment (Table 7) uses data on all commercial banks for an out-of-sample forecast for the first three quarters of 2010. In the first quarter of 2010, there are complete data for 7186 individual commercial banks, 37 of which failed. In the second quarter, there are complete data for 7112 banks, 40 of which failed. In the third quarter, there is complete information for 7014 commercial banks, 32 of which failed. Thus, there are 21312 observations in the out-of-sample period, the first three quarters of 2010.

The model is able to accurately forecast failure with all three information sets. The estimate that uses only one-quarter data is able to predict failure accurately for 96.33% of failed banks, the best combined accuracy. All three estimates have an overall out-of-sample forecasting accuracy of at least 92% for the all bank out-of-sample forecasts. This is significant because it indicates that the proxies for the CAMELS variables can be successfully applied to the entire commercial banking industry.

Table 7a: 2-Year Probit Model with All Commercial Banks out-sample

	In-sample			Out-sample		
	Survive	Fail	Total	Survive	Fail	Total
% Correct	97.55%	86.12%	93.74%	94.15%	98.17%	94.17%
% Incorrect	2.45%	13.88%	6.26%	5.85%	1.83%	5.83%
# of Obs.	980	490	1470	21203	109	21312

Figure 7b: 1-Year Probit Model with All Commercial Banks out-sample

	In-sample			Out-sample		
	Survive	Fail	Total	Survive	Fail	Total
% Correct	97.96%	86.73%	94.22%	92.13%	99.08%	92.16%
% Incorrect	2.04%	13.27%	5.78%	7.87%	0.92%	7.84%
# of Obs.	980	490	1470	21203	109	21312

Figure 7c: 1-Quarter Probit Model with All Commercial Banks out-sample

	In-sample			Out-sample		
	Survive	Fail	Total	Survive	Fail	Total
% Correct	98.16%	87.35%	94.56%	95.61%	96.33%	95.62%
% Incorrect	1.84%	12.65%	5.44%	4.39%	3.67%	4.38%
# of Obs.	980	490	1470	21203	109	21312

Both probit and logit models produce similar forecasting results. However, because of the large number of surviving institutions, even a small decrease in prediction accuracy for survival could lead to a large number of banks being incorrectly predicted to survive.

One change that could dramatically affect the forecasting accuracy of binary models is the selection of a critical value for failure. According to Barr and Siems, a change in the critical value used in the model could significantly alter the number of Type I and Type II prediction errors. Table 8 demonstrates how a change in the critical value alters the predictive accuracy during the forecasting period. Increasing the critical value for the probability of failure makes it more likely that some predicted values will be classified as a survivor, and vice versa. Since the number of surviving institutions vastly outweighs the number of failed institutions, it may be appropriate to increase the critical value in order to increase the total predictive accuracy of the model.

Overall predictive accuracy does not decrease even when the critical value is set to 0.8. However, the one quarter model's ability to predict failure accurately does fall below 91% when the critical value is raised to 0.8. It appears that there are several ideal critical values that could be used depending on the researcher's objectives.

Table 8: Forecasting Accuracy when Altering the Critical Value

Critical Value		2-year Model			1-Year Model			1-Qtr Model		
		Survivals	Failures	Combined	Survivals	Failures	Combined	Survivals	Failures	Combined
0.5	% Correct	94.15%	98.17%	94.17%	92.13%	99.08%	92.16%	95.61%	96.33%	95.62%
	% Incorrect	5.85%	1.83%	5.83%	7.87%	0.92%	7.84%	4.39%	3.67%	4.38%
0.6	% Correct	95.12%	97.25%	95.13%	93.18%	98.17%	93.20%	95.61%	96.33%	95.62%
	% Incorrect	4.88%	2.75%	4.87%	6.82%	1.83%	6.80%	4.39%	3.67%	4.38%
0.7	% Correct	96.00%	94.50%	95.99%	94.12%	98.17%	94.14%	96.41%	96.33%	96.41%
	% Incorrect	4.00%	5.50%	4.01%	5.88%	1.83%	5.86%	3.59%	3.67%	3.59%
0.8	% Correct	96.86%	88.99%	96.82%	95.04%	98.17%	95.05%	97.26%	90.83%	97.23%
	% Incorrect	3.14%	11.01%	3.18%	4.96%	1.83%	4.95%	2.74%	9.17%	2.77%

V. Estimates and Forecasts for the Entire Commercial Banking Industry

In the previous section the management quality measure has an unexpected positive effect on commercial bank failure. In addition, several of the other variables examined there are not significant at conventional levels. Because previous researchers have found some of these variables to be significant, the dataset was expanded to include all commercial banks that provide complete information in the call reports. When including all available commercial banks, the continuing, though reduced, accuracy of the model still stands out and indicates its usefulness.

The call reports offered by the FDIC are again used. This time, all commercial banks with complete data given by the FDIC in the bulk call reports are used. In-sample data for banks during the periods 1990Q1-1993Q4 and 2008Q3-2009Q4 are used. Survivals and failures in the first three quarters of 2010 are forecast, as in the out-of-sample period used in section IV.4. The in-sample data contain 234,109 quarterly records, 552 of which failed during the period, which amounts to a 0.236% rate of failure. Using a probit model, the included variables are restricted to those that are significant with at least two of the information sets.

Above, it is found that CAMEL measures were the most significant variables, while several variables including age of the institution, foreclosure inventory, total assets, the currency ratio, and the lagged real price of energy were found to be insignificant with at least two of the three information sets. Among non-CAMEL variables, only the interest rate spread was significant in the smaller sample. Now, in the all bank sample, it is not, but AGE and the foreclosure inventory usually are.

All variables in this reduced model are significant at well beyond the 1% level. All variables also have the expected effect on failure. Due to the larger sample size, FOREC_INV and AGE, which in the original smaller sample were found to be insignificant at conventional levels, now have the expected relationship with failure and usually are significant at the 0.01 level.

Table 9: Reduced Probit Model Using CAMELS Variables for All Banks

Variable	2 Yr Model		1 Yr Model		1 Qtr Model	
	Coefficient	Prob.	Coefficient	Prob.	Coefficient	Prob.
C	-1.8841	0.0000	-2.0504	0.0000	-1.8714	0.0000
C01	-6.6871	0.0000	-6.7083	0.0000	-7.8080	0.0000
A	14.7397	0.0000	12.5458	0.0000	8.6469	0.0000
M	-0.6812	0.0000	-0.4114	0.0000	-0.5981	0.0000
E	-13.3198	0.0000	-7.8796	0.0000	-5.0945	0.0000
L	4.58E-05	0.0000	3.77E-05	0.0000	4.04E-05	0.0000
FOREC_INV	0.0599	0.0005	0.0335	0.0072	0.0288	0.1439

AGE	-0.0035	0.0000	-0.0033	0.0000	-0.0026	0.0000
McFadden R-squared	0.3609		0.3976		0.4612	

This means that banks operating in states with more foreclosure inventory are more likely to fail. The AGE of institution, which was only significant at the 0.05 level in the original one quarter model, now has a significant negative effect on bank failure with all three information sets. Older institutions are less likely to fail. The pseudo-R^2 measure for all three models is between 0.36 and 0.47, with the one quarter model having the highest R^2.

While the R^2 for the full industry is lower than in the estimates presented in Tables 2 and 3, this appears to be due to more variability in the data, with several extreme outliers skewing the estimates. The commercial banking industry is extremely diverse. Numerous institutions currently operate both as commercial banks and investment banks or insurance companies. Because of this diversity and the number of surviving institutions relative to failed ones, probabilities of failure in the model become extremely skewed to the left.

Table 10 shows the out-of-sample prediction accuracy of the CAMEL variables with foreclosure inventory and age during the first three quarters of 2010. The out-of-sample forecast data includes 21,312 observations over the first three quarters of 2010 including 109 bank failures. This gives a 0.51% failure rate during the forecast period. The table also shows the results for a variety of values of the critical probability value used to classify bank failure.

At the standard value of 0.5, the probit model is unable to predict a high share of failures, but at this value the best combined accuracy of estimates is achieved, and for all three information sets. This is because of the extremely high accuracy of the survivor forecast resulting in very few failures that were erroneously forecast to be survivors.

Table 10: Out-of-Sample Forecasts For the All Bank Model With Various Failure Criteria										
Critical Value		2-year Model			1-Year Model			1-Qtr Model		
		Survivals	Failures	Combined	Survivals	Failures	Combined	Survivals	Failures	Combined
0.5	% Correct	99.97%	8.26%	99.50%	99.81%	24.77%	99.43%	99.95%	16.51%	99.53%
	% Incorrect	0.03%	91.74%	0.50%	0.19%	75.23%	0.57%	0.05%	83.49%	0.47%
0.2	% Correct	99.54%	33.03%	99.20%	99.14%	61.47%	98.95%	99.66%	45.87%	99.39%
	% Incorrect	0.46%	66.97%	0.80%	0.86%	38.53%	1.05%	0.34%	54.13%	0.61%
0.1	% Correct	98.78%	55.96%	98.56%	98.27%	82.57%	98.19%	99.16%	74.31%	99.03%
	% Incorrect	1.22%	44.04%	1.44%	1.73%	17.43%	1.81%	0.84%	25.69%	0.97%
0.05	% Correct	97.39%	82.57%	97.31%	96.97%	88.07%	96.92%	98.28%	88.99%	98.24%
	% Incorrect	2.61%	17.43%	2.69%	3.03%	11.93%	3.08%	1.72%	11.01%	1.76%
0.025	% Correct	95.03%	90.83%	95.01%	95.19%	97.25%	95.20%	96.94%	97.25%	96.94%
	% Incorrect	4.97%	9.17%	4.99%	4.81%	2.75%	4.80%	3.06%	2.75%	3.06%
0.01	% Correct	90.03%	96.33%	90.07%	91.15%	98.17%	91.19%	94.01%	99.08%	94.04%
	% Incorrect	9.97%	3.67%	9.93%	8.85%	1.83%	8.81%	5.99%	0.92%	5.96%
0.005	% Correct	83.40%	97.25%	83.47%	86.05%	99.08%	86.11%	90.53%	99.08%	90.07%
	% Incorrect	16.60%	2.75%	16.53%	13.95%	0.92%	13.89%	9.47%	0.92%	9.93%

Accuracy is more balanced around a critical value of 0.025, where all three information sets yield exceptional accuracy of better than 95 percent for combined accuracy and, with the exception of the failure forecast with two-year information, for both survivors and failures.

If one gives equal weight to minimizing erroneous forecasts of survival in catching failures, the combined out-of-sample forecast is exceptionally good and the 0.5 setting for the critical value is best. If one can adjust the critical value, it appears that this model for only failure forecast accuracy still performs well for the entire industry. All three information sets produce similar prediction results between the in-sample and forecasting sample, but the one-quarter information set has the best in-sample fit and the combined accuracy for the out-of-sample forecast is also the best for the one-quarter information set. A decrease in the critical value is shown to increase the incidence of Type I error among survivors, with the benefit of smaller forecast errors for failures. However, the trade-off is not favorable to overall forecast accuracy. The 0.5 critical value is exceptionally accurate for survivors (99.8 percent to 99.97 percent) and for the combined accuracy. If one desires highly accurate forecasts for the failures only, setting the critical value at 0.025, five times the actual probability of failure in the period, produces balanced and accurate results, including over 90 percent forecast accuracy for failure forecasts only.

VI. Conclusions

In this study, proxies for bank-specific CAMELS variables are combined in a model with data on economic conditions. The goal of the paper is to test the overall accuracy and robustness of the model for the commercial banking industry of the United States. This model, which is similar to that of Barr and Siems (1999) and incorporates the use of the DEA model proposed by Haslem (1999), is found to accurately fit and forecast failure using both probit and logit models with two years, one year, or one quarter of data for measures of the independent variables. The results presented here support those of Barr and Siems (1999). DEA efficiency, used as a proxy for management in the CAMEL approach, is found to have a significant negative relationship with failure in the larger commercial banking industry sample. While it has the "wrong" sign in the smaller sample of about 1500 banks, it contributes significantly to the in-sample fit and forecast accuracy of the model. All CAMEL variables, and age and foreclosure inventory, are significant in the larger sample and have the expected relationship with failure. Older banks and banks in states with relatively lower foreclosure inventory are also found to be less likely to fail in the all bank sample. Other economic variables, such as the real price of energy, total assets, the interest rate spread and currency ratio are not found to be significant in at least two of the models with the larger sample.

The model developed in this study has strong forecasting accuracy in both the in-sample and out-of-sample forecasts. More importantly, the out-of-sample forecasts for the first three quarters of 2010 using data for all U.S. available commercial banks (or for the smaller sample) outperform the in-sample estimates. The criteria used here follows the usual practice of using a Bayesian prior of a 50 percent probability of failure. Experiments with the original sample show that this can be improved upon by moving the probability up to 70 percent, and for the out-of-sample forecasts using all bank data, it is suggested that Type 1 and Type II errors can be minimized by independent variation of the probability of failure toward 80 percent to minimize type II error, though the 50 percent criterion works extremely well for minimizing errors in failure forecasts.

When all available data on commercial banks are used (Section V) the high number of surviving banks relative to failures affects the overall forecast quality, resulting in an optimal 0.5 setting for the critical value, but the forecast accuracy for failures alone could be dramatically improved by lowering the critical value closer to the relatively percentage of failures observed ex post. It seems advisable that the critical value should be set prior to forecasting by examining in-sample prediction accuracy and selecting an ideal critical value that best satisfies the objectives of the forecaster with regard to survivor, failure or combined forecasts.

The model is identified very well in the estimates here because of the use of data from the two most severe periods of bank failure in the United States since the early 1930s. Future work will focus on extending the out-of-sample-forecast window. Also a more decisive effort to sort out the most useful window of information on independent variables would be beneficial, though results here favor the use of very recent data only. Generally the one-quarter information set dominates the others in both the in-sample fits and in the out-of-sample forecasts.

References

Barr, R. and T. Siems (1999). "Bank Failure Prediction Using DEA to Measure Management Quality." R. Barr, R. Helgason and J. Kennington, eds., Computer Science and Operations Research: Advances in Metaheuristics, Optimization and Stochastic Modeling Technologies

Barr, R. S., L. M. Seiford, et al. (1993). An envelopment-analysis approach to measuring the managerial efficiency of banks. Third European Workshop on Efficiency and Productivity Measurement, Louvain, Belgium.

Bell, T. B. (1997). "Neural nets or the logit model? A comparison of each model's ability to predict commercial bank failures" International Journal of Intelligent Systems in Accounting, Finance and Management **6**: 249-264.

Berger, A. N. and L. K. Black (2011). "Bank size, lending technologies, and small business finance." Journal of Banking & Finance **35**: 724–735.

Belogna, Pierluigi (2011). "Is There a Role for Funding in the Recent Bank Failures," IMF Working Paper 11/180, July.

Boyacioglu, M. A., Y. Kara, et al. (2009). "Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey." Expert Systems with Applications **36**: 3355-3366.

Canbas, S., S. Canbas, et al. (2005). "Prediction of commercial bank failure via multivariate statistical analysis of financial structure: The Turkish case." European Journal of Operational Research **166**: 528-546.

Cole, R. (1998). "The Importance of Relationships to the Availability of Credit." Journal of Banking & Finance **22**: 959-977.

Cole, R. and J. Gunther (1995). "Separating the Likelihood and Timing of Bank Failure." Journal of Banking and Finance **17**: 1073-1089.

DeGryse, Hans, Luc Laeven and Steve Ongena (2009). "The Impact of Organizational Structure and Leading technologies on banking Competition, " *Review of Finance,* Oxford University Publishing for the European Finance Association, 13 (2) : 225-259.

Demyanyk, Y. and I. Hasan (2009). "Financial Crises and Bank Failures: A Review of Prediction Methods." Omega 38 (5) 315-324.

Gerken,William and Stephen Dimmock (2011), "Predicting Fraud by Investment Managers," Networks Financial Institute Working Paper 2011-WP-09, May. Forthcoming in the *Journal of Financial Economics.*

Halling, M. and E. Hayden (2007). Bank Failure Prediction: A 2-Step Survival Time Approach. Proceedings of the International Statistical Institute's 56th Session.

Hanweck, G. A. (1977). "Predicting Bank Failure." Research Papers in Banking and Financial Economics (Financial Studies Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System)(November).

Haslem, J. A., C. A. Scheraga, et al. (1999). "DEA efficiency profiles of U.S. banks operating internationally." International Review of Economics and Finance **8**: 165-182.

Kao and Liu (2004). "Predicting bank performance with financial forecasts: A case of Taiwan commercial banks." Journal of Banking and Finance **28**: 2353-2368.

Kolari, J., D. Glennon, et al. (2002). "Predicting Large U.S. Commercial Bank Failures." Journal of Economics and Business **54**: 361-387.

Lee, K., D. Booth, et al. (2005). "A Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Firms." Expert Systems with Applications **29**: 1-6.

Min, J. H. and Y. C. Lee (2005). "Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters." Expert Systems with Applications**28**: 603-614.

Ohlson, J. A. (1980). "Financial Ratios and the Probabilistic Prediction of Bankruptcy." Journal of Accounting Research **18** (11 Spring).

Olmeda, I. and E. Fernandez (1997). "Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction." Computational Economics **10**: 317-335.

Shumway, T. (1999) "Forecasting Bankruptcy More Accurately: A Simple Hazard Model", Journal of Business, 74, 101-124.

Vapnik, V. N. (1995). "The Nature of Statistical Learning Theory." New York: Springer-Verlag.

Whalen, G. (1991). "A Proportional Hazards Model of Bank Failure: An Examination of Its Usefulness as an Early Warning Tool." Economic Review (Q 1): 21-31.

Wheelock, D. C. and P. W. Wilson (2000). "Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions." The Review of Economics and Statistics **82**(1).

Wu, C.-H., G.-H. Tzeng, et al. (2007). "A real-valued genetic algorithm to optimize the parameters of 3 support vector machine for predicting bankruptcy." Expert Systems with Applications **32**(2): 397-408.