

## Modeling Banks' Probability of Default

Xiaoming Tong\*

MBA Student, 310 Greenwich St. Apt 14G New York, NY 10013 USA

\*Corresponding author: Dr. Xiaoming Tong, MBA Student, 310 Greenwich St. Apt 14G New York, NY 10013 USA, Tel: +374 10 23-72-61; E-mail: [katextong@gmail.com](mailto:katextong@gmail.com)

Received date: April 17, 2014, Accepted date: December 18, 2014, Published date: December 28, 2014

Copyright: © 2014 Tong X. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

The unprecedented financial crisis of 2008-2009 has called attention to limitations of existing methods for estimating the default risk of financial institutions. To address this need, I built and tested a time-adaptive statistical model that predicts the default probabilities of banks. The model inputs are a set of financial ratios suggested in the literature, and subsequently verified, to be effective in forecasting future bank failures. The model provides estimates of banks' cumulative default probability profiles from one to thirty years out, albeit with decreasing accuracy. The model was validated through out-of-sample testing regarding its ability to accurately predict the defaults of U.S. depository institutions between 1992 and 2012. This method provides out-of-sample testing as well as best mimics how the model will be used in practice. The model performed well at separating potential defaulting banks from non-defaults over one-year horizons. Although performance drops monotonically when predicting defaults over longer horizons, the model performs significantly above chance for time periods as long as five years from the scoring date.

**Keywords:** Modeling banks; Probability; Financial crisis; Financial ratios

### Introduction

Since exposure to credit risk continues to be the leading source of problems in financial institutions, banks need to be able to identify, measure, monitor and control credit risk in order to ensure they hold adequate capital against the default risks. Regulators and financial institutions have placed a great deal of emphasis in recent years on the importance of models for credit risk measurement and management. Generating accurate model-based estimates of default probabilities (PDs) for financial firms has proven difficult over the past decade. Some reasons for this are financial institutions' high levels of leverage, the relative opacity of their assets and liabilities, potential support from governments, extreme risk of "tail events", and regulatory changes.

Since the financial crisis of 2008–2009 and the subsequent downgrade of many financial firms, investors are increasingly interested in better assessing and managing their credit exposure to financial institutions. In addition, the U.S. Office of the Comptroller of the Currency (OCC), in accordance with the Dodd–Frank Act, has published final rules<sup>1</sup> that remove references to credit ratings from its regulations pertaining to investment securities, securities offerings, and foreign bank capital equivalency deposits.<sup>2</sup> Amid this backdrop, the development of accurate models for assessing bank credit risk appear critical both for managing exposure to financial firms and for compliance with Federal regulations.

The empirical literature indicates the Merton-type structural models do not appear to be a sufficient statistic in predicting default probabilities as these models under-predict spread on corporate bonds.<sup>3</sup> I have constructed and tested an adaptive non-linear regression model to estimate default probabilities for U.S. banks using information only from their financial data as reported by the U.S. Federal Deposit Insurance Corporation (FDIC). The model is a logistic regression whose input variables are selected based on their past effectiveness at predicting bank failures and whose inclusion in the model and weights are to be updated quarterly. Model performance at discriminating between defaults and non-defaults was evaluated for horizons of one to five years using a sequence of ten-year annual forward out-of-sample tests. I tested the ability of the model to predict absolute default rates out to five years. The model performed well at estimating the annual bank default rates except that it underestimates the high bank default rates during the financial crisis. The model performs favorably at predicting default risks with a 97% accuracy ratio (AR) at one year before default, and decreasing, but still above-chance predictive power out to five years. For example, the top 10% banks with the highest risk scored from the model contain 94% of the banks that defaulted within the following year. Although performance drops monotonically when predicting defaults over longer horizons, the model performs significantly above chance for time periods as long as five years from the scoring date.

### Modeling Banks' PD

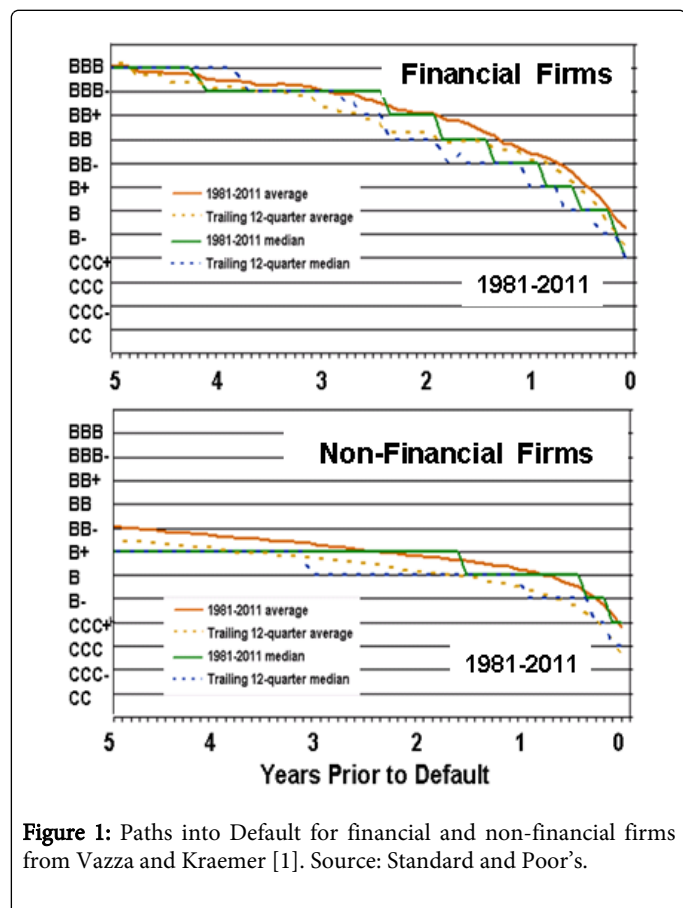
The numerous bank failures amid the financial crisis of 2008-2009 and the subsequent ratings downgrades of many financial firms have

<sup>1</sup> Department of the Treasury, 2012

<sup>2</sup> Section 939A of the Dodd–Frank Act requires federal agencies to review regulations that require the use of an assessment of creditworthiness of a security or money market instrument and any references to, or requirements in, those regulations regarding credit ratings. Section 939A then requires the agencies to modify the regulations identified during the review to substitute any references to, or requirements of, reliance on credit ratings with such standards of creditworthiness that each agency determines to be appropriate.

<sup>3</sup> For example Huang and Huang (2000) and Eom, Helwege, and Huang (2003).

highlighted limitations of agency credit ratings and current credit models to anticipate defaults for financial firms. During the crisis, many banks went from apparent solvency to default in a very short period of time presumably reflecting the particular sensitivity of financial institutions and insurance companies to sudden declines in investor confidence. Although the credit ratings of financial firms are concentrated in the investment grade range, results from Vazza and Kraemer [1] in Figure 1 demonstrate that despite their higher credit ratings, financial firms have a faster and steeper path to default than their non-financial counterparts<sup>4</sup>.



**Figure 1:** Paths into Default for financial and non-financial firms from Vazza and Kraemer [1]. Source: Standard and Poor's.

There are several challenges to measuring credit quality of financial institutions. First, financial firms operate differently from most non-financial corporates, running highly levered balance sheets financed by short-term borrowing, thereby having greater exposure to market risk and funding risk.<sup>5</sup> Differences also include the relative opacity of banks' assets and liabilities, potential support from governments, extreme risk of "tail events," and exposure to regulatory changes. I note that it is a general perception that Merton-type structural models, including Hybrid Probability of Default Model (HPD), fare relatively poorly at estimating PDs for financial firms. There are several reasons for this including:

- Financial firms typically use short-term borrowing to finance long-term obligations, thereby carrying much higher leverage than similarly risky non-financial firms. Because leverage is an important source of risk in structural models, Merton-type structural models typically over-estimate PDs for financial firms, relative to similarly risky non-financials. To correct this, structural models often embed adjustments to leverage and/or volatility for financial firms, but these can cause other problems;

- Although several available structural models differentiate between financial firms and industrial ones (emphasizing short-term liabilities more), there are relatively few defaults by publicly traded financial firms, posing difficulties for model calibration. Thus, the financial models are calibrated to ratings upgrades and downgrades which tend to trail perceptions of risk as indicated by credit. This introduces uncertainty in mapping model outputs (a ranked risk measure) to historical default rates;

- The asset quality of financial firms is often opaque (e.g., Lehman Brothers), making it more difficult to assess the credit quality of financial firms from an examination of their financial statements.

- Although we find that the HPD model does well at identifying rich and cheap bonds for commercial and industrial firms, it does not work as well for financial firms.

Modeling considerations aside, it is important to note that there are many banks and most of which are privately held firms and often unrated by major agencies. Since private firms have no publically traded equity, Merton-type default models such as KMV cannot score those firms. As of March 2013, there were 7,019 depository institutions in the U.S. reporting to the FDIC with total liabilities of \$12.8 trillion.<sup>6</sup> For investors with broad exposure to the banking sector, it is difficult to analyze a large number of banks using fundamental analysis. Thus, I chose a statistical approach to estimating risk of bank failure. As described below, I use an adaptive logistic regression function on information contained in banks' financial statements as published by the Federal Deposit Insurance Corporation (FDIC). The inputs to the model are financial ratios found to be effective in forecasting future bank failures and the outputs are predictions of annual cumulative PDs for each bank from one to 30 years.

$$PD = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i^* x_i\right)}}$$

Regression Function, the Input Variables and Coefficients.

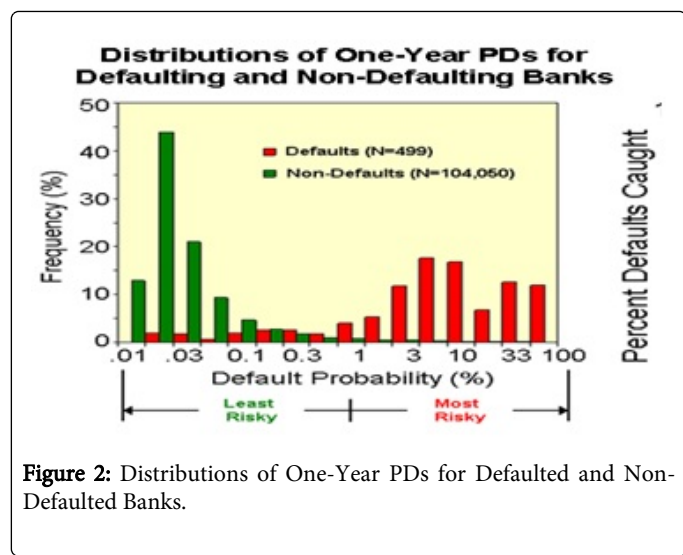
I back-tested the model's ability to predict defaults of U.S. depository institutions between 1992 and 2012. The testing was conducted using an annual walk-forward procedure to best simulate how the model will be used in practice. The testing dataset included 16,520 distinct banks and 604 default events<sup>7</sup>. That is, to make PD predictions for any year, I only use information before that year to select model variables and calibrate the model coefficients.

<sup>4</sup> For example, Vazza and Kraemer [1] report that in 2011 only 20% of all financial firms had speculative grade ratings.

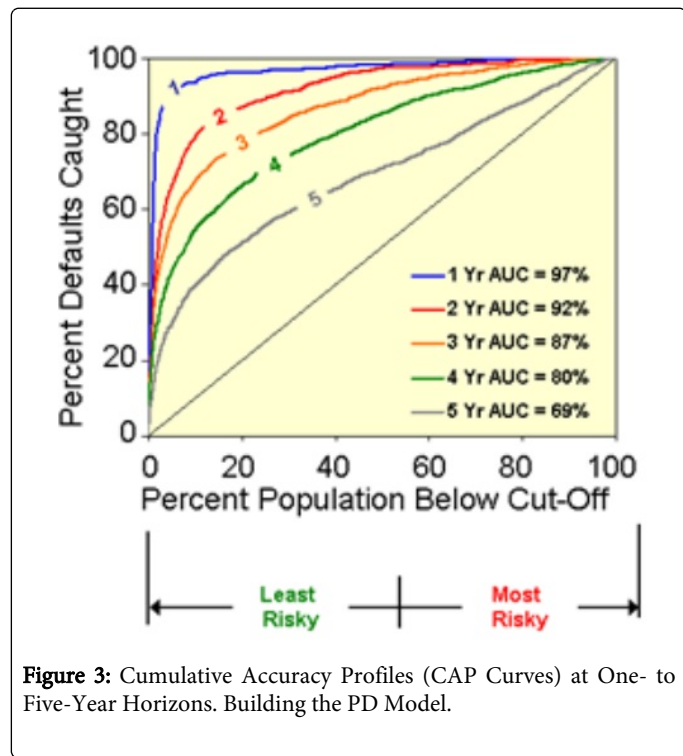
<sup>5</sup> As of March 2013, the average liability-to-equity ratio of the US banking sector was 7.8.

<sup>6</sup> Information about aggregate bank sector size obtained from the FDIC "Statistics on Banking," which is accessible online at <http://www2.fdic.gov/SDI/SOB/>.

<sup>7</sup> The difference between these numbers of defaults and firms shown in Figure 1 is that 105 defaulted banks were included in the first development sample and were thus unavailable for out-of-sample testing.



**Figure 2:** Distributions of One-Year PDs for Defaulted and Non-Defaulted Banks.



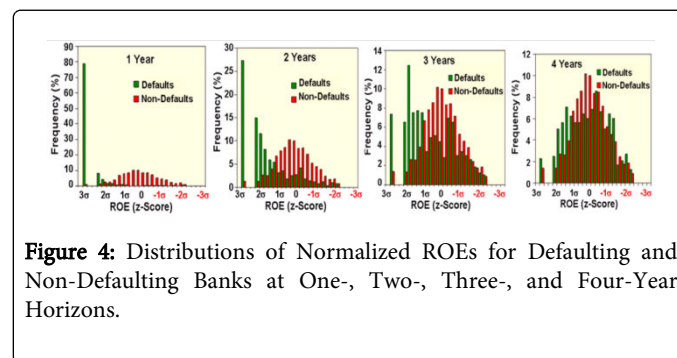
**Figure 3:** Cumulative Accuracy Profiles (CAP Curves) at One- to Five-Year Horizons. Building the PD Model.

Estimates of PDs from the model displayed a high degree of accuracy in out-of-sample back testing, both at predicting relative default risk and absolute default probabilities. Both aspects of performance are important for different applications as discussed further below. The model performed well at separating potential defaulting banks from non-defaults over one-year horizons (Figure 2). The banks in highest 10th percentile of PDs include 94% of the defaulting banks within the next year.<sup>8</sup> As expected, the predictive power of the model decreases as prediction horizon is extended but the model still performs well above chance multiple years out (Figure 3).

Also, the number of defaulters within the 10% of banks having the largest PDs declines from 94% to 80%, 68%, 55%, and 40% at two, three, four, and five year horizons out, respectively.

### Selecting Financial Variables

Since the pioneering work of Beaver [2] and Altman [3], financial modelers have realized that certain financial ratios are highly predictive of a firm's future default. The same is true for banks. For instance, banks with low, especially negative, return on equity (ROE) are much more likely to default. Intuitively, banks with low or negative profitability will likely struggle to pay their liabilities on time and will have difficulty finding additional funding. To illustrate this effect Figure 4 displays normalized distributions of ROEs for defaulting and non-defaulting banks. Distributions are shown for one-, two-, three-, and four- year horizons in successive panels.<sup>9</sup> Inspection of Figure 4 reveals that banks with low ROE are much more likely to default than those with high ROEs. Also, the predictive power of the ROE as regards default decreases with increases in the time horizon. That is, the distributions of ROEs from defaulting and non-defaulting banks are clearly apart from each other at one- and two-year horizons, but those differences narrow, becoming very small at four years out.



**Figure 4:** Distributions of Normalized ROEs for Defaulting and Non-Defaulting Banks at One-, Two-, Three-, and Four-Year Horizons.

A similar testing procedure as illustrated in Figure 4 for ROE revealed other financial ratios that are useful for default prediction. These include firms' leverage ratios, ratios of non-performing to performing loans, and net loans to bank capital, to name a few. A challenge in predicting default is to select an appropriate set of variables and combine them appropriately in a multivariate model. To do this, I employed a walk-forward logistic regression technique. The logistic regression function (described in the following section) is commonly used for predicting variables with binary outcomes, particularly when the inputs are non-linearly related to the desired output. The walk forward method constructs a new model each year from the candidate variable set, while adding the data from the previous year to the development sample. For variable selection in each new model, I use an automated procedure called forward stepwise selection, which is explained in detail below.

### Logistic Regression

Logistic regression is the sum of linear functions of multiple input variables put through a non-linear transformation before output. It has similarities to the more familiar multiple linear regression method,

<sup>8</sup> If the model were performing at chance, only 10% of the defaulters would be included in the top 10% of the population ranked by PDs from the model.

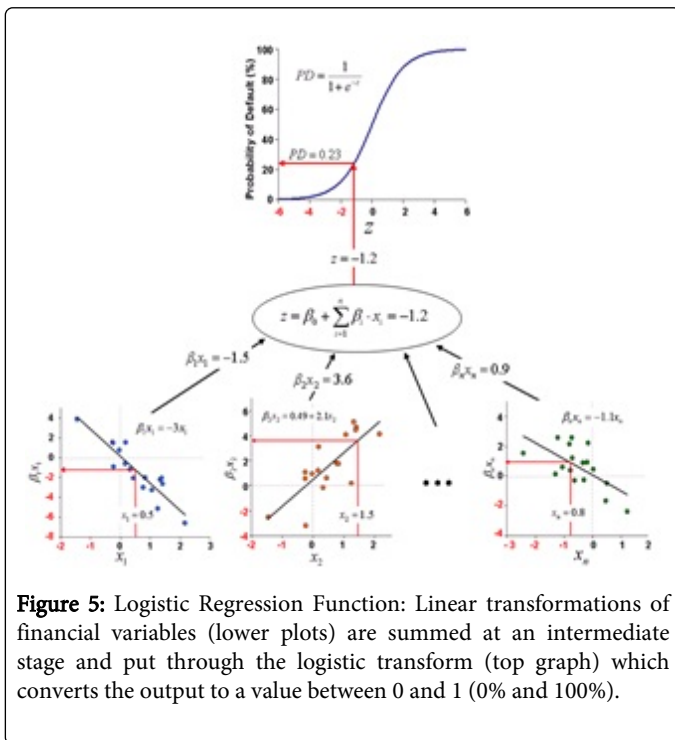
<sup>9</sup> Because financial ratios such as ROEs can have very dispersed distributions, I converted firms' ROEs into standard normal distributions before plotting. This transformation does not change the ordering of firms on the ROE axis.

but involves an extra step, the logistic transform. I illustrate this graphically for a set of hypothetical input variables in Figure 5. The application begins with selection of a set of candidate financial variables, denoted  $x_i, i=1, \dots, n$ . The inputs,  $x_i$ , could be financial ratios or other quantities. The lower portion of Figure 5 depicts how values of hypothetical input variables (the circles in each plot) are fit by functions, of the form

$$f(x_i) = \alpha_i + \beta_i x_i \quad (1)$$

to derive constants,  $\alpha_i$ , and coefficients,  $\beta_i$ , for each input variable.

Then, for a given set of inputs, each  $x_i$  is put through its linear transform in Equation 1. For variable  $x_1$  for the example in Figure 5, the constant  $\alpha_1=0$  and the coefficient  $\beta_1=-3$ . Thus, if  $\alpha_1=0.5$  as shown in the figure,  $f(x_1)=-1.5$ . Hypothetical functions and outputs for  $x_2$  and  $x_n$  are also shown in Figure 5.



**Figure 5:** Logistic Regression Function: Linear transformations of financial variables (lower plots) are summed at an intermediate stage and put through the logistic transform (top graph) which converts the output to a value between 0 and 1 (0% and 100%).

Linear functions of the input variables are summed at an intermediate stage in the model whose output is transformed using the logistic function. The resulting outputs of the first stage of the logistic regression, the values  $f(x_i)$ , are summed at an intermediate stage whose output  $z$  can be represented as

$$z = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (2)$$

Where

$$\beta_0 = \sum_{i=1}^n \alpha_i \quad (3)$$

For the example in Figure 5, the resulting value of  $z$  is assumed to be  $-1.2$ .<sup>10</sup> The value of  $z$  from Equation 2 is then put through the

logistic transform that serves to constrain the output of the regression to a value between 0 and 1. For example, for the default model, the resulting PD is given as:

$$PD = \frac{1}{1 + e^{-z}} \quad (4)$$

Where the resulting value of PD for  $z=-1.2$  is 0.26 or a 26% probability of default over the time frame in question.

### Forward Stepwise Selection of Input Variables

I use a forward stepwise selection procedure to choose input variables for the logistic model. Note that the overall plan is to derive a new model each year, incorporating into the learning sample the data from each successive year's defaulting and non-defaulting firms. Because the factors that influence defaults and their relative contributions may change over time, I chose to use an adaptive procedure for selecting variables for each annual model. I first assembled a set of 20 candidate financial ratios that have been shown to be predictive of subsequent default. Because the distributions of different financial ratios can vary widely, I chose to standardize all input variables via transformation into standard normal distributions before testing their usefulness as inputs to each annual model.

Variables for each annual model are chosen via an iterative procedure, whereby variables are prioritized with respect to their predictive power. The process of model construction begins with only the logistic function and no variables chosen for inclusion. Then, for each candidate input variable, I build a logistic default model by selecting values of  $\alpha_i$  and  $\beta_i$  for each variable that enables the best prediction of default on the development sample. That is for each input variable  $x_i$ , I solve for  $\alpha_i$  and  $\beta_i$  in the following equation for PD:

$$PD = \frac{1}{1 + e^{-(\alpha_i + \beta_i x_i)}} \quad (5)$$

The variable with the greatest predictive power with respect to default is chosen as the first input variable. As described in further detail below, I chose the Bayesian Information Criterion (BIC) developed by Schwartz [4] as my measure of predictive power. The BIC measures how well the model fits the data, but also imposes a penalty for having too many variables, thereby guarding against over fitting the data. After selection of the first variable, the process repeated to select a second variable, and so on, until model performance ceases to improve. Once all the variables for the model are selected, the value of the constant  $\beta_0$  and coefficients  $\beta_i (i=1, \dots, n)$  for each of the variables are refit to minimize the error in the logistic regression equation:

$$PD = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i^* x_i\right)}} \quad (6)$$

Successive variables are chosen until no further improvement in performance is achieved. An illustration of the results of variable selection is presented in Figure 6. The top portion of the left panel displays the logistic regression equation, with the table below it listing the input variables to the model in the order in which they are selected. That is, variables are listed in descending order of their predictive

<sup>10</sup> Note that this value of  $z$  cannot be deduced from the values shown in Figure 3 as it is assumed to have contributions from variables  $x_3$  to  $x_{n-1}$  that are not given in the figure.

power. The BIC values resulting from inclusion of each variable are also displayed. The right portion of Figure 6 is a plot of the BIC values that result from the inclusion of each variable. For instance, the model starts with only a constant term whose BIC value is 7,459. The variable selection procedure determined that banks' return on equity (ROE) provides the largest predictive power of all candidate variables, and its inclusion in the model achieves a BIC of 4,092.

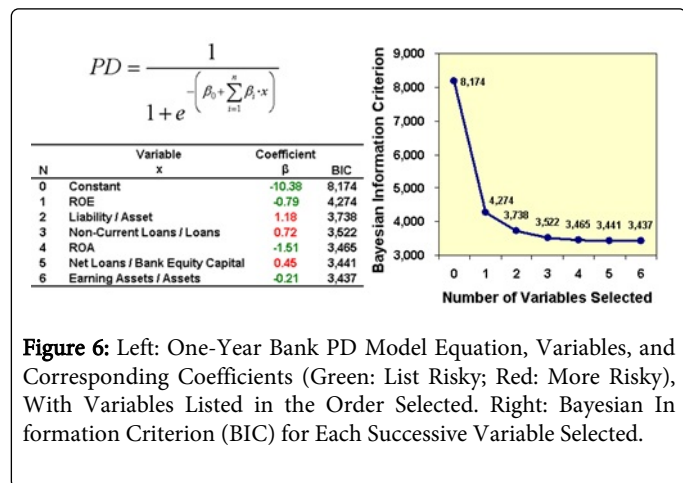


Figure 6: Left: One-Year Bank PD Model Equation, Variables, and Corresponding Coefficients (Green: List Risky; Red: More Risky), With Variables Listed in the Order Selected. Right: Bayesian Information Criterion (BIC) for Each Successive Variable Selected.

After selection of the ROE, the procedure is run again, picking the Liability/Asset ratio as the best of the remaining candidate variables, bringing the BIC down to 3,568. This procedure continued until the BIC could no longer be decreased. At that point, six variables had been selected and their corresponding coefficients appear in the left table of Figure 6.

### The Term Structure of Bank PDs

The method I have described can be used to predict defaults over one- to five-year horizons. However, some applications (e.g., long-term investment portfolios) require estimation of the term structure of PDs over longer periods. My approach to extending the term structure of bank PDs for terms beyond five years is to use long-term annual average marginal default rates determined from historical data on bank defaults.

Construction of PD term structures begins by using the set of five logistic regression models, each developed for the marginal default rate between successive years over a period from one to five years. That is, let PD<sub>t</sub> denotes the model designed to predict bank defaults t years from now, conditional on the given banks surviving to year t-1. That is, for years t =1,..., 5, PD<sub>t</sub> is the conditional logistic regression model where

$$PD = \frac{1}{1 + e^{-\left(\beta_t + \sum_{i=1}^n \beta_{t,i} \cdot x_{t,i}\right)}} \quad (7)$$

Then, for each bank j, the probability of default in year t assuming survival to year t-1 is given by

$$PD = \frac{1}{1 + e^{-\left(\beta_t + \sum_{i=1}^n \beta_{t,i} \cdot x_{t,i,j}\right)}} \quad (8)$$

Note that because I fit a separate model for each year, the variables selected and the coefficients β<sub>j,t</sub> will, in general, be different for each year. Let, CPD<sub>t,j</sub> be the cumulative probability of default for bank j from time t=0 to t years. Then, the cumulative probabilities for bank j over horizons from t=1 to T years can be determined from their annual PDs as:

$$CPD_{1,j} = P_{1,j}$$

$$CPD_{2,j} = CPD_{1,j} + (1 - CPD_{1,j}) \cdot P_{2,j}$$

$$CPD_{T,j} = CPD_{T-1,j} + (1 - CPD_{T-1,j}) \cdot P_{T,j}$$

The procedure for calculating marginal PDs beyond five years is illustrated in Figure 7. First, I construct a map between one-year PDs and Standard & Poor's rating categories. This is made possible using a map that I derived between average probabilities of default for commercial and industrial firms from HPD model [5,6] and their corresponding agency ratings<sup>11</sup>.

For example, the left panel of Figure 7 illustrates a mapping between one-year PDs from the HPD model to rating categories calibrated using data of all U.S. banks between 1982 and 2012. Using this map, I can assign an implied rating to each bank that corresponds to its current one-year PD from the logistic regression model. Then, for a given bank, I combine its term structure of cumulative default rates from one to five years with the marginal annual default rates reported by Moody's from its imputed credit rating from six to thirty years. That is, I assume each bank's conditional PD beyond five years follows the long-term historical values for its implied rating category. A resulting set of stylized bank annual cumulative default rates by implied whole letter rating categories appear in the right panels of Figure 7. The top panel shows cumulative default rates on a linear PD scale, whereas the lower plot shows those same data in logarithmic PD units [7-10]. Notice that, as expected, average cumulative default rates for any given tenor increase with decreasing rating categories.

I back-tested the model by constructing an annual series of models of the bank models using all available US bank data from 1992 to 2012. The number of non-defaulting banks and defaulting banks in the sample by year are given by the green bars (left axis) and red bars (right axis) in Figure 8. Notice that there were roughly 14,000 banks in the sample in 1993, but that number declined to around 7,000 by 2012. Also, there are three apparent waves of defaults: one in the early 90s, a small one around the year 2000, and a surge of bank failures during the recent financial crisis.

In order to determine out-of-sample performance of the model, I used a walk-forward procedure as illustrated in Figure 9 for the one-year model [11-13]. The test set is sufficiently large, with a total of 499 defaulters out of 11,114 distinct banks, to provide a strong test of model performance. Because the model needs a minimum number of years of data for development, data from the years 1992 through 1999 were used to construct the first annual model (select variables and calibrate the weights) for each horizon for one to five years. The one-year model for 1999 was then used to score all non-defaulting banks at

<sup>11</sup> That rating map is constructed using PDs from the HPD model for non-bank corporate firms. Then firms are ranked with respect to their model PDs and assigned to rating categories that replicate the number of firms in each rating category in the sample. Finally, implied ratings for U.S. banks are assigned based on their inclusion within PD boundaries determined for each rating category.

the beginning of 2000 and its ability to predict defaults in 2000 was determined. Models for year two through five used only banks that had survived to the model year to score for prediction. Thus for the two-year model, firms surviving until 2001 were scored with its 1999 model, and so forth for the longer horizons.

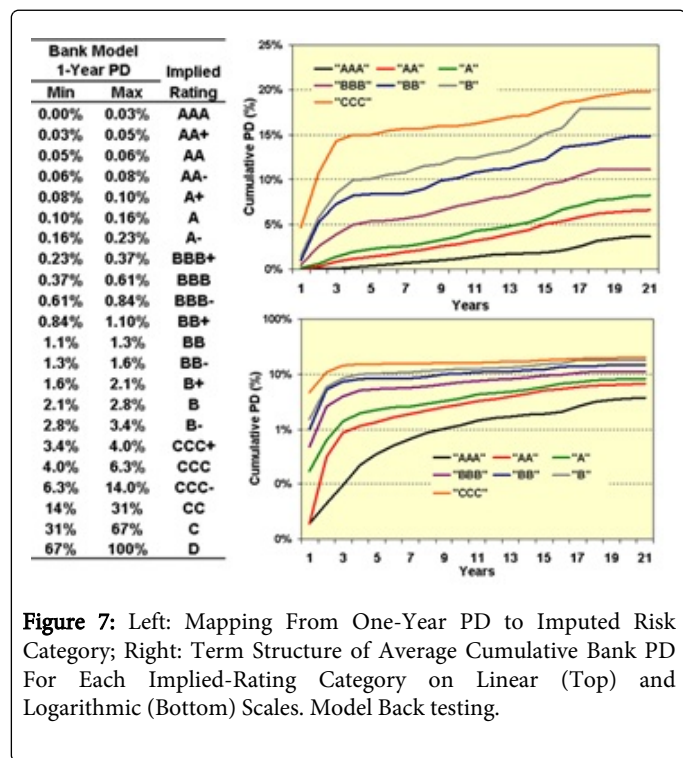


Figure 7: Left: Mapping From One-Year PD to Imputed Risk Category; Right: Term Structure of Average Cumulative Bank PD For Each Implied-Rating Category on Linear (Top) and Logarithmic (Bottom) Scales. Model Back testing.

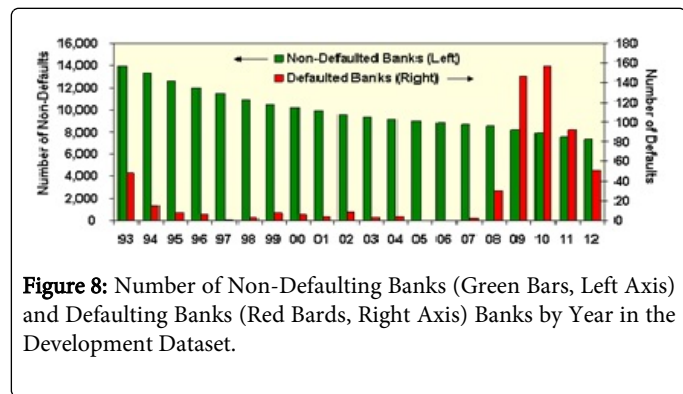


Figure 8: Number of Non-Defaulting Banks (Green Bars, Left Axis) and Defaulting Banks (Red Bards, Right Axis) Banks by Year in the Development Dataset.

To generate the set of models for year 2000 (i.e., used to predict defaults in 2001 to 2005 for one- to five-year models), I added the data from year 2000 to the set from 1992 to 1999. Variables were selected and coefficients determined and the model was tested on the corresponding test sample for the given horizon. That procedure was repeated annually until 2012. Of course, from models at horizons longer than one year, testing was only able to be done to year 2012 minus the horizon year. I adopted the walk forward procedure because it most realistically estimates the performance of the model as it will be deployed in practice.

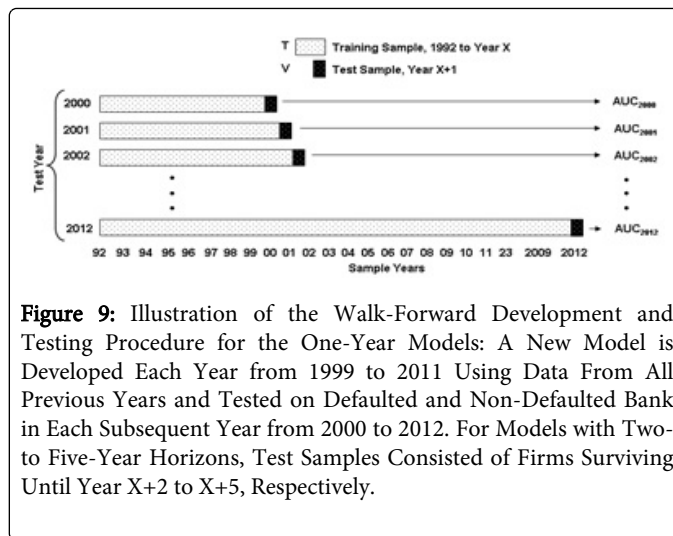
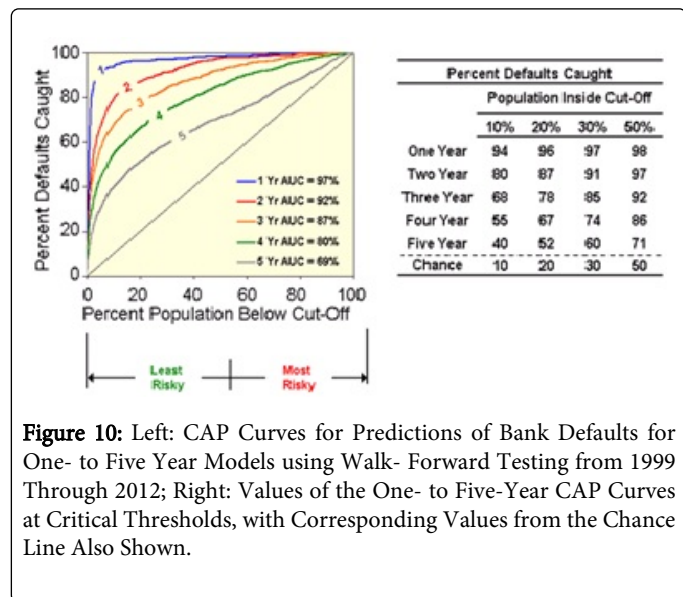


Figure 9: Illustration of the Walk-Forward Development and Testing Procedure for the One-Year Models: A New Model is Developed Each Year from 1999 to 2011 Using Data From All Previous Years and Tested on Defaulted and Non-Defaulted Bank in Each Subsequent Year from 2000 to 2012. For Models with Two- to Five-Year Horizons, Test Samples Consisted of Firms Surviving Until Year X+2 to X+5, Respectively.

To evaluate model performance at separating banks that will default from non-defaulters, generated Cumulative Accuracy Profile (CAP) Curves for the one- to five-year model horizons. The cumulative resulting CAP curves for test years 1999- 2012 are displayed in Figure 10. For example, to generate the one year-curve (blue line), I first rank all banks over the entire 13-year test period from highest to lowest by their one-year PDs from the models. Then, for successive intervals in the ranked population I calculate the cumulative fraction of defaulting banks contained within that interval. The interpretation of CAP curves is straightforward; for any criterion, the fraction of defaulters caught above the population percentile is measures the discriminatory power of the model. For example, the CAP curve for the one-year model at the 10% population criterion caught 94% of the banks that defaulted within the following year over the period from 1999-2012. The higher and steeper the CAP curve over the diagonal chance line, the better the model is at discriminating defaulters from non-defaulters. The table at the right in Figure 10 displays values of the CAP curves for each of the model horizons for various values of the population cut-off. The left-most values in the table show that the 10% of banks ranked riskiest by the one- to five-year models capture 94%, 80%, 68%, 55%, and 40% of the defaulting banks, respectively. Not surprisingly, those data reveal that the power of the models decline as the horizon extends beyond one year, but even the five-year model is performing well above chance, capturing 40% of the banks that default in the fifth year after model development and scoring. Finally, it is important to note that even though the models are only regenerated on an annual basis, the financial data from the banks is available to update bank default scores on a quarterly basis and that is how the model will be used in practice.

From a risk management perspective, the most relevant horizon for prediction is at one year. Thus, if a bank survives for that one year, the next year's model can be used to assess its subsequent risk. Still, there are applications for which multi-year estimates of losses and portfolio relative value are of interest. These include buy-and-hold portfolios of bank obligations, such as structured products. For example, if one holds a portfolio of bank TRUPS (trust preferred securities) with five years of remaining maturity, they may wish to estimate five year portfolio losses. For this type of application, it is important that the absolute PD levels be accurate. The CAP curves, because they rank

PDs, assess only the relative accuracy of the models<sup>12</sup>. Indeed, the models do specify absolute PD levels and I can assess their accuracy using the reliability plots in Figure 11. To construct the plots in Figure 11 I separated all banks into bins by 5% PD increments, and plot each bin's average predicted PDs on the horizontal axis and the realized rate of defaults on the vertical axis. The interpretation of reliability plots is as follows.

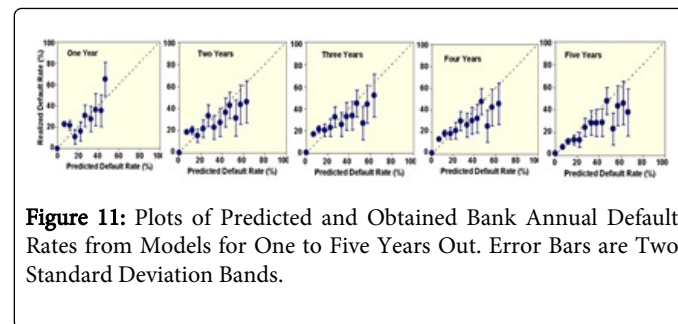


**Figure 10:** Left: CAP Curves for Predictions of Bank Defaults for One- to Five Year Models using Walk- Forward Testing from 1999 Through 2012; Right: Values of the One- to Five-Year CAP Curves at Critical Thresholds, with Corresponding Values from the Chance Line Also Shown.

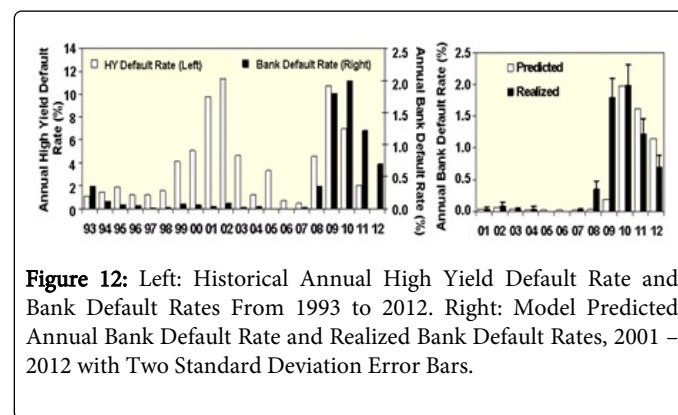
For example, the one-year plot includes the point (27% predicted, 31% obtained), which means for all the banks assigned one-year PDs between 25% and 30%, 31% of them actually defaulted within the following year. A perfect model would have all points falling on the diagonal line for which predicted PD and realized default rates match exactly. Error bars at two standard deviations for the realized default rates are also shown in each plot.

The plots in Figure 11 indicate that the default probabilities generated by the model are reasonably accurate at predicting default rates for banks over multi-year horizons. With respect to the two standard deviation bars, most data predictions do not differ significantly from the diagonal “perfect model” line. However, a notable exception is that the bank model typically underestimates the default rates for the second- and third-highest bins (i.e., the high default 60%-70% bins). Further analysis revealed that the model under-predicted the sudden surge of defaults during the financial crisis of 2008 and 2009. Consider the left panel Figure 12 which displays the historical annual high yield corporate default rates (left axis) and U.S. bank default rates (right axis) from 1993 through 2012. Notice that the high yield default rates varied substantially over the period, with high rates early in the century. The banks had been relatively safe before 2008, with an average annual default rate of only 0.06% and even the maximum during that period is only 0.34%. The right panel of Figure 12 plots average predicted and realized annual default rates from the one-year bank model. The bank default models that are constructed annually did not predict well the overall bank default rate in 2008 and 2009, the years of high bank defaults. More generally, the plot reveals that PD levels from the bank model tend to trail observed annual PD rates by one year. Note that the financial data for U.S. banks are

published quarterly by the U.S. Federal Deposit Insurance Corporation (FDIC). Thus, in practice, I plan to update the model quarterly, potentially minimizing the lag in accurately predicting annual default rates.



**Figure 11:** Plots of Predicted and Obtained Bank Annual Default Rates from Models for One to Five Years Out. Error Bars are Two Standard Deviation Bands.



**Figure 12:** Left: Historical Annual High Yield Default Rate and Bank Default Rates From 1993 to 2012. Right: Model Predicted Annual Bank Default Rate and Realized Bank Default Rates, 2001 – 2012 with Two Standard Deviation Error Bars.

### Summary

In this paper, I develop a dynamic measure to overcome limitations of the Merton-type structural models in predicting default probabilities for financial firms. I built and tested adaptive statistical models to estimate default probabilities for U.S. banks. As described in detail, the models are logistic regression whose input variables are selected and calibrated based on their past effectiveness at predicting bank failures. Selection of variables in the model and their weights were updated yearly using a “walk-forward” procedure. The model predicts defaults at annual horizons from one to five years. Performance of the models at discriminating between defaults and non-defaults was evaluated for horizons of one to five years using a sequence of annual walk-forward out-of-sample tests from 1992 to 2012. I also measured the ability of the models to predict absolute default rates from one to five years and, except for underestimating the high bank default rates during the financial crisis, the models performed well at estimating the annual bank default rates. In general, the models perform favorably at predicting defaults, with a 97% accuracy ratio (AR) at one year prior to default, and decreasing, but still above-chance predictive power out to five years. The models are designed to be updated on an annual basis, but updated financials for inputs to the model are available from the FDIC on a quarterly basis.

<sup>12</sup> For example, if one multiplies all PDs by 10 the CAP curves will not change, but the absolute PD levels implied by the models will be too large.

---

## References

1. Vazza D, Kraemer N (2012) Default, Transition, and Recovery: 2011 Annual Global Corporate Default Study and Rating Transitions, Standard and Poor's.
2. Beaver WH (1966) Financial Ratios as Predictors of Failure. *J of Accounting Res* 4: 71-111.
3. Altman E (1968) Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *J of Financ* 23: 589-609.
4. Schwarz GE (1978) Estimating the Dimension of a Model. *Annals of Statistics* 6: 461-464.
5. Sobehart J, Keenan S (2002) Hybrid Contingent Claims Models: A Practical Approach to Modeling Default Risk, *Credit Ratings: Methodology, Rationale and Default Risk*, M. Ong (ed.), Risk Books.
6. Sobehart J, Keenan S (2003) Hybrid Probability of Default Models: A Practical Approach to Modeling Default Risk, *Quantitative Credit Analyst* 3: 5-29.
7. Benzschawel T (2012) Comments on Model-Based Default Probabilities (PDs) for Financial Firms.
8. Benzschawel T, Adler D (2002) Empirical Analyses of Corporate Bond Spreads, *Quantitative Credit Analyst* 1, Salomon Smith Barney.
9. Department of the Treasury, Office of the Comptroller of the Currency, 12 CFR Parts 1, 5, 16, 28, and 160 [Docket ID OCC-2012-0005] RIN 1557-AD36, Alternatives to the Use of External Credit Ratings in the Regulations of the OCC, *Federal Register* 77 (114), June 13, 2012.
10. Eom Y, Helwege J, Huang J (2003) Structural Models of Corporate Bond Pricing: An Empirical Analysis, *Forthcoming: Review of Financial Studies*.
11. Huang J, Huang M (2000) How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk? A New Calibration Approach, Working Paper, Cornell University.
12. Kealhofer S (1999) Credit Risk and Risk Management, AIMR Conference Proceedings. 80-94. Moody's Investors Service, Corporate Default and Recovery Rates, 1920-2009, February, 2010 Modeling Banks' Probability of Default.
13. Vasicek O (1984) Credit Valuation, KVM White Paper, KVM LLC.