Aalto University
School of Economics

Predicting Failures of Large U.S. Commercial Banks

Economics

Master's thesis

Markus Vilén

2010

Department of Economics
Aalto University
School of Economics

AALTO UNIVERSITY SCHOOL OF ECONOMICS
Department of Economics
ABSTRACT

Master's Thesis
22.10.2010

PREDICTING FAILURES OF LARGE U.S. COMMERCIAL BANKS

Since 2007 several banks have fallen into bankruptcy in the U.S. What is historically notable in this situation is the amount of assets lost in bankruptcies that are already measured in hundreds of billions of dollars. Since financial institutions magnitude to the current economic system is crucial, bank failures have a dramatic impact also on real economy. Therefore, feasible bank failure prediction models can also diminish the real economy problems.

The purpose of this thesis is to study how accurately recent U.S. commercial bank failures can be predicted with logistic regression model utilizing financial statement variables. With the overall predictability of the model, also the statistical significance of the independent variables is studied. In addition, it is tested how the prediction accuracy reduces as the timeframe from the bankruptcy prolongs from one quarter to three years.

The evolution and history of bankruptcy prediction models and banking crises is also studied in order to develop the base for the bank failure model. It is also noted that several bankruptcy prediction models contain financial statement variables. In addition, by studying the history of the banking crises it is noticed that traditional banking factors such as liquidity, credit risk, and profitability have a substantial impact on failures of individual financial institutions. It is also argued that banks' exposure to the subprime related securities might deteriorate the solvency of the financial institutions.

The data of the analysis in gathered from the FDIC database. It contains bank-specific variables from 2004 to 2009 including 124 commercial banks with total assets worth more than 500 million dollars. In the empirical part of the thesis it is noted that 25 independent variables are statistically significant for the bank failure prediction. On the other hand, several of these explanatory variables correlate significantly with each other, erasing the possibility to include all the statistically significant variables into the same model. Therefore, 72 potential models are constructed, which are then studied with the help of logistic regression. After short and long-term analysis it can be noticed that the most accurate failure prediction is provided by the model having nonaccrual rate, loan diversification, return on equity, capital growth, tax exposure, CMO ratio, uninsured deposits, risk free securities, dividend rate, loan growth, assets variation, and liquid assets as the explanatory variables. The accuracy of the model is measured by correctly classified (CC) percentage figure. The model in question can predict 95.16% of failures correctly one quarter prior the bankruptcies. CC figure is 82.26% one year, 72.58% two years, and 71.77% three years prior the bank failures.

The analysis recovers, however, that only two of the explanatory variables, nonaccrual rate and risk free loans, are both statistically significant in the long-term and consistent with time. Although the model with only two explanatory variables tends to lose some of its predicting accuracy, it permits a precise analysis of the independent coefficients. It can be confirmed that only one percentage point increase in nonaccrual rate at least doubles the bankruptcy probability. On the other hand, the more the bank ties its investments to risk free securities the greater the probability that the bank survives. Empirical analysis proves also that the logit model is slightly more suitable for bank failure prediction than the corresponding probit model.

Keywords: Bankruptcy prediction, bank failure, logistic regression, off-site analysis.

AALTO YLIOPISTON KAUPPAKORKEAKOULU TIIVISTELMÄ
Taloustieteen laitos
Pro gradu -tutkielma 22.10.2010

SUURTEN YHDYSVALTALAISTEN PANKKIEN KONKURSSIENNUSTAMINEN

Vuodesta 2007 lähtien lukuisat yhdysvaltalaiset pankit ovat vajonneet konkurssiin. On historiallisesti merkittävää, että hävittyjen pääomien summa määritellään jo sadoissa miljardeissa dollareissa. Koska pankit ovat nykyisessä talousjärjestelmässä hyvinkin merkittävässä asemassa, heijastuvat pankkikonkurssit voimakkaasti myös reaalitalouteen. Näin ollen konkurssien ennustaminen ei ainoastaan tarjoa arvokasta informaatiota pankkisektorille, vaan on merkittävässä asemassa myös reaalitalouden ongelmien vähentämisessä.

Tämän tutkielman tarkoituksena on tarkastella kuinka tarkasti pankkikonkursseja voidaan ennustaa logistisella regressiolla pankkien tunnuslukuja hyväksi käyttäen. Kokonaismallien ennustavuuden lisäksi tutkielma testaa myös yksittäisten muuttujien tilastollista merkitsevyyttä. Tämän lisäksi tutkitaan kuinka ennustustarkkuus heikkenee, kun aikajänne konkurssista pitenee yhdestä kvartaalista kolmeen vuoteen.

Tutkimuksessa perehdytään myös konkurssimallien ja pankkikriisien kehitykseen ja historiaan, minkä pohjalta luodaan malli empiiriseen analyysiin. Huomataan, että useat konkurssimallit sisältävät tunnuslukuaineistoa. Lisäksi aiempien merkittävien Yhdysvaltain pankkikriisien tutkimisella saadaan selville pankkitoiminnan kannalta perinteisten tekijöiden kuten likviditeetin, luottoriskin ja pankin kannattavuuden heijastuvan yksittäisten pankkien konkursseihin. Myös niin sanottuun subprime-kriisiin liittyvien arvopaperisijoitusten merkitsevyyttä viimeisimpiin pankkikonkursseihin tarkastellaan empiirisessä analyysissä.

Tutkielman aineisto on koottu FDIC:n tietokannasta vuosien 2004 ja 2009 väliseltä ajalta, ja sisältää kaikkiaan 124 pääomaltaan yli 500 miljardin dollarin kokoista Yhdysvaltalaista liikepankkia (eng. commercial bank). Empiirisessä analyysissä todetaan 25 selittävän muuttujan olevan tilastollisesti merkitseviä pankkikonkurssien ennustamisessa. Toisaalta useat näistä tunnusluvuista ovat vahvasti korreloituneita keskenään, mistä syystä kaikkia merkitseviä tunnuslukuja ei voida sisällyttää samaan malliin. Näin ollen saadaan luotua kaikkiaan 72 erilaista mallia, joiden tarkkuutta pankkien konkurssiennustuksessa tutkitaan logistisen regression avulla. Lyhyen ja pitkän aikavälin analyysit osoittavat mallin, jossa selittävinä muuttujina ovat korkoa kerryttämättömät lainat, lainan diversifikaatio, oman pääoman tuotto, oman pääoman kasvu, verorasite, CMO suhdeluku, vakuuttamattomat talletukset, riskittömät sijoitukset, osinkoaste, lainojen kasvu, pääoman vaihtelu ja likvidi pääoma ennustavan parhaiten konkursseja. Ennustustarkkuutta mitataan oikein ennustettujen pankkien luumäärällä suhteessa otoksen kokoon (CC, eng. correctly classified). Kyseinen malli pystyy ennustamaan pankkikonkurssit oikein 95.16% tapauksista neljä kuukautta ennen tapahtumaa. Vastaava CC-luku on 82.26% vuosi, 72.58% kaksi vuotta ja 71.77% kolme vuotta ennen konkurssia.

Tarkempi analyysi kuitenkin paljastaa, että vain kaksi selittävää muuttujaa, korkoa kerryttämättömät lainat ja riskittömät sijoitukset, ovat tilastollisesti merkitseviä ja johdonmukaisia pitkällä aikavälillä. Vaikka kahden muuttujan malli menettääkin jonkin verran ennustustarkkuutta, antaa se mahdollisuuden tarkempaan analyysiin kyseisten selittävien muuttujien kohdalla. Huomataan, että vain yhden prosenttiyksikön kasvu korkoa kerryttämättömissä lainoissa kasvattaa konkurssin todennäköisyyttä useita kymmeniä prosentteja. Toisaalta mitä suurempi osa sijoituksista on sidottu riskittömiin kohteisiin, sitä epätodennäköisemmältä konkurssi vaikuttaa. Analyysi osoittaa myös, että käytetty logit malli ennustaa hieman paremmin pankkikonkursseja kuin vastaava probit malli.

Avainsanat: Konkurssiennustaminen, pankkikonkurssi, logistinen regression, tunnuslukuanalyysi.

Contents

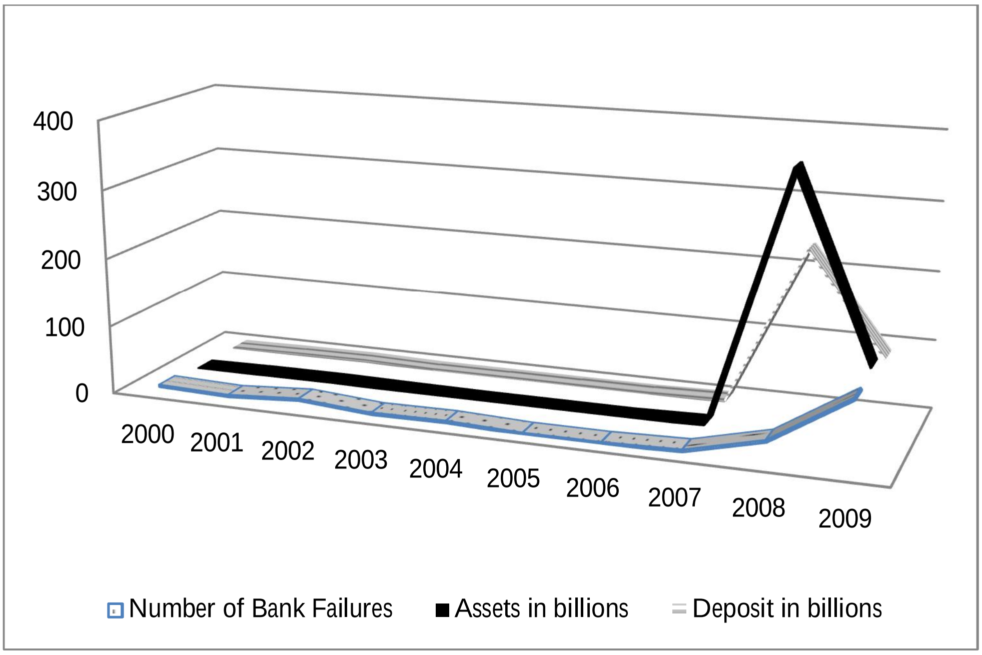
Figures

Tables

1 INTRODUCTION

The year 2008 will be remembered as one the most shocking in the U.S. financial sector history. After many solid years of growth, banks started to fail with increasing speed. There were no commercial bank failures in 2005 or 2006 and only three during 2007. However, in 2008 already 25 financial institutions failed giving foretaste for the 142 bankruptcies in year 2009. One of the most alarming sign was the failure of Washington Mutual Bank, together with its subsidiary Washington Mutual Bank FSB, having assets worth over 300 billion dollars at the time of failure. (FDIC, 2009) Indeed, not only did the number of commercial bank failures dramatically increase in 2008, but also the actual value of bankruptcies.

Figure 1. U.S. Commercial Bank Failures from 2000 through 2009



Source: FDIC

The commercial banks were not the only ones to suffer, as also many investment banks and insurance companies failed or fell into troubles. For example, Bear Stearns, being one of the most prestigious investment banks, crashed in March 2008 and was acquired by J.P. Morgan (Landon, 2008). Additionally, another famous investment bank, Lehman Brothers, filed for bankruptcy protection in September 2008. To give perspective to the incident, the failure of Lehman Brothers acts as the largest corporate failure in the entire history of U.S. bankruptcy court (Tkaczyk, 2009).

One will never know how many bankruptcies there would have been without U.S. government interventions. Certainly, the U.S. government has bought out significant amounts of so called toxic assets, mainly subprime related securities, in order to support financial sector and avoid total financial crash. The bailout of American International Group (AIG), worth of 85 billion dollars, is considered as the most notable one. The bailout happened only two weeks after the Treasury took over the federally chartered mortgage finance companies Federal National Mortgage Association (Fannie Mae) and Federal Home Loan Mortgage Corporation (Freddie Mac), being government's the most radical intervention to private business in the U.S. history. (Andrews, de la Merced, & Williams Walsh, 2008). Thus it goes without saying that 2008 was indeed one of the worst years in the history of U.S. financial sector.

On contrary to the baffled reactions, the financial crisis came as no surprise. As banks' balance sheets are loaded with debt, the banking system itself can be considered as highly volatile. Debt levels are generally stretched to risky limits as banks attempt to remain in competitive business. Naturally, intensive risk taking can debase banks' ability to defend itself against negative shocks and surprises. In addition, banks are closely tied together through interbank lending, which tends to reflect insolvent banks' problems to solvent ones. Once the domino effect is on, the financial crisis is inevitable. Therefore, it is crucial to notice the early signs of potential bank failures.

Additionally, it needs to be noted that the assets and deposits lost in bank failures are not only the money people have lost in crises, but also the value of houses have crashed, consumer credit defaults increased, the bond issuers downgraded, value of commercial property diminished, and corporate defaults exploded (Wolf, 2008). Consequently, the crises extend to real economy, which makes the banking problems even more severe than it might seem at the first glance.

The current crisis is nothing unforeseen as there has been two rather similar periods in U.S. financial history. The first wave of bank failures happened during the Great Depression in the 1930s and the second during Savings and Loan (S&L) crisis in the 1980-90s. Although, the study at hand concentrates on the determinants behind the latest commercial bank failures, the history of U.S. banking difficulties will be briefly covered in order to enlighten the common factors behind the three financial crisis and individual bank failures.

In academic discussion, three main topics relating to the banking sector difficulties can be found: currency crisis prediction, banking panic forecasting, and individual institutions failure mapping.

In this study the focus is on individual commercial bank failure prediction as it can be argued that the number of financial institution bankruptcies is strongly related to depth of the overall crisis. Therefore, by minimizing the number of bankruptcies on the financial sector, the severity of the crisis can be reduced. Naturally, banks tend to fail when overall economic condition deteriorates. Nevertheless, here the focus will be on detecting the troubled banks early on, so that systemic risk and contagion could be minimized and finally potential crisis avoided.

The main objective for the research is to find the most accurate model for bankruptcy prediction. Academics worldwide have invented numerous models and methodologies for bankruptcy forecasting. Logistic regression, in other words logit model, seems to be commonly accepted tool for bank failure prediction due to its accuracy and easiness to use. The logit model is therefore applied also in this study in order to reliable predict the possible bankruptcies.

Thus, the focus will be on commercial bank failures, more accurately on FDIC covered institutions. Since the data related to subprime mortgage instruments and the overall exposures of the banks are virtually impossible to gather, the concentration is on bank-specific variables, in other words, in different financial ratios. Therefore, systemic risk, contagion, and interbank lending will not be covered in the analysis. Previous studies indicate that at least capital ratio as well as nonperforming loans to assets will forecast bankruptcies as early as one or two years prior the actual failure. Hence they will be analyzed also in this study alongside with new independent variables relating to the subprime loans and securities that are relatively new instruments in the lending market. Although, the overall subprime exposure is not possible to detect, proxies of how much assets banks are tied to subprime securities, such as MBS to total securities, can be provided. By combining traditional independent variables with modern ones, the prediction accuracy can be improved and more suitable bank failure forecasting model can be generated.

The thesis consists of five sections. After the introduction, an overview of the existing literature concerning the bank failure predicting is covered. Here also the key concepts of the study and the main elements behind the bank failures will be provided. The literature overview focuses on the Great Depression, U.S. banking crisis from 1980s through the 1990s, and the subprime crisis. The issues around the banking failures are emphasized, while the real economy problems will be left uncovered. Finally, also the failure predicting models are illuminated in the end of the section.

In the third section the methodology and data of the study are presented. As mentioned, the logistic regression method is used in this thesis. Thus, the background and function of the model is explained alongside with the reasoning behind selecting the method. Additionally, the data of the study and the reasoning for choosing the particular financial ratios are covered at the end of the third section.

Section four provides the empirical analysis and the results of the study. As the main purpose of the thesis is to test whether some of the financial ratios predict the bank failures, the procedure of choosing specific variables in the final predicting model and the accuracy of the model are carefully presented. Finally, in section five, the study is summarized and concluding remarks are provided.

2 LITERATURE SURVEY OF BANKRUPTCY PREDICTION

In this section different elements relating to bankruptcy prediction are presented. In the beginning, the key concepts are explained in order to grasp a clear view of the bankruptcy phenomenon. Thereafter, the different parameters relating to bankruptcy prediction are illustrated. Then the glance of the U.S. bank failure history is presented. The focus is on the factors behind the earlier bank failures and the joint features between previous and current bankruptcies. Finally, in the last part of the section, different failure prediction models, in other words early warning systems (EWSs), are illustrated.

2.1 Key Concepts

Banks, like any other firms, try to maximize shareholders value by generating earnings that exceed expenses. There are, however, some factors that differentiate banks from regular companies. When adopting the traditional view towards commercial banking, banks are seen as fixed-value deposit takers. Deposits are the banks' principal liability and can be withdrawn at very short notice. Banks lend long-term money to different parties such as industrial companies, individual consumers, or even other banks. Loans, therefore, can be seen as banks' principal assets. When law of large numbers applies, only small portion of assets needs to be held in liquid reserves to meet deposit withdrawals. However, here lies a liquidity problem if for one reason or other exceptionally high withdrawals occur. This certainly is a valid risk although banks might be mainly in good condition. Therefore, one can argue that the health of a bank not only depends on its success in picking profitable investment projects for lending but also on the confidence of depositors in the value of the loan book and in their confidence that other depositors will not run the bank. To sum up, the key elements of bank problems are mainly dependent of loan characters, liquidity, the quality of borrowers and investments, and exogenous shocks.

EWSs are predicting tools for individual bank failures or for detecting the financial distress of a complete banking system. As already stated, the focus here is on individual bank failure prediction, not on discovering the distress level or depth of the whole banking sector. Therefore, every time EWS is referred, financial institution failure forecasting is indicated. When a bank faces problems, it either falls or stays on market. Thus, bankruptcy is defined as a binary event in contrast to the index of banking system distress that can take a continuum of values.

The definition of bank failure is obviously one of the most crucial definitions in the thesis. Financial institution is categorized as failed if it has fallen either into Chapter 7 or Chapter 11 in U.S. bankruptcy code. Chapter 7 means straight bankruptcy, or in other words liquidation, where the bank would not be operating anymore. Chapter 11, also called rehabilitation bankruptcy, is much more complicated than chapter 7 as it allows a firm, or in this case a bank, the opportunity to reorganize its debt and try to re-emerge as a healthy organization. (Bankruptcy Basics, 2009)

The bankruptcy models or EWSs can be divided into two categories: on-site and off-site assessments. On-site assessment is done in the premises of a bank, by examining bookkeeping records, business books, subsidiary ledgers, and other records and accounts in order to evaluate bank's financial soundness and compliance with laws and regulatory policies. The aim here is also to assess the quality of its management and to evaluate the systems of internal control. Off-site analysis, on the other hand, can be made by using publicly available information only. That information includes annual and quarterly reports that banks are obligated to compile to the regulators. Although on-site assessment is inclusive and precise, off-site analysis takes less effort and can be done frequently, which makes its valuable tool for regulators. Cole and Gunther (1998) even argue that off-site examination can be more accurate than on-site assessment. Those positive aspects of the off-site examination with the difficulty to visit numerous banks in the U.S. make the decision to focus on off-site models quite obvious. Therefore, the models used in the analysis will contain only financial statement figures.

Although this study is not about the subprime crisis, the subprime related loan instruments can be considered as having an impact on bank failures. As mentioned already, subprime is relatively new loan instrument having developed under the blessing of Bush administration, and being targeted to low-income consumers that do not fulfill the criteria for prime loans. Moreover, subprime mortgages were packed together to form mortgage-backed securities that were then sold to investors and other banks. Thus, financial institutions were not only buying and selling securities, but also guaranteed them. This complex structure of subprime loans makes it very difficult to detect what is the total subprime exposure of a specific bank, but it is commonly agreed that subprime securities were one of the main reasons behind the current banking problems. Hence, also subprime loans are discussed in more detailed in the upcoming sections.

2.2 Elements of an Early Warning Systems and Bank Failures

Now that definitions are covered it is time to turn the attention to the elements used in constructing early warning systems. Thereby, the focus is on creating a model that predicts individual bank failures. The choice of explanatory variables for the bank failures is guided by economic theory, more accurately, by the recognized sources of financial fragility arising from the banking industry. Next the main fragility factors are listed.

Liquidity. Banks are suppliers of liquidity, since they transfer illiquid assets into liquid liabilities. This function naturally makes banks vulnerable to liquidity crises, and hence the set of explanatory variables needs to include measurement of liquidity risk. Examples of liquidity risk ratios are total security holdings to total assets and valuable time deposits to total time deposits.

Credit risk. Banks' customers have diverse backgrounds and thus can be seen as unequal debtors. Fortunately, banks can pool the risk of different investment projects. Nevertheless, this does not erase the problem of credit risk. Indeed, variables that proxy the credit risk must be included to the analysis. Such ratios are, for example, total loans and leases to total assets and assets in nonaccrual status to total assets.

Profitability and taxes. Banks, as many other companies, cannot stay in business unless they make profits. Therefore, ratios including profitability need to be added into the model. Classical ratios such as return on assets and return on equity are good proxies for profitability as well as cash dividends to total assets. It is also notable that solvent banks tend to pay higher taxes than banks facing severe profitability problems. Thus, applicable income taxes to total assets ratio should be included to the analysis.

Size and growth. If the financial institution is sizeable and growing fast, the probability of the bank failure decreases. Therefore, figures such as total assets and change in total equity from previous year to current year cannot be ignored.

Loan mix. It is also crucial to study whether the type of a loan matters. In other words, it is important to examine if the real estate lenders have higher probability to fail than banks lending for example to businesses. Furthermore, the liquidity of the loan portfolio might include vital information. For example, ratio of time deposits to total deposits should be included to the analysis.

Securities. Although the business idea of a bank is to buy and sell assets, it still invests to securities at some level. As mentioned, the banks' total exposure to subprime related instruments is hard to detect, but the picture of riskiness of banks' investment portfolio can be studied by dividing the portfolio into pieces and analyzing the nature of failure banks' investments. Ratios such as mortgage-backed securities to total securities and collateralized mortgage obligations to total securities should therefore be included to the analysis.

Instability. A bank close to failure can be expected to have some degree of volatility in its financial performance. As sudden changes in amounts of assets or equity prior the collapse of a bank can be seen as a proxy of failure variables, instability parameters must be added into the analysis. Total assets to mean of total assets and equity to mean of equity are example of ratios worth of closer look.

Elements presented here are only ceilings to the specific data used in the study. Ratios that are crucial for the analysis will be explained more carefully in the data section.

2.3 Background of the U.S. Bank Failures

The history overview of the U.S. banking starts from the foundation of Federal Reserve System (FED) in 1913. FED was founded as an attempt to bring stability to financial markets after the Panic of 1907 exposed weakness of an uncontrolled system (FDIC, 2006). It is vital to understand the lessons learnt from the Great Depression and the S&L crisis as they reveal crucial issues behind a bank operating logic. By studying the history of banking problems, it is possible to predict the critical factors behind the future financial institutions bankruptcies. The regulatory details will not be discussed in detail as only the main regulatory development behind the recent bank failures will be clarified.

2.3.1 The Era of the Great Depression

In the era of 1929-33 U.S. financial system experienced very difficult and chaotic period. Numerous bank failures culminated to the shutdown of the entire banking system in March 1933 (Bernanke, 1983). As Bernanke (1983) points out there were two main reasons behind the financial collapse in the 1930s: the loss of confidence in financial institutions, mainly commercial banks, being the one, and the widespread insolvency of debtors being the second. As topic of this paper is in bank failure prediction, both of the elements are considered relevant.

The failures of financial institutions were substantial during the Great Depression. Many different financial institutions were swept away, but commercial banks were the ones to take the heaviest hit. The percentage of operating banks that failed in each year from 1930 to 1933 was 5.6, 10.5, 7.8 and 12.9, respectively. Due to the failures and also mergers in banking industry, the number of operating banks at the end of 1933 was a little bit above half of the number that existed in 1929. Also those banks that survived faced severe losses during the depression. (Bernanke, 1983)

There are several reasons behind the bank failures in the 1930s. One explanation is the substantial number of small banks in the U.S. at the time with low capacity to recover from any kind of shock. One can even argue that some of the bank failures where desirable and were due to ‘natural causes’, as the legal barriers to entry in banking were very low compared to the other big western countries such as Canada, U.K., and France. However, not all the failures were ‘natural’ as panic in financial markets explains great amount of the failures. When bank’s customers are withdrawing their assets nearly simultaneously, the bank quickly needs to sell its illiquid assets or even dump part of them in order to survive. Other investors soon realize the problem causing bank panic, where substantial amount of people are withdrawing their funds simultaneously. This is driving otherwise solvent banks into troubles. In the end, this will lead to a situation, where solvent banks fail mainly because of illiquidity of their assets and the panic in the market. (Bernanke, 1983) Therefore, the expectation of failure, by the mechanism of the run, tends to become self-conforming. The phenomenon is also referred as a sunspot (Cass & Shell, 1983).

As Bernanke (1983) states, another reason causing financial crisis was the insolvency of debtors. During the Great Depression deflation made the loan payback burden unbearable as the loan contract where written in nominal terms. Especially households, small companies, and farmers faced severe problems due to deflated food prices. Thus, borrowers were unable to meet their payment obligations causing major problems to already struggling commercial banks.

In addition to Bernanke's (1983) explanation, Friedman and Schwartz (1963) have their own explanation to the issue. According to them the first banking crisis after October 1930 might be a consequence from poor loans and investments made in the twenties. During that time the quality of loans granted was not high enough and substantial amount of risk was taken. Also the level of bank reserve requirements was low allowing banks to grand loans to poorer and less solvent customers. (ibid.) Consequently, the original sin in the 1920s and 1930s seems to be the same than what it was in the 2000s: lack of liquid assets, low bank reserves, and considerable amounts of risky lending. Therefore, liquidity, credit risk, and loan mix are important variables to include to the empirical analysis of the thesis.

2.3.2 The Banking System Problems from 1980s to early 1990s

After the Great Depression the banking industry in the U.S. grew steadily many decades without significant problems. However, the 'problem free era' ended in the 1980s, when banks started to fail with increasing speed. During the 1980s and early 1990s more than 1,600 commercial or savings banks insured by the FDIC failed. The peak was in years 1987 and 1989, and overall failing rate reached the level of nine percent measured both in total number and total assets of banks compared to the figures in the end of 1979 including all the banks founded during the subsequent 15 years. (FDIC, 1997) Although there was an unquestionable costly problem in the U.S. banking industry, the episode did not necessarily fulfill all the components of a crisis at least when compared to the episode of the Great Depression.

The special feature of the S&L crisis was the concentration of the bank failures on a few states. The most affected states were Texas, Oklahoma, Louisiana, New Hampshire, Massachusetts, Connecticut, and California. In addition, bank failures seemed to occur in different periods. (González-Hermosillo, 1999) For the purpose of this study these problem regions are divided into three sub areas, the Southwest, the Northeast, and California, in order to provide more precise explanation of the specific problems related to the episode.

The substantial part of the bank failures in the Southwest (Texas, Oklahoma, and Louisiana) occurred during 1986-92, peaking in 1988-89. The main reason behind the banking problems was the significant drop in oil price after 1981, and its final collapse in 1986, giving a severe shock for the energy-producing Southwest states. As a result of the bloom of energy market in the 1970s, also the real estate prices had ballooned. Therefore, as oil prices began to fall, the real estate markets started to struggle as well leaving damaged banks behind. Moreover, the weak agricultural price level toughened the situation deepening the banking problems. (González-Hermosillo, 1999)

However, not all the banks failed during the energy price crisis. Interestingly, the main sorting factor between surviving and failing banks was the level of commercial and industrial loans relative to the total loans. Banks with more aggressive business loan ratio took the hardest hit, since a large part of the industry loans belonged to energy companies that were struggling, and a substantial portion of commercial loans belonged to real estate sector both facing major problems. (FDIC, 1997) As a conclusion, it can be argued that a severe exogenous shock was the starting point of the banking problems in the Southwest.

If the exogenous shocks caused the problems in the Southwest, the Northeast suffered mainly due to fall of the real estate markets that boomed during the 1980s. The most of the bank failures happened during 1991-92, just after the end of the cold war due to decline in computer industry that was heavily concentrated to New England. Again the banks with high level of commercial and industrial loans relative to total loans suffered more severe losses than the banks with more conservative loan ratio. (González-Hermosillo, 1999) California had similar problems than the Northeast, but in addition it was also a major recipient of Japanese investments. Therefore, the recession in Japan in the 1990s was strongly reflected to the financial sector of California in years 1992 and 1993. (González-Hermosillo, 1999)

Although many banks in the U.S. failed after 1985, the actual cause originated from 1960s. As mentioned, couple of severe macroeconomic shocks had a huge impact on banks, but one of the common factors behind the problems was the deregulation of the banking sector from 1960s through the 1980s. For example, in 1967 state of Texas approved major liberalization of so called savings and loan association or thrift powers. Since then, the property development loans were allowed up to 50% of the net value. This was only a starting point of series of new regulations. Depository Institutions Deregulation and Monetary Control Act (DIDMCA) enacted in March

1980 and Garn - St Germain Depository Institutions Act in December 1982. Those acts basically increased the size and diversity of loans granted. In addition, in September 1981 Federal Home Loan Bank Board permitted troubled thrifts to issue so called income capital certificates that were purchased by Federal Savings and Loan Insurance Corporation (FSLIC). (FDIC, 2002) The granted certificates made insolvent institutions look virtually solvent and actually deleted the downside risk of borrowers' possible failures.

As a consequence of the deregulation, S&L institutions faced though competition in lending markets aggravating banks opportunity for profit making. Therefore, banks realized an opportunity for notable profit margins when junk-bond market started to blossom in the 1980s. Many S&L institutions invested in junk bonds, which led into troubles when house and oil prices started to fall. In 1991 The Financial Reform, Recovery, and Enforcement Act (FIRREA) contributed to the problems in the junk-bond market to some extent. The main function of FIRREA was to reverse part of the deregulation made earlier in the 1980s. Thrifts were now denied to invest in junk-bonds, and a minimum of 70 percent of thrifts' assets were required to hold in residential mortgages and mortgage-backed securities. Additionally, one major objective of the act was the bailout of insolvent institutions. (Wolfson, 1986)

Finally, in the beginning of the 1990s, the interest rates declined causing sharply upward-sloping yield curve that improved the value of bank security portfolios and raised net interest margins on new loans, diminishing the number of bank failures. (FDIC, 1997) As a summary, it can be argued that deregulation with macroeconomic shock is an unsettling combination and might conclude severe problems to banking industry if banks lack cushion and conservative loan ratios and are investing heavily on risky assets. Thus liquidity, credit risk, and loan mix are the factors that should be covered in the upcoming empirical analysis.

2.3.3 The Subprime Crisis

The year 2008 was one of the darkest years in the history of U.S. banking. One of the most shocking incidents was the failure of Lehmann Brothers in September 2008 revealing the seriousness and depth of the current crisis (Epiq, 2009). As a consequence, FED injected substantial amounts of money into the banks to keep the other financial institutions alive. Suddenly people were talking about a new depression. But what were the actual reasons behind the crisis and, more precisely, bank failures? Why it all happened such a storming pace and depth? This chapter provides an answer to these questions.

2.3.3.1 Housing Market Boom

Kiff and Mills (2007) argue that one of the main initial factors behind the crisis was the increasing inflation in housing market that made a house purchase exceptionally expensive in the 2000s. Due to inflated real estate prices consumers were forced to apply sizeable mortgages. Kiff and Mills (2007) continue that people that normally had not afforded to own a house were now able to buy one as they were offered subprime mortgages, a household loans targeted to individuals that do not fulfill the criteria of prime loans. The assessment of the subprime borrower is usually done by using three different standards: borrower's statistically calculated credit score or so called FICO score, debt service-to-income (DTI) ratio, and the mortgage loan-to-value (LTV). DTI is defined as the percentage of a consumer's monthly gross income that goes toward paying debts. LTV, on the other hand, expresses the amount of a first mortgage lien as a percentage of the total appraised value of real property. Normally, borrowers with low FICO scores, below 620, DTIs above 55%, and/or LTVs over 85% are likely to be considered subprime (Foote, Gerardi, Goette, & Willen, 2008 and Demyanyk & Van Hemert, 2008).

There were also loans falling in between prime and subprime loans. So called Alt-A loans were borrowed by the consumers that usually covered all the criteria mentioned, but failed to provide complete income documentation. (Demyanyk & Van Hemert, 2008). Kregel (2008) illuminates that Alt-A and subprime loans cover relatively small part of the total population of mortgages as they are rather new instruments in financial market.

The procedure for sorting potential borrowers has changed dramatically. Approximately a decade ago banks were still institutions that made risk assessments based on individual track record and trust. Now, however, this task has been transferred to rating companies such as Fair Isaac. What they do is they basically convert a person into numbers and employ algorithms to create individual's risk level. (Kregel, 2008) This, of course, has increased efficiency but has also had major downturn, namely the undervalued risks. The level of risk culminated in subprime and Alt-A loans, which were not guaranteed by a government sponsored entity such as Fannie Mae or

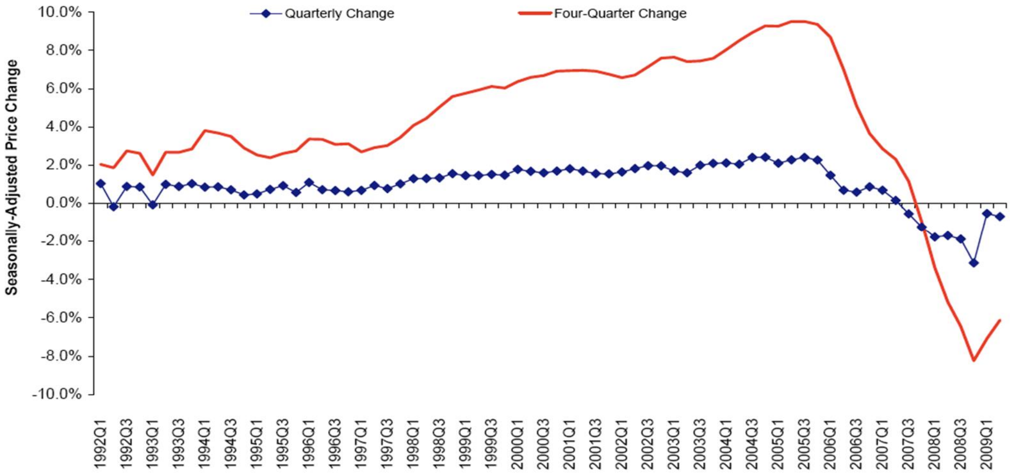
Freddy Mac. Therefore, banks that were lending subprime loans or investing in subprime securities were carrying the full credit risk.

Subprime mortgages in general have higher interest rates than prime loans, but the loan terms are easier to meet. This can be seen as one of the main reasons for 5% increase in homeownership rate from 1995 to 2005 as poor people were granted mortgages. There have also been speculators obtaining loans on the basis of expected collateral appreciation. Naturally, increased homeownership rate combined with opportunistic behavior pushed the real estate prices to higher level and therefore forced banks to lend larger mortgages, to enable Americans to buy a decent size home. (Kiff & Mills, 2007) Mortgage tax deductions can also be seen as one incentive increasing oversized real estate loans as it has made it possible to use part of the mortgage for personal consumption (Klyuev & Mills, 2007).

Due to the booming house prices mortgage lenders ended up in difficulties as they were unable to find borrowers to meet the conforming loan terms such as DTIs and LTVs although the FICO score requirements were met. Therefore, increasing amount of borrowers fell into Alt-A or subprime class. (Kiff & Mills, 2007) The market share of the subprime mortgages basically exploded between 2001 and 2006 from around 8 percent to 20 percent, and securitized share of the subprime mortgage market grew from 54 to 75 percent. Especially, the high-LTV borrowers became increasingly risky compared to the low-LTV borrowers. This also showed in the interest rates paid by the households as the above-average LTV borrower's premium compared to an average LTV borrower grew from 10 basis points to 30 basis points. In many respects the subprime market experienced a classic boom-bust scenario with rapid market growth, loosening underwriting standards, and worsening loan performance. This all was disguised by the raising house prices. (Demyanyk & Van Hemert, 2008) Thus, in the analysis part also the impact of increased mortgage risk is covered by studying, for example, the ratio of total real estate loans to total loans.

Furthermore, the loan granting procedure in banks were rather loose couple of years ago as so called ninjas (no income, no job, and no assets) grew heavily during 2000-06. This was also the period when brokers received huge upfront paid commissions from each new loan. It has been an effective way not only to swell brokers' salary but also to increase banks' risk levels due to reckless loan granting. (Blackburn, 2008) The reason behind the loose loan granting system has been so called 'ownership society' envisioned by the Bush administration. As the press release of the White House reveals President Bush believed that “homeownership benefits individual families by building stability and long-term financial security” (The White House, 2004). Although being a noble goal, it was not very realistic as also the people unable to pay mortgage got one. As we understand it now this kind of system cannot live forever, and after a couple of years with teaser rate mortgages interest rates in house loans skyrocketed at the same time with falling house prices leaving many distressed home owners no other option than default as prepayment and refinancing options were not realistic with little or no housing equity. (Blackburn, 2008) Foote et al. (2008) share the same view and stress that there are clear causality between falling real estate prices and subprime borrowers foreclosures. The dramatic fall of the house prices is illustrated in the Figure 2 below. As the main question of this study is how the inflated loan exposure effected on bank failures, the explanatory variables answering the question need to be added into the data set.

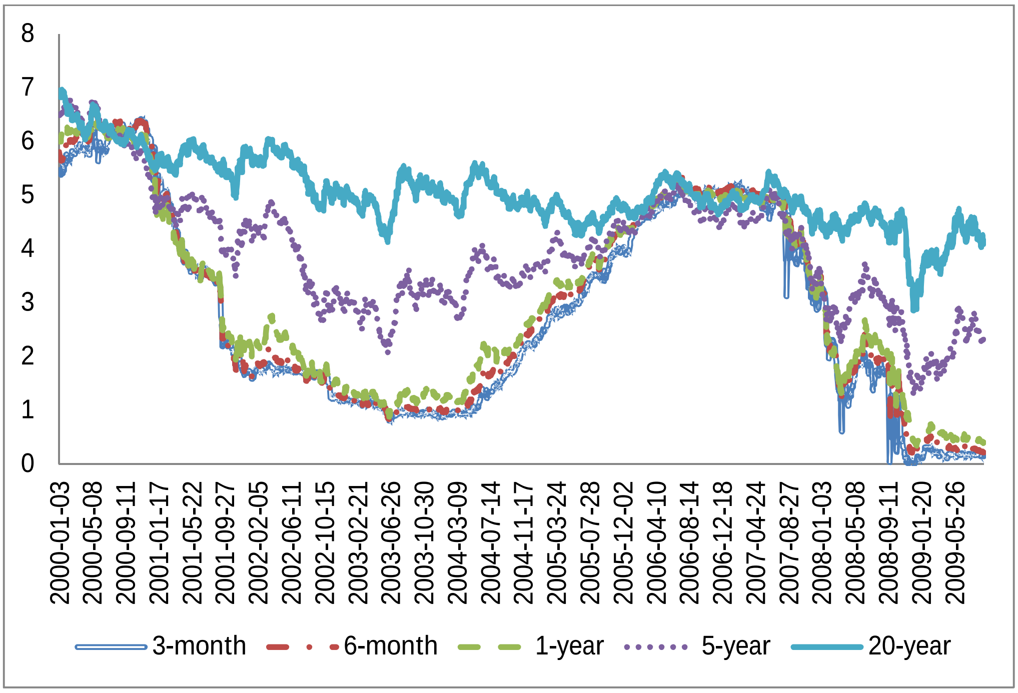
Figure 2. FHFA House Price Index History for U.S.



Source. Russell & Mullin, 2009

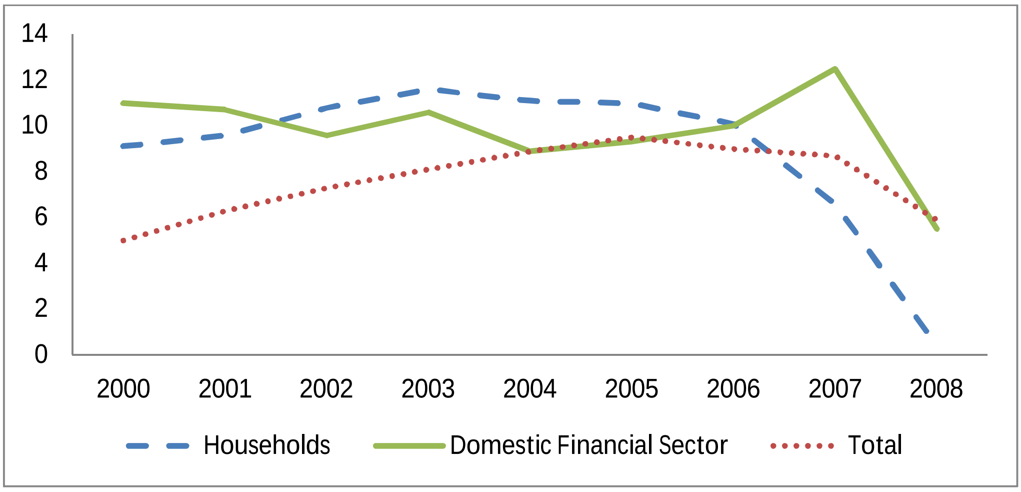
The base for the crisis was built between 2001 and 2005, when interest rate levels were at the low level. During that period U.S. increased its debt burden enormously especially in the households and financial sector. Such a low interest rates and loose monetary policy can be explained by soft landing from the share bubble and national security after the 9/11 attacks, since U.S. wanted to show its economic power by increasing purchasing power of its citizens (Blackburn, 2008). Americans consumed with debt while banks and regulators encouraged for that. Figure 3 and 4 exhibit the interest rate movements and increased debt burden.

Figure 3. Market yield on U.S. Treasury securities



Source: The Federal Reserve Board

Figure 4. Seasonally Adjusted Debt Growth



Source: The Federal Reserve Board

There were also illegal speculators with appétit for quick profits who supplied poison assets to the market by using so called predatory lending where borrowers were deceived to take an unfair loan. Usually some teaser rate was included to the deal as bait and when the originator granted the loan it was almost simultaneously repacked and sold to the investors as securities. Predatory lending is naturally against the law, but it took a while before authorities realized the problem whilst the damage was already done. (Blackburn, 2008)

As Blackburn (2008) stresses teaser rates and predatory lending seem to have a crucial role in subprime mortgage crisis. However, as Foote et al. (2008) point out that might not be the case. They show that the so called teaser rate was actually rather high around 8 percent and was well above 1-year prime ARM rate. Therefore, it can be argued that the teaser rate was not the issue behind the subprime crisis, although existing during that time. Also the dramatic increase in interest rate levels after two or three-year period is misleading according to Foote et al. (2008). They stress that the usual way of defining interest rate after the initial period was to use 6-month LIBOR as a base rate and then add around 6 percentage points' margin on the top of it. For example, when in 2006 the 6-month LIBOR was around 5% and the margin was 6 percentage points, the mortgage rate set up at the level of 11 %. Indeed, three-percentage point increase to the high initial rate is hardly explosive growth as some authors state it.

Increasing defaults among U.S. subprime mortgage holders in the last quarter of 2006 and early 2007 can be seen as the trigger for the credit crunch. It is crucial to notice that it is not only poor individuals with bad credit history that took too much loan. On contrary, foreclosure figures state that subprime borrowers covered only minor part of the defaults as in the U.S. prime loan borrowers were responsible at the majority of the foreclosures. (Foote, Gerardi, Goette, & Willen, 2008) At the time of the foreclosures, interest rates climbed to protect the falling dollar. As the first signs of the crisis were revealed the world's central banks tried to pump huge amounts of liquidity into the global financial system. The impact was, however, temporary and the banks remained unwilling to lend to another. Again, liquidity seems to be related to the failure of financial institutions with the already mentioned credit risk variables. Therefore, it is obvious that liquidity should be one of the factors under the study of the thesis.

2.3.3.2 *Mortgage Securitization*

Second major factor behind the current financial crisis according to Kiff and Mills (2007) is the increased demand of so called mortgage-backed securities (MBS). In the eyes of investors these securities appeared as relatively low risk trading instruments with high-expected rate of return. In MBSs mortgages were combined together and then sold to investor mainly during 2005-06. High demand was a good incentive for lenders to sell as much loans as possible, since the risk taker in the case of default was the MBS guarantor. This is possible when financial institutions are both investing and guarantying MBSs. This was done in order to limit informational asymmetry and create faith among investors. In addition, issuing banks felt that MBSs and other backed securities had desired risk-return and diversification profile. Therefore, when the depositors started to default, banks started to make huge write downs when realizing they had quite a bit of 'trash' in their balance sheets. (Gibson, 2004) The subprime crisis generated write-downs and credit losses worth of approximately 1.5 trillion dollars during 2008-09. (Tan, 2009)

Also the former Vice President of Lehman Brothers bank, Larry McDonald, stated in the BBC's HARDtalk (2009) that Lehman Brothers' biggest mistake was to go into the risky 'storage' business instead of staying purely at trading business. The bank now tied huge amounts of money in mortgage related securities instead of doing its core business by taking fees from trading transactions. McDonald continues that the reason why Lehman Brothers was able to avoid failures in the past was its pure concentration on being a financial broker. The same applies with commercial banks, which main business is to lend and borrow money. One may argue that the sin of tying assets into the risky securities deteriorated banks situation so that failures occurred.

The securitization was usually made by so called special purpose vehicle (SPV) or special purpose entity (SPE). The reason of creating such a stand-alone entity was banks' need to transfer loans and the credit risk related to them from their balance sheets to the SPV, which is a legally independent financial institution that issues its own liabilities in order to acquire the assets originated by the bank. However, the liabilities that were sold to the institutional investors needed to carry an investment-grade rating from a nationally recognized rating organization. (Kregel, 2008) As the reputation of prestigious originating banks helped to create thrust among rating companies, MBSs carried risk rating that seemed low and reliable to the investors. As can be imagined, the shock was significant when investors realized that the risk was estimated terribly wrong.

As mentioned, SPVs absorbed different kinds of assets. These assets have been ranked so that Level 1 assets could be valued in stock and included liquid types of securities. The value of Level 2 assets were based on a model, which relate them to an index or similar traded assets. The Level 3 values of assets could be determined by using different kinds of mathematical and theoretical models with no directly traded element. Naturally, Level 3 assets were the most illiquid ones and hardest to price accurately. It is important to note these levels as the level structure plays an important role, since a SPV needs to set a market price to its assets from time to time, in other words mark to market. As most of the MBSs fell from Level 2 to Level 3 in 2007 many financial entities ended up having plenty of gravely priced Level 3 assets in their possession, creating an enormous possibility for a valuation error. (Blackburn, 2008) This naturally made the situation of the SPVs rather difficult for the banks and investors, since the level of risk increased dramatically when major part of the assets were more illiquid than earlier. As one can notice, the liquidity issue can be seen again quite crucial matter in the case of bank difficulties.

Although only MBS have been presented until now, it is not the only security that includes subprime loans and mortgages. Asset-backed securities (ABSs) may also include subprime loans, although they do not include mortgages. On the other hand, collateralized debt obligation (CDO), or collateralized mortgage obligations (CMO), which includes different kinds of banks' receivables such as loans and mortgages is the most complex instrument even including assets that were backed twice. It is also important to notice that while creating for example CDOs different tranches were formed. The purpose of different mortgage and asset tranches was to build a pool of assets with different level of quality and risk. Generally, tranches were divided into three different classes: senior, mezzanine, and equity. Senior tranches had the highest quality and lowest expected rate of return and it was used as a base for MBSs and CDOs. Usually, securities included approximately 70 percent of senior assets. Since the senior tranches' expected return was rather low compared to other tranches, mezzanine and equity tranches were developed to tempt investors. Second highest quality assets were mezzanine assets that offered higher expected return than senior tranches. Bottom of the chain were equity assets that in a case of default would lose its value first. Obviously, equity tranche offered also the highest expected return. (Blackburn, 2008)

One should notice that there were also so called single-tranche CDOs. In those securities only one tranche of the capital structure is sold to an investor. What it basically means is that issuers were able to sell only tranches that faced highest demand. Also the overall demand of CDOs increased when single-tranche CDOs came out, since investors were able to buy only an instrument that they believed would serve their investment portfolio the best. (Gibson, 2004) One more reason why CDOs were offered with so relentlessly was the commission that the underwriters received (Blackburn, 2008). In other words banks made decent money just by repacking the loans.

Gorton (2008) points out that the asymmetric information among different parties in loan markets had also notable impact on the subprime crisis. In other words, it is impossible to predict for example future house prices while lending and borrowing mortgages. He stresses that house prices and mortgage performance information arrives with a lag, and thus they cannot trigger the crisis. Gorton (2008) claims that different kinds of ABSs were traded and priced with the issuers' reputation in stake. Therefore, when complex securities were not understood, buyers took the risk rate and expected rate of return as given and blindly trusted the issuer's prestige bank reputation. Moreover, he states that current crisis is after all caused by a banking panic. Although the structured securities and capital markets have been the eye-catcher this time, the fundamental base case is the short term liability holders' refusal to fund banks due to the fear of losses, thus the logic is the same than in the earlier panics in the U.S. This time, nevertheless, the SPVs played a crucial role and eventually were the reason why liquidity of repo market dried up.

As a conclusion one may argue that lending to risky subprime-individuals was reckless decision for many banks. Also loan ratios were not as conservative as they should have been in order to maintain solvency. Again roughly same mistakes were made than in the earlier crises creating the possibility to implement the traditional ex ante prediction tools for forecasting current bank failures with traditional financial ratio variables. Now, however, relatively new subprime mortgage related instruments should have the focus they deserved in the analysis. Therefore, some weight is also put on quality of mortgage lending and banks' investments.

2.4 Failure Prediction Models

After historical glance of the banking crises and elements related to them, the next step is to illustrate the evolution of bankruptcy prediction models. As stated earlier, bank failure models are divided into on-site and off-site models. On-site models include different kinds of applications of so called CAMEL analysis employed by several regulators. Off-site analysis is mainly done with help of statistical tools and financial statement information. In this chapter, the on-site models and other systems used by U.S. regulators are firstly covered and thereafter the mainly academic off-site models presented.

2.4.1 Regulators' models

The traditional approach to assess financial difficulties of individual banks is closely related to the work of supervisors of the banking system and rating agencies. The most knowing rating system is so called CAMEL the acronym for the criteria: capital adequacy, assets quality, management, earnings, and liquidity. The score of individual performance of an institution is compared to all other institutions, generating a bank-specific rating index. The original CAMEL system is the purest form of the on-site model. In 1997 the sixth component, sensitivity to market risk, has been added to the system and thus the model is currently called CAMLES (Lopez, 1999). Both CAMEL and CAMELS work the same way and include a rating for each individual component of a scale from 1 (best) to 5 (worst). The rating is based on broad and general on-site evaluation of both qualitative and quantitative information of the bank. From the individual component ratings the overall index is computed. The supervisor running the analysis weighs the ratings of different components into the composite index. Then the overall information gathered is used to decide if the specific action or tighter supervisory is needed. Normally, if bank's index is less than two, the financial institution is considered to be a high-quality bank, whereas institutions with scores four or five are rated to be insolvent (Curry, Elmer, & Fissel, 2009).

Although the rating systems such as CAMELS are effective measures of the current condition of banks, and therefore can be seen essential tools for banking supervisory, there has been an increasing debate about the limits of the approaches. CAMEL(S) models parallel the condition of the bank only the time of the examination, while variables in the systems are highly responsive to changes in the economic conditions and the bank performance. Moreover, the fact that the measurement is ex-post in nature does the reacting to the insolvency in time difficult. (Gaytán & Johnson, 2002) Nevertheless, CAMEL(S) systems are still rather popular among regulators.

One way to predict bank failures is to use off-site data in order to estimate ratings that institutions gain from on-site examinations such as CAMELS. U.S. Federal Reserve's System of Estimate Examination Ratings (SEER) is one example of that kind of model using limited dependent regression techniques in order to determine the historical relationship between a set of variables. The SEER has an indicator function, which takes a value 1 if the dependent variable belongs to a predetermined interval and zero otherwise. Variables in the SEER are proxies for credit risk, leverage risk, and liquidity risk. The results given by SEER are used for a periodic estimation of ratings. Although the current estimation mainly reflects the present condition of the bank, the history of the analysis performance can show any deterioration in the condition of the institution. Moreover, the model can create an ex-ante indicator of insolvency, since it allows estimating the probability of rating downgrade of a whole financial institution, and the specific areas responsible for the downgrade. The rating for SEER goes form 1 (best) to 5 (worst) the same way than for CAMEL(S). (Gilbert, Meyer, & Vaughan, 2000) FDIC's SCOR is another example of the similar on-site bank failure predicting tool. Extension for SCOR, FDIC created its own CAEL system, which works the same way than CAMELS, but its quantitative part of the analysis is mainly based on off-site data. (Jagtiani, Kolari, Lemieux, & Shin, 2003)

There still has been demand for models not only predicting failures but also the timing of the failure. That kind of information is naturally very valuable for regulators and banks' managers. One example of the model detecting also the timing of the possible failure is called Canary and it has been created by the U.S. Office of the Comptroller of the Currency (OCC). It estimates the likelihood of the failure and the probability that the bank survives beyond a two-year period. The set of explanatory variables includes nonperforming loans, provisions for loan, and capital-asset ratios as well as economic indicators such as interest rates, wages, and unemployment rates. Those variables are then compared against the historical benchmarks, which state the probability of the potential bankruptcy of the bank. (Hawke, 2000)

Not all periods are covered with bankruptcies, and therefore models for predicting failures have been difficult to build. For example, the French Banking Commission's Support System for Banking Analysis (SAAB) estimates not straight the possible failure but the potential future losses of the bank. However, these kinds of models are beyond the scope of this study, since in the concentration is on complete bank failures and ways of predicting them.

2.4.2 Academic models

There is a long history of predicting bankruptcies in the finance and economics literature using only the off-site data. In this chapter the evolution of the bankruptcy predicting models is covered. First, earlier models are explained and second, more modern and sophisticated models are illustrated.

2.4.2.1 Earlier models

Paired-sample technique generates major part of the bankruptcy prediction models. One part of the data is gathered from firms or banks that eventually failed, whereas the other part includes data from solid companies at that same period of time. Then, traditionally, a number of plausible financial ratios are studied from the financial statements that were published before the failure. Next, an equation that best discriminates between companies that failed and companies that remained solvent is developed by using either single ratio or combination of ratios.

Beaver (1966) is one of the first researchers implementing financial ratio analysis for bankruptcy prediction. Unlike subsequent studies, Beaver uses broad definition of bankruptcy in his research, defining a firm to be failed if it faces bankruptcy, bond defaults, overdrawn bank accounts, or has missed preferred dividends. The broadening of the definition does not seem to have any substantially impact on empirical results. In his study Beaver finds three individual financial ratios, which are well suited to predict financial failure: cash flow to total assets, net income to total debt, and cash flow to total debt. It should be noted that each of these ratios consist of a flow variable involving earnings or cash flow divided by a stock variable. In his research Beaver derives cut-off points for each ratio so that companies with ratios above the cut-off point are classified as possible non-failures, while those with lower ratios are considered as potential failures. These cut-off points are derived from an original sample, but are then used to classify firms also in a holdout sample. As a result, he concludes that single ratios can predict bankruptcies rather well.

Altman presented in 1968 similar idea than Beaver in 1966 that firms with certain financial structure have greater probability to fail within the next period compared to companies with opposite characteristics. Altman, however, felt that basing bankruptcy prediction only on a single ratio is too simplistic to capture the complexity of financial failure, although Beaver’s model gives surprising accurate predicting results. Altman’s tool for analyzing failures is multiple discriminant analysis (MDA), which is a statistical technique used to classify a categorical dependent having more than two categories, and using it as predictors for number of independent variables. In his study, MDA is used to construct a predictive algorithm based on five key financial ratios: working capital to total assets, retained earnings to total assets, earnings before interest and taxes (EBIT) to total assets, market value equity to book value of total debt, and sales to total assets. With help of those variables so called Z-index or Z-score is calculated.

The Z-index is one of the first statistical off-site models for predicting bankruptcies. Although it is built for companies and not for banks the model is still worth of a bit deeper analysis. The MDA computes the discriminant coefficient, while the independent variables are the actual values. As was mentioned, Altman chooses five key ratios that include the most of the financial information. One major reason for that is the statistical limitation of the Z-index; there is a reason to believe that some measurements have a high degree of correlation or collinearity with each other. This naturally forces the researcher to keep the predictive variables in limited numbers.

As a first notable bankruptcy prediction model, Altman's Z-index has been disputed greatly. Academic discussion seems to disagree what are the key ratios that should be used in the analysis. For example, Boritz (1991) finds as many as 65 different financial ratios used as predictors in bankruptcy studies. Moreover, Hamer (1983) argues that ratios selected to the analysis do not have notable effect on the model's ability to predict failure. Karels and Prakash (1987) suggest quite opposite encouraging researchers carefully to select the financial ratios to include to the model, in order to improve prediction accuracy.

Not all the models are based on financial ratios deriving directly from financial statements. For example, cash flow models are based on fundamental financial assumption that company's value equals the net present value of its expected future cash flows. The company files for bankruptcy if it does not have sufficient cash available to service debt outflows as they become due. If assumption that current cash flows predict future financial status of the company, then it can be argued that past and present cash flows are good indicators of both company's value and eventually probability of bankruptcy. Gentry, Newbold, and Whitford (1985a and 1985b) and Aziz, Emanuel, and Lawson (1988), among others, use this assumption while building cash flow model of bankruptcy. In the model, the value of the firm is derived from the ratio of streams of the discounted cash flows to and from operations, government, lenders, and shareholders. In their studies, matched bankrupt and non-bankrupt firms are compared, and it has been found that the group means for operating cash flows and cash taxes paid differ significantly in all five years prior to bankruptcy. The presented findings seem intuitively reasonable, since the company that is close to failure has made reckless investment decisions and operational efficiency can be expected to be weak. In addition, while all the companies are minimizing their tax payments, distressed companies with low level of earnings have practically no tax liabilities. If accuracy of the two models, Altman's Z-index and cash flow models is compared, Aziz, Emanuel, and Lawson (1989) conclude that cash flow model is superior to the Z-index.

One should note that Altman, Haldeman, and Narayanan used Z-index to create an updated bankruptcy prediction tool called ZETA in 1977. The new model, however, is a proprietary effort from researchers, which is the reason why it does not fully disclose the parameters of the market. This may be the explanation why the whole upgrade of the Z-index has been occult from the academic discussion. Only Scott (1981) has studied the model and concluded that it is well suited for bankruptcy prediction.

Gambler's ruin theory is a bankruptcy prediction model worth mentioning, although it turns out to be too simple for the failure forecasting. Wilcox and Vinso in their studies in the 1970s based bankruptcy prediction models on that theory (for example Wilcox 1973 and 1976, and Vinso 1979). The simple idea of the theory is that the gambler begins the game with a random amount of money. He wins a dollar with a probability p or loses a dollar with a probability 1 – p. The game goes on until the gambler or his component loses all the money. In financial applications the firm is viewed as the gambler, and bankruptcy occurs when the firm's net worth falls to zero. Naturally, the model used in bankruptcy prediction is well developed from the example presented; the logic remains nevertheless the same.

Beaver (1968) is one of the first researchers analyzing relationship between bankruptcy probability and stock returns. He finds that equity returns predict bankruptcy earlier than financial ratios in general, which is consistent with market efficiency hypothesis. Altman and Brenner (1981) and Clark and Weinstein (1983), conclude that the stock market indicates bankruptcy at least a year before it happens. In addition, Clark and Weinstein (1983) find that bankrupt firms lose approximately 26 % of their capital during the two-month period surrounding the bankruptcy declaration date. One dimension to the stock return analysis is given by Aharony, Jones, and Swary (1980) while suggesting a bankruptcy prediction model based on the variance of market returns. They conclude that the volatility of firm-specific returns increases as bankruptcy converges. Some level of differences in behavior of variances in returns between failed and non-failed companies can be seen as early as four years prior the failure. Although the approaches presented give a useful tool to predict bank failures, they can be employed only for firms and banks quoted in stock exchange. This naturally cause severe limitation to the models.

Black-Scholes (1973) option pricing model is one addition for bankruptcy studies. In its framework the value of a company's equity is viewed as an option, which is valuable at the time the company's debt matures only if the debt can be paid in full. The assumption of the model states that the market value of the firm follows a diffusion-type stochastic process with known parameters. In their papers Black and Scholes (1973) and Merton (1974) make an assumption that the firm's debt consists of a single pure discount bond. To ease this assumption in order to make the model more realistic Schwartz (1977) provides methods that can be used to obtain a solution in a situation when a firm is expected to have several debt issues that must pay interest and principal on different date. Schwartz's model differs somewhat from the model created by Black and Scholes (1973), but the logic behind it remains the same. In general, the option pricing methodology can be therefore considered as useful bankruptcy prediction tool. However, it does not contain any earnings or cash flow variables causing limitations to the model.

In order to cover losses companies can sell securities alongside with its assets. Two kinds of models have been created around this idea. The first model, created by Scott (1976), assumes perfect access to external capital. This is an opposite view compared to gambler's ruin theory, where after losing the company's money the firm falls into bankruptcy. Another model developed also by Scott (1981) lies somewhere between Scott's previous model (1976) and gambler's ruin theory stating that a firm has a limited access to security markets. Following assumptions change the perfect-access model into a model with imperfect access: a firm may face flotation costs when it sells securities, there may be a personal tax system that favors corporate investments that are internally-funded, or systematic imperfections in the market's pricing of securities may limit corporate access to external capital. As can be noticed, here the model with imperfect access seems more realistic one and is therefore also more reliable model for bankruptcy prediction. Although pure external capital access models have not been presented since Scott's paper in 1981, elements from these models can be found from bankruptcy studies.

One of the most important extensions to the bankruptcy literature lies in the area of banking. The first study falling into this field is from Mayer and Pifer (1970). They use a similar approach as Altman in 1968 and find that financial ratios can predict the failure of the banks rather accurately with a lead-time of one or two years. The methodology Mayer and Pifer employ is a simple regression although they state that the distributional assumptions of discriminant analysis are appropriate while using the bank data. In addition, one interesting application for statistical EWSs was presented by Batte et al. (2007), who created a model where predictive rules jointly use traditional accounting ratios and factors deriving from the long-term debt repayment schedule of the firm. The benefit of taking long-term debt schedule information into account is the improvement in the efficiency of the prediction especially when focusing on long, over three years, time horizons.

Santomero and Vinso (1977) used gambler's ruin model while detecting bank failures. As mentioned earlier, the attempts to apply this model for company bankruptcy prediction have been somewhat insufficient. One reason being the assumption that cash flow results from series of independent trials, regardless the management action. The bank failure probabilities estimated in the study are once more rather low, making the model rather unconvincing. In addition, researchers have not provided a test for the model.

2.4.2.2 Modern models

In addition to the methodologies mentioned, different kinds of logit and probit models are also commonly used in failure predicting studies. Zavgren (1985), Keasey and McGuinness (1990), and Thomson (1991) among others, conclude that reliable results can be derived from financial ratios by using logistic regression. In logit model the failure indicator is a binary variable (zero-one), estimated using a set of explanatory variables such as financial ratios as Altman (1968) in his study. By using the logit estimation, the predicted outcomes are restricted to lay in the unit interval, and are considered as the probability of failure. With this methodology, it is possible to evaluate the explanatory contribution of each independent variable, which can be seen as an advantage of the model. One can also find probit models in couple of bankruptcy studies, but as Maddala (1983) argues the unequal frequency of the failed and non-failed samples suggests the use of logit model rather than probit estimation, since the logistic regression is not as sensitive to the uneven sampling frequency problem as probit model. The logit model used in this study is explained more precisely in section three, where the methodology of the thesis is presented.

In the field of bank failure predicting, researchers have realized that bankruptcy models need to be slightly different for emerging markets than for developed banking industries. The major differences between developing and western banking sectors can be seen in their levels of liquidity, accounting deficiencies, and supervisory framework. For example, González-Hermosillo, Pazarbasioglu, and Billings (1996) as well as Dabos and Sosa Escudero (2004) examines bank failures in Latin America by using two-step survival or hazard analysis and duration models. Although the valid data set for bankruptcy studies may be different for Latin America and U.S., the failure studying methodology is nevertheless reliable.

The survival analysis is one way of predicting bank failures and it has been mainly used in medical and biological studies. Lancaster brought the analysis for the use of unemployment studies in 1990, where it moved to bank failure prediction later on. The model itself is quite complex and only the main points are illustrated here. The procedure employed by González-Hermosillo, Pazarbasioglu, and Billings (1996) in their bankruptcy study has two steps. First, the probability of failure and the factors affecting the likelihood of bankruptcy is determined. In this case the regulators intervention is a discrete variable, which can take value of one (intervention) or zero (no intervention). The way of estimating the failure probability is done by using logit model in panel data context. Second step in survival analysis is to determine the factors explaining the duration of a state of solvency. The main issue survival models emphasize is the conditional probability, which is the likelihood that the event will end in 'the next period' given that it has not already ended. Intuitively, the question best answered through the survival analysis is "what is the probability that bank will fail during 'the next period' if it has survived so far?" For more precisely explanation of the methodology please see for example Kiefer (1988).

With the two-step-modeling one can derive both the probability and timing of the potential bank failure. The data used in the González-Hermosillo, Pazarbasioglu, and Billings' (1996) research is gathered quarterly, not annually, which improves the accuracy of the prediction. Nevertheless, the model has couple of limitations too. First, the model creation requires substantial amount of data, and can be therefore difficult to derive if the banking sector have absorbed new features or instruments affecting the banking business and only couple of banks have failed since then. Second, using it as EWS requires further elaboration until the results can be considered to point signals of distress. Finally, the data for the model is gathered around the banking crisis period, which makes the implication of the results rather challenging for more stable periods.

Machine learning techniques can be seen as the newest branch of bankruptcy prediction models. The artificial intelligence models require specified computer hardware and software, which makes the methodology implementation rather challenging. However, a couple of interesting application can be found from the finance literature. Tam and Kiang (1992) applied so called neural network learning algorithm while making a comparison between failed and non-failed banks and argue that neural networks offer better predictive accuracy than discriminant analysis. A neural network consists of a number of interconnected processing units, where each unit is a solitary computation device. The behavior of the unit is modeled by using the following logic: one specific unit obtaining input signal from other units, thus combining those signals based on an input function, and finally producing an output signal based on an output or transfer function. The signal generated is then transferred to other units as directed by the topology of the networks. The result of the final output unit of the route defines whether the bank is predicted to fail or not.

Other methods from the artificial intelligence field such as evolutionary computation have been rarely used for the bankruptcy prediction modeling. However, some authors, such as Kim and Han (2003) and Brabazon and Keenan (2004), have used genetic algorithms (GA) either on its own or as a hybrid method with neural networks for insolvency prediction problems. In addition, most of the approaches from the evolutionary computation field use genetic programming (GP) again as its own or as a hybrid method with neural networks. GAs provide a stochastic search procedure based on principles of neural genetics and survival of the fittest. The main difference between GA and GP is the GA's more primitive role in decision making, whereas GP can be seen as an extension to the GA. In GA the structures in the population are fixed-length character strings that encode candidate solutions to a problem, whereas in GP structures are programs that, when executed, are the candidate solutions to the problem. One important advantage of the GP approach to bankruptcy prediction is it capability to yield the rules relating the measured data to the likelihood of becoming bankrupt (Alfaro-Cid, Sharman, & Esparcia-Alcázar, 2007). Thus, for example, a financial analyst can see which factors and functions are important for predicting bankruptcy.

Fuzzy set theory can also be counted to the field of artificial intelligence models. Fuzzy set is originally proposed by Zadeh (1965), but different modern applications from the model have been constructed since then. In short, it is a theory of graded concepts. Fuzzy logic uses human experiential knowledge to model domain. In prediction problems fuzzy logic implements the knowledge of the domain expert and utilizes fuzzy mathematics to come out with fuzzy inference systems. Michael et al. (1999), among others, proposed fuzzy rule generator method for bankruptcy prediction and compared it with numerous other predicting models. They conclude that fuzzy set is well suited for bankruptcy prediction and actually outperforming all other methods.

Kolari et al. (1996) brought the so called trait recognition (TR) analysis to the field of failure prediction in commercial banking. The model is first applied in forecasting probability of earthquakes and the existence of mineral deposits, and has been used in bank failure prediction quite rarely. One reason might be the newness of the model and the lack of computer programs that can run the analysis. The model is, however, somewhat more straightforward than other artificial intelligence models.

The reasons why Kolari et al. (1996) test this non-parametric model in bank failures, is the limitations of the dependent regression widely used in EWSs. They point out three main drawbacks of the earlier EWS. First, it is a demanding task to determine which explanatory variables are the most useful in predicting the bank failure. The result of the dependent regression analysis indicates only effectiveness of the independent variable in discriminating between failed and non-failed banks. The same problem has been noticed by Boritz (1991). Second, the dependent regression analysis lack information about how each explanatory variable affects Type I and II errors per se. Third, the previous EWS models are not appropriate for examining relations between variables. Finally, Lanine and Vander Vennet (2006) state that one major disadvantage of the parametric approaches is their dependence on distributional assumptions for the explanatory variables. Naturally, TR has also limitations as it requires substantial amounts of discretionary judgment, which may create estimation bias to the model.

After inclusive introduction of the bank failure models, more detailed explanation of the logit model is provided in the next section with the dataset gathering process illustration. The main focus is on function of the logit model as well as the bank sample and independent variables used in the empirical analysis.

3 EMPIRICAL METHODOLOGY AND DATA

This section provides reasons why logit model has been selected as the methodology for the study. Also more precise explanation about the model is presented. The end of the section is used to illustrate the data applied in the analysis.

3.1 Logit Model

As explained in previous section logistic regression has been used considerably lot in bank failure prediction. It gives accurate estimates and is user-friendly tool for analyzing bankruptcies. For example, Jagtinali el al. (2003) argue that quite simple logit model can beat much sophisticated and complex trait recognition model with the same data. They continue that the quality of the logit model should be tested under different economic conditions and therefore different period of time. As the data Jagtinali et al. (2003) used in their study dates back to 1980s, and the context of U.S. banking industry has changed since then, it is beneficial to apply the data from 2000s.

Another reason why logistic regression is preferable compared to other accurate predicting models is its easiness to use as statistical software for running the logit model is available. In this thesis Stata software is employed, but naturally all the other generally exercised statistical programs could be used. In addition, the feature of providing explanatory power of all the independent variables can be seen as an advantage of the logit model.

Logistic regression lies on an assumption that the probability of a failure depends on a vector of independent variables. Using the logit estimation, the predicted outcomes are limited to lie in the unit interval, and are construed as the probability of an event. The logit model has the statistical property of not assuming multivariate normality among the independent variables, contrary to the probit model that does assume a normal distribution of the data. This can be seen as an advantage when analyzing banking data, as it generally is not normally distributed. The dependent variable *p* is logarithm of the financial institution's, *i*'s, probability of capital inadequacy versus capital adequacy as shown in equation below

1 $$\log\left[\frac{p_i}{1-p_i}\right] = a_1 + a_2 X_{i1} + a_3 X_{i2} + \ldots + a_n X_{in}$$

where $a = a_1, a_2, \ldots a_n$ is a vector of regression coefficient for forecasting variables X_i, where $i = 1, 2, \ldots, n$. (Maddala & Lahiri, 2009) The purpose of the forthcoming analysis is to find the most reliable vectors and estimation values for every *a* of the study.

One major disadvantage of the logit model, as any parametric model, is that it is not well suited for exploring interactions between variables as Kolari et al. (1996) state. The reason behind the problem is the loss in degree of freedom, when the variable set increases. In order to decrease the dilemma, variables can be computed by multiplying two variables, which unfortunately tends to reduce the information. According to Aldrich and Nelson (1984) logit model also needs at least 50 observations per parameter in order to produce unbiased logit test statistics. Moreover, Stone and Raps (1991) present logit sample size requirements by creating a simulation study in order to assess the effect of the overall sample size, disparity of response group size, and the number, skewness, and distribution of predictor variables. They state that in a case of four to six predictors and skewed data, as with accounting data, sample size of 200 or more will be needed to guarantee that logit test statistics will be properly calibrated. In this regard, when sample sizes are smaller, logit test statistics is moderately miscalibrated. However, Stone and Rasp (1991) argue that most of the increase in classification errors was due to variable skewness rather than small sample sizes. This is a crucial point to bear in mind, when analyzing the results in section four, since the data used in the study will cover only 124 banks. Therefore, the test results logit model presents need to be diagnosed with great deliberation.

3.2 Data

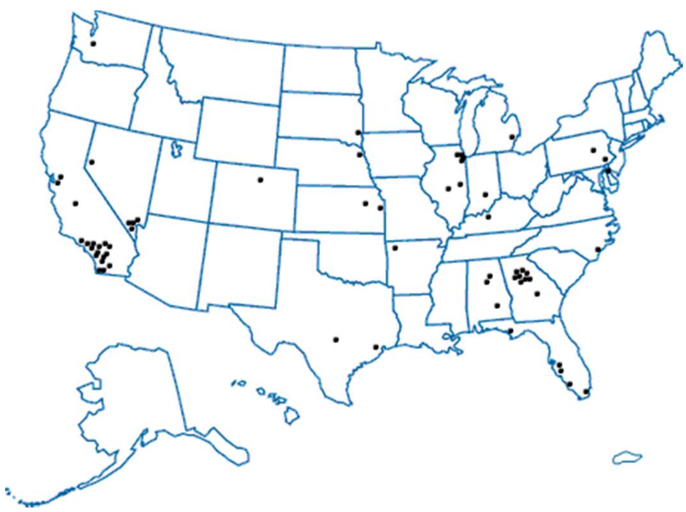
In this chapter the data of the study is carefully presented. The sample banks, independent variables, and the reasons for choosing exactly those financial institutions and parameters are provided. First, the procedure of the bank sampling is explained and thereafter explanatory variables are illustrated.

3.2.1 Bank Samples

In order to construct a reliable model for predicting commercial bank failures the sample banks need to be limited only for large financial institutions. Thus the sample consists of the commercial banks that failed in the U.S. during 2007 and 2009 and whose asset value was at the time of failure worth more than 500 million dollars. As smaller banks have slightly different earning logic and because their significance for the economy is somewhat smaller, they are not included to the sample.

Failed banks are defined as insured institutions for which disbursements by the Federal Deposit Insurance Corporation (FDIC) required following to their closure. By limiting the sample banks to include only FDIC guaranteed institutions, the data gathering became substantially easier as the data is compiled from the FDIC database. 62 failed FDIC guaranteed commercial banks with assets worth over 500 million dollars were then selected. As can be assumed, there are thousands of active commercial banks in the U.S. even after the crisis period. Regarding to FDIC database there are currently 7,941 active FDIC guaranteed commercial banks in the U.S. In order to avoid the biasness and miscalibration of the data the matched-pair sample need to be created. Matching is done by searching a pair for all 62 failed financial institutions. The active bank is considered to be a match with a failed one when its asset size is close to the bankruptcy bank at the quarter of failure. The matched-pair procedure could also be done more accurately by choosing active banks sharing not only the same value of assets than failed banks but also the same state. As the banking crisis is, however, nationwide, as Figure 5 below illustrates, that kind of selection would not probably have an impact on the forthcoming analysis in any substantial way. The total bank sample covers 124 financial institutions after the matched-pair procedure.

Figure 5. Bank Failure Locations



Source: FDIC database

Most of the EWSs measure whether the model can predict the failure one or two years prior the bankruptcy. In this study the data gathering will start from three years, in other words 12 quarters, before the failures and will end up to one quarter prior the bankruptcy. Years are divided into quarters in order to improve to predicting accuracy. It would be beneficial to rank the banks based on their value of assets and then divide them equally to both the original and holdout samples by using every other matched pair. The original sample would then be used to build the bank failure forecasting model and the holdout sample to study the predictive power of the model created. As the sum of failed over 500-million-dollar banks is limited to 62, that kind of division would decrease the degree of freedom too much as Aldrich and Nelson (1984) and Stone and Raps (1991) explain. Therefore, the data is limited to original sample, which is carefully analyzed is order to create a reliable failure predicting model.

Next the independent variables applied in the study are presented. The variables used in the study are liquidity, credit risk, profitability and taxes, growth, loan and deposit mix, securities, and instability as was mentioned already in section two.

3.2.2 Independent Variables

In order to create an accurate bank failure prediction model several fragility factors and independent variables need to be included to the analysis. Following studies have worked as a benchmark while choosing explanatory off-site variables: Beaver (1966), Altman (1968), Zmijewski (1984), Thompson (1991), Kolari et al. (1996), Jagtiani et al. (2003), Dabos and Escudero (2004), and Lanine et al. (2006). As was covered in theoretical part of the study the most commonly used financial ratios can forecast potential failures rather well. In addition, a couple of infrequently appeared ratios are added into the variable set, since the banks' decision of engaging to risky securities needs to be analyzed. That kind of selection concludes to 32 different financial ratios. The importance of every financial fragility elements was explained in section two. However, a detailed discussion concerning the chosen explanatory variables still needs to be presented. Definitions provided by FDIC (2008) are used in the explanation of independent variables.

> Liquid assets. Tradable securities practically equals to notes and coins as they can be sold at very short notice. Therefore, in case of banking panic, banks can sell

them quickly in order to respond to the increasing withdrawals. Moreover, when a bank ties its assets to liquid securities it improves the bank's cushion in the case of crisis. In this study, securities are defined as total investment securities excluding securities held in trading accounts. The used ratio is determined as total security holdings to total assets.

Uninsured deposits. FDIC guarantees time certificates of deposits and open-account time deposits with balances of 100,000 dollars. If value of loss exceeds 100,000 dollars, there is no insurance for that money and the losses need to be covered by the bank or the depositor. This is naturally risky for the depositors, and the higher the ratio of time deposits of 100,000 dollars or more to total time deposits, the lower the credibility and trust of the bank in the case of crisis.

Loan exposure and loan funding. Loan exposure is determined as the ratio of total loans and leases to total assets. Total loans and leases are defined as total loans and lease financing receivables, net of unearned income. As can be assumed the more the bank has loans and leases the more vulnerable it is for the default of its customers. Loan funding illustrates similar risk, but the dominator of the ratio is total deposits instead of total assets.

Nonaccrual rate. The ratio is determined as assets in nonaccrual status to total assets. Asset can be stated as nonaccrual, when its principal and interest is unpaid for at least 90 days and is no longer accruing interest. Obviously, imprudent lending decisions inflate the nonaccrual asset ratio and expose the bank for difficulties.

Past due loan rate. Assets at least 90 days deferred, but still accruing interest are considered to be past due loans. Therefore, past due assets can be seen to belong to less terminal loan class than nonaccrual assets. The ratio is determined as assets past due 90 or more days to total assets.

Loan loss allowance and provision rate. Loan loss allowance is a valuation account with a running balance of the allowances for loan losses established to report loans receivable at their net realizable value. The allowance for loan losses is reduced when a loan or a portion of a loan is written off as uncollectible. The allowance for loan losses is increased when a provision for loan losses is established. The provision for loan losses is the periodic expense for loan losses established in the current period. Loan loss allowance ratio is determined as loss allowance to total loans and leases. Provision rate, on the other hand, is defined as provision for loan and lease losses to total assets.

Loss rate. The banks are required to perform charge-offs when the actual loan losses occur. The loss rate equals to the net loan and lease financing receivable charge-offs to total loans and lease financing receivables. The loss rate demonstrates similar way than loan loss allowance and provision rate the level of credit risk the bank is exposed. As can be concluded, the high level of net charge-offs is an alarming sign of banking problems.

Capital ratio. The rate is determined as equity capital to total assets. Banks with substantial amounts of equity in their balance sheets are considered to be more solid ones as they possess better cushion for the unexpected predicament.

ROA and ROE. Return on assets (ROA) measures how profitable and, therefore, solvent the bank is. It is defined as net income after taxes and extraordinary items to total assets. Similar ratio, return on equity (ROE), is also providing information about the bank's profitability. Now, however, the denominator is total equity instead of total assets.

Dividend rate. It can be argued that when a company or a bank is profitable and in stable stage it tends to distribute generous dividends. Dividend rate is defined as cash dividend paid on common and preferred shares to total assets.

Net interest margin and net operating margin. Both of the variables continue to measure how profitable the bank is. Net interest margin is defined as total interest income less total interest expense to total assets. Net operating margin, on the other hand, is determined as net operating income to total assets.

Tax exposure. Solid and solvent banks tend to pay higher taxes than banks facing severe profitability problems. Applicable income taxes to total assets is a ratio that measures the tax exposure of a bank.

Capital and loan growth. Banks with growing capital levels are typical in good condition. The capital growth rate is determined as the change between the levels of total equity in previous and current periods to current level of total equity. Loan growth can, on the other hand, be considered to be a signal of increased level of risk of the bank. Loan growth ratio is also defined by comparing current and previous levels of total loans and leases to current level of total loans and leases.

Loan mix. Different kinds of loan types do not necessarily share the same risk profile. Loan mix ratios are testing the difference in level of risk between commercial, real estate, agricultural, and credit card loans. The risk ratios are created by dividing the different loan types to total loans and leases. It is also studied if the risk of failure can be reduced by diversifying the loans granted evenly between different loan types. Loan diversification is measured by sum of squared proportions of the four loan mix ratios.

Demand deposit mix and time deposit mix. Demand deposits mix is determined as demand deposits to total deposits. By definition demand deposits can be withdrawn at any time without a notice. Time deposits, however, include deposits that cannot be withdrawn for a certain period of time. In the case of financial crisis banks with more demand deposits instead of time deposits seem to have a better cushion against the failure. Time deposit mix is determined as time deposits to total deposits.

Securities. Banks can also invest in securities and actually in the subprime crisis it was stated that banks should have avoided risky instruments such as ABSs, MBSs, or CMOs. Instead risk free government debt securities should have been favored in order to limit the risk. The ratios are defined as MBSs to total securities, ABSs to total securities, CMOs to total assets, and government debt securities to total securities.

Instability. It can be argued that size of solvent banks' assets do not vary substantially between periods, in other words asset variation is minimal. The ratio measuring the variation is defined as current level of total assets to year-to-date average of total assets. Similar variation ratios is also determined to total loans and leases as well as equity capital.

As mentioned in theoretical part of the study, big financial institutions tend to be more solvent than smaller ones. In this thesis, however, the banks' asset size is applied in matched-pair procedure erasing the possibility of exploring the impact of financial institutions size on failures. Therefore, independent variables such as number of employees or total deposits are not included to the dataset.

As can be seen many of the variables, also provided in Table 2 next page, are considerably related to one another. This raises questions whether one can benefit the variables in the same model without facing the problem of multicollinearity. Multicollinearity refers to situation where two or more predictor variables are highly, but not perfectly, correlated. In this situation, the coefficient estimates may change erratically in response to small changes of the model or the data. Multicollinearity, however, does not reduce the predictive power or reliability of the model as a whole; it only affects calculations regarding individual predictors. (Pindyck & Rubinfeld, 1997) Therefore, it is crucial to minimize the problem of multicollinearity in order to receive reliable values for coefficients. This can be achieved by measuring correlation between variables and creating several alternative models that do not include explanatory variables strongly correlating with each other. The accurate explanation of the procedure is provided in the next section.

Table 1. Definitions of Explanatory Variables

Financial ratios	Definition
Liquidity	
Liquid assets	Total security holdings / Total assets
Uninsured deposits	Time deposits of $100,000 or more / Total time deposits
Credit Risk	
Loan exposure	Total loans and leases / Total assets
Loan funding	Total loans and leases / Total deposits
Nonaccrual rate	Assets in nonaccrual status / Total assets
Past due loan rate	Assets past due 90 or more days / Total assets
Loan loss allowance	Loss allowance / Total loans and leases
Provision rate	Provision for loan and lease losses / Total assets
Loss rate	Net charge-offs / Total loans and leases
Capital ratio	Equity capital / Total assets
Profitability and taxes	
Return on assets	Net income / Total assets
Return on equity	Net income / Total equity
Dividend rate	Cash dividends / Total assets
Net interest margin	Net interest income / Total assets
Net operating margin	Net operating income / Total assets
Tax exposure	Applicable income taxes / Total Assets
Growth	
Capital growth	(Total equity_t - Total equity_{t-1}) / Total equity_t
Loan growth	(Total loans and leases_t - Total loans and leases_{t-1}) / Total loans and leases_t
Loan and deposit mix	
Commercial loan risk	Commercial and industrial loans / Total loans and leases
Real estate loan risk	Total real estate loans / Total loans and leases
Agricultural loan risk	Total agricultural loans / Total loans and leases
Credit card loan risk	Credit card loans / Total loans and leases
Loan diversification	Sum of squared proportions of the four loan mix ratios for each bank
Demand deposit mix	Demand deposit / Total deposits
Time deposit mix	Time deposits / Total deposits
Securities	
MBS ratio	Mortgage-backed securities / Total securities
ABS ratio	Asset-backed securities / Total securities
CMO ratio	Collateralized mortgage obligations / Total securities
Risk free securities	Government debt securities / Total securities
Instability	
Assets variation	Total assets / Mean of total assets
Loans and leases variatio	Total loans and leases / Mean of total loans and leases
Equity variation	Total equity / Mean of equity capital

Soure: Beaver (1966), Altman (1968), Zmijewski (1984), Thompson (1991), Kolari et al. (1996), Jagtiani et al. (2003), Dabos and Escudero (2004), and Lanine et al. (2006)

4 EMPIRICAL ANALYSIS

In this section the analysis procedure and the results of the study are provided. As the purpose of the thesis is to construct reliable bank failure prediction model, the first step of the analysis is to reveal the explanation power of the independent variables. Then a correlation between the most reliable explanatory variables is studied leaving great variety of potential models with slightly different independent variables. The next step then is to test all the models in order to find the most accurate and reliable one. This procedure is followed by the discussion of credibility of the most statistically significant explanatory variables. Finally, the comparison between logit and probit models is provided in order to confirm the hypothesis that a logit model is more suitable for bank failure prediction than a probit model.

4.1 Analysis of the Independent Variables

The test of relevance of the independent variables is done in two ways. First, the mean between active and failed banks' financial ratios is studied for all 12 quarters. The validity of the variables is studied by using the Student's t-test at the 10 percent significance level. In order to gain strong explanation power, the variable's mean between failed and active banks need to be statistically significant at least in all three quarters before the failure. The variable also needs to be fairly consistent with time. In other words, explanation power should decrease as time from the bankruptcy lengthens. With that kind of limitation, 17 financial ratios stand out: liquid assets, nonaccrual rate, loan loss allowance, provision rate, loss rate, capital ratio, return on assets, dividend rate, net interest margin, net operating margin, tax exposure, loan growth, real estate loan risk, loan diversification, time deposit mix, CMO ratio, and loans and leases variation. If the difference in means is statistically significant at least in five quarters out of twelve the variables are considered to have a moderate explanation power. These variables are uninsured deposits, loan exposure, return on equity, capital growth, commercial loan risk, risk free securities, asset variation, and equity variation summing up to eight financial ratios. Remaining seven explanatory variables are considered to be irrelevant for the prediction model. The table illustrating the overall t-test results is provided in appendices.

The second way to test the fitness of the variables is to explore how well one variable at the time predicts the probability of a bank failure. This is done by using a logit model and the data closets to forthcoming failures. Results from the logit analysis and the first quarter's t-test are provided in Table 2.

Table 2. Test of Explanatory Variables

Label	Mean							
	Failed	Active	t-test	Wald chi2(1)	Prob > chi2	Pseudo R2	CC	ROC
Liquidity								
Liquid assets	0,1028	0,1603	0,0016	4,88	0,0272	0,0614	68,55 %	0,7098
Uninsured deposits	0,2273	0,1877	0,0497	3,46	0,0628	0,0229	61,29 %	0,6015
Credit Risk								
Loan exposure	0,7460	0,7074	0,1281	2,20	0,1380	0,0140	52,42 %	0,5825
Loan funding	0,9406	0,9367	0,9309	0,01	0,9289	0,0000	45,16 %	0,4342
Nonaccrual rate	0,1074	0,0199	0,0000	15,84	0,0001	0,4778	87,90 %	0,9121
Past due loan rate	0,0056	0,0031	0,2967	0,27	0,6044	0,0070	60,48 %	0,5905
Provision rate	0,0344	0,0031	0,0000	12,90	0,0003	0,3218	79,03 %	0,8650
Loss rate	5,0199	1,1880	0,0000	12,69	0,0004	0,2892	79,84 %	0,8585
Loan loss allowance	4,0315	1,7596	0,0000	11,81	0,0006	0,3602	84,68 %	0,8970
Capital ratio	3,2753	9,6546	0,0000	10,01	0,0016	0,4286	86,29 %	0,9305
Profitability and taxes								
Return on assets	-6,8633	0,1870	0,0000	19,28	0,0000	0,4717	86,29 %	0,9199
Return on equity	-132,4739	-0,7230	0,0025	0,69	0,4075	0,0803	80,65 %	0,8824
Dividend rate	0,0005	0,0029	0,0627	1,12	0,2910	0,0587	70,16 %	0,7089
Net interest margin	2,2561	3,6376	0,0000	18,32	0,0000	0,2456	76,61 %	0,8226
Net operating margin	-6,1698	0,3312	0,0000	20,89	0,0000	0,4742	85,48 %	0,9095
Tax exposure	-0,0012	0,0013	0,0043	4,65	0,0310	0,0534	70,16 %	0,6933
Size and growth								
Capital growth	13,2180	0,1211	0,3731	5,85	0,0156	0,0057	51,61 %	0,2053
Loan growth	-0,0449	-0,0009	0,0093	0,71	0,4000	0,0572	76,61 %	0,8226
Loan and deposit mix								
Commercial loan risk	0,1049	0,1362	0,1041	1,63	0,2010	0,0159	58,87 %	0,6446
Real estate loan risk	0,8345	0,7456	0,0143	3,26	0,0708	0,0375	60,48 %	0,6891
Agricultural loan risk	0,0116	0,0139	0,7967	0,05	0,8297	0,0004	54,84 %	0,6491
Credit card loan risk	0,0166	0,0325	0,4996	0,75	0,3850	0,0028	47,58 %	0,5018
Loan diversification	0,7661	0,6558	0,0014	8,30	0,0040	0,0606	60,84 %	0,6746
Demand deposit mix	0,0563	0,0689	0,3946	0,27	0,6027	0,0045	66,94 %	0,7019
Demand and time deposit mix	0,7155	0,5368	0,0000	9,61	0,0019	0,1607	75,81 %	0,8049
Securities								
MBS ratio	0,5648	0,5487	0,7646	0,09	0,7639	0,0005	49,19 %	0,5479
ABS ratio	0,0174	0,0345	0,4068	0,41	0,5218	0,0042	54,03 %	0,5265
CMO ratio	0,2573	0,1649	0,0543	4,53	0,0332	0,0221	55,65 %	0,5035
Risk free securities	0,6332	0,7057	0,1751	1,96	0,1616	0,0109	54,03 %	0,4992
Instability								
Assets variation	0,9543	1,0160	0,0000	9,92	0,0016	0,1642	75,81 %	0,7903
Loans and leases variation	0,9478	1,0025	0,0000	10,86	0,0010	0,2001	72,58 %	0,8156
Equity variation	0,4966	0,9931	0,0000	6,51	0,0107	0,2334	82,26 %	0,8660

Source: Author's calculations

The purpose of running the single variable logit analysis is to examine whether the same seven financial ratios still lack statistical significance. Since the logit model is used at the first time, the headers of Table 2 should be explained. Wald chi2, Prob > chi2, and Pseudo R2 are all from the statistical program Stata and are used for studying measurement of fit of logit and probit models. Wald chi2 is "the value of a likelihood-ratio chi-squared for the test of the null hypothesis that all of the coefficients associated with independent variables are simultaneously zero. The p-value is indicated by Prob > chi2. The number in parentheses is the number of coefficient being tested." (Long & Freese, 2003) In other words, smaller the Prob > chi2 figure, the more reliable the model, and in this case the explanatory variable, is. Pseudo R2, on the other hand, is a measure of fit also known as McFadden's R^2 and is defined as

2 $$R^2 = 1 - \frac{\ln L(M_{Full})}{\ln L(M_{Intercept})}$$

where M_{Full} is the model with predictors, $M_{Intercept}$ is the model without predictors, and L is the estimated likelihood. In other words, ln $L(M_{Full})$ presents the log-likelihood value for the fitted model and ln $L(M_{Intercept})$ is the log-likelihood value for the null model excluding all explanatory variables. Pseudo R2 falls between 0 and 1, the same way than regression R^2, with 0 indicating that the explanatory variables fail to increase likelihood and 1 indicating that the model perfectly predicts each observation. If one compares two models with different explanatory variables, pseudo R2 would be higher for the model with the greater estimated likelihood. It is notable that pseudo R2 can be kept as a reliable measure of fit between models with different data but same methodology. (Baum, 2006) In the case of comparing two different methodologies such as a logit model and a trait recognition analysis some other measurement of fit need to be applied. In other words, the superiority of the models cannot be compared with the help of pseudo R2.

As it is desirable to be able to compare the empirical results of the analysis between the results given by other studies, CC percentage and ROC figure should be used. CC is an abbreviation from the words correctly classified. It provides a reliable measure of fit between the models and is therefore used for the same purpose than pseudo R2. The logic behind the method is somewhat different, since CC calculates the percentage of correctly classified instances. As the logit model provides an equation for predicting binary outcomes there are four possible outcomes for that equation. If the instance, in this case a bank, is active and is classified as active, it is counted as a true positive; if it is classified as failed, it is counted as a false negative (Type II error). In addition, if the bank is failed and is classified as failed, it is counted as a true negative; if it is classified as active, it is counted as a false positive (Type I error). In other words the higher the CC percentage, the smaller the number on Type I and II errors and more accurate the model. Receiver operating characteristic (ROC) uses the same logic than CC and presents therefore similar results than CC. ROC is a technique for visualizing and selecting classifiers based on their performance. A ROC graph is a two-dimensional graph in which true positive rate, equivalent to sensitivity, is plotted on the Y axis and the false positive rate, equivalent to (1 – specificity), is plotted on the X axis. The best possible prediction model would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% sensitivity, in other words no false negatives, and 100% specificity, no false positives. (Fawcett, 2006) In this thesis ROC figure presents the area under the ROC line having 1.0 when the model perfectly predicts bank failures and approximately 0.5 when the model cannot predict the failure. It is crucial to note that ROC and CC can falsely give an impression that the model is predicting well. In a binary model, such as this one, it is possible by chance to predict correctly at least 50% of the cases even without knowing independent variables. Therefore, the interval of the model's predictability should be set between 0.7 and 1.0 for ROC, and 70% and 100% for CC.

When all the methods explained are used as a part of the analysis, it can be stated from Table 2 that loan funding, past due loan rate, agricultural loan risk, credit card loan risk, demand deposit mix, MBS ratio, and ABS ratio remain irrelevant for the final model. There are also other explanatory variables that seem to be insignificant for the study. However, in order to maintain all the potentially valuable information in the long-term, all the variables that are given a strong or a moderate status in the Student's t-test will remain in the analysis.

Before entering the multicollinearity analysis one needs to note that independent variables from all the fragility factors stated in section two can be considered to be statistically significant for the bankruptcy prediction. Both variables of liquidity, liquid assets and uninsured deposits, have either strong or moderate explanation power, which proves that it is crucial for banks to maintain liquidity assets in their accounts in order to avoid possible bankruptcy. Credit risk, on the other hand, seems to be one of the most important prediction factors although loan funding and past due loan rate variables are statistically insignificant for the bank failure prediction. Nonaccrual rate, provision rate, loss rate, loan loss allowance, and capital ratio have strong and loan exposure moderate explanation power, which proves that banks need to be extremely cautious in their lending processes. Careless lending decisions can be seen as inflated level of credit risk, which seems to expose banks to great risk of bankruptcy.

Profitability and taxes is the other element of bankruptcy prediction that seems to be considerably important to include to bank failure forecasting models as return on assets, dividend rate, net interest margin, net operating margin, tax exposure variables receives strong and return on equity moderate explanatory power status. Therefore, the hypothesis that banks are required to preserve the profitability cannot be rejected. Growth can also be considered to be an important factor behind the bank failure prediction. Loan growth has strong and capital growth moderate explanatory power both indicating that growing banks have enhanced likelihood to stay in the business.

Loan and deposit mix give rather interesting results. Whereas real estate risk and loan diversification variables have strong explanatory power, agricultural loan risk and credit card loan risk parameters stand out as insignificant variables for bank failure prediction. Commercial loan risk and risk free securities variables, on the other hand, have moderate explanation power. This indicates that the amount of loans granted to farmers and credit provided to consumers do not have an effect on banks' bankruptcy likelihood. Time deposits mix, however, can be as substantially significant independent variable, whereas demand deposit mix remains indifferent explanatory variable for the bank failure prediction. Therefore, one can argue that failed and active banks seem to have approximately same quantity of demand deposits in their accounts, whereas value of time deposits varies substantially.

Finally, it can be stated that security holdings and financial institutions potential instability have an impact on potential bank failures. Whereas CMO ratio and risk free securities support the hypothesis that banks should invest government debt securities in order to avoid failures, ABS and MBS ratios seem to be indifferent for bankruptcy prediction. Therefore, it can be argued that failed and active banks invest approximately same amount of money to ABS and MBS securities, while there is a difference on how financial institutions invest in risk free and CMO securities. Instability variables, asset, loans and leases, and equity variations, seem to improve bank failure prediction as they have either strong or moderate explanatory power. Therefore, hypothesis that failing banks will present some level of volatility prior the potential failure cannot be rejected.

As mentioned earlier, many of the statistically significant explanatory variables are closely related to each other. Therefore, potential multicollinearity needs to be detected and measured. The next chapter explores the procedure of minimizing the problem of multicollinearity while creating reliable bank failure detecting models.

4.2 Multicollinearity Minimizing and Plan Analysis

As already mentioned, multicollinearity may cause reliability problems for coefficients when studying the independent variables. In order to minimize the problem, correlation between potentially important financial ratios is explored. When the correlation between variables rises above the absolute value of 0.5 both of the variables cannot be included to the same model. Because the overall prediction power can be seen as important as the relevance of the independent variables, the aim is to find a model with high overall prediction power with financial ratios that comport tolerably well together.

By using the correlation analysis for all 25 independent variables, 72 different models can be created. Those 72 models are titled as plans in order to attain a difference between the final failure predicting model and sub-models that are used to derive the final model. Many differences between the plans are only between one or two variables. However, even minor changes can make a considerably difference between statistical significance of the plans. One of the most notable issue considering the correlation analysis is the fact that nonaccrual rate, net interest margin, net operating margin, and return on assets variables cannot be used together, since the smallest absolute value of correlation between mentioned variables is as high as 0.554. This is the base of the predicting models giving four different starting points, or plans, by using only one of the four financial ratios at a time. Provision rate is next under an examination showing strong correlation between all the other three variables stated above except the net interest margin variable. Therefore, provision rate cannot be applied simultaneously with nonaccrual rate, net operating margin, and return on assets variables. Also loss rate and loan loss allowance cannot be used together correlation being 0.583 between these two variables. The same way both of the variables are also strongly correlated with nonaccrual rate, net operating margin, return on assets, and provision rate. Again, net interest margin is the only variable having moderate correlation, around -0.300, with both of the loss rate and loan loss allowance ratios. This provides us six base plans as listed on next page.

Base plans:

1. Nonaccrual rate
2. Net operating margin
3. Return on assets
4. Net interest margin and provision rate
5. Net interest margin and loss rate
6. Net interest margin and loan loss allowance

In addition, capital ratio is strongly correlated with all the base plans and thus needs to be left out of the analysis. The next step doubles the number of plans as loans and leases variation and assets variation variables are alternately added to the base plans. These two independent variables are highly correlated between each other and cannot therefore be used simultaneously. Due to high level of correlation, the decision to use either loan diversification parameter or commercial loan risk and real estate loan risk variables need to be made. In order to avoid losing any potentially relevant information, there will be models including both of the possibilities, although not simultaneously. This naturally doubles the number of plans once more summing the total number of plans to 24.

Time deposit mix is strongly correlated with many variables already build into plans. The only base combinations with absolute correlation below 0.500 with the time deposit mix parameter are the plans with net operating margin and return on assets variables. As a result, time deposit mix ratio is added to all plans that have either return on assets or net operating margin as a base. Explanatory variables liquid assets and loan exposure are also strongly correlated between each other and cannot therefore be used simultaneously in the model. In order to add both of the variables to the plans, the doubling needs to be applied once again. Consequently, the number of plans has already risen to 48.

Return on equity is not correlated in any notable way with any of the plans illustrated so far. It cannot, however, be used simultaneously with equity variation, which in fact also strongly correlates with net operating margin and base plans of number four and six. In order to sustain all the information that could be valuable for the final bank failure predicting model, return on equity and equity variation variables are applied to the plans they seem to suite. After this procedure the amount of plans ascends to final number of 72. Rest of the explanatory variables can be added to all 72 plans at the same time. These variables are capital growth, tax exposure, CMO ratio, uninsured deposits, risk free securities, dividend rate, and loan growth. The table of the correlation between all 25 variables is somewhat extensive and can, therefore, be found from the appendices.

The next step is to run the logit analysis for all of 72 plans created. Data used in the analysis is from one quarter prior the forthcoming failures and includes therefore the most accurate figures to predict bankruptcies. The accuracy of the plan is tested with help of chi square distribution and pseudo R2 methods as well as CC and ROC figures. Since the purpose of the analysis is to maximize the number of correctly classified banks simultaneously minimizing the Type I and II errors, the main focus is on CC characteristic. Other statistical figures remain on background and are studied more carefully in the case where CC cannot provide separation between the plans. Since the table of results from the plan analysis is rather extensive, it can again be found form the appendices.

As Appendix 3 illustrates plans 52, 61, and 64 receive CC percentage of 95.16 %, which is slightly superior to other plans. Pseudo R2 for the plans in question is 0.6560, 0.6459, and 0.6488, and ROC values are 0.9688, 0.9659, and 0.9672, respectively. Mean for all 72 plans is 0.6529 for pseudo R2, 0.9641 for ROC, and 91.59 % for CC. The goodness of fit of the key plans is also provided in Table 3 next page.

Table 3. Explanation Power of the Key Plans

Variable	Plan 52	Plan 61	Plan 64
Nonaccrual rate	-77,9327**	-81,5075**	-81,2238**
Loans and leases variation		5,7942	
Assets variation	7,2871		6,2504
Loan diversification	1,7358	1,5947	1,2182
Return on equity	0,0017*	0,0017*	0,0018*
Liquid assets	4,5987		
Loan exposure		-0,7494	0,2065
Capital growth	0,0105	0,0111	0,0112
Tax exposure	19,2633	10,2456	22,3697
CMO ratio	-2,9660**	-2,5886**	-2,5508*
Uninsured deposits	5,1491	4,7258	4,4917
Risk free securities	4,6480***	4,7729**	4,9809***
Dividend rate	-18,7979	-34,2668	-32,7537
Loan growth	7,5084**	5,4798	6,2280*
Cons.	-8,4924	-5,6533	-6,7260
Wald chi2(12)	35,70	35,06	36,46
Prob > chi2	0,0004	0,0005	0,0003
Pseudo R2	0,6560	0,6459	0,6488
CC	95,16 %	95,16 %	95,16 %
ROC	0,9688	0,9659	0,9672

Dependent variable is defined as 0 in the case of failure and 1 otherwise.

Index stars illustrate the significance level of the variables. One star presents 10% level of significance, whereas two and three stars state 5% and 1% significance levels, respectively.

Source: Author's calculations

It is crucial to note that all three models share the same base plan, which includes nonaccrual rate variable. The similarity between plans continues as loan diversification, return on equity, capital growth, tax exposure, CMO ratio, uninsured deposits, risk free securities, dividend rate, and loan growth variables are present in all of the key plans. However, plan 61 includes loans and leases variation variable, whereas plans 52 and 64 contain assets variation ratio. Another difference between the plans can be seen between liquid assets and loan exposure variables. Plan 52 applies liquid assets as one of the explanation variables whereas plans 61 and 64 hold the loan exposure variable.

To verify that the problem of multicollinearity is absent the variance inflation factor (VIF) is studied for all of the three plans mentioned. VIF is defined as

3 $$VIF(\beta_i) = \frac{1}{1-R_i^2}$$

where R_i^2 is the squared multiple correlation coefficient between x_i and the other explanatory variables. VIF can be interpreted as "the ratio of the actual variance of β_i to what the variance of β_i would have been had x_i been uncorrelated with the remaining xs. In ideal situation is considered to be one where the xs are all uncorrelated with each other and the VIF compares the actual situation with an ideal situation." (Maddala & Lahiri, 2009) If VIF, however, reaches the level of 5.0 problem of multicollinearity can be considered to be high. As one can realize R^2 is not provided in the logit analysis and therefore logit model needs to be transferred to a linear regression model in order to get VIF figures. The procedure does not have an impact on predictability of the model since the final results are studied from the logit model analysis, not from the linear one. Linear regression is simply provided in order to study how severe the multicollinearity problem is, since the correlation between independent variables does not illustrate the level of multicollinearity, which is, after all, a property of the predictors not of the model. Table 4 below presents the results from the VIF analysis and confirms that there is no explanatory variable with problem of collinearity. Therefore, all of the plans can be considered to be suitable for bank failure predicting.

Table 4. VIF Figures

Variable	Plan 52	Plan 61	Plan 64
Nonaccrual rate	1,75	1,70	1,73
Loans and leases variation		1,75	
Assets variation	1,31		1,41
Loan diversification	1,38	1,50	1,51
Return on equity	1,22	1,22	1,23
Liquid assets	1,25		
Loan exposure		1,44	1,55
Capital growth	1,09	1,08	1,09
Tax exposure	1,18	1,15	1,17
CMO ratio	1,15	1,16	1,14
Uninsured deposits	1,32	1,33	1,33
Risk free securities	1,09	1,09	1,09
Dividend rate	1,10	1,15	1,10
Loan growth	1,13	1,50	1,22
Mean VIF	1,25	1,34	1,30

Source. Author's calculations

As can be seen from Table 3 and 4, plan 52 seems to be slightly superior compared to other plans. Although CC, 95.16%, is same for all key plans, pseudo R2 and ROC characteristics as well as mean VIF of the plans are faintly better for plan 52 compared to plans 61 and 64. However, long-term analysis still needs to be provided until plan 52 can be considered to be the most accurate predicting model.

4.3 Long-term Analysis

Although all three plans seem to work well in short-term prediction, it is unclear how well they work in the long-term. Therefore, the next step is to run logit analysis for all of the 12 quarters prior the bankruptcies. The main focus is on models' overall predictability not in statistical significance of the independent variables. Table 5 below illustrates the results.

Table 5. The Key Plans in the Long Run

Plan 52	Wald chi2	Prob > chi2	Pseudo R2	CC	ROC
Q1	35,70	0,0004	0,6560	95,16 %	0,9688
Q2	43,57	0,0000	0,5153	87,10 %	0,9272
Q3	27,27	0,0071	0,5707	88,71 %	0,9417
Q4	31,64	0,0016	0,4025	82,26 %	0,8866
Q5	38,76	0,0001	0,3807	81,45 %	0,8790
Q6	28,36	0,0049	0,2736	75,00 %	0,8223
Q7	28,81	0,0042	0,2752	78,23 %	0,8252
Q8	21,60	0,0423	0,2471	72,58 %	0,8236
Q9	26,39	0,0094	0,2243	74,19 %	0,8054
Q10	29,30	0,0036	0,3171	79,84 %	0,8541
Q11	32,43	0,0012	0,2845	75,00 %	0,8254
Q12	27,93	0,0057	0,2432	71,77 %	0,8171

Plan 61	Wald chi2	Prob > chi2	Pseudo R2	CC	ROC
Q1	35,06	0,0005	0,6459	95,16 %	0,9659
Q2	43,84	0,0000	0,5045	87,10 %	0,9222
Q3	36,53	0,0003	0,5436	83,87 %	0,9308
Q4	32,06	0,0014	0,4063	83,06 %	0,8837
Q5	37,94	0,0002	0,3883	82,26 %	0,8803
Q6	29,22	0,0037	0,2811	75,00 %	0,8260
Q7	29,30	0,0035	0,2768	78,23 %	0,8275
Q8	21,55	0,0428	0,2485	73,39 %	0,8252
Q9	24,40	0,0180	0,2236	73,39 %	0,8072
Q10	31,69	0,0015	0,3153	79,84 %	0,8520
Q11	34,04	0,0007	0,2776	76,61 %	0,8249
Q12	27,74	0,0060	0,2426	70,97 %	0,8156

Plan 64	Wald chi2	Prob > chi2	Pseudo R2	CC	ROC
Q1	36,46	0,0003	0,6488	95,16 %	0,9672
Q2	44,45	0,0000	0,4986	87,10 %	0,9170
Q3	27,27	0,0071	0,5707	88,71 %	0,9417
Q4	33,64	0,0008	0,4066	83,06 %	0,8837
Q5	35,97	0,0003	0,3715	81,45 %	0,8702
Q6	29,88	0,0026	0,2810	75,00 %	0,8286
Q7	35,22	0,0004	0,2960	75,81 %	0,8392
Q8	17,45	0,1334	0,2450	75,81 %	0,8280
Q9	26,11	0,0104	0,2111	70,16 %	0,7893
Q10	33,69	0,0008	0,2936	74,19 %	0,8358
Q11	35,30	0,0004	0,2761	76,61 %	0,8182
Q12	26,46	0,0092	0,2421	70,97 %	0,8202

Mean					
Plan 52	30,98	0,0067	0,3659	80,11 %	0,8647
Plan 61	31,83	0,0138	0,3618	79,50 %	0,8616
Plan 64	31,95	0,0066	0,3628	79,91 %	0,8634

Source: Author's calculations.

As could be foreseen, all the plans act similar way and provide results that are significantly similar. However, if the mean of the characteristics from all 12 quarters are examined, it can be noted that all measurement of fit indexes pseudo R2, CC percentage and ROC figure are again slightly superior for plan 52 compared to plans 61 and 64. At this point it seems that plan 52 is the best option for the final bank failure predicting model.

The next step is to take a careful look at the explanatory variables of the plan. As was mentioned plan 52 encloses 12 independent variables: nonaccrual rate, assets variation, loan diversification, return on equity, liquid assets, capital growth, tax exposure, CMO ratio, uninsured deposits, risk free securities, dividend rate, and loan growth. Table 3 illustrates that many of the variables are not statistically significant in the short-term. It is, however, crucial to examine how the situation changes when the time line from the bankruptcies is prolonged. The focus of the analysis is on both the statistical significance of the independent variables and the signs and values of the discriminant coefficients. Table 6 next page summaries the analysis.

As could be expected nonaccrual rate stands out as a considerably reliable and stable explanatory variable. It seems, however, that all the other explanatory variables than CMO ratio and risk free securities are exceptionally unstable as the sign and value of the coefficients changes dramatically from period to period. This naturally diminishes the credibility of the variable although it might be statistically significant at many quarters. Therefore, more precise explanation is provided only for those three explanatory variables.

Table 6. Plan 52's Independent Variables in the Long Run

Variable	Q1	Q2	Q3	Q4	Q5	Q6
Nonaccrual rate	-77,93269**	-65,76568***	-47,91318***	-60,29815***	-94,35348***	-122,8807***
Assets variation	7,287129	-7,011643	-3,198023	0,918106	-10,31932**	1,087165
Loan diversification	1,735756	0,7094394	-0,3438951	0,4960994	0,456459	-0,154095
Return on equity	0,0017253*	-0,0000815*	0,044645***	0,030071	0,0049657	-0,0192315
Liquid assets	4,598661	4,822927	12,53156***	4,322823	1,354269	1,293459
Capital growth	0,0105052	0,0446541	8,437697**	3,064864	5,069754***	1,160437
Tax exposure	19,2633	89,47886	-78,29423	-9,912962	-39,99143	-85,38045
CMO ratio	-2,966044**	-1,41391	-1,085159	-0,5226233	-0,7467394	-0,9734643
Uninsured deposits	5,149105	4,882139	3,988529	-2,354424	-2,678055	-2,084122
Risk free securities	4,648002***	2,851474***	2,409734*	1,855624	2,043018	2,447537***
Dividend rate	-18,79789	41,63586	52,15787	-11,71733	-2,155777	-0,1022386
Loan growth	7,508367**	3,12273	9,649704	-2,913252	9,125941*	-3,845614
Cons.	-8,492397	5,944122	1,569046	-0,9927121	10,91415**	-0,4564288
Wald chi2(12)	35,70	43,57	27,27	31,64	38,76	28,36
Prob > chi2	0,0004	0,0000	0,0071	0,0016	0,0001	0,0049
Pseudo R2	0,6560	0,5153	0,5707	0,4025	0,3807	0,2736
CC	95,16 %	87,10 %	88,71 %	82,26 %	81,45 %	75,00 %
ROC	0,9688	0,9272	0,9417	0,8866	0,8790	0,8223

Variable	Q7	Q8	Q9	Q10	Q11	Q12
Nonaccrual rate	-95,72772**	-80,48797	-86,86967*	-178,4092***	-152,0902***	-112,4604**
Assets variation	-3,211138	7,620794*	-11,87174*	-18,06264*	-3,137022	-1,340421
Loan diversification	-2,145035	-3,593077**	-2,590752**	-3,00306**	-3,542729***	-2,467921*
Return on equity	-0,0355807	-0,0330717	-0,0704591**	0,0121248	-0,0961256**	-0,0648677*
Liquid assets	1,801145	0,2429351	-0,9707748	-2,245419	-3,187236	-0,8272695
Capital growth	8,980748*	-3,53215	0,7809656	-27,47916**	-7,118194	1,105367
Tax exposure	105,9084	79,15493	164,9866*	228,9971**	124,5383	28,82668
CMO ratio	-0,6870669	-1,615901	-1,471475	-1,596867	-1,50752	-2,018302**
Uninsured deposits	-1,587984	-1,36386	-2,803611	-4,6968**	-5,138157***	-4,800971**
Risk free securities	2,064974**	1,033945	2,073221*	1,764351	2,263554*	1,822466
Dividend rate	40,22499	23,2639	93,13827	-46,59644	138,8638*	128,9773**
Loan growth	-11,92131	-17,59788***	11,81632	-2,048338	-0,6539087	-1,171042
Cons.	4,604397	-4,435874	13,65452*	21,85833**	7,479117	4,273029
Wald chi2(12)	28,81	21,60	26,39	29,30	32,43	27,93
Prob > chi2	0,0042	0,0423	0,0094	0,0036	0,0012	0,0057
Pseudo R2	0,2752	0,2471	0,2243	0,3171	0,2845	0,2432
CC	78,23 %	72,58 %	74,19 %	79,84 %	75,00 %	71,77 %
ROC	0,8252	0,8236	0,8054	0,8541	0,8254	0,8171

Dependent variable is defined as 0 in the case of failure and 1 otherwise.

Index stars illustrate the significance level of the variables. One star presents 10% level of significance, whereas two and three stars state 5% and 1% significance levels, respectively.

Source: Author's calculations

As was stated nonaccrual rate can be seen as one of the measurements of credit risk. Asset receives the nonaccrual status when its principal and interest is unpaid for at least 90 days and is no longer accruing interest. Different kinds of loans such as real estate, installment, and commercial loans can be considered to include banks' assets. As it was noticed already in the independent variable analysis, nonaccrual rate is one of the best explanatory variables for the bank failure prediction. Also authors Kolari, Caputo, Wagner, Jagtiani, Lemieux, Shin, Lanine, and Vennet have successfully included nonaccrual rate variable in their bank failure prediction models in the 1990s and 2000s. Implicitly, it is logical that financial institutions face problems when their assets no longer accrue interest, since banks make their profit mainly by lending and borrowing money. Therefore, it is evident that the more nonaccrual assets there are in the banks' accounts, the greater the probability that the bank will fall into bankruptcy.

The explanatory variable of risk free securities is a ratio of U.S. government debt securities and total securities. By definition U.S. Government securities includes U.S. Treasury securities as well as U.S. Government agency and corporation obligations. It contains also U.S. government issued or guaranteed mortgage-backed securities. (FDIC, 2008) As it seems, the more the bank has tied its securities to risk free investments, the smaller the probability of bankruptcy. On the other hand, the risky CMO securities tend to increase the probability of the failure. Collateralized mortgage obligations are defined as "mortgage-backed securities held-to-maturity at amortized cost and available-for-sale at fair value which are either issued or guaranteed through FNMA (Fannie Mae) or FHLMC (Freddie Mac), or privately issued and collateralized by mortgage-backed securities issued or guaranteed by FNMA, FHLMC, or GNMA (Ginnie Mae) and all other privately-issued." (FDIC, 2008) Consequently, it seems that a bank, which keeps its investments in risk free securities instead of risky CMOs has better prospects to avoid the failure.

4.4 Simple Model Analysis

Before the final model can be verified, one more step needs to be taken. It seems that CMO ratio has dramatically low statistical significance in the long-term analysis and therefore only two independent variables tend to be essential for the bankruptcy prediction. Consequently, the model with only nonaccrual rate and risk free rate variables is created and tested. In order to analyze the reliability of the model all 12 quarters are been studied. The results of the analysis are provided in Table 7 next page.

Table 7. Simple Model's Independent Variables in the Long Run

Variable	Q1	Q2	Q3	Q4	Q5	Q6
Nonaccrual rate	-69,65873***	-67,9112***	-67,38556***	-75,451***	-97,68132***	-108,0926***
Risk free securities	4,33758***	3,328931***	3,057069***	2,298365***	2,197093***	2,263842***
Cons	0,4560483	0,3635777	-0,0849264	0,1112892	0,0834568	-0,3266339
Wald chi2(2)	12,13	23,63	37,80	26,56	24,36	19,34
Prob > chi2	0,0023	0,0000	0,0000	0,0000	0,0000	0,0001
Pseudo R2	0,5776	0,4701	0,4054	0,3239	0,3022	0,2382
CC	91,94 %	83,87 %	82,26 %	82,26 %	79,84 %	76,61 %
ROC	0,9534	0,9110	0,8770	0,8379	0,8348	0,8044

Variable	Q7	Q8	Q9	Q10	Q11	Q12
Nonaccrual rate	-120,3101***	-112,0131**	-109,7247***	-127,3356***	-100,1563***	-97,34172***
Risk free securities	2,263932***	1,670808**	1,59823**	1,410856*	1,60204**	1,804193**
Cons	-0,4912399	-0,3901536	-0,5726825	-0,44387	-0,7655963	-0,9387143*
Wald chi2(2)	14,72	8,60	9,12	11,69	9,70	11,07
Prob > chi2	0,0006	0,0136	0,0104	0,0029	0,0078	0,0039
Pseudo R2	0,1916	0,1188	0,0807	0,0787	0,0564	0,0624
CC	72,58 %	68,55 %	66,13 %	62,90 %	62,10 %	60,48 %
ROC	0,7942	0,7518	0,6857	0,6550	0,6322	0,6397

Dependent variable is defined as 0 in the case of failure and 1 otherwise.

Index stars illustrate the significance level of the variables. One star presents 10% level of significance, whereas two and three stars state 5% and 1% significance levels, respectively.

Source: Author's calculations

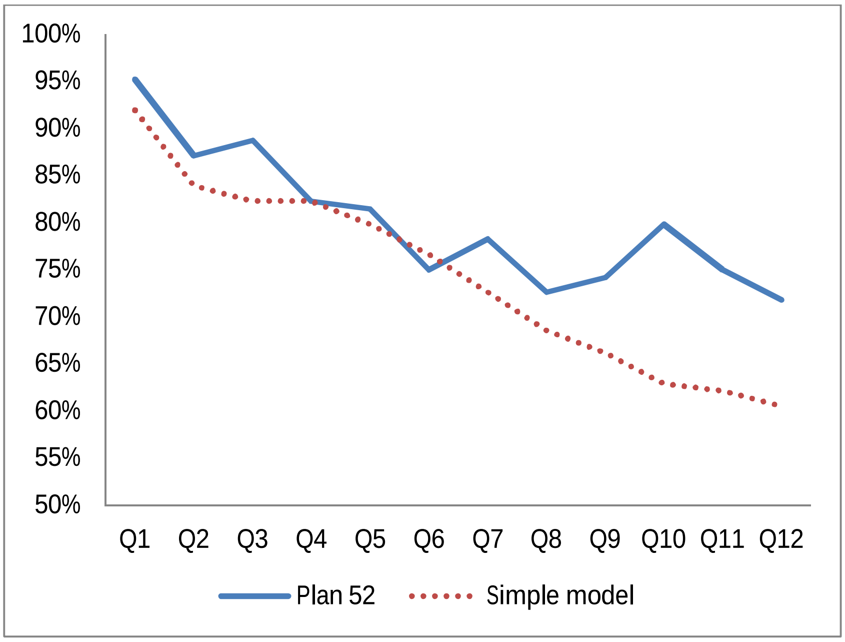
The logit analysis indicates that the simple model seems to lose some of the predictability compared to plan 52. The simple model's CC percentage is above required 70% in only seven quarters out of twelve, whereas plan 52 manages to break 70% cut-off point at every 12 quarters. However, in the simple model all explanatory variables are statistically significant through the three-year period contrary to plan 52. This can be seen as a valid strength of the model.

As can be seen from Table 7, even slight changes in nonaccrual rate have a major impact on possibility of failure of the bank. For example, one quarter prior the bankruptcy the mean of the nonaccrual rate is 6.36%, stating that about 6% of banks' assets have nonaccrual status. If the rate rises only by one percentage point, the chance to fail doubles. On the other hand, if the nonaccrual rate decreases from six to five percent, the bankruptcy probability goes down by 50%. The importance of nonaccrual rate variable seems to increase with time, since five quarters before the potential failure one percentage point change in nonaccrual rate increases the probability of bank failure by 166%. On the contrary, 100 basis point (bp) decline in nonaccrual rate lowers the bankruptcy probability over 60%. Seven quarters prior the possible failure 100 bp increase in nonaccrual rate inflates the bankruptcy probability by 233%, whereas the probability of bank failure reduces by 70% if nonaccrual rate decreases by one percentage point. If one goes further from bankruptcy, the value of CC percentage goes under 70% and the predictability of the model comes rather undependable. Therefore, one can argue that only slight changes in nonaccrual rate have a tremendous impact on a bank's probability to fail. However, it is crucial to note that 100 bp change in nonaccrual rate reflects approximately 15% change from its average. Therefore, it might also be appropriate to test 10bp change in nonaccrual rate. For the first quarter 10 bp increase in nonaccrual rate would enhance the bankruptcy probability by 7% and 10bp decrease would reduce the probability by the same 7%. Naturally, the effect will become greater as time from the failure elapses.

The other explanatory variable that needs to be studied is the risk free securities. One quarter before the potential bankruptcy one percentage point increase in risk free securities decreases the bank failure probability by 4%. The impact is same size, but to different direction when a bank ties one percentage point less money for government related securities and places the money into some riskier securities instead. As the time elapses, the significance of the variable diminishes. From four to seven quarters prior the potential bankruptcy 100 bp change in risk free securities have approximately only two percent impact of bankruptcy probability. Therefore, one can argue that risk free securities variable does not have as dramatic impact on bankruptcy prediction as nonaccrual rate seems to have. In addition, the impact seems to abate in the long-run. However, 100 bp change in risk free securities corresponds to 10 bp change in nonaccrual rate as the mean of risk free securities is roughly ten times greater than the mean of nonaccrual rate. Bearing this in mind, the difference of effect between these two variables is not as striking as it looks at the first glance.

It is also crucial to note that CC rates as well as all the other measurement of fit characteristics seem to be rather inconsistent with time. For example, quarter 6 prior the failure receives CC percentage of 75, but at quarter 7 the value has improved to 78%. The same pattern appears between quarters 8, 9, and 10. The trend of CC percentage should point down as time from the failure lengthens and the model's predictability decreases. This is not, however, the case. Conversely, CC rates are 72.58%, 74.19%, and 79.84%, respectively. This naturally deteriorates the credibility of plan 52. On the other hand, simple model works perfectly when the consistency of bank failure prediction is used as a criterion. From period to period the CC percentage lowers or remains the same, but does not increase. Therefore, it can be argued that independent variables, included to the plan 52 but not to the simple model, seem to behave somewhat irrationally along with time. This can be explained by the change in the economic state during 2008 as dramatic escalation of macroeconomic conditions occurred within a year. However, the change in economic conditions does not seem to have an impact on predictability of nonaccrual rate and risk free securities. As it appears, difficult decision between plan 52 and the simple model needs to be made. Whereas plan 52 provides substantial predictability values, the simple plan presents high level of model's stability and statistically significant explanatory variables. Figure 6 below illustrates how CC percentage performs in the long-term for both of the models.

Figure 6. Performance of CC in the Long Run



Source: Author's calculations

4.5 Comparison between Logit and Probit Models

The final step is to provide an analysis between logit and probit models. In order to make a decision of which one of the methods suites better for the bank failure prediction, probit model analysis needs to be executed for both the plan 52 and the simple model. The reliability of the method for all 12 quarters is studied with the help of Akaike information criterion (AIC), which is defined as

4 $$AIC = 2k - 2\ln(L)$$

where k is the number of parameters in the statistical model and L is the maximized value of the likelihood function for the estimated model. AIC can be viewed as measures that combine fit and complexity. Given two models fit on the same data, the model with the smaller value of the information criterion is considered to be better. (Maddala & Lahiri, 2009)

As can be seen from Table 8 the logit model seems to outperform the probit model for the first year prior the failure. Thereafter the models go basically hand in hand. The number of independent variables does not seem to affect the results of the analysis. Therefore, one can argue that logit model suites slightly better for the bank failure prediction, although the difference is subtle between the models.

Table 8. AIC Analysis

Plan 52	Logit	Probit
Q1	85,1342	90,8782
Q2	109,3232	110,6194
Q3	99,8001	100,824
Q4	128,7131	129,6915
Q5	132,4569	132,4451
Q6	150,8769	150,6753
Q7	150,5887	150,6873
Q8	155,4190	156,5059
Q9	159,3369	159,6112
Q10	143,3932	143,1964
Q11	148,9939	148,4407
Q12	156,1005	156,6046

Simple model	Logit	Probit
Q1	78,6045	84,4408
Q2	97,0892	98,7780
Q3	108,2096	108,4689
Q4	122,2223	122,8455
Q5	125,9565	126,3666
Q6	136,9538	137,2490
Q7	144,9650	145,8584
Q8	157,4860	158,3402
Q9	164,0198	164,1358
Q10	164,3668	164,2748
Q11	168,1991	168,1002
Q12	167,1820	167,0453

Source: Author's calculations

4.6 Conclusion of the Empirical Analysis

In the beginning of this section all the independent variables were compared. The analysis was conducted by two ways: first the Student's t-test was applied in order to study differences between the means of failed and active banks. The data of 12 quarters was used resulting strong or moderate explanation power for 25 independent variables. The test was then verified by the logit analysis and the data closets to forthcoming failures. Independent variables were used in the logistic regression one at a time.

The next step was to minimize the problem of multicollinearity by studying the correlation between all statistically significant explanatory variables. With help of the correlation table 72 different plans were created in order to include all the potential information and without having severe correlations between the variables. After running the logit analysis for all the plans, three of them were distinguished the criterion being the value of CC percentage. The long-term analysis was then executed for plans 52, 61, and 64 confirming that plan 52 is the most accurate plan to predict the bank failures. All the plans offered over 70% prediction accuracy for all 12 quarters and had CC of 95.16% one quarter prior the bankruptcy. The closer examination revealed, however, that majority of the independent variables of plan 52 was rather unstable and inconsistent with time. For this reason one more analysis was executed. The last model, namely the simple model, showed worsened explanation power as a whole, but offered a tool for closer analysis of two most stable and statistically significant independent variables; nonaccrual rate and risk free securities. In addition, simple model seems to be very consistent with time as prediction accuracy weakened when the time from the bankruptcy elapsed.

Due to statistically strong explanatory power that nonaccrual rate and risk free securities provided, the sensitivity analysis could be executed. It proved that nonaccrual rate is substantially sensitive for changes and therefore needs to be retained under careful examination. As the value of assets with nonaccrual status increases, the bank's probability to fail rises simultaneously. Risk free securities variable, however, does not seem to be as sensitive to changes as nonaccrual rate, although the bankruptcy probability increases as banks invest more heavily on risky securities instead of government bills, notes, and bonds. It is notable that the level of investments tied to risk free securities can be easily changed whereas nonaccrual rate is the consequence of the negligent investment decisions. Therefore, banks should carefully analyze the size and type of the loan and the quality of the potential borrower before making the lending decision.

In the final part of the empirical analysis comparison between logit and probit models was provided. Logit models seem to be slightly superior compared to probit models, but the difference is faint. However, as Aldrich and Nelson (1984) state, there exists a requirement of 50 observations per parameter in order to produce unbiased logit test statistics. Moreover, Stone and Raps (1991) state that in a case of four to six predictors and skewed data, as accounting data is, sample size of 200 or more will be needed to guarantee that logit test statistics will be properly calibrated. Consequently, one can ask if the logit model of plan 52 with 12 explanatory variables and only 124 observations responds to these concerns. In this regard, when the sample sizes are small, logit test statistics can be moderately miscalibrated. Therefore, it can be argued that the simple model with only two explanatory variables can be seen as a more suitable model for the bank failure prediction when the sample size is as small as 124 observations. Relative low degree of freedom can also explain the inconsistent CC percentages received from plan 52 as the time line form the bankruptcies elapses.

5 CONCLUDING REMARKS

Since 2008, bank failures have caused general anxiety in the U.S. as several commercial banks have fallen into bankruptcy between 2007 and 2009. Although there is substantial amount of research around bankruptcies, a study including data from the current wave of U.S. bank failures was missing. Therefore, the purpose of this thesis was to study how accurately recent U.S. commercial bank failures can be predicted. The time frame chosen to the study started from one quarter and ended to 12 quarters prior the failure, while the logistic regression was used as a methodology of the analysis. The thesis balanced between predictability of the model as a whole and the statistical significance of the independent variables. Also the comparison between logit and probit models was provided in order to confirm the hypothesis that the logit model is more suitable for bank failure prediction than the probit model.

At the beginning of the study, bankruptcy literature and theory was specifically illustrated. Liquidity, credit risk, profitability and taxes, size and growth, loan and deposit mix, securities, and instability can be seen as causing the bank failures creating a basis for the independent variables applied in the empirical analysis of the thesis. Historical overview of the banking crises was also covered in order to get clearer picture of the key issues behind the bank failures. The main finding of the section revealed that the traditional prediction tools for forecasting bank failures, with classic financial variables such as return on assets, seem to offer an excellent base for bankruptcy prediction analysis. However, the most recent crisis being referred as a subprime crisis, also proxies concerning banks' mortgage related instruments and overall security exposures needed to be included to the analysis.

After the historical glance of the banking crises, the evolution of bankruptcy prediction models was explored. As illustrated, great variety of bankruptcy prediction models and methodologies can be found since the year 1966 when Beaver published his groundbreaking study. One of the most well-known, reliable, and accurate method seems to be a logistic regression with the financial statement variables. It is essential to note that there are also several other proper bankruptcy prediction models than the logit model. As the results suggest, majority of those models and methods would, however, require great variety of specialized software and hardware. This naturally sets limitations for the appliance of the models. For example, trait recognition analysis, neural networks, and genetic programming belong to the branch of machine learning techniques.

The definition and more precise explanation of the thesis' methodology were provided in section three as well as the data gathering procedure. The data covered all 62 failed banks from 2007 to 2009 with total assets worth more than 500 million dollars at the time of failure resulting sample size of 124, when also the match-pairs were included.

The research question being how accurately recent bank failures can be predicted multiple steps had to be performed. The first step was to study the explanation power of all of the 32 independent variables so that the main elements behind the bank failures could be explored. After performing the Student's t-test and single variable logit analysis, the number of statistically significant independent variables diminished to 25. Thus, loan funding, past due loan rate, agricultural loan risk, credit card loan risk, demand deposit mix, MBS ratio, and ABS ratio variables were considered to be insignificant for the bank failure prediction.

Due to the problem of multicollinearity, 72 different models were constructed in order to sustain all relevant independent variables. The short-term logit analysis revealed that three of the models were strongest in their explanatory power performing 95.16 % prediction accuracy of the bank failures. After examining the pseudo R2 and ROC statistics, one of the models, named as plan 52, came to stand out when compared to the two other models. The independent variables of plan 52 are nonaccrual rate, loan diversification, return on equity, capital growth, tax exposure, CMO ratio, uninsured deposits, risk free securities, dividend rate, loan growth, assets variation, and liquid assets covering all financial fragility factors. Therefore, it can be argued that the findings of the previous researches can be applied also to the 21st century bank failure studies. On the other hand, it is interesting to notice that subprime related CMO ratio actually improves the predictability and accuracy of the model. As a result it can be stated that banks with substantial amounts of assets tied to risky CMO securities are more likely to fail.

In order to strengthen the hypothesis that the model in question is actually the most suitable model for bank failure prediction, long-term analysis was provided. As a result it confirms that plan 52 really is slightly superior to the other models. Maybe the most significant result of the long-term analysis is its finding that although the time frame of the analysis is lengthened to 12 quarters, the percentage of correctly classified (CC) banks remains at the level of 70. In addition, nearly 90 % prediction accuracy is stated although the analysis is performed three quarters prior the bankruptcies. As a result, it can be argued that the logit model estimation of forthcoming bankruptcies is rather reliable even as early as three years prior the failures.

Although the overall predictability of the model is rather substantial, the closer examination of the independent coefficients reveals that the majority of the explanatory variables are quite unstable and inconsistent with time. This can be explained by the changes in the economic state during the year 2008 as dramatic escalation of macroeconomic conditions indeed occurred during that year. However, the changes in economic conditions do not seem to have an impact on predictability of nonaccrual rate and risk free securities. Therefore, one more model with only these two independent variables was constructed.

After testing the model, it became clear that the overall explanatory power of the model was not as strong as plan 52. The simple model, however, seems to be very consistent with time as the overall prediction accuracy reduces when the time from the bankruptcy lengthens. The situation is rather opposite with plan 52 as its prediction accuracy varies inconsistently with time. For example, six quarters prior the failure CC percentage is 75, but at quarter seven the value has improved to 78 percent. The same pattern appears between quarters 8, 9, and 10, when the trend of CC percentage points upwards as the time from the failure lengthens. This naturally deteriorates the credibility of plan 52. On the other hand, the simple model's CC percentage decreases or remains the same, as the time from the bankruptcies elapses.

Nevertheless, the simple plan did provide valuable tools for closer analysis of the causation between the independent variables and the bank failures. It was confirmed that the nonaccrual rate is more sensitive for changes and therefore needs to be kept under careful examination. Risk free securities ratio, however, does not seem to be as responsive to changes as nonaccrual rate. On the other hand, it is crucial to understand that the ratio of government debt securities to total securities can be easily increased in order to reduce bankruptcy probability, whereas inflated nonaccrual rate should be seen as a long-term consequence of the careless lending decisions. Therefore, the bank needs to manage its lending processes to keep the nonaccrual rate level low and to invest mainly in risk free securities in order to avoid bankruptcies. The outcome is logical and supports the theory of bank failures.

Finally, it was tested if logit model is more suitable for bank failure prediction than probit model. According to the hypothesis, the accounting data used in the empirical analysis is not usually normally distributed suggesting that the logistic regression is preferable methodology for bankruptcy prediction. As was presented the differences between the models seem to be rather faint, but logit models nevertheless stand out as slightly superior compared to probit models.

For the future research more sophisticated bank failure prediction methods than logistic regression should be used. It would be interesting to study whether there is a difference between the results of this thesis and the results of a research made with, for example, neural network learning algorithm or genetic programming. It would also be fascinating to test the model constructed here for the future bank failures. In addition, it would be interesting to explore banks' total subprime exposure and include it to the dataset in order to improve the accuracy of the prediction model. As was mentioned, the task is practically impossible at the moment so new kinds of regulations are required to do so.

Bibliography

Aharony, J., Jones, C. P., & Swary, I. (1980). An Analysis of Risk and Return Characteristics of Corporate Bankruptcy Using Capital Market Data. *The Journal of Finance , 35* (4), pp. 1001-1016.

Aldrich, J., & Nelson, F. D. (1984). *Linear Probability, Logit and Probit Models* (Vols. No. 07-045). Beverly Hills, CA: SAGE Publications, Quantitative Applications in the Social Sciences Series .

Alfaro-Cid, E., Sharman, K., & Esparcia-Alcázar, A. (2007). A Genetic Programming Approach for Bankruptcy Prediction Using a Highly Unbalanced Database. (M. G. al., Ed.) *EvoWorkshops 2007, LNCS 4448* , pp. 169-178.

Altman, E. H., & Narayanan, P. (1977). ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance , 1* (1), pp. 29-54.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance, Vol. 23, No. 4* , 589-609.

Altman, E. I., & Brenner, M. (1981). Information Effects and Stock Market Response to Signs of Firm Deterioration. *The Journal of Financial and Quantitative Analysis , 16* (1), pp. 35-51.

Andrews, E. L., de la Merced, M. J., & Williams Walsh, M. (2008, September 16). *Fed's $85 Billion Loan Rescues Insurer*. Retrieved September 21, 2009, from The New York Times, Business: http://www.nytimes.com/2008/09/17/business/17insure.html

Aziz, A., Emanuel, D. C., & Lawson, G. H. (2007). Bankruptcy Prediction - An Investigation of Cash Flow Based Models. *Journal of Management Studies , 25* (5), pp. 419 - 437.

Bankruptcy Basics. (2009). *A Publication of the Bankruptcy Judges Division, Administrative Office of the United States Courts* .

Batte, J. A., Philosophov, L. V., & Philosophov, V. L. (2007). Predicting the Event and Time Horizon of Bankruptcy Using Financial Ratios and the Maturity Schedule of Long-Term Debt. *EFA 2005 Moscow Meeting Paper* .

Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata.* College Station, Texas: A Stata Press Publication.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Empirical Research in Accounting, Selected Studies* .

Beaver, W. H. (1968). Market Prices, Financial Ratios, and the Predicition of Failure. *Journal of Accounting Research , 6* (2), pp. 179-192.

Bernanke, B. S. (1983). Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression. *The American Economic Review, Vol. 73, No. 3* , 257-276.
Blackburn, R. (2008). The Subprime Crisis. *New Left Review* .

Bongard, M., Vaintsveig, M., Guberman, S., & Izvekova, M. (1966). The Use of Self Learning Programs in the Detection of Oil Gontaining Layers. *Geology Geofiz , 6*, pp. 96-105.

Boritz, J. E. (1991). *The "going concern" assumption: accounting and auditing implications.* Toronto, Canada: The Canadian Institute of Chartered Accountants (CICA).

Brabazon, A., & Keenan, P. B. (2004). A hybrid genetic model for the prediction of corporate failure. *Computational Management Science 1* , pp. 293-310.

Cass, D., & Shell, K. (1983). Do Sunspots Matter? *The Journal of Political Economy, Vol. 91, No. 2* , 193-227.

Clark, T. A., & Weinstein, M. I. (1983). The Behavior of the Common Stock of Bankrupt Firms. *The Journal of Finance, Vol. 38, No. 2, Papers and Proceedings Forty-First Annual Meeting American Finance Association New York, N.Y. December 28-30, 1982* , 489-504.

Cole, R. A., & Gunther, J. W. (1998). Predicting Bank Failures: A Comparison of On- and Off-Site Monitoring Systems. *Journal of Financial Services Research 13:2* , 103-117.

Curry, T. J., Elmer, P. J., & Fissel, G. S. (2009, September 11). *Using Market Information to Help Identify Distressed Institutions: A Regulatory Perspective*. Retrieved November 11, 2009, from FDIC Banking Review: http://www.fdic.gov/bank/analytical/banking/2003sep/article1.html

Dabos, M., & Escudero, W. S. (2004). Explaining and Predicting Bank Failure Using Duration Models: The Case of Argentina After the Mexican Crisis. *Revista de Análisis Económico, Vol. 19, No 1* , 31-49.

Demyanyk, Y., & Van Hemert, O. (2008). Understanding the Subprime Mortgage Crisis. *Federal Reserve Bank of Cleveland and New York University - Department of Finance, Working Paper Series* .

Epiq. (2009). *Lehman Brothers Holdings Inc. (Chapter 11)*. Retrieved September 15, 2009 from Epiq Systems: http://chap11.epiqsystems.com/clientdefault.aspx?pk=de7ced2b-52e7-4172-92e1-9ec425933bd0&l=1

Fawcett, T. (2006). An intorduction to ROC analysis. *Pattern Recognition Letters* , 861-874.

FDIC. (2010, July 16). *Failed Bank List*. Retrieved July 26, 2010, from FDIC: http://www.fdic.gov/bank/individual/failed/banklist.html

FDIC. (2009, September 21). *Failures and Assistance Transactions*. Retrieved September 21, 2009, from Federal Deposit Insurance Corporation: http://www2.fdic.gov/hsob/SelectRpt.asp?EntryTyp=30

FDIC. (1997). *History of the Eighties - Lessons for the Future.* Washington, DC: Federal Deposit Insurance Corporation.

FDIC. (2006, May 2). *Learning Bank*. Retrieved July 8, 2009, from Federal Deposit Insurance Corporation (FDIC): http://www.fdic.gov/about/learn/learning/when/index.html

FDIC. (2008, February 15). *Statistics on Depository Institutions*. Retrieved August 20, 2010, from FDIC: http://www2.fdic.gov/sdi/main.asp

FDIC. (2002, December 20). *The S&L Chrisis: A Chrono-Bibliography*. Retrieved July 9, 2010, from Federal Deposit Insurance Corporation (FDIC): http://www.fdic.gov/bank/historical/s&l/

Foote, C. L., Gerardi, K., Goette, L., & Willen, P. S. (2008). Subprime Facts: What (We Think) We Know about the Subprime Crisis and What We Don't. *Public Policy Discussion Papers, No. 2* .

Friedman, M., & Schwartz, A. J. (1963). A Monetary History of the United States, 1867-1960. *Princeton University Press, New Jersey.*

Gaytán, A., & Johnson, C. A. (2002). A Review of the Literature on Early Warning Systems for Banking Crises. *Central Bank of Chile Working Papers, No. 183* .

Gentry, J. A., Newbold, P., & Whitford, D. T. (1985a). Classifying Bankrupt Firms with Funds Flow Components. *Journal of Accounting Research , 23* (1), pp. 146-160.

Gentry, J. A., Newbold, P., & Whitford, D. T. (1985b). Prdicting Bankruptcy: If Cash Flows Not the Bottom Line, What Is? *Financial Analysts Journal , 41* (5), pp. 47-56.

Gibson, M. S. (2004). Understanding the Risk of Synthetic CDOs. *Ferderal Reserve Board, Finance and Economics Discussion Series, No. 36* .

Gilbert, A. R., Meyer, A. P., & Vaughan, M. D. (2000). *The Role of a CAMEL Downgrade Model in Bank Surveillance.* St. Louis: Federal Reserve Bank of St. Louis.

González-Hermosillo, B. (1999). Determinants of Ex-Ante Banking System Distress: A Macro-Micro Emiprical Exploration of Some Recent Episodes. *IMF Working Paper* .

González-Hermosillo, B., Pazarbaşioğlu, C., & Billings, R. (1996). Banking System Fragility: Likelihood Versus Timing of Failure - An Application to Mexican Financial Crisis. *IMF Working Paper* .

Gorton, G. (2008). The Subprime Panic. *Yale International Center for Finance Working Paper No. 25* .

Hamer, M. M. (1983). Failure prediction: Sensitivity of classification accuracy to alternative statistical methods and variable sets. *Journal of Accounting and Public Policy , 2* (4), pp. 289-307.

Hawke, J. D. (2000, September 17). Remarks by John D. Hawke, Jr Comptroller of the Currency Before the American Bankers Association. Washington D.C. Retrieved July 8, 2009, from: http://www.occ.treas.gov/ftp/release/2000-71a.txt

Jagtiani, J., Kolari, J., Lemieux, C., & Shin, H. (2003, Q3). Early warning models for bank supervision: Simpler could be better. *Economic Perspectives* , pp. 49-60.

Jones, S., & Hensher, D. A. (2004). Predicting Firm Financial Distress: A Mixed Logit Model. *The Accounting Review , 79* (4), pp. 1011-1038.

Karels, G., & Prakash, A. (1987). Multivariate Normality and Forecasting of Business Bankruptcy. *Journal of Business, Finance and Accounting* (14), pp. 573-593.

Keasey, K., & McGuinness, P. (1990). The Failure of UK Industrial Firms for the Priod 1976-1984, Logistic Analysis and Entropy Measures.

Kiefer, N. M. (1988). Economic Duration Data and Hazard Functions. *Journal of Economic Literature, Vol. 26, No. 2* , 646-679.

Kiff, J., & Mills, P. (2007). Money for Nothing and Checks for Free: Recent Developments in U.S. Subprime Mortgage Markets. *International Monetary Fund, Working Paper* .

Kim, M. J., & Han, I. (2003). The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. *Expert Systems with Applications , 25*, pp. 637-646.

Klyuev, V., & Mills, P. (2007). Is Housing Wealth an ''ATM''? The Relationship Between Household Wealth, Home Equity Withdrawaland, and Saving Rates. *IMF Staff Papers, Vol. 54, No. 3* .

Kolari, J., Caputo, M., & Wagner, D. (1996). Trait Recognition: An Alternative Approach to Early Warning Systems in Commercial Banking. *Journal of Business Finance & Accounting* , 1415-1434.

Kolari, J., Glennon, D., Shin, H., & Caputo, M. (2002). Predicting large US commercial bank failures. *Journal of Economics and Buisness* (54), pp. 361-387.

Koza, J. R., & Poli, R. (2006). Genetic Programming. In E. K. Burke, & G. Kendall, *Search Methodolgies, Introductory Tutorials in Optimization and Decision Support Techniques* (Vol. Ch. 5, pp. 127-164). Springer US.

Kregel, J. (2008). Minsky's Cushions of Safety: Systemic Risk and the Crisis in the US Subprime Mortgage Market. *The Levy Economics Institute of Bard College, Public Policy Brief, No. 98* .

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press.

Landon, T. J. (2008, March 14). *JPMorgan and Fed Move to Bail Out Bear Stearns.* Retrieved September 21, 2009, from The New York Times, Business: http://www.nytimes.com/2008/03/14/business/14cnd-bear.html?_r=1

Lanine, G., & Vander Vennet, R. (2006). Failure Prediction in the Russian Bank Sector with Logit and Trait Recognition Models. *Expert Systems with Applications* , 463-478.

Long, S. J., & Freese, J. (2003). *Regression Models for Categorical Dependet Variables Using Stata.* College Station, Texas: A Stata Press Publication.

Lopez, J. A. (1999, June 11). *Using CAMELS Ratings to Monitor Bank.* Retrieved July 22, 2009, from Federal Reserve of San Francisco, Economic Research and Data: http://www.frbsf.org/econrsrch/wklyltr/wklyltr99/el99-19.html

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* New York: Cambridge University Press.

Maddala, G. S., & Lahiri, K. (2009). *Introduction to Econometrics.* West Sussex, England: John Wiley & Sons Ltd.

McDonald, L. (2009, September 7). HARDtalk. *HARDtalk.* (S. Sacur, Interviewer) BBC. New York.

Meyer, P. A., & Pifer, H. W. (1970). Prediction of Bank Failures. *The Journal of Finance, Vol. 25, No. 4* , 853-868.

Michael, S., Georgios, D., Nikolaos, M., & Constantin, Z. (1999). A fuzzy knowledge-based decision aiding method for the assessment of financial risk: The case of corporate bankruptcy prediction. *European Symposium on Intelligent Techniques* .

Pindyck, R. S., & Rubinfeld, D. L. (1997). *Econometrics Models and Economic Forecasts.* Irwin/McGraw-Hill.

Russell, C., & Mullin, S. (2009). Home Prices Decline Slightly in Second Quarter. *Federal Housing Finance Agency, News Release* .

Santomero, A., & Vinso, J. (1977). Estimating the probability of failure for firms in the banking system,. *Journal of Banking and Finance , 1* (2), pp. 185-205.

Scott, J. (1981). The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. *Journal of Banking and Finance , 5,* pp. 317-344.

Stone, M., & Rasp, J. (1991). Tradeoffs in the Choice between Logit and OLS for Accounting Choice Studies. *66* (No. 1, pp. 170-187).

Tam, K. Y., & Kiang, M. K. (1992). Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science , 38* (7), pp. 926-947.

Tan, A. (2009, July 10). *DBS Sued by Investors for Losses Tied to Lehman Notes (Update 1)*. Retrieved July 13, 2009, from Bloomberg.com: http://www.bloomberg.com/apps/news?pid=newsarchive&sid=ayRIru1HUMZM

The White House. (2004, August 9). *Fact Sheet: America's Ownership Society: Expanding Opportunities*. Retrieved September 18, 2009, from The White House, President George W. Bush: http://georgewbush-whitehouse.archives.gov/news/releases/2004/08/20040809-9.html

Thomson, J. B. (1991). Predicting Bank Failures inthe 1980s. *Economic Review* .

Tkaczyk, C. (2009, June 1). *The 10 largest U.S. bankruptcies*. Retrieved October 20, 2009, from CNN Money, Fortune: http://money.cnn.com/galleries/2009/fortune/0905/gallery.largest_bankruptcies.fortune/index.html

Wilcox, J. (1976). The gambler's ruin approach to business risk. *Sloan Management Review* .

Wilcox, J. W. (1973). A Prediction of Business Failure Using Accounting Data. *Journal of Accounting Research, Vol. 11, Empirical Research in Accounting: Selected Studies* , pp. 163-179.

Vinso, J. D. (1979). A Determination of the Risk of Ruin. *The Journal of Financial and Quantitative Analysis , 14* (1), pp. 77-100.

Wolf, M. (2008, February 20). *America's economy risks mother of all meltdowns*. Retrieved September 21, 2009, from Financial Times, Economists' Forum: http://blogs.ft.com/economistsforum/2008/02/americas-economy-risks-mother-of-all-meltdowns/

Wolfson, M. H. (1986). *Financial crises: Understanding the postwar US experience.* Notre Dame, Indiana: M. E. Sharpe, Inc.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control 8* , pp. 338-353.

Zavgren, C. V. (1985). Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. *Journal of Business Finance & Accounting, 12(1)* .

Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research , 22,* pp. 59-82.

Appendix 1. Student's t-test of Independent Variables between Failed and Active Banks

	Liquid assets	Uninsured deposits	Loan exposure	Loan funding	Nonaccrual rate	Past due loan rate	Loan loss allowance
Q1	0,0016	0,0497	0,1281	0,9309	0,0000	0,2967	0,0000
Q2	0,0037	0,1053	0,3559	0,6777	0,0000	0,4279	0,0000
Q3	0,0288	0,0803	0,1801	0,2995	0,0000	0,4050	0,0000
Q4	0,1503	0,0514	0,0619	0,1699	0,0000	0,1866	0,0000
Q5	0,2272	0,0378	0,0754	0,0691	0,0000	0,6643	0,0003
Q6	0,2533	0,0600	0,0337	0,0576	0,0000	0,6649	0,0402
Q7	0,2508	0,0672	0,0317	0,0621	0,0000	0,8013	0,1131
Q8	0,2890	0,0661	0,0724	0,1297	0,0007	0,6325	0,5527
Q9	0,4548	0,0064	0,1925	0,1479	0,0059	0,7996	0,5874
Q10	0,3049	0,0001	0,1201	0,1330	0,0034	0,4391	0,4686
Q11	0,2471	0,0001	0,0847	0,0894	0,0245	0,3978	0,6039
Q12	0,2304	0,0000	0,0442	0,1066	0,0281	0,6019	0,5027

	Provision rate	Loss rate	Capital ratio	Return on assets	Return on equity	Dividend rate	Net interest margin
Q1	0,0000	0,0000	0,0000	0,0000	0,0025	0,0627	0,0000
Q2	0,0000	0,0000	0,0000	0,0000	0,4496	0,0822	0,0000
Q3	0,0000	0,0000	0,0000	0,0000	0,0000	0,0565	0,0019
Q4	0,0000	0,0000	0,0002	0,0000	0,0000	0,1695	0,1549
Q5	0,0001	0,0012	0,0027	0,0000	0,0000	0,3694	0,4425
Q6	0,0559	0,1242	0,0636	0,0184	0,0235	0,8707	0,6917
Q7	0,0602	0,3637	0,1440	0,0454	0,0660	0,6086	0,6583
Q8	0,1028	0,9194	0,3595	0,5584	0,6511	0,6784	0,8704
Q9	0,7456	0,8556	0,5897	0,8932	0,5213	0,3156	0,8037
Q10	0,8003	0,6988	0,5632	0,9404	0,6355	0,5750	0,9349
Q11	0,9203	0,5610	0,6605	0,9578	0,6555	0,1100	0,7967
Q12	0,5097	0,9405	0,4415	0,9454	0,5103	0,1249	0,6112

	Net operating margin	Tax exposure	Capital growth	Loan growth	Commercial loan risk	Real estate loan risk	Agricultural loan risk
Q1	0,0000	0,0043	0,3731	0,0093	0,1041	0,0143	0,7967
Q2	0,0000	0,0000	0,8787	0,0035	0,1945	0,0182	0,8519
Q3	0,0000	0,0000	0,0000	0,0000	0,0992	0,0140	0,9878
Q4	0,0000	0,0000	0,0008	0,0346	0,1327	0,0163	0,9512
Q5	0,0000	0,0002	0,0001	0,0651	0,1507	0,0188	0,9862
Q6	0,0132	0,0717	0,4038	0,5754	0,1633	0,0195	0,9021
Q7	0,0240	0,0989	0,0324	0,1007	0,0806	0,0083	0,8794
Q8	0,3795	0,4191	0,4609	0,0061	0,0476	0,0050	0,9949
Q9	0,8054	0,8010	0,3475	0,1559	0,0466	0,0081	0,9064
Q10	0,7446	0,3289	0,0084	0,2631	0,0434	0,0073	0,9078
Q11	0,7510	0,2967	0,0119	0,1170	0,0354	0,0056	0,9814
Q12	0,7540	0,7327	0,2281	0,8277	0,0307	0,0053	0,9142

	Credit card loan risk	Loan diversification	Demand deposit mix	Time deposit mix	MBS ratio	ABS ratio	CMO ratio
Q1	0,4996	0,0014	0,3946	0,0000	0,7646	0,4068	0,0543
Q2	0,4861	0,0020	0,0703	0,0000	0,5310	0,5998	0,0460
Q3	0,4666	0,0030	0,0540	0,0000	0,3185	0,9146	0,0104
Q4	0,5331	0,0030	0,1251	0,0002	0,3836	0,7801	0,0308
Q5	0,5041	0,0044	0,4486	0,0013	0,1661	0,8467	0,0148
Q6	0,5294	0,0038	0,2473	0,0055	0,1468	0,8841	0,0291
Q7	0,4783	0,0023	0,1113	0,0107	0,1897	0,9150	0,0765
Q8	0,5644	0,0009	0,2040	0,0081	0,1963	0,8100	0,1188
Q9	0,7370	0,0006	0,3295	0,0018	0,1812	0,7767	0,1225
Q10	0,6951	0,0004	0,2147	0,0013	0,1880	0,7628	0,2433
Q11	0,6580	0,0004	0,4667	0,0011	0,2010	0,8116	0,2279
Q12	0,7320	0,0003	0,5592	0,0014	0,3740	0,7918	0,1129

	Risk free securities	Asset variation	Loans and leases variation	Equity variation
Q1	0,1751	0,0000	0,0000	0,0000
Q2	0,1128	0,1231	0,0029	0,9319
Q3	0,0647	0,0004	0,0000	0,0000
Q4	0,0839	0,0234	0,0305	0,0000
Q5	0,0539	0,9879	0,3445	0,0001
Q6	0,0337	0,8395	0,8530	0,2895
Q7	0,0501	0,2554	0,0412	0,4444
Q8	0,0348	0,5016	0,3882	0,7009
Q9	0,0724	0,9160	0,3264	0,8703
Q10	0,1615	0,0039	0,0765	0,0041
Q11	0,0931	0,0590	0,2234	0,0446
Q12	0,0416	0,2940	0,8340	0,0846

Appendix 2. Correlation between the Key Independent Variables

	nonacc~e	netope~n	return~s	netint~n	provis~e	lossra~e	loanlo~e	capita~o	loansa~n
nonaccrual~e	1,000								
netoperati~n	-0,696	1,000							
returnonas~s	-0,5916	0,8973	1,000						
netinteres~n	-0,6296	0,5791	0,554	1,000					
provisionr~e	0,5503	-0,8088	-0,684	-0,293	1,000				
lossrate	0,589	-0,8199	-0,703	-0,349	0,822	1,000			
loanlossal~e	0,657	-0,630	-0,5574	-0,3189	0,6072	0,583	1,000		
capitalratio	-0,602	0,661	0,7233	0,6739	-0,4959	-0,4846	-0,463	1,000	
loansandle~n	-0,341	0,427	0,4379	0,2262	-0,4125	-0,4388	-0,423	0,515	1,000
assetsvari~n	-0,3754	0,4682	0,448	0,197	-0,419	-0,418	-0,319	0,3835	0,6278
loandivers~n	0,4063	-0,2494	-0,189	-0,275	0,219	0,286	0,2875	-0,3187	-0,1387
realestate~k	0,366	-0,266	-0,2212	-0,4794	0,1133	0,1001	0,035	-0,436	-0,019
commercial~k	-0,2621	0,128	0,067	0,195	-0,052	-0,158	-0,1275	0,2863	0,187
timedepos~x	0,539	-0,467	-0,4238	-0,534	0,2991	0,3323	0,440	-0,470	-0,293
liquidassets	-0,265	0,308	0,2067	0,0762	-0,2996	-0,1607	-0,160	0,121	0,0223
loanexposure	0,2585	-0,2329	-0,152	0,043	0,2544	0,126	0,106	-0,059	0,132
returnoneq~y	-0,3942	0,4421	0,198	0,221	-0,459	-0,456	-0,2505	0,2759	0,2507
equityvari~n	-0,4345	0,5303	0,461	0,324	-0,583	-0,518	-0,2672	0,5062	0,3154
capitalgro~h	0,23	-0,1148	-0,094	-0,161	0,078	0,130	0,0145	-0,1111	-0,0413
taxexposure	-0,2387	0,154	0,140	0,244	-0,321	-0,128	-0,2832	0,2271	0,0774
cmoratio	0,0259	-0,0178	-0,061	-0,166	0,059	-0,014	0,0159	-0,2048	0,0093
uninsuredd~s	0,2837	-0,3086	-0,231	-0,3217	0,346	0,241	0,2157	-0,2197	-0,0354
riskfreese~s	0,1059	-0,0254	0,011	-0,1176	-0,028	-0,070	-0,0517	0,0605	0,0204
dividendrate	-0,1602	0,1667	0,166	0,2075	-0,134	-0,140	-0,1503	0,3075	0,2806
loangrowth	-0,1465	0,1354	0,179	0,229	-0,092	-0,178	-0,2214	0,2977	0,5351

	assets~n	loandi~n	reales~k	commer~k	timede~x	liquid~s	loanex~e	return~y	equity~n
assetsvari~n	1,000								
loandivers~n	-0,136	1,000							
realestate~k	-0,0993	0,7177	1,000						
commercial~k	0,137	-0,626	-0,4874	1,000					
timedepos~x	-0,1588	0,4442	0,398	-0,2323	1,000				
liquidassets	0,1379	-0,2715	-0,273	0,0888	-0,122	1,000			
loanexposure	-0,266	0,382	0,393	-0,1004	0,109	-0,793	1,000		
returnoneq~y	0,256	-0,165	-0,157	0,1897	-0,2274	0,1312	-0,045	1,000	
equityvari~n	0,346	-0,203	-0,1926	0,131	-0,2238	0,1559	-0,108	0,787	1,000
capitalgro~h	-0,013	0,104	0,0775	-0,075	0,1135	-0,0465	0,038	-0,143	-0,093
taxexposure	-0,053	-0,196	-0,2518	0,120	-0,1691	0,1169	-0,061	0,026	0,067
cmoratio	0,071	0,051	0,0607	0,102	0,1033	0,1778	-0,197	-0,022	-0,230
uninsuredd~s	-0,026	0,380	0,423	-0,184	0,3657	-0,1806	0,206	-0,119	-0,192
riskfreese~s	-0,023	-0,055	0,1224	0,183	0,1278	-0,0357	0,125	-0,035	0,094
dividendrate	0,163	-0,172	-0,142	0,143	-0,1736	-0,0428	0,000	0,067	0,073
loangrowth	0,205	-0,059	-0,0151	0,072	-0,2647	-0,1532	0,247	0,121	0,166

	capita~h	taxexp~e	cmoratio	uninsu~s	riskfr~s	divide~e	loangr~h
capitalgro~h	1,000						
taxexposure	0,000	1,000					
cmoratio	-0,068	-0,086	1,000				
uninsuredd~s	-0,010	-0,296	0,165	1,000			
riskfreese~s	0,080	-0,013	-0,215	-0,051	1,000		
dividendrate	-0,020	0,112	-0,102	-0,170	0,070	1,000	
loangrowth	-0,038	0,086	-0,081	-0,052	0,073	0,093	1,000

Appendix 3. The Plan Analysis

	Wald chi2	Prob > chi2	Pseudo R2	CC	ROC
Plan 1	34,64	0,0005	0,6682	92,74 %	0,9670
Plan 2	47,95	0,0000	0,6463	91,13 %	0,9602
Plan 3	46,37	0,0000	0,6243	92,74 %	0,9592
Plan 4	51,21	0,0000	0,6412	90,32 %	0,9584
Plan 5	74,23	0,0000	0,6994	91,94 %	0,9716
Plan 6	51,37	0,0000	0,5802	91,94 %	0,9472
Plan 7	36,90	0,0002	0,6683	92,74 %	0,9677
Plan 8	49,39	0,0000	0,6470	91,13 %	0,9625
Plan 9	45,39	0,0000	0,6352	92,74 %	0,9592
Plan 10	56,64	0,0000	0,6412	90,32 %	0,9584
Plan 11	67,09	0,0000	0,7104	91,94 %	0,9742
Plan 12	47,82	0,0000	0,5994	91,94 %	0,9511
Plan 13	47,21	0,0000	0,6810	91,94 %	0,9727
Plan 14	66,43	0,0000	0,6500	91,13 %	0,9631
Plan 15	48,72	0,0000	0,6284	91,94 %	0,9605
Plan 16	51,21	0,0000	0,6412	90,32 %	0,9584
Plan 17	66,77	0,0000	0,7280	93,55 %	0,9774
Plan 18	55,63	0,0000	0,5916	89,52 %	0,9502
Plan 19	44,26	0,0000	0,6794	93,55 %	0,9727
Plan 20	70,45	0,0000	0,6510	91,13 %	0,9633
Plan 21	50,06	0,0000	0,6393	92,74 %	0,9631
Plan 22	69,65	0,0000	0,6431	90,32 %	0,9586
Plan 23	66,74	0,0000	0,7352	91,13 %	0,9789
Plan 24	55,05	0,0000	0,6088	91,94 %	0,9581
Plan 25	34,06	0,0007	0,6631	92,74 %	0,9664
Plan 26	46,40	0,0000	0,6195	88,71 %	0,9553
Plan 27	41,19	0,0001	0,6308	92,74 %	0,9602
Plan 28	49,91	0,0000	0,6329	89,52 %	0,9563
Plan 29	66,43	0,0000	0,7050	92,74 %	0,9714
Plan 30	43,34	0,0000	0,5929	92,74 %	0,9527
Plan 31	34,37	0,0006	0,6639	93,55 %	0,9677
Plan 32	48,90	0,0000	0,6176	89,52 %	0,9542
Plan 33	40,21	0,0001	0,6193	92,74 %	0,9545
Plan 34	56,26	0,0000	0,6331	90,32 %	0,9560
Plan 35	69,78	0,0000	0,6975	91,94 %	0,9719
Plan 36	38,98	0,0002	0,5854	91,13 %	0,9514

	Wald chi2	Prob > chi2	Pseudo R2	CC	ROC
Plan 37	50,07	0,0000	0,6785	91,13 %	0,9727
Plan 38	75,58	0,0000	0,6282	88,71 %	0,9592
Plan 39	47,80	0,0000	0,6418	92,74 %	0,9641
Plan 40	65,68	0,0000	0,6372	89,52 %	0,9584
Plan 41	52,23	0,0000	0,7304	92,74 %	0,9784
Plan 42	67,36	0,0000	0,6079	90,32 %	0,9592
Plan 43	48,21	0,0000	0,6784	91,94 %	0,9729
Plan 44	83,87	0,0000	0,6267	87,90 %	0,9584
Plan 45	41,61	0,0001	0,6258	92,74 %	0,9620
Plan 46	76,26	0,0000	0,6376	91,13 %	0,9592
Plan 47	58,21	0,0000	0,7155	91,94 %	0,9774
Plan 48	43,27	0,0001	0,5953	91,13 %	0,9581
Plan 49	34,04	0,0007	0,6533	94,35 %	0,9680
Plan 50	47,00	0,0000	0,6485	91,94 %	0,9599
Plan 51	59,33	0,0000	0,6579	91,13 %	0,9644
Plan 52	35,70	0,0004	0,6560	95,16 %	0,9688
Plan 53	47,94	0,0000	0,6488	91,13 %	0,9615
Plan 54	59,48	0,0000	0,6878	91,94 %	0,9701
Plan 55	40,96	0,0001	0,6698	91,94 %	0,9729
Plan 56	65,55	0,0000	0,6530	91,13 %	0,9641
Plan 57	56,94	0,0000	0,6903	91,94 %	0,9722
Plan 58	35,02	0,0008	0,6699	93,55 %	0,9729
Plan 59	68,89	0,0000	0,6536	91,13 %	0,9638
Plan 60	59,59	0,0000	0,7159	90,32 %	0,9758
Plan 61	35,06	0,0005	0,6459	95,16 %	0,9659
Plan 62	45,43	0,0000	0,6243	87,90 %	0,9568
Plan 63	57,44	0,0000	0,6782	91,94 %	0,9677
Plan 64	36,46	0,0003	0,6488	95,16 %	0,9672
Plan 65	47,37	0,0000	0,6227	88,71 %	0,9563
Plan 66	59,95	0,0000	0,6774	92,74 %	0,9672
Plan 67	44,67	0,0000	0,6651	92,74 %	0,9716
Plan 68	70,91	0,0000	0,6340	88,71 %	0,9594
Plan 69	49,96	0,0000	0,7104	91,94 %	0,9750
Plan 70	42,02	0,0001	0,6663	92,74 %	0,9729
Plan 71	79,77	0,0000	0,6325	88,71 %	0,9581
Plan 72	55,31	0,0000	0,6984	91,13 %	0,9727