UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

PREDICTION OF BANK FAILURES: STATISTICAL AND MACHINE LEARNING TECHNIQUES

RELATORE:

CH.MA PROF.SSA CINZIA BALDAN

RELATORE:

CH.MO PROF. FRANCESCO ZEN

LAUREANDA: GIULIA MARCATO

MATRICOLA N. 1203538

ANNO ACCADEMICO 2019 – 2020

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Giulia Marcato

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Bank failures happen when the institutions are not able to fulfill their obligations, requiring the intervention of the regulator. These failures have a strong negative effect on the economic and financial systems, this is the reason why investors and supervisors created different models to control the institutions.

One of the biggest reasons why the bank failures have this negative consequence is the contagion effect; the default of these institutions is damaging because it spreads through the banking system leading other banks to fall insolvent.

Banking system is hence strongly regulated, banks need to satisfy prudential requirements either nationally or internationally decided but sometimes the regulations are not efficient and lead to counterproductive behavior (increase of the probability of failures and increasing the costs connected to them).

The failures of the banking system are driven by different factors: high presence of Non-Performing Loans on banks' balance sheets, distortive managerial incentives that lead to take a high level of risk, issuance of loans to individuals which are not creditworthy (i.e. low FICO score) are some of them. Both internal and external causes are fundamental to understand why banks can fail and why some of them are weaker than others.

There are different tools to analyze if a bank is healthy or not, the main distinction between these methodologies is on-site and off-site.

Under FDICIA (Federal Deposit Insurance Corporation Improvement Act) of 1991 new rules have been set for the on-site examinations. At least once a year (once every 18 months for small and sound institutions), the regulators should visit the offices of the bank to conduct an on-site analysis. It is an expensive analysis in terms of time and money and it cannot be frequently conducted, often offering old and not updated data.

Nevertheless, as soon as they are conducted, on-site examinations are highly accurate and allow to deeply understand if the bank needs to change something in its operations and if tighter requirements are needed. It also allows to examine the quality of managers' work which is else difficult to estimate.

One of the most common used on-site examination in the U.S. is the CAMELS (Capital, Assets, Management, Earnings, Liquidity, and Sensitivity) which gives as output a composite rating for the institution that goes from 1 (sound bank) to 5 (failing bank). The regulator imposes tighter regulations for the banks rated 3, 4, and 5.

Off-site analysis is mainly based on data publicly available (financial statements and regulatory reports). The results from the analysis are less accurate than the ones the regulator retrieves from the on-site examinations but they give a better understanding of what it is happening to the bank in the period of time between the on-site analyses; indeed, the frequency of supervision is more frequent.

U.S. government implemented different models for the off-site analysis: the Uniform Bank Surveillance Screen, CAEL System, SCOR System and SEER. All these systems mainly use statistical techniques to evaluate the banks' conditions.

The costs associated with the off-site monitoring tools are lower if compared to the on-site analyses and usually these analyses also are less time consuming.

This work is based on the analysis of different types of off-site examinations to understand how accurate they can be to predict the failure of U.S. large commercial banks. The predictors used for this research are all retrieved from the FDIC database for the period of time that goes from 2005 to 2015; the ratios have been selected on the basis of their relevance in predicting the failure of the banks.

Both statistical and machine learning techniques are used in order to evaluate in a more precise way which is the best model to predict the failure of the banks.

In particular the statistical techniques that are implemented are Logistic Regression and Linear Discriminant Analysis.

Machine learning algorithms learn from the data which are given as input. Support Vector Machines, k-Nearest Neighbors, Random Forest, and Backpropagation Neural Network are the methods implemented.

Type I and Type II errors and other measures for the accuracy and prediction ability are to define which method can be more reliable as an off-site examination tool.

Once the best off-site tool is found between the ones implemented, it is used to verify how the prediction power behaves as the time from failure increases (up until 8 quarters prior to failure).

The work is presented as follows: Chapter 1 is focused on the analysis of the reasons of the bank failures, including a section on the importance of the prediction of the failure. In Chapter 2 on-site and off-site analyses are showed and compared. Different off-site examinations are presented to understand the strengthen and weaknesses point of each of them. Chapter 3 focuses on the different statistical and supervised machine learning approaches, highlighting their mechanisms and their ability of prediction through the examination of some literature. Chapter 4 presents the empirical analysis: the tools explained in Chapter 3 are implemented in order to compare them and understand which of the tools is a better in predicting the failure of the analyzed banks.

1. REASONS OF BANKING FAILURE

The aim of this chapter is to discuss some of the most important causes that affect the stability of the banking system and lead some banks to insolvency. In a world in which the resources of an economy are fully-employed, the bankruptcy of the bank comes from the misallocation of these resources (Meyer, 1970); the authorities need to adopt measures in order to minimize the risk of failure and also to increase the policies to avoid the misallocation of the assets. From a social point of view, a voluntary liquidation or any action focused in removing the losses at an early date, are always preferable to a banking failure. In this chapter there will be an analysis of some of the major reasons that lead commercial banks to fail.

The crises of the banking system become worse when the macroeconomic environment is weak (low GDP growth) and when the inflation is high. Moreover, they are a frequently studied topic because they can damage the functioning of the payment system and lead to a lack of confidence in the domestic banks. The regulators of each country should implement some well-designed regulations to decrease the fragility of their banking system. Nevertheless, these rules can have several drawbacks which are not easy to forecast or which cannot be deleted (Detragiache, 1998). Moreover, the operations that the governments implement in order to save the banking system can also have the effect of weaken the managerial incentives and force healthy banks to bear the losses. Policies that prevent the event of systemic banking problems are hence necessary. There are different events that can cause fragility in the banking system such as: loss of foreign exchange reserves, high interest rates, low output growth and decline in stock prices.

The weak macroeconomic environment is not the only cause that lead to financial instability, other elements will be analyzed in this chapter.

One of the biggest problems regarding the banking system is the ***contagion effect***; the linkages between different banks determine an interconnection that can be an issue during period of financial crisis. The contagion effect can be defined as the transmission of the idiosyncratic risk (Hasman, 2013) that firstly affects one bank of a group of banks and then is transferred to other institutions in the economic sector. It is used to describe the spillover of the effect of a shock from one or more firms to other (Kaufman, 1994). The banking crises underlined how the banking sector can amplify the troubles and how interconnected are these financial institutions.

Since the contagion seems to be a relevant exposure, some prudential regulations are needed either in the form of capital and/or liquidity requirements which allow banks to be more resistant to shocks on the market conditions. Sometimes the regulations that require adequate levels of capitalization and/or liquidity create some perverse incentives that need to be fixed in order for the whole mechanism to work smoothly.

The linkages among some of the institutions in the financial markets create concern while thinking about the financial contagion, the domino effect will propagate the shock borne by one agent to other agents (Hasman, 2013). The chain of contagion starts from a bank which is not able to settle the payment obligations and it propagates to other banks and institutions; this contagion is mainly due to the fact that the banks are interconnected and they maintain deposit balances in other banks to facilitate the clearing of payments under normal circumstances. These deposits have the characteristic of not being collateralized. At the same time, the crisis can be strong enough to propagate also to banks which are illiquid in that moment but they would be otherwise solvent.

The network between institutions works well also meanwhile some banks are facing troubles but with a bigger shock the connection can disrupt value of the entire system and spread the systemic risk also to the other sectors.

This threat explains in an efficient way why the failure of one bank may be worse than a failure of a firm outside the banking system.

The Section 1.1 will include some of the main reasons of the banking failure and the reasons why this topic is relevant, whilst in the Section 1.2 there will be an analysis of the distortions created by inefficient regulations.

1.1 Why do banks fail?

The cost of bank failure can be large and it may cause instability, this is the main motive why the focus should be on the events that lead to default. A first main division regarding reasons that can cause the failure of banks can be done between two factors: internal and external (Meyer, 1970). The local economy conditions and the general economy are the exogenous factors whilst the quality of management and the integrity of employees are the endogenous ones. Considering all the other conditions equal, banks established in areas with high income and fast growth rate are less likely to fail if compared to banks located in areas with lower rates of growth and a poorer economy. At the same time, also the structure of the banking system in the country can be a relevant factor while considering the performance of the institutions.

One of the main factors that play an important role in the investigation of the banks' default is the asymmetry of information; the banking crisis mainly arise when there is panic between the depositors whom believe that their bank is not safe anymore. There can be different sources of panic that are mainly due to a lack of information between the parties. The depositor may not discern if the individual bank is solvent or not but he can observe the impact of the shock on the bank's portfolio (Santonu, 2003).

In many cases, the crises of the banking system lead the customers to either have a panic view or an information-based view of the situation (Hasman, 2013). The first one was largely described by Diamond and Dybvig; it is a situation in which there is a problem of coordination among the depositors which start believing that the banks' assets are not safe anymore even if the view does not reflect the reality. The banks are defenseless while a run occurs and there aren't enough regulations and lead them to liquidate their assets prematurely. Under normal conditions of financial stability, the banks estimate their monthly liquidity needs and they can decide to loan their excess reserves to other institutions that are facing liquidity shortage. This creates a certain level of credit risk between them.

The information-based view instead focuses on the problems coming from the uncertainty and the asymmetry of information between the parties about the stability of the banks and the conditions that may provoke a run of the depositors. This kind of view (information-based) is important in this context because suggests that the failure of the banks can mainly come from two factors: from a bad management and from a macroeconomic risk.

The investors build their own idea of the level of riskiness of the bank on the basis of their perception of the risk; rational investors are more cautious while evaluating the exposure of a bank in a context of financial crisis. They are aware of the effects that has the interconnection of the institutions in the system; when the creditors lose confidence they demand immediate payment of the existing loan and the banks may be forced to liquidate their assets too early and this creates the incentive on more depositors to withdraw their money creating an exacerbation of the original effect.

The failure of the banking system is a relevant topic to analyze since, as mentioned above, the risk of contagion in is high if compared to other sectors. The bank contagion is particularly delicate if the adverse shock (a default or near-default) of one bank is going to have a big impact on other banks and also beyond the financial system. Banks are unique and need regulations applicable only for them. The contagion in the banking system is a concern mainly because of the following causes listed by Kaufman in 1994.

Occurs faster: in the case in which a large number of depositors simultaneously choose to transfer their deposits to another bank because they feel it unsafe, a run occurs. These phenomena are at the basis of a failure spread over the whole banking system. In order to satisfy the depositor who would like to withdraw all their money, the banks generally have to sell some assets in a quickly way and/or borrow funds from other banks. The liquidity problems are hence likely to arise in the case of a bank run; these issues can become worse and cause insolvency if the losses are large enough. In the banking system the contagion is faster since the participants are strictly interconnected hence, a run happening in one bank, is going to affect in a fast way the banks which are creditors.

Is wider within the industry: there can be identified two types of contagion in the banking contest (Kaufman, 1994). The pure industry specific contagion happens whenever the information regarding one or more banks in the industry affect all the other firms including the sound banks with few characteristics in common with the failed banks than being in the same industry.

The second kind of contagion happens when the characteristics that the failed and non-failed banks have in common are plenty and the problems that arise are due to a common aspect that the two banks share. The higher the similarity regarding the size, location, and the market served, the higher the probability that the intensity of contagion will be more severe.

The first kind of contagion is the most unpredictable and it is what mainly allows to state that it is wider within the industry. Moreover, the depositors usually have less information regarding the health condition of their own bank; the reasons are that many small depositors stated that an evaluation of the banks are costly (time and money), some assets the institutions have are unique and don't really have a market and if they have one, their value can rapidly change due to different external conditions. Hence, the depositors in many cases cannot distinguish if their bank is solvent or not and thus, they prefer to withdraw their money if they think it is not safer to keep them deposited. The process is quick because the deposits are considered a short-term liability for banks. At the same time the depositors can prefer to keep their money deposited unless there are clear evidences of weakness since it can be costly to rebuild a relationship with the new bank.

As stated before, one of the main reasons at the base of bank contagion happens after runs of depositors. They decide to move their money to a bank they feel to be safer or they can move them to currency. Usually, the higher the number of the banks that the depositors feel unsafe, the higher the possibility to choose to move to currency. Generally it is in the interest of other banks to assist institutions facing problems of liquidity (but not solvency); what they can do is purchasing the bank's assets at an equilibrium price rather than fire-sale price (Kaufman, 1994). The problems arise when the depositors flee to currency because the deposits are not redistributed among banks but they are lost; this creates a reduction in the money supply and a disruption in the payment system. (Kaufman, 1996)

Results in a higher number of failures: one of the main roles of the capital held by banks is the possibility to absorb the losses and avoid that the creditors will be affected by an eventual shock. It is clear that a smaller capital ratio leads to a higher possibility of default for the bank.

As it will be discussed later, it can happen that banks may tend to hold lower amount of capital since there is the insurance on deposits which creates incentives to bear more risk for the bank. During 1990s the level of banking failure has been high also because the legal and regulatory structures allowed them to increase the risk since there has been imposed a reduction on the geographical diversification.

It extends beyond the banking industry adversely affecting other financial industries and the macroeconomy: the crises that affect the banking sector can be seen as more severe than the ones that involve other industries. The main reason is the role of the banks in the whole system; they provide financial intermediation services and an interruption of the institution can lead to an increase of the cost of credit to the real sector causing a credit crunch. The failure of banks can extend beyond their sector; a reduction in the amount of deposits and the creation of losses for depositors create an overall reduction in the aggregate wealth. The failure of the banks can also lead other banks to choose less risky projects to finance creating a reduction of loans.

The fragility of the banking system is not a signal that the failure rate is higher in this sector than in other ones. The fragility means that there is the need of more attention and regulation (Kaufman, 1996). The introduction of the government regulation in the U.S. are focused on the protection against the fragility of the banks.

1.1.1 Credit Risk

Usually, the assumption underlying the loans that the banks issue to the customers is the expectation that the borrowers will repay in the future the principal and the interest on the debt and thus, the borrower's future rate of return will be sufficiently high. It is clear that the lender does not have the certainty that this will happen.

Credit Risk and Credit Standard are two elements that influence the operation of the banks. It is difficult, and almost impossible, for the banks to forecast in advance the magnitude of a potential default. The asset side of the banks' balance sheet strongly depends on the loan issued to the clients. To avoid or reduce the risk of failure, the banks need focus on the uncertainty of the repayment of the loans. To have some guarantees, the banks often ask for collaterals or some other form of security which are alternative to the cash repayment. The values of these alternative methods of payment need to be uncorrelated with the value of the project for which the loan has been issued and should have a stable value (not subject to strong fluctuations).

A fall in the value of the collateral can have great impact on the stability of the banks; in the case in which the borrower is not able to repay and the value of the collateral is deteriorated, there will be a loss for the bank with a magnitude not definable a priori.

The Credit Risk arises in the moment in which the borrowers are not able to secure the entire loan through the security standards; it substantially measures the amount of the position which is exposed to the risk of default (unsecured position). The main issue that emerges is the impossibility for the banks to secure all the loans they have in their balance sheets. Some authors argue that the phenomenon became even more intense with the financial liberalization and the worsening of the competitivity (Santonu, 2003). In order to limit the Credit Risk, the banks need to compute a Risk Analysis in which they estimate the attitude toward risk of their clients that has the drawback of being sometimes independent from the pecuniary position. The risk comes from the possibility of the actual rate of return on the investment to be lower than the expected return. Looking at the past performances it does not always conduct to significant and accurate results regarding the expected interest rate.

Banks are required to make some provisions against the Credit Risk for example incorporating the amount while computing the interest rate (Risk Premium) which can be computed on the basis of the bank's forecasts of number the future default which is uncertain. This uncertainty coming from this Risk leads banks to the necessity of follow some Credit Standards in order to avoid to lend to individuals that are more willing to default. Indeed, since the value of the banking assets can drop due to the fact that the borrower cannot repay back their loan, the Credit Risk can be reduced by screening the loan applicants and diversifying the loan portfolio by lending to borrowers that are subject to risks coming from different positions. The discrimination between borrowers it is not always easy; in the U.S. the banks decide if to issue the loan based on the FICO score which is an ex-ante observation (the project can always fail):

It is hence clear that the level of the Credit Risk that the banks undertake does not always reflect their preferences but depends also on many endogenous and exogenous factors including also the managers' behavior and the competitive structure discussed in the next paragraphs.

1.1.2 Managerial incentives

One of the most common cause of the banks' decline is the poor quality of the assets that can lead to an erosion of the banking capital. The causes that can lead to this condition can be both internal and external (Clarke, 1988); in this Section there will be presented an analysis of the internal elements.

The management has a fundamental role on the success of the bank performance; a poor management reflects into the balance sheet and into the income statement of the institution. The quality of the board of directors depends mainly on their experience, capability and integrity; the management is responsible for the strategies that should be carried on, for the stability and for the profitability. The long-term health of the institutions depends on the capability of the managers to be involved in the bank's affair.

One of the issues that can lead to bank instability is uniformed or inattentive board of directors or managers: many failed banks have the common characteristic of having directors that lack the necessary banking knowledges or are not well informed regarding the situation of the bank. These deficiencies can also come from the fact that the management does not inform rapidly the board of directors. The management has many important roles within the bank and it should be responsible of monitoring the operations and ensuring adequate internal controls. Also the board of directors has plenty of relevant roles to maintain stability of the bank: it should guarantee that the loan policies are correctly followed, avoid any inadequacy in the system to ensure the compliance with the policies and in the supervision of the key bank officers or departments. It frequently happens that failed banks don't have enough developed internal policies and control systems to guide the entire staff, this is reflected in an instable profitability.

There are situations in which the management and/or the board of directors are overly aggressive. This behavior is focused on an intense growth of the bank and it is not necessarily a weakness. It becomes an issue when the policies of the bank are not well established and there is a lack of the required control. Growth minded policies have to be sustainable if compared to the external environment in which the bank is working and to the capability of the management.

The inappropriate lending policies (liberal repayment terms or low credit standard), excessive loan growth with respective to the ability of management to supervise the sources of funding and strong reliance on the volatile liabilities and a poor choice of not sufficiently liquid assets as source of liquidity are some of the motives for which the banking failure can be considered strictly connected to the quality of the management and of the board of directors. In particular, if the bank does not require minimum standard to borrowers, the asset quality will decrease. Indeed, an over-lending (a high amount of loan if compared to the ability of repayment of the borrower) or an insufficient collateral required by the bank can cause a fast deterioration of the assets and this can lead to an instable situation for the bank.

Other elements that have to be taken into account regarding the managerial role in the bank failures are the insider abuse and the fraud. The first behavior arises when the managers/directors, for example, make inappropriate or unauthorized transactions. Both the actions lead to a higher probability of failure (Clarke, 1988).

It is important to mention that not only the management and the board of directors have a fundamental role while analyzing the reasons of failure. All the other employees have a big impact on the institutions; their honesty determine the success of a bank (Meyer, 1970). Unfortunately, this factor is not observable; indeed, an objective measure of honesty and commitment is difficult and almost impossible to find.

1.1.3 Banks concentration

The level of competition in a certain market depends on the number of players that are operating in it. The level of concentration of banks in a market can be measured by the HHI index (Herfindahl-Hirschman index).

Focusing on the specific case of the U.S., there can be found differences between two levels of concentration in the local market and on the regional market. The reduction of the banks from 1990s to 2010s did not increase the average concentration of the local banking market due to the antitrust policies and the DOJ regulations which denied the merger that would have led to too much power. At the same time, the regional concentrations during the same years increased.

Competition and banking failure are somehow related and they can be inversely or directly correlated on the basis of different variables that regard the regulatory policies, national institutions, ownership structure of the bank, macroeconomic and financial condition, capital regulation, economic growth, level of economic development, and so on.

The correlation between competition and banking failure have been largely studied but the different empirical analysis that have been conducted during these years still highlight how there is not possible to have a unique and general conclusion for all the cases observed and expected.

Some analysis highlighted how the concentration of the banks in one area increase the risk of bankruptcy; the reduction in competition results in higher deposit rates, the bank profits going up and in general less risk for the financial institution. A highly concentrated banking system creates fragility; the large banks can receive some subsidies (when they are considered too big to fail) which can create incentives in increasing the risk. Moreover, the possibility of monitoring a large number of small banks is easier with respect to monitor large banks that have more power and can have complicated situation and more factors to look at. Additionally, if a bank has a larger market power, it will charge for a higher interest rate which lead to a greater risk.

Tighter restrictions on the entry of new banks in the market and more severe regulations involve bank fragility; entry barriers bring a destabilizing effect on the banking system increasing the probability of a shock; the stricter regulation on the banking system lead banks to decrease the diversification outside the traditional business increasing the riskiness of their portfolios. Countries with institutions that promote the competition are less likely to suffer from the systemic banking crises.

Another point of view underlines that the risk of failure declines when the competition in the banking system decreases. If the deposit market is more concentrated, the banks can use their market power to increase the profitability; they are less interested in seeking outcomes with a low probability of success but with higher returns.

Banks that can have larger positioning in the market leading to less competition can have the advantages of reducing the transaction costs and hence, lowering the default and increasing the profitability of the bank.

With high competition the banks seem to relax their credit standards generally increasing the instability on the loans on their balance sheets. From the loan perspective, the rates will increase and the borrower profits going down (Boyd, 2005). In some countries a highly competitive banking sector creates a level of instability for which there is the need of more regulations.

A large competition among the banks lead to more difficulties while attracting both the borrowers and the depositors and require interest rates on the loans which will compensate the credit risk (Santonu, 2003), the banks try to retain the market share they have or to expand it. An increase in the competition gives the possibility to borrowers to be less dependent upon banks since it can be easier to get funds for the projects.

The excessive competition can lead to some undesirable outputs from a social point of view such as banking failures, runs and panic (Boyd, 2005).

Less concentrated banking systems (more competitive ones) with many banks is more prone to financial crises than a banking sector with few larger banks. Large banks can better diversify their portfolios so the banking system will be characterized by a few big banks but less fragile than systems with many smaller banks (Beck, 2003). Moreover, a concentrated banking system can boost the profits and reducing the fragility; indeed, higher profits work as a buffer against the shocks also reducing the incentives for banks to take excessive risks.

Lastly, some authors argued that a concentrated market allows to monitor in an easier way the banks; the corporate control in such situations are less complicated and are more effective (Beck, 2003).

The concentration can increase when the active banks decide to merge with or acquire other banks that are facing trouble. This is what happened in the U.S. during the 1980s and 1990s: the financial crisis led to a decrease in the number of the commercial bank present in the market; this number was lowered even more by the merger of non-failed banks and troubled banks, the same effect can be observed analyzing the number of banks in the market after 2010. The Federal Reserve during the 1990s prohibited the banks to obtain more than (Detragiache, 1998) of the total U.S. deposits and more than 30% of the single state's total deposits by acquiring other banks (Wheelock, 2011). Also the antitrust policy was preventing the merger between non-failed bank but it was nor regulating the acquisition of failed and neither putting any limits on the market concentration.

1.1.4 Comparing internal and external causes

Both the internal and external factors are relevant while analyzing the reasons behind the failure of the banks and/or the whole banking system. The banks can work in depressed conditions which usually come from the deterioration in the different sector of the economies.

It is important to highlight that there are some situations in which internal and external cause are interconnected. If the economy is affected by some shocks it is not always true that the banks will have a distress; the ability of the managers can help to minimize the situations in which the bank is affected by external environment. At the same time, with a poor management the institutions will be more involved in a collapse of some sectors of the economy. The economic conditions are rarely the primary factor in determining the banking condition (Clarke, 1988). If there is not a sufficiently competent management and necessary internal development, the banks benefit less from the improvement of the economy and suffer more from the external shocks. Some banks are hence more likely to fail than others.

On the other side it is also interesting to analyze which are the factors that help the banks to recover from a distress period. First of all, a change in the management (if the old one was not able to handle the situation) plays a crucial role. Other aspects are the improvement of banking factors, changes in banking philosophy, improvement of the assets held and of the capitalization but also the improvement of the external conditions is necessary in order to improve the stability and the profitability.

Focusing on the capitalization aspect, the capital works as a buffer between the operating losses and the insolvency; a higher level of capital allows banks to be more resistant to losses.

The banks need to put effort in order to change internally (knowledge, philosophy, strategy) to see some improvement and to rehabilitate. A strong economy helps the jobs of the management and promote the recovery but the management and the board of directors should act positively to boost it.

1.2 How to prevent banks to fail

Banks have an important strategical role mainly regarding the impact they have on many sectors. These institutions are more susceptible to contagion because: they tend to have low capital-to-assets ratio that doesn't allow for large losses, they have low cash-to-asset ratios which oblige banks to sell off earning assets if they want to meet the obligation with the depositors and they have high demand debt and short-term debt-to -total debt ratios that can lead banks to sell fast the assets in order to pay the depositors who decided to withdraw their funds (Kaufman, 1996).

The lending activity of the banks leads to the possibility of default risk coming from the event in which the borrowers cannot repay back the loan (principal and interest). Another exposure that the bank face is the funding one; these institutions have to borrow short-term in order to finance their assets (loans to customers). Moreover, as mentioned above, the systemic risk is another exposure that the banks face that is largely due to the lack of confidence from the market's participants coming from a disruption in liquidity and/or from the decline of the asset price.

To avoid or reduce all the exposures, there is the need of one or more regulators. In the U.S both the federal and the state governments regulate the financial markets. They work independently and their aim is to ensure the financial stability of the market. The main goals are the prevention of fraud and the maintenance of efficiency in the banking system.

Some of the federal regulatory bodies that have been established in U.S. to maintain efficiency in the banking system are: the Federal Reserve Board (also responsible for the implementation of the monetary policy), the Office of the Comptroller of the Currency (mainly supervising and regulating charters of banks in the U.S. and responsible for ensuring the efficiency of the banking system), the Federal Deposit Insurance Corporation (providing insurance for the funds of the depositors in the bank up to $250,000), the Office of Thrift Supervision, the Commodity Futures Trading Commission (mainly regulating the futures and options) and the Financial Industry Regulatory Authority.

As previously mentioned, the U.S. regulatory system is dual; both the federal and the state are in charge of overseeing the banking system. The most important authorities at State level are: the State Bank Regulators (have powers and goals comparable to the OCC but at state level), State Insurance Regulators (mainly protect the consumers regarding the insurance on their deposits) and State Security Regulators.

The U.S. federal banking agencies have the power to regulate both the banks and the bank holding companies which are under their jurisdiction. There is a hierarchy regarding the regulation in the U.S.; the most important are the Federal statutes and legislative mandates which authorize the agencies to set the minimum capital requirement and the capital adequacy standards, the regulations and reporting requirements that establish the capital adequacy rules and the policy statements, interpretations and supervisory guidance and manuals (Branson, 2014).

All these bodies should help the banking system to avoid fragility and be resistant to shocks and guarantee the fairness for the customers.

1.2.1 Analysis of the regulations

All the public policies regarding the banking system have to be focused on the stability of it. The regulators should prevent the fragility and consequently to reduce the failure rate. The systemic risk is an exposure that arises because the banks are interconnected and should be one of the main points of the banking regulations.

One of the most important regulations to which the banks in the U.S. are subject to is the Basel III (Basel I and Basel II in the past years) that requires the institutions to maintain some capital requirements.

Basel III is an international regulatory framework which consist in an agreement between the central banks and the bank supervisory authority in order to standardize the bank capital requirements (Bjorksten, 2014). This regulation, approved by the U.S. in 2010, defines the regulatory capital and increases the capital holding requirements for banks

Basel I was fully implemented in the U.S. in 1992 and the regulatory framework was including the following major requirements. **Tier 1** capital component composed by common shareholder's equity, disclosed reserves, retained earnings, and preferred stocks. The banks should have had at least the minimum required Tier 1 capital risk-weighted asset ratios (Tier 1 capital divided by bank assets weighted according to their likelihood of default). **Tier 2** capital includes set aside for expected loan losses and the allowances for loan and lease losses; these allowances are adjusted quarterly and the reserves for loan losses have to come from the earnings. When the allowances are higher than 1.25% of the total risk weighted assets, the excess is not counted as Tier 2 capital. Basel I also provided an obligation of conducting a stress test in order to define whether the bank is able to be not highly affected by losses coming from a recession or from systemic risk and if it is able to be still well capitalized after these events.

Basel II was implemented because some institution did not meet some standards especially regarding the asset risk weighting system. Basel II.5 was developed after the 2007-2009 financial crisis in order to create more rules to avoid the credit risk especially in the trading book of the bank (the securities that the banks will not hold until maturity and are accounted at current market value). Making the division between assets held until maturity and the ones which are sold before is not easy; Basel II.5 was focused on finding and also prevent an inappropriate placement of the securities that would give banks favorable accounting treatments but lead to miss capital buffer.

Basel III increases the amount of regulatory capital that the bank must hold. The rules also implement the provisions made by the Dodd-Frank Wall Street Reform and Consumer Protection Act (2010) that already regulated the capital reserve requirements for the banks.

Basel III is focused on the creation of a new regulatory framework for the bank risk-based capital requirements. The reforms that the U.S. regulators already did before the new Basel framework, highlight how they were significant since the banks were already complying with the Basel regulation. Basel III was introduced to mitigate the risk of the international banking sector through the requirement of adequate leverage ratios and reserve capital. A new requirement introduced by this new framework is about the cyclical changes; in period of credit expansion the banks have to keep aside more capital whilst during period of contraction of credit, the banks can relax the capital requirement. More leverage and liquidity requirements have been introduced in order to avoid excessive borrowing allowing banks to have the liquidity they need.

The U.S. agencies are making progresses regarding the introduction of and strengthening the requirements for banks. The U.S. requirements meets the Basel minimum standards with some deviations.

As mentioned in the previous sections, the interconnection between agents lead to the fragility of all the banks if one of them faces troubles and insolvency. The public policies have the power to mitigate both the likelihood and the severity of the systemic risk (Kaufman, 1996); the regulations should increase the macroeconomic stability and avoid the bubbles in the asset values. Some discretionary powers can be delegated to bank regulatory agencies in order to provide a safety net and to avoid unnecessary fire-sale losses from the asset sale by banks affected by runs on the banking system. The regulators should avoid the incentive incompatible policies and be focused on the reduction of the moral hazard behavior of the banks. The incentive to diminish this kind of behaviors the banks are asked to keep a sufficient level of capital and they are subject to sanctions in case they do not respect these requirements.

In order to prevent the runs on the banks, the government can decide to offer an insurance to the small depositors up to a certain amount. These explicit insurances are helpful since it allows the depositors to reduce the expenses they would bear if they had to control alone for the safety of their bank. The insure can also avoid the situations in which the depositors are more likely to run into currency and cause systemic risk.

In the U.S., all the lending institutions that accept the federally insured deposits are the insured depository institutions and have to comply with regulations regarding the safety and soundness regulations (Bjorksten, 2014) that mainly consist in holding an adequate level of capital. A bank is considered solvent if it has enough capital above the threshold set by the Basel III. Under these rules, the banks' assets have to grow proportionally if compared to the growth of capital.

Policies that lack of consistency or lead to wrong incentives for the banks can exacerbate the moral hazard and thus create more principal-agent problems. What the regulators should look at when deciding for new regulations are the systemic risk but also the non-systemic bank failures due to inadequate policies.

There is the possibility of reduction of the likelihood and the cost of banking failure without encountering strong moral hazard problems; this can be primarily done by the creation of effective system of structured early intervention and resolution (Kaufman, 1996).

1.2.2 Inconsistent regulations: what are the effects?

In absence of regulations or with less tight rules, the ability of the banks to maintain a minimum credit standard often erodes. As mentioned before, the minimum credit standard is needed in order to decrease the overall fragility of the banking system. A too high exposure to the credit risk increases the probability of default on loans causing a higher rate of default of banks.

Banking crises are more likely to occur in a liberalized financial system (Detragiache, 1998). The external environment has a great impact on the financial stability: if it is strong, the fragility of the banking system is less intense. The financial liberalization leads the countries to ease the interest rate ceilings, lower capital reserves and also the entry barriers. Moreover, the government interference in the decision of capital allocation are reduced. During 1980s and 1990s the financial liberalization in the led to more fragility in both developed and developing countries. The benefits that can come from the financial liberalization should be weighed against the cost of the fragility coming from these policies.

In countries in which the financial system is controlled in a tighter way, the banks have a ceiling on the interest rate they can charge and this allows to reduce the possibility for them to finance projects with a high level of risk. With the financial liberalization what commonly happen is that the interest rate ceiling is lifted up and thus, the riskier projects will be financed creating benefits for some customers that under other condition would have not received the financing. Hence, since the financial liberalization gives more freedom to the intermediaries, the regulations in these conditions should be more effective. Moreover, the liberalization leads to more competition (that, as discussed before, can lead to a higher failure rate) and often to lower profits creating a distortion on the level of risk that the bank should take.

There are also conditions in which the financial liberalization does create distortions that are fully compensated by the benefits that the costumers are receiving and where the governments can implement regulations that avoid the market failures.

It is important to notice that the effectiveness of the regulation is strictly connected to the environment in which the banks work. A consistent and adequate policy in one country can be totally a wrong fit for other banks. The regulators play a fundamental role in how a rule is implemented (Lucca, 2014).

As previously mentioned, in the U.S. the regulatory power is given to both the state and the federal authority that implement different rules and give distinct incentives. This dual system allows to keep under control what is happening both at a local level (state supervision) and at a national level (federal supervision). The conformation of the organization leads in some occasions to a competition that generate political interference giving the choice to the banks to choose the less strict regulation. Additionally, the dual system can cause a production of regulatory arbitrage that results in regulatory laxity and coordination problems as well. Generally, the federal regulator imposes more strict rules with respect to the state regulators. If the banking system is affected by a fragility it is necessary to analyze where it comes from. There can happen that the federal policies are too tight and impose costs which are not necessary or it can also happen that the state regulations are too lenient causing also a delay in the implementations of the corrective policies and having the effect of increasing the failure probability (Lucca, 2014).

Another important instrument that in the U.S. helps to failures given to the runs on the banks is the deposit insurance that has the drawback of working for banks as an incentive to internally take the failure risk possibly without constraints. This comes from the payoff structure in which the banks can make large gains leading the government to face large losses. This distortion can be reduced or eliminated by providing the shareholders a stake in the firm which is sufficiently high so they have incentives which are aligned with the ones of the depositors and the insurer. Moreover, a sudden withdraw cannot happen when there are high inflows of foreign capital and the domestic interest rate decreases while the foreign interest rate increases. In this way the domestic banking system can become illiquid if it was dependent from the foreign deposits. (Brinkmann, 1995)

The higher level of capital required by Basel III can reduce the insolvency risk which is costly for the FDIC but at the same time can imply more expensive bank credit for borrowers.

Moreover, bank capital reserves can have a restricted effectiveness in reducing the systemic risk since there is a large amount of lending which is outside the regulated banking system.

Some of the Basel III components are assessed as non-compliant; the two components that fall in this category are the securitization framework and the Standardized Approach for market risk. The first one is defined as "Materially non-compliant" (Branson, 2014) since there have been found divergencies between the U.S. rules and the standards required by Basel; indeed, the U.S. regulation lead to a lower securitization RWA outcomes if compared to the requirements of Basel III. Also the Standardized Measurement Method for Market Risk is not compliant since the U.S. rules implement only some of the required provision of Basel standards.

Concluding, the causes that can lead to a failure of the banking systems are numerous. The regulations in the U.S. that help to reduce them can be more or less efficient also depending on the macroeconomic factors. In the next chapter there will be analyzed the importance of the on- and off-site monitoring to check the soundness of the banks.

2. PREDICTION OF FAILURES: ON-SITE AND OFF-SITE EXAMINATIONS

The asymmetry of information which characterize the commercial banking world is driven by the impossibility for the market participants to have a clear view about what is happening to an institution in an accurate way. This leads to uncertainty and to the necessity of conveying more information even if it is costly (Flannery, 1999); it is important to mention that the most accurate source is available only to government which has the possibility of operating on-site examinations that are not accessible to private rating agencies, stock and bond investors and to depositors which are trying to detect the soundness of the institution.

Banks can be defined as opaque financial institutions (Berger, 1998) so the information that is acquired is private and costly for whoever is interested in the financial condition of the bank. Hence, the examination process should be directed towards this goal: gather information about the quality of the bank which otherwise would have been hidden and that can help investors and depositors to make the right decisions on the basis of their risk aversion. It also should help regulators whenever on-site examinations are not possible.

Although regulators can have an advantage in reaching confidential information, they can face issues in being the best monitors and this role can sometimes be taken by institutional investors and bond rating agencies. Indeed, the private sector can find more interesting information given the incentive the participants may have; in these cases, the private sector gathers more accurate information even without an examination process as formal as the one implemented by the government. Clearly, the information available to government and to private sector can be really different and if the examination process gives the regulators an advantage in terms of information, this ends up to favor the regulatory discipline whilst if the process helps the private sector, it will be end in favor of the market discipline.

This chapter is based on the importance of finding information about the wealth of the financial institution, and on whether the data acquired are good enough to be used to predict the failure of a bank. There are different methods used by market participants to find interesting news about the banks that quite always end up in an analysis of the financial ratios regarding different characteristics of the institution.

The Section 2.1 is focused on the ways in which the status of the bank can be analyzed; in particular, the paragraph 2.1.1 focuses on the strength and weakness points of the on-site examination, and the paragraph 2.1.2 instead, focuses on some particular off-site examinations. These off-site analyses can be accurate but it will be seen that there are some drawbacks associated with them too. Lastly, the ection 2.2 contains a comparison between on-site and off-site examination, discussing also the possibility of the on-site analysis to help consistently the off-site one.

2.1 How to analyze the status of a bank?

The analysis of the safety and soundness of the banks is necessary to understand if there will be failing institutions in the short, medium, and long term (whenever possible).

As previously mentioned, the banking system is permeated with hidden information and therefore, the participants in the market need to retrieve costly information with the instruments available for them.

It can be said that the main purpose for the bank examination is the information acquisition (Berger, 1998); there can be identified three types of information retrieved by the examinations:

1. Auditing information
2. Regulatory discipline information
3. Private information about bank conditions

A more accurate analysis of the types of information will be discussed in the last section, explaining the relationship between on- and off- site examinations.

When an analysis of the banks is done, it is necessary to understand if it have led to a good source of data or if the evaluation is not satisfying. The most important issue is to obtain sufficiently accurate information about the condition of the bank that affect in an important way the efficiency of the corporate governance (Berger, 2000).

In order to analyze how the status of a bank can be detected, an initial distinction can be done between the on- and off- site examinations.

The on-site examinations are exclusively conducted by the supervisors which have the right of analyzing all the details of an institution by going to its the headquarter and study several aspects which are otherwise not observable.

The off-site examinations are mainly based on the information disclosed by the institutions through the Call Reports and hence, on the financial ratios. These ratios can be a really sophisticated tool that can be used whenever an on-site examination is not possible.

2.1.1 On-site examinations: difficulties and effectiveness of CAMEL

The on-site examination is the approach with which the regulators can assess the probability of default of a banking institution. With this kind of method, the supervisor is able to detect difficulties of the individual banks in an accurate way. Indeed, the on-site supervision is a really effective and efficient tool through which the regulator can verify the current condition of the banks.

The bank supervision, as seen in the previous chapter, is a crucial issue in order to avoid losses connected to the failure of a bank or, at least, to reduce the costs incurred. In order for the government to check the soundness of the banks, in 1991 was created a set of rules under the name of Federal Deposit Insurance Corporation Improvement Act (FDICIA). This Act obliged the U.S. banks to submit an examination (either federal or state) every 12 months or every 18 months (for the small and well capitalized banks). The examinations mainly focus on some components that help to understand if the institution is safe and sound; the most common used variables regard 5 different characteristics of the banks: capital protection, asset quality, management competences, earning strength, and liquidity risk exposure.

This analysis is the so-called CAMEL which is considered the purest form of the on-site examination. In 1997 a sixth component was added to the analysis giving more importance also to the sensitivity to the market risk (CAMELS).

These two approaches work in the same way, they include ratings for each individual component that goes from 1 (best rating) to 5 (worst rating). The evaluations assigned to each category depend on the on-site evaluation and they take into account both quantitative and qualitative characteristics. Once the rating is assigned to all of the 5 components (6 in the CAMELS case), an overall index is then computed through a weighting system. The result of these analysis allows the regulator to decide whether is better to take an action, i.e. by tightening the supervision or by allowing for more liberty. An index which results to be less or equal than 2 is considered to detect high quality banks, whereas an index bigger or equal than 3 is assigned to institutions that are not properly safe and sound. The CAMELS index is considered one of the most accurate existent methods to study the financial situation of a bank but it has some drawbacks too that will be discussed later in this section.

In order to understand the CAMEL system, it is important to analyze what component ratings and composite ratings are.

Component ratings: the regulators assign a value between 1 and 5 to the single components (Todd, 1996) that they believe to be the most important to detect the different risks (Credit, Interest Rate, Liquidity, Transaction, Compliance, Reputation and Strategic). The components are outlined below.

Capital: the bank is required to hold a sufficient amount of capital in order to avoid to become not sound. When the regulators proceed to an on-site analysis, they take into account different aspect regarding the "capital" component. One of the first feature which is considered is the level of capital but also the quality of it; the regulators also look at the ability of the management to add more capital when it is needed. The capital level must be compliant with the regulatory requirements (Basel Committee requirements); the past experiences of the bank, the composition of the balance sheet and the growth plans are other elements which are taken into account. The adequacy of capital is measured also on the basis of the credit and market risk faced by the institution.

After all the evaluations of the variables that affect the capital are completed, a rating is assigned. A rating of 1 determines that the bank has a strong level of capital if compared to the risk exposure; a rating of 2 means that the institution is holding a satisfactory amount of capital. The ratings between 3 and 5 are the ones that should worry the regulators. Indeed, a rating of 3 means that an improvement has to be done since the amount of capital is less that what is considered satisfactory; a rating of 4 and 5 underlines a deficiency in the level of capital held by the bank which threatens the stability of the institution and which requires immediate actions of the shareholders and other sources which can be external.

Asset Quality: the examination should also highlight excessive levels of risk associated to a poor quality of the assets (loans, investment portfolios, real estate, etc.). The management has to be able to detect Non-Performing Loans and in general all the loans that will be not repaid in full that create a deterioration in the quality of the assets for the bank. The regulators evaluate different factors before giving a rating; they check the nonperforming assets, the adequacy in the level of the allowances for losses in loans and leases, the level of the diversification of the assets in general, the adequacy of the policies regarding the investments, but also the ability of the management to implement internal controls on the assets and to handle the information system.

After all the investigation is completed, the rating is selected; the ratings of 1 correspond to an asset quality considered strong that does not create concerns for the regulators. A rating of 2 is given to the institutions that have a quality of the assets which is considered satisfactory and with really small weaknesses. The ratings of 3, 4, and 5 lead the regulators to worry about the asset quality of the banks. In particular, the 4 and 5 ratings require immediate measures and additional control in order to avoid the deterioration of the value of the bank.

Management: is the third component analyzed in the on-site examination. The assessment of management capability is not an easy task but it is fundamental to create a stable financial institution. A good management rating is assigned whenever the managers are actively involved in the life of the institutions and make decisions about the adequate levels of risks (credit, market, reputation, strategic, legal, liquidity, and so on). The oversight by the board of directors and by the management are considered extremely important in order for the bank to be sound. Internal decisions should be taken with an appropriate frequency taking into account all the internal and external risks that the bank is facing and it is supposed to face in the future.
The management is also considered good whenever it supervises the adequacy of the internal controls; moreover, it should be focused on the promotion of policies which are effective and on the compliance to the regulations. Additionally, the managers have to demonstrate the willingness to maintain the bank safe and sound with their decisions.

The on-site examination helps to reduce the asymmetries of information regarding the incentives of the managers; after the analyses are done, the regulator assigns the rating. The rating 1 is assigned in the occasion in which the management and the board of directors almost perfectly control the risk exposures. A rating of 2 is given when there are some minor weaknesses that do not largely affect the safety and soundness of the bank; the management is considered to operate in a satisfactory way and it is able to identify and monitor the major exposures. The 3, 4, and 5 ratings are assigned in cases in which the management lacks the instruments to observe the risks and to create efficient policies to maintain the bank safe and sound. In the worst cases, the regulators may require that the either the management or the board of directors is substituted or strengthened.

Earnings: the fourth component of the CAMELS method regards the earnings. Specifically, when the on-site analysis is done, the regulators don't necessarily look only at the trend and the amount of earnings but also at the quality of them and, in particular, to the nature of the factors that allow to register the earnings. If the ability of making positive returns is driven by elements which have a high possibility to deteriorate, the rating assigned will be high even if the earnings are extremely positive. Unstable earnings can be reported in cases in which the credit risk is not well managed or cases in which the market risk creates uncertainty given by the instability of the interest rates. There are many components that affect the quality of the earnings, some of them are: favorable tax effects, wrong forecast of the future operating expenses, inefficient business plans, etc.

The rating of 1 means that the earnings are enough to support the continuity of the operation and are generated thanks to factors with a low level of risk that create stability for the bank. A 2 rating is given in cases of earnings considered enough to allow the continuity of operations and to satisfy the capital requirements and whenever there is also the presence of either static or declining earnings. A 3 rating indicates the presence of earnings that are generated by other factors that do not fully sustain the stability of the operations. Ratings of 4 and 5 are given whenever the earnings are not sufficient to maintain the continuity of the operations and to earnings which are subject to instable trends (or negative trends). The regulators are concerned of these rating since the institution could find in a situation in which also the capital will be eroded.

Liquidity: the level of liquidity should be enough to timely satisfy the needs of the bank and its financial obligations (i.e. bank runs issue). The analysis of the liquidity of the banks should take into account the unexpected events but at the same time, the management has to create a strategy for which the liquidity held is not going to create large costs. The banks have to hold in their balance sheets assets that are readily convertible in cash in case it was needed. It is important for the institution to have an easy access to the money market or to other sources of funding in case of necessity. What is taken into account while analyzing the liquidity level is also the reliance that the bank has on the short-term sources of funds which usually are volatile (i.e. borrowings) in order to fund the long-term assets. The liquidity in the banking system strongly depends on the amount of deposit held by the banks. Therefore, the trend and also the stability of the deposits will affect the rating given by the regulators.

The ratings assigned depend of the strengthen of the liquidity levels. A rating of 1 is given when the level of liquidity is considered strong and when the bank has easy access to the resources to get liquidity almost immediately. A rating of 2 indicates that the level of liquidity is considered satisfactory but some weaknesses have been detected regarding the ability of the management to handle the sources of liquidity. A rating of 3, 4, and 5 lead the regulator to take some measures since they may need help from external sources in order to be safe and sound and avoid liquidity shortage.

Sensitivity to Market Risk: it is included only in the CAMELS analysis and it helps the regulators to understand how the changes in components as the interest rate, the prices of commodity, the foreign exchange rates, and the equity prices can affect the banks' stability.

The elements that have to be taken into account during these analyses are mainly the sensitivity of the banks to the previously mentioned variables but also the ability management is to keep under control the risks coming from unexpected changes in the market components. The management has to be aware of how much the institution can be affected by a change in the overall market risk. The sensitivity to market risk can be generated by the nature of the interest rate risk exposure coming from a nontrading position or from the presence of foreign operations in the balance sheet.

The regulators give a rating also to this component. As for the other analyses, the rating of 1 is assigned when the market risk is under the accurate supervision of the management who can reduce the risks and who is able to minimize the potential reductions in the capital amount and the earnings. The rating of 2 is given when the sensitivity to market risk is not strong but there is a possibility that the amount of capital and the performance will be affected negatively by the increasing or unexpected market risk. The ratings 3, 4 and 5 are assigned whenever the institutions are sensitive to the market risk and the management actions are not satisfactory or even deficient; in these cases, the risk for the banks is to face a shrinkage in the capital held and a decline in the quality of earnings. In the worst cases, the regulators take action and can decide to make decisions on the management.

Composite rating: as mentioned above, the CAMELS analysis ends up in a rating which is called "composite rating". The number assigned to each institution depends on the ratings assigned to each component. The composite ratings depend on evaluation made on the basis of the managerial, operational, financial and compliance performance (Todd, 1996). The ratings go from 1 to 5; in particular:

- *Composite 1*: this rating is given to institutions that have the characteristic of being sound in each and every aspect. This means that the rating assigned to the CAMELS components are really good (usually of one or two), highlighting a bank which is safe and sound in all of its areas. The weaknesses that can arise after the analysis or that are found during the examination are considered minor and can be easily handled by the management. The institutions with a 1 composite rating are resistant to most of the external shocks (i.e. financial instability) and they comply with all the regulations. Form these banks it can be expected a strong performance and the ability of manage efficiently the risk given that the management is aware of the size of the institution, and the complexity.
- *Composite 2*: the banks that are fundamentally sound in every aspect but have some minor weaknesses receive a composite rating of 2. The institution is a candidate to

receive this rating if it does not receive a component rating worse than 3. As said before, the institutions have really few weaknesses that can easily solved by the management or by the board of directors. They comply with the laws in a satisfactory way and so, the regulators do not have concerns.

- *Composite 3*: the institutions to which a composite rating of 3 is assigned are considered not fully safe and sounds. The regulators have concerns about some components of these banks. The weaknesses that are found during the examination can be either moderate or severe. Usually, the worst component rating should not be worse than 4 to end up with a composite rating of 3. In these cases, the institution is characterized by a management which can be not fully prepared to timely reduce the shocks and to create strategies to reduce the overall risk. These institutions are hence more vulnerable if compared to the institutions that have a composite rating of 1 or 2. One issue connected to these banks is the noncompliance with the regulation that lead the regulators to be more focused on them and sometimes to operate more supervision. The banks with a composite rating of 3 rarely fail but they need to strengthen some characteristics.
- *Composite 4*: a composite rating of 4 is given to institutions that result to be not safe and sound from the examination. Their problems come from the way in which they are managed or from the financial situation in which they are. The bank is then considered to have severe criticalities and neither the management nor board of directors are able to handle them in a satisfactory way. These institutions usually do not comply with several regulations and they are not able to face the shocks and risks coming from the market. Indeed, the risk management strategies adopted usually do not fit the size and the complexity of the banks. The regulators often require for more regulations and enforcement actions since the lack of a plan could result in a failure or major weaknesses.
- *Composite 5*: the composite rating of 5 is the worst the institutions can get after an examination is ended. These banks are considered extremely unsound and unsafe and they have plenty of issues regarding the performance. They can receive this rating if the management is not able to keep under control the risk taken and it does not decide a strategy sufficiently coherent with both the size and the complexity of the bank. The problems that arise are too severe to be controlled by the management, hence there is the need of immediate financial help from external sources and of more tight supervision from the regulator. The regulators are aware that the institutions with a composite rating of 5 are willing to fail with a high probability, creating also risks for the deposit

insurance fund. The composite rating of 5 usually highlights banks that are willing to fail in the next 12 months.

The institutions that result to be weaker are subject to more frequent examination to reduce the probability of failure and to assure that the management will take a coherent level of risk with respect to the financial situation of the bank.

The rating which are given by the examiners after the on-site analysis lead the regulators to some choices: if a bank which was previously rated 1 or 2 and is downgraded to 3 in the last analysis, there will be concern for the supervisors which can decide to tighten the requirements, to change the policies, and to proceed with a more accurate examination to understand where the weaknesses come from. These institutions that are downgraded to a rating of 3 will be classified with the same index for a period that ranges from 6 months to few years before receiving an upgrade.

The downgrading to 4 and 5 ratings generates even more concern between regulators. These indexes detect institutions that have a high probability of failure and which are facing serious problems; the institutions which have been downgraded to 4 or 5 require timely actions and also an intensive monitoring from the regulators.

It is important to mention that the modern techniques to analyze the institutions can also start with an off-site examination: the periodic on-site analyses of safety and soundness begin with off-site "pre-exam" reviews of the data (Curry, 2003). The analysis proceeds with a more accurate on-site review that helps to check if the information was right and accurate. The on-site examinations help to acquire data that are not present on the reports used for the off-site examinations. Indeed, it allows the regulators to be well informed since the quarterly financial data can be not enough to understand if a bank is safe and sound. Some of these data can be the percentage of nonperforming loans compared to the total loans the bank has in its balance sheet, and the level of adequacy of the provisions for the loan losses expected. In the previous chapter it has been explained the difficulty to retrieve information about the ability of the management to govern a bank; the on-site examiners can have a more detailed view about the role of the management and they are able to give a rating to it.

Additionally, the CAMELS ratings are more accurate since they include not only the financial condition of the banks but also their compliance with the laws and the regulations and their internal policies. The CAMELS method has not been created to exclusively study the probability of failure of institutions but it results to be a good approach to do so.

Although on-site examination can be really useful to study in a detailed way the institutions in all of their aspects, this method also has some drawbacks. The most important shortcoming that has been found while considering the on-site examination is the lack of ability of the examiners to have a wide view of what it is happening during all the months between one analysis and another one. Indeed, the rating system allows to clearly study the bank in that precise moment but does not allow to include the changes that some variable will have in the case of changes of the economic and financial conditions.

The main issue that has to be cited is the speed with which the information content of the CAMELS rating decays (Cole, 1998). Since the financial conditions can be subject to rapid changes, the supervisory ratings given to an institution through the on-site examination can result to be not applicable for real long period of times. The FDICIA requires these examinations every 12 or 18 months. However, in many situations they are not enough accurate. Indeed, the most recent ratings can come from examinations that have been done days, moths or years earlier. For this reason, the CAMELS ratings which are based on exams that have been conducted close in time with respect to the exam period are expected to give a good forecast about the possibility of survive.

The possibility that the information retrieved from the supervisory examination decays, leads the regulators to look for other sources of data that can be less accurate than a proper on-site examination but more accurate in case of which the on-site examination becomes old and no more trustable. For this reason, in some cases the off-site surveillance can become a powerful tool in order to predict the failure of the banks.

2.1.2 Off-site examinations: are they viable?

The aim of the examinations conducted by the regulators is primarily to reduce the probability of failure at the minimum possible level by detecting the institutions that are facing critical situations. As seen before, the on-site surveillance systems may be the most accurate ones but in some cases the regulators decide to fill the gaps with the off-site systems. These methods are also used by the other market participants that don't have necessary power to discover all the information behind the rating given by an on-site exam.

Many monitoring systems have been developed in order to reach the highest possible level of accuracy. As mentioned before, the output of these examinations can also be used as starting point for an on-site examination.

Hereafter, some of the most known off-site systems that have been implemented to predict the failure of the financial institutions in general.

The Uniform Bank Surveillance Screen:

The financial ratios are one of the best ways through which the market participants and the government can analyze off-site the situation of the institutions. During the mid 1980s, the Fed system decided to adopt the Uniform Bank Surveillance Screen (UBSS) as the primary system of surveillance. This UBSS system was used by the supervisor until 1993 when it was replaced by another system.

UBSS is a method that uses the financial data retrieved from the compulsory reports in order to identify the institutions with deteriorated financial ratios with respect to a peer group, composed by institutions that have similar characteristics (Cole, 1995).

The UBSS takes into account 6 financial ratios that are computed using the quarterly financial data presented in the Call Reports. After all the ratios for every peer group are computed, they are rated from best to worst (for each group); the ranks are then summed up with each ranking obtaining the same weight. In this way, the resulting score helps to detect the institutions which need a higher level of analysis. Indeed, the banks with the highest composite ratings are then subject to a deeper off-site analysis. The reports that are created with the off-site surveillance can be used as analytical tools that include the effects of the decisions of the management, the economic condition of the bank, the performance, and also the balance sheet composition.

When this kind of off-site information is used by the regulator, it allows to take decisions about the capital adequacy, the asset quality, the earnings, the liquidity, and the management of the bank itself. All the data that are used for this off-site monitoring are in the form of ratios, percentages or dollar amount. The UBSS allows to understand also how an institution is ranked if compared to the peer it belongs to. This method allows whoever conducts the analysis to combine data retrieved from different time periods for both the single bank and the peer group averages leading to the possibility to evaluate the current situation of the institution analyzed but also to investigate the trend of the performances.

If the supervisors notice that the banks' conditions have worsened since the last on- or off-site analysis, they can decide to take actions to respond to this situation also asking to the banks for more information and explanations about the performance.

CAEL System:

During mid-1980s, the FDIC developed another surveillance system which is known as CAEL (Cole, 1995) which is similar to the UBSS system. It is an off-site examination is mainly based on the quarterly data collected from the Call Reports. The CAEL system rates give rating on a scale between 0.5 and 5.5.

The CAEL method allows to compute a rating which tries to imitate the CAMEL rating of an institution. Moreover, the CAEL scores are computed in a similar way if compared to the UBSS indexes; it is important to notice that the CAEL system is more complicated to implement with respect to the UBSS since it involves more financial ratios and more calculations.

The CAEL examination starts with the division of the banks analyzed into peer groups based, as for the UBSS, on the asset size. The rankings are then calculated for a group of 4 financial ratios which correspond to the 4 components of the CAEL. To each component is then assigned a rating (similar to the component rating of the CAMEL system) which is computed as the weighted average of the ratios previously computed.

The procedure is concluded with the computation of a composite rating which is the weighted average of the 4 CAEL rankings. All the weights and the financial ratios that are used in the analysis are determined at the beginning of the examination by bank examiners.

CAEL has been used for a long period as a benchmark for accuracy (Collier, 2003). The CAEL system helps mainly to rate 4 components: Capital, Asset quality, Earnings, and Liquidity; it does not produce any information regarding the management since it is not easy to find appropriate ratios to identify the quality. The weighting system of the CAEL is more complicated than the one used to implement the UBSS.

Limitation of UBSS and CAEL systems:

The UBSS and CAEL surveillance systems are really helpful in order to understand the situation of the banks from different points of view. The scores which result from the analyses can also allow the regulators to identify the financial condition of an institutions in periods in which is not possible to compute an on-site examination.

Despite these methods seem to work smoothly, there are some limitations that should be underlined. The ratios used to compute the component ratings are subjective, meaning that they are not able to study the institutions completely resulting in an incomplete analysis. Whoever computes these kind of off-site examinations selects the ratios from a really large set of variables that can be studied and which are all correlated to the financial performance of a bank. The ratios which are chosen may be not inclusive and so they do not produce an accurate assessment of risk (Cole, 1995); there can exist other ratios that result more accurate.

Another limitation of these methods regards the choice of the weights through which the final composite rating is computed. There are cases in which the weights are chosen without a statistical analysis; moreover, UBSS uses the same weight to each financial ratio even if some of them can be more valuable than others. At the same time, CAEL uses a fixed system of weights which is determined by a panel of bank examiners, this method can create issues in the case of shocks faced by the banks: the CAEL method does not allow for a timely adjustment of the weights.

Another drawback that can be identified is the dependence of the result on the initial peer group analysis. Both the methods considered have the common characteristic of dividing the banks into different per groups but they don't change if there are variations in the asset size or in the performances. Consequently, an increase of rating of a single institution could not affect in a coherent way the performances of the peer group to which it belongs to.

SCOR System:

In the late 1990s the FDIC introduced another powerful tool to examine the condition of the banks. This method is still widely used by the supervisors as a valid off-site tool. The system was created in order to help the FDIC to detect the banks experiencing financial distress from the previous on-site examination. One of the primarily aims of the SCOR model is to identify all the institutions that have been considered safe and sound on the last on-site examination but that could have been downgraded since that time (mainly 1 and 2 rated institutions that have the characteristic of a 3 rated bank). Hence, the aim is to understand with are the institutions that will receive ratings of 3 or 4 in the next on-site examination. A possible downgrade to these ratings creates concerns because these institutions require more supervision and even more expenses in terms of insurance on the deposits.

The accuracy with which the SCOR system can identify institutions that are willing to receive a downgrade strongly depends on the Type I and II errors. Type I errors are the so called false negative; they are due to the inability of the examiners to detect institutions that were financially deteriorating and would have been eventually downgraded. Type II errors are false positive: they represent the percentage of banks that were safe and sound but have been wrongly considered subject to deterioration.

A Type II error of 100% implies a Type I error of 0%; this means that all the banks are considered to be downgraded. It is clear that the supervisors need to resolve a trade-off between these two kinds of errors. Both of them create costs which are unnecessary; in particular, the Type I error slows down the reaction of the supervisors to a possible downgrade, delaying the supervision. The Type II error instead leads to supervision costs that are not needed.

If the SCOR system is chosen as an off-site analysis method by the FDIC, a case manager is assigned to every bank and if the bank is identified as deteriorated, the manager reviews all the information about that institution and decides if more actions are needed.

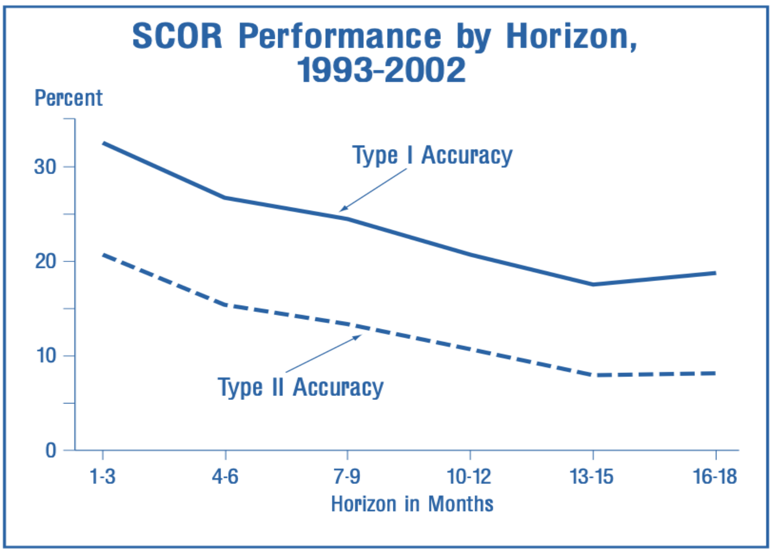
SCOR is a statistical method that compares the examination ratings with the financial ratios, it identifies which are the ratios that are more correlated the ratings given after the examination and that could help to forecast which will be the next on-site rating. It is important to notice that the SCOR model is also focused in detecting any changes in the relationship between ratios and ratings in order to be as much accurate as possible, also implying changes in the coefficients.

The SCOR rating is analyzed every quarter and is built in such a way for which the ratios and the rating can be taken for recent data; it also allows to eliminate the ratios whenever their correlation with the rating is not significant.

The aim of the model is to identify the probability for a bank to receive a specific rating and the probability of being downgraded. Only the "flagged" banks are then deeply analyzed by the case manager; if the flagged institution are too many to be analyzed, the flag is decreased. Under normal circumstances, the flag is assigned when the supervisors believe that the institution is going to be downgraded with a probability higher than 35%.

The accuracy of the model strongly depends on the time horizon. As showed in **Figure 1** the accuracy decreases when the time horizon increases. At the same time, the SCOR method still has a positive accuracy even for 16-18 months. Usually, the forecasts for long periods of time are done for safe and sound banks and not for the institutions that are likely to face deterioration in the short period.

Figure 1



Source: (Collier, 2003)

The SCOR model is entirely based on financial ratios, this implies that if the reasons for the downgrade are of financial nature, the model can be accurate, whilst if the reasons are of operational nature the model will not be highly accurate. This limitation of the SCOR model implies that it is a useful off-site tool but it cannot fully substitute the on-site examinations which can detect problems in the banks even before they affect the financial conditions of the institution. Moreover, the on-site analyses can reveal distort financial statements.

SCOR system was introduced to obtain ratings more understandable than what was obtained by the CAEL method. Indeed, SCOR examination helps to determine some areas that lead the bank to be weaker also using weights for every component that allows to identify the cause of the low ratings. The model produces ratings for all the 6 components analyzed in the CAMELS examination (with some issues regarding the managerial area) and it also computes an overall rating through the system of weights.

This particular system allows to understand which are the areas that create more problems to the institutions. The weight is used to make a comparison between a 2-rated benchmark institution and the bank analyzed. If the weight given is 10% it means that the component examined accounts for the 10% of the difference that can be identified with respect to the SCOR rating of the benchmark institution.

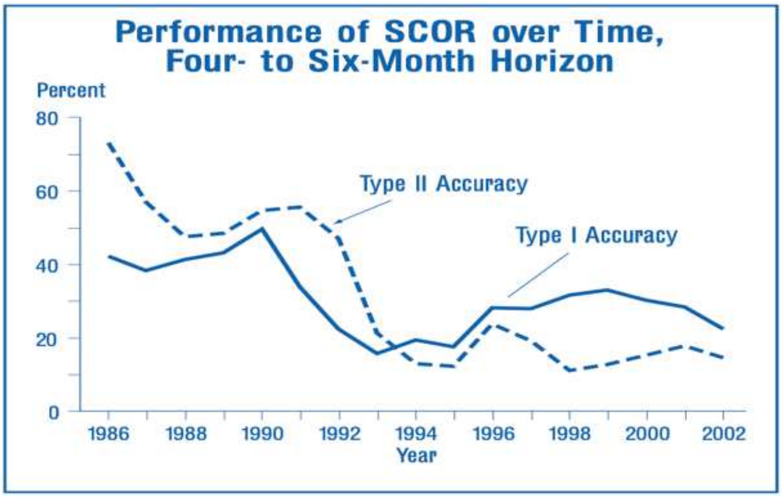
Limitations of the SCOR system:

As previously mentioned, the SCOR system has some issues in the computation of the management ratings. They cannot be associated to any financial ratio available to the off-site examiners. Indeed, several factors can only be studied with an on-site analysis since they are not disclosed. The examiners that conduct the SCOR analysis use ratios to understand the management condition of the bank, since they are the only available tools. For example, they look at the past due loans (30-89 days).

Other issues in correctly identifying a rating come from other factors that are not disclosed in the Call Reports. The Asset Quality can be an example: the rating for this component is given without really knowing the loans that are willing to default with a higher probability (default loans). The Capital rating is not fully accurate since it would require information which can be collected only with an on-site examination.

Additionally, as shown in **Figure 2** the SCOR system is not accurate when the economy is prosperous but it is a useful tool in periods in which there is a recession. In any case, the SCOR system performs better than the CAEL in both recession and expansion periods.

Figure 2



Source: (Collier, 2003)

SEER system:

In the 1990s another surveillance system was created to analyze the condition of the banking system: the SEER (or Financial Institutions Monitoring System, FIMS). As stated above, the on-site examinations are the most effective tools for this purpose but they are bot costly and difficult to fulfill (Gilbert, 2002). The off-site examinations discussed since now in this chapter use the financial ratios taken from the quarterly financial statements as the primary element to use in the analysis.

The SEER model was created in order to use the econometrics inference to analyze the financial ratios; the main objective is to reduce the error caused by the subjective choices (i.e. about weights to assign to each ratio). The econometric model is extremely helpful in order to identify banks which could be downgraded and it allows to allocate the resources of the regulator in a more efficient way. Nevertheless, the econometric models have a limitation: they can be modified only after the new risk has been detected and has produced a series of effects on the safety and soundness that can allow to specify again the model.

The Fed implemented two different off-site econometric tools that belong to the same system called SEER (System for Estimating Examination Ratings). These two models are:

- SEER *risk rank model*: it focuses on the risk of failure. It uses the financial data retrieved from the last Call Report to forecast which banks will default within a time period of two years. This model is used as off-site monitoring system because it helps to find the financial variables that predict in a consistent way the risk of default. It is a fixed coefficient model that allows to take into account the reasons of the changes over time of the financial performance of a bank. The SEER risk rank model is a probit regression that uses the information in the Call Repost to identify all the institutions that will fail in the next two years or institution that will have a tangible capital below the threshold of 2% with respect to the total assets. The model has the characteristic of being frequently updated to detect any important chances on the status of health of banks. The updates regard the financial ratios included in the analysis. The SEER risk rank model outputs either 0 or 1 which respectively indicate willing-to-fail and survival banks (Cole, 1995).

- SEER *rating model*: it focuses on the risk of receiving a high rating (3, 4 or 5). It also uses the last Call Report to create a "shadow CAMELS rating" (Gilbert, 2002) for every institution that will be examined on-site. An estimation of the CAMELS rating is computed looking at the financial statement available every three months. This method allows to study the relationship between the financial performance today and the condition of the institution in the future allowing any changes in the aspects that create lack of safety and soundness. The SEER rating model is a multinomial logit regression used to forecast which will be the future CAMELS rating given the information available on the Call Report. Hence, it tries to forecast both which will be the next rating but also which the rating would have been if the on-site examination was done on that date. In order to be as much coherent as possible, the SEER rating model uses constantly updated data: whenever a new report is disclosed, the new shadow CAMELS is computed. The model also takes strongly into account the results of the on-site examinations computed by the supervisors. Usually the SEER rating model is implemented using the data from the two most recent quarters in order to minimize the error (Cole, 1995). Since the model tries to estimate a CAMELS rating, it outputs a number between 1 and 5.

The financial ratios that are used in order to complete these econometric analyses are chosen on the basis of the most important ratios that affect the CAMELS model.

The main goal is to be as more accurate as possible avoiding to misclassify the institutions. The SEER system has been created in order to give to the supervisors the best possible off-site estimation tool. This method presents several advantages with respect to the ones proposed earlier. First of all, the model is really accurate while estimating the financial conditions of an institution and its probability of default. SEER also guarantees the possibility for the regulators to gain objective measures regarding the health of the banks (Cole, 1995). Another advantage of the SEER model regards the fact that the outputs obtained from the examinations are consistent and timely.

An important difference with respect to other models as the UBSS and CAEL that can be underlined is the ability of SCOR to include new variables whenever they are considered useful tools to examine the financial condition of the banks. The UBSS and CAEL systems use a set of fixed financial variables that can be difficultly changed.

To conclude, SEER offers a consistent, timely, accurate, and objective result that can be trusted by the regulators. It allows to use the resources in a more efficient way and to avoid with a higher probability the failure of the banks while it is not possible to compute an on-site examination.

2.2 Comparing on-site and off-site examinations

The asymmetries of information that characterize the banking sector is due to data that cannot be disclosed by the institutions since they are private and cannot be shared with the market participants. For this reason, depositors, rating agencies and investors interested in understanding the condition of a bank have to rely sometimes on less accurate analyses.

As seen before, the government can find more accurate information since it has the power to operate on-site examinations in which private information are analyzed.

On-site examinations can be computed only by the government which has the power of controlling the status of banks every 12 months or 18 months in case in which the banks are small and are well capitalized. This is a really useful tool for the regulators to detect those banks that need stricter rules since there are weaknesses in one or more areas evaluated. The possibility of seeing the institution so in depth helps the regulators to avoid the failure of the banks which show deficiencies. The most frequently used system is the so-called CAMELS; where the supervisors have the possibility of talking to managers, director and personnel.

The two most relevant examinations that can be done are: the risk focused supervision and the full scope examination. The first one is mainly focused on the level risk that the institution is facing and if that level is coherent with the size and the characteristics of the bank. In order to maintain the safety and soundness requirements, the financial institutions have to choose an exposure to risk which can be maintained for a long period avoiding the default of important weaknesses. The examiners identify all the risks (liquidity, credit, marker, etc.) and can decide to require stricter rules in order to lower the level of exposure of to maintain it under control.

The full scope examination is focused on the assessment of the adequacy of the management of risk but it also concentrates on the IT, Anti-Money Laundering and compliance with the requirements (FDIC, 2019).

The on-site examinations are really helpful to the regulators but they are not available to all the other participants which create their own off-site surveillance systems. Although these methods seem to be useful for whoever does not have access to the on-site examination documents, they are widely used also by the government. The reason is that in the months between two on-site examinations several issues can arise for an institution. The accuracy of the on-site examination in the exact moment in which it is computed is high but it lowers with time.

Off-site analyses can be more or less accurate depending on the choices made by the examiners; there are cases in which every variable chosen is subjective (financial ratios and weights) and this can lower the truthfulness of the information retrieved. The econometric models can be sometimes more difficult to carry on. The off-site examiners have different documents that can be useful such as: Call Reports, Wall Street Analyses, other Reports, etc. If the market participants can interpret the public information in an efficient way, it can also happen that the off-site examinations can complement the on-site monitoring system (Curry, 2003).

Other useful tools that can be used to improve the off-site analyses are the market signals; indeed, they can improve the prediction accuracy of these methods. These variables are: the return volatility and the trading volume. In particular, an increase in the return volatility happens when the institution is willing to default. The analysis computed by Cole in 2003 showed how the stock market variables are effective only for the distressed institutions that have obtained a rating of 4 or 5 on the last on-site examinations. The prediction value decreases for the wealthier banks.

In the recent years, an increasing attention has been given to modern machine learning techniques which, in some cases, can help the off-site examiners to predict the failures. These techniques will be discussed in the next chapter and will be the focus of the empirical analysis.

In the next paragraph it will be studied if the on-site analysis can be helpful to improve the off-site examination or if, instead, the off-site examinations do not use the information retrieved from the CAMELS analysis in order to be completed.

2.2.1 Can CAMEL method help the off-site examination?

On- and-off site examinations are focused on detecting the institutions with a weak structure that can cause instability of the overall banking system.

There are two points of view regarding the correlation between on- and off-site examinations: one clearly states that the on-site surveillance can implement the accuracy of the

information that are retrieved from the off-site analyses. The other one, contrarily, states that the data that the market participants get from the off-site examinations are not dependent on the CAMELS analyses.
The aim is to understand if the investors interpret the financial statements of the banks in a different way if the bank has been examined and so if the on-site examination affect the value of the banks.

The first point of view is supported by the fact that, even if the information contained in the on-site analysis is private and confidential, the on-site regulators can also decide to ask for the disclosure of more accounting documents that will be available to the public implying a more accurate off-site analysis. The market interprets the information in a different way when the banks was subject to a recent examination.

Another point of view states that the on-site surveillance system does not help the off-site examination. This is particularly true during periods in which there is an economic expansion when the banking industry is living a profitable period and is characterized by a good quality of the assets. In these cases, it also can be seen that the on-site downgrade models outperform the off-site surveillance systems (Gilbert, 2002).

After the examination of the ways in which the on-site information affect the off-site examinations; it is also needed to consider in which way the information can reach the market: there are cases in which the information reaches the participants in an illegal way and there are other cases in which the market just responds to data which are disclosed in public documents. As mentioned at the beginning of the chapter, the types of information that can be identified are three: the auditing, the regulatory and the private information.

- Auditing information can be really efficient for the bank to disclose useful information to the market but it is costly to produce (compliance costs). Moreover, other costs that are faced by the banks are the ones associated to the possibility that the regulator will ask the institution to change its behavior since the risk taken is increasing.
- Regulatory discipline information is another kind of information that can be transmitted to the market. It usually is the information that is disclosed when a bank has been downgraded and has to follow stricter regulations. In cases of upgrades the information disclosed is not useful for market participants.
- Private information can also be transmitted by the banks to the market in order to let the public to know the situation of the bank. This information can regard the ability of the borrowers to repay the loans or any other information that is not disclosed with the quarterly financial statements.

This information tends to affect more the off-site examinations whenever the institutions are downgraded to ratings of 3, 4 or 5.

Given the importance of the off-site examinations shown in this chapter, in the next chapter it will be discussed more in detail these kinds of analyses applying both statistical and machine learning methods, trying to find if there is evidence if one is a better predictor than the others.

3. MACHINE LEARNING AND STATISTICAL TECHNIQUES

Last chapter highlighted the importance for the regulators and other market participants to use off-site analyses and early warning systems in order to be aware of the health of the banks. In order to understand if there are methods which are more accurate than other, a comparison between the statistical and the machine learning techniques is needed.

It is fundamental to detect the banks that need tighter regulations in order to avoid failures. The FDIC reported that during the period 2008-2014 the number of banks that were declared failed was more than 500. The main problem is that the cost of default for each dollar of failed bank asset is high and it is forecasted to increase. If the bankruptcy rate of the financial institution increases, the costs for resolving the after-failure events. At the end of the 2013 an estimation made by the FDIC highlighted that the cost to the fund of deposit insurance in order to resolve the defaulted banks was higher than \$ 30 billion (Le, 2016).

During the period 2007-2010 the increase in the in the bank failures also led to an increase of the costs for the FDIC. In 2007 the cost was \$0.21 billion, in in 2008 the cost raised to \$ 19.86 million, and in 2010 the cost reached \$ 37.35 billion.

The creation of Early Warning Systems can be useful supervisory tool (Sinkey, 1975) since they can allow the banking system agencies to:

- allocate the resources in a more efficient way reducing the costs
- use in a more efficient the preexisting data
- provide some information in order to assess the deposit insurance premiums
- include some objective criteria in the valuation

The majority of the studies that are focused on the analysis of the bank performance and the prediction of the failure are based on the financial ratios as explanatory variables. The ratio analysis takes into account several indicators; the usefulness of these indicators must be tested with respect to the purpose of the prediction of banking failure (Beaver, 1966). In this chapter will be studied and compared only methods which involve the financial ratios since they showed to be the best predictors

The prediction of banking failure is an important topic since many years and the methods to detect the insolvent institutions are many. From 1960s the interest in this field increased and lead to some of the most relevant methods which are still used nowadays (with some small changes).

The cruciality of a bank failure leads the creditors, stockholders, auditors, etc. to find a way to analyze the future of the institution. The health of a bank depends on different drivers and on the environment in which it is working in; as studied by Ravi Kumar et al. (2007), the possibility for the institution to be competitive in the industry depends upon:

- Financial solvency;
- Possibility to create in an efficient way a stream of cash directly coming from the continuous operations;
- Ability to quickly have access to capital;
- Capability of maintaining the competitive position even in case of unexpected cash shortfalls.

In order to build a model that can be useful in different scenarios, the different techniques which are applied should take into account some aspects such as: the source through which the data are retrieved, the choice of different financial ratios (their correlation with the dependent variable and their significance level), the numbers of years prior to failure to consider in the study, etc.

Hence, in this chapter will be done a review of the most relevant literature concerning the prediction of banking failure analyzing both the statistical techniques and some of the machine learning techniques. The techniques that will be studied are all classification methods which are supervised. In particular, in the Section 3.1 the statistical methods will be presented whilst Section 3.2 focuses on the machine learning techniques.

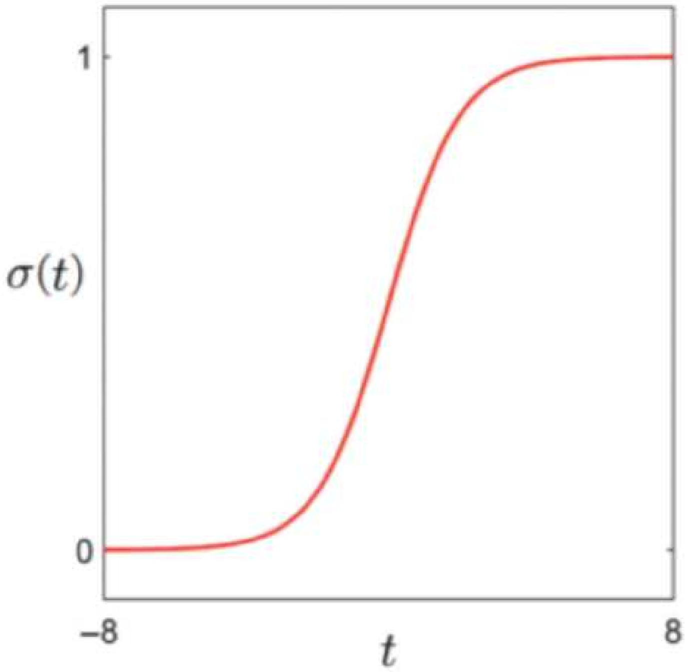
3.1 Statistical methods

In order to get accurate and useful off-site analyses, one of the most powerful instruments that has been widely used is the statistical tool. With the help of the financial statement information, the bankruptcy prediction models can be implemented; the choice of the financial ratios and/or other variables is extremely important during the whole process. In general, the two most important choices that has to be taken are: the technique to use and the explanatory variable to gather from financial statements or other sources.
Different statistical methods have been applied in order to predict the banking failures; some of them are listed hereafter.

3.1.1 Logistic regression (or logit):

Logistic regression is a classification algorithm which is able to produces a binary output. It is an adaptation of the linear regression model which is used when the output variable is 0 or 1 (binary). It is implemented when there is a need of a prediction of an event and the historical data are given. This method comes from an adjustment of the ordinary linear regression, the process consists in the creation of a linear combination of the variables used as inputs and then on the application of a function which maps the number between 0 and 1 which allows to use the logit as a classification method. The function which is used to map is the logistic function (a special type of sigmoid function).

Figure 3



Source: (Altman, 1968)

Where the sigmoid function is:

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

This method has been widely used in the banking sector as an off-site method to predict the banking failures. The majority of the studies that use the logistic approach (but also other approaches) gather information about failed banks from their financial statements and then compares these data with the ones collected for the healthy banks. The analyses can be done also 1-2 years prior to failure but, as it will be seen later and as it was stated in chapter 2, the accuracy of the models usually decreases with time. Another method that can be used to predict the failure of the banks is the creation of a priori probabilities which are independent from the historical values of the actual default. Additionally, other approaches to predict the failure consist in the analysis of the deviation of the individual bank's ratios from a peer group of similar institutions. The usefulness of the models used and implemented strongly depends on how well the data fit on the estimated ones but also on the ability of the model to be stable as time increases.

Martin (1977) understood that the banking system needed some statistical techniques that could have been applied to the financial statements of the banks. The early warning system had to be able to forecast the probability of default of the banks. He gave a definition of “bank failure” which is a situation in which the net worth becomes negative and/or the institution is no more able to continue its operation with large losses. The study completed by Martin (1977) was one of the first applications of logit model to the prediction of the bank failures, using as explanatory variables the financial ratios and other macroeconomic indicators. The author focused on the time period 1970-1976 and created a small subset of financial ratios (asset risk, liquidity, capital adequacy, and earnings). The classification accuracy coming from this analysis shows that the logit model implemented using these ratios is useful.

Ohlson (1980) used a dataset derived from 10-K statements from the period 1970-1976. The most important findings are 2:

- The possibility of identifying the factors which are statistically significant in order to predict the banking failure (dimension of the bank, measures of performance, measures of current liquidity, etc.)
- One of the biggest threats of the models is to overstate the predictive power, it is needed to test the ratios and the model.

The logistic regression has some advantages with respect to other methods that will be discussed later. Regarding the distribution of the predictors, the variance-covariance matrix doesn’t need to be the same for both the groups (failed and non-failed). The output coming from the logistic model is easily interpretable (0 or 1) and helps to avoid the misclassification.

Thomson (1991) used the logit model in order to predict the bank failures and reduce all the costs correlated to that. The correctly implemented EWS are extremely useful for regulators in order to take decisions that may result in really high costs for the FDIC and the taxpayers. The FDIC while controlling the banks have some troubles in identifing all the important aspects due to different causes (imperfect information, budget constraint, principal-agent conflicts, administrative constraints, etc.). The analysis showed that the most two important indicators are the solvency and liquidity of a bank which are able to predict the failure up to 30 months before the failure. Other elements that demonstrated to be relevant for this purpose are: the asset quality, earnings, and management. With this model the overall estimation error (computed as weighted sum of Type I and Type II error) ranges from 6.86% (when the failure happens in 6 to 12 months) to 18.56% (when failure happens in 36 to 42 months). The Type. I, Type II and overall error increase with time, demonstrating that accuracy decreases as time increases. The author applied the logistic regression technique also to test the out-of-sample data; also in this case the results were satisfactory and accurate.

Jagtiani et al. (2003) showed the importance of the capital adequacy in the banking system, underlying that the satisfaction of the capital requirements are fundamental for the health of a bank. In their analysis they hence focused on the capital adequacy, assigning 0 to the capital inadequate banks and 1 to the capital adequate ones. The institutions with a low capital adequacy are different from the other banks looking at their financial statements. The study is interesting since the authors compare two different early warning models: a simple logit model and a more complex model. The former takes into account only the capital ratio and the lagged change in that ratio; it showed an overall accuracy of prediction of 79.76% (with a Type I error of 21.05%). In particular, the model predicted in an accurate way 30 of the 38 capital inadequate banks. The latter, including more than 40 explanatory variables, predicted accurately the failure of 29 banks over 38 and resulted in a Type I error of 76.32%).
The result of this study showed that that Early Warning Systems effectively work even if they are simple; if they use the right predictors, the accuracy is going to be high. The authors also underline how the marginal cost of developing new ratios is not high if compared to the cost of a bankruptcy. Together with the EWS it is also needed to examine the data and to be aware of the two types of error produced by the model.

The logistic regression approach was also applied by Taha (2013), that showed how this method could be adopted to discriminate between two types of banking groups: health banks from the institutions which are in a difficult situation. His study included a set of account ratios taken from the individual financial statements of the banks chosen as targets. The ratios were selected to show the behaviors of the banks, and their macroeconomic activities; they were used as explanatory variables. In particular, the author finds ratios that can be associated to 5 groups (liquidity, management, activity, profitability, and vulnerability). The result of this study showed that the most important ratios to determine the failure of the banks are the bank profitability (a decrease of this indicator should alarm the regulators) and the ability to repay debt.

Zavgren (1985) applied the logistic regression model to a sample of failed and non-failed industrial firms. His study can also be applied to the banking system. The aim of the study was to implement a logit technique in order to verify how accurate it is when time from the failure event increases (she used a period of time of 5 years). The results of this research demonstrated that the information that can be retrieved from the model is significant also 5 years prior to failure.

Many of the studies that implemented the logit approach in order to predict the failure of the banks highlighted the importance of selecting the right database and the right ratios. The decision can be sometimes too subjective and lead to less accuracy.

3.1.2 Multiple Discriminant Analysis (or MDA)

Multiple Discriminant Analysis is a statistical technique that has been widely used for classification purposes. The observation is classified in one of the groups which has been identified a-priori. In order to implement this method, the first step that has to be done is the selection and creation of the groups, once this process is done, the data can be collected. The MDA tries to find a linear combination of the characteristics which best discriminate between groups (Altman, 1968).

The Multiple Discriminant Analysis technique has the power of considering the interactions between the characteristics taken into account. Furthermore, MDA allows to reduce the space dimensionality; if the analysis' goal is to discriminate between bankrupt and non-bankrupt institutions, the whole analysis will be transformed in one dimension.

The discriminant function has the form:

$$Z = v_1x_1 + v_2x_2 + \ldots + v_nx_n$$

$v_1, \ldots, v_n$ = Discriminant coefficients

$x_1, \ldots, x_n$ = Independent variables (ratios)

It transforms the variables into a discriminant score (Z) that is then used to classify the object. In particular, the greater the probability for a bank to default, the lower the discriminant score. MDA has the objective of detect all those variables that help to better discriminate between groups and which instead, are similar within the group.

Many studies applied the Multiple Discriminant Analysis to bankruptcy problems. One of the most relevant study has been conducted by Altman (1968), after the analysis conducted by Beaver (1966), who implemented the method to predict corporate bankruptcies. As seen in the previous chapter, the banks are different from other entities but some of the studies that have been conducted to predict the corporate failures are a good fit for banking system too. The author focused on the importance of the choice of the financial ratios for the analysis, using a multiple discriminant analysis to avoid the drawbacks coming from the univariate models. In particular, the univariate analysis can lead to erroneous interpretations and to confusion. The ratios have been selected through a procedure that included: the observation of the significance level of different indicators, the study of the intercorrelation between variables, the analysis of the prediction accuracy, and a final judgement made by the analyst. The a-priori groups need to be statistically different. The ratios that have been selected are:

- Working Capital / Total Assets
- Retained Earnings / Total Assets
- EBIT / Total Assets
- MV Equity / BV of Debt
- Sales / Total Assets

The final analysis underlined the high accuracy and predictive power of the Multiple Discriminant Analysis, showing that the predictive power is high up to two years prior the failure (after the second year, the accuracy rapidly decreases).

Sinkey (1975) focused on the time period 1969-1973; the banks have been chosen based on different characteristics in order to be part of the sample (location, total amount of deposit held, number of branches, and Fed membership). In his analysis he took into account several ratios but also some factors not so easy to define: the quality of management and the honesty of the employees. The ratios have been extracted from the year-end financial statements in such a way to cover some of the most relevant areas: liquidity, loan operations, asset composition, deposit composition, efficiency, profitability, and capital adequacy (Sinkey, 1975). Thanks to the power of the discriminant analysis, the group mean vectors and dispersion matrices are significantly different for all the years that have been considered. The Type I error (which consists in the wrong classification of a problem bank) decreased over time (46.36% in 1969 to 28.15% in 1972).

Tai (1986) focused on the question “Who is next?”; the ability of understanding which will be the future bank to fail, as mentioned, is the focus of many researches. Six financial ratios were used in the analysis to discriminate between fail and non-failed banks. The study showed that the Multiple Discriminant Analysis is an accurate model (90.91% of classification accuracy with an overall classification rate of 9.09% and a 18.18% Type II error).

Cox and Wang (2014) demonstrated how the Type II error can be reduced thanks to a Multiple Discriminant Analysis. The authors also discussed other types of discriminant analysis such as the Linear Discriminant Analysis and the Quadratic Discriminant Analysis.

LDA is a method that it is used also for classification purposes to separate two or more classes of events. It uses the Bayes Theorem in order to retrieve the class probability from the probabilities of predictors. The model requires some assumptions on the data: it requires that the independent variables are normally distributed and it also requires that in each class the features must have same var-covar matrix (but different means). As MDA, LDA also allows to reduce the space dimension.

LDA applies the Bayes Theorem to gather the class probability from predictors probabilities. The aim is to decide the class of the subjects given the values of the features. The decision rule for the Bayes classifier is to assign X=x to the class k for which the discriminant function is the largest (where the discriminant function is a statistical function that classifies the unknown observations and gives the probability of being classified in a certain class). LDA approximates the Bayes classifier using estimates for the mean, variance, and also for the probability. Here the discriminant function is given by the equation:

$$\tilde{\delta}_k(x) = x\frac{\tilde{\mu}_k}{\tilde{\sigma}^2} - \frac{\tilde{\mu}_k^2}{2\tilde{\sigma}^2} + log(\tilde{\pi}_k)$$

Where:

$$\tilde{\mu}_k = \frac{1}{n_k}\sum_{i:y_{i=k}} x_i$$

and

$$\tilde{\sigma}^2 = \frac{1}{n-K}\sum_{k=1}^{K}\sum_{i:y_{i=k}} (x_i - \tilde{\mu}_k)2$$

Where $\tilde{\pi}_{k\,=} {}^{n_k}/_{n}$ is the estimated probability that an observation belongs to the class k. Hence, the LDA classifiers uses the estimates and then assigns the observation X=x to the class for which the discriminant function $\tilde{\delta}_k(x)$ is the largest.

Quadratic Discriminant Analysis can be considered a generalization of the LDA model in which the discriminant function is quadratic; this method does not require for all the variances to be equal. The choice between LDA and QDA is mainly done on the basis of the bias-variance tradeoff: QDA needs to estimate more parameters, it has a low bias but a high variance.
The results obtained by Cox and Wang (2014) show that LDA is a better predictor for the banks which are willing to survive (overall accuracy which is in the range 70.69%-94.92%) and QDA better predicts the banks which are willing to fail (the prediction accuracy varies from 62.66% to 79.72%).

In order to construct an appropriate prediction model, a PCA can be conducted in order to understand which are the most important variables that capture the highest level of variance. In the case of the prediction of bank failures the aim is to detect the most relevant ratios that can explain the variation in the health of the banks. PCA (Principal Component Analysis) helps to clarify which is the underlying pattern of the relation between the explanatory variables which have been chosen.
Canbas et al. (2005) created an EWS including the use of PCA and finding three main components that explains the largest variance of the model. These components have then been used as independent variables in order to implement other models. A discriminant model was chosen ad predictor technique with the aim of maximizing the variance between groups and minimize the variance within the groups. The resulting model showed a high prediction accuracy and a low overall error.

As seen before, the MDA is helpful to reduce the space dimensionality, to derive a linear combination of the features that best discriminate the dependent variable, and to take into account the interactions between the relevant characteristics (Altman, 1968). Moreover, one of the biggest advantages of MDA is that it is capable of analyzing the important features in a simultaneous way instead of a sequentially way.

3.2 Machine Learning methods

The banking crises of late 1980s and early 1990s, the broadening of the regulations, and the needs of the customers lead the regulators to find a new prediction method or to improve the existing ones in order to avoid the economic instability. New prediction models have been studied and have been compared to the older approaches to understand if the prediction power could be increased.

Machine learning techniques have the advantage of both learning and adapting. They learn from the input data (specifically the ones used as training) and then improve their performance. This allows to reduce the costs and the risk associated with a wrong implemented model.
The previous section highlights that the prediction accuracy of the statistical models strongly depends on the initial decisions regarding many factors. In this section there will be analyzed if the machine learning methods outperform or underperform the other techniques.

Table 1

PROS	CONS
ML methods are able to fit the relationships which are not linear	ML techniques are often compared to Black Boxes since the middle layers cannot be interpreted.
Can result in a high accuracy if implemented with the appropriate set of variables	Can be too sensitive to the outliers and can lead to overfitting problems.

Source: (Leo, 2019)

All the techniques that will be cited can be more or less accurate than other methods; this does not only depend on the technique itself but especially on how it is implemented.

3.2.1 Artificial Neural Network (ANN)

The Artificial Neural Network technique is inspired by the biology and it is extremely helpful for the classification purposes but also for clustering and pattern recognition issues. It has been created following the procedure that the human central nervous system follows; it is a complex system of highly interconnected neurons.
The neuron can be described as a microprocessor that is able to receive and is able to associate the signals that come from other neurons. The Artificial Neural Network is composed by different elements that allow to process the inputs to transform them in outputs. The inputs and the outputs are connected thanks to some elements which are placed between them.
There is also a system of weights; every connection is multiplied by a weight (comparable to the synaptic strength on neural connection) (Bell, 1997).

The Neural Network technique is mainly based on two aspects:

1. Learning Process. It is an iterative process that allows to improve the weights and the procedures in order to get a better result.
2. Recall. Uses the weights which are obtained during the learning phase.

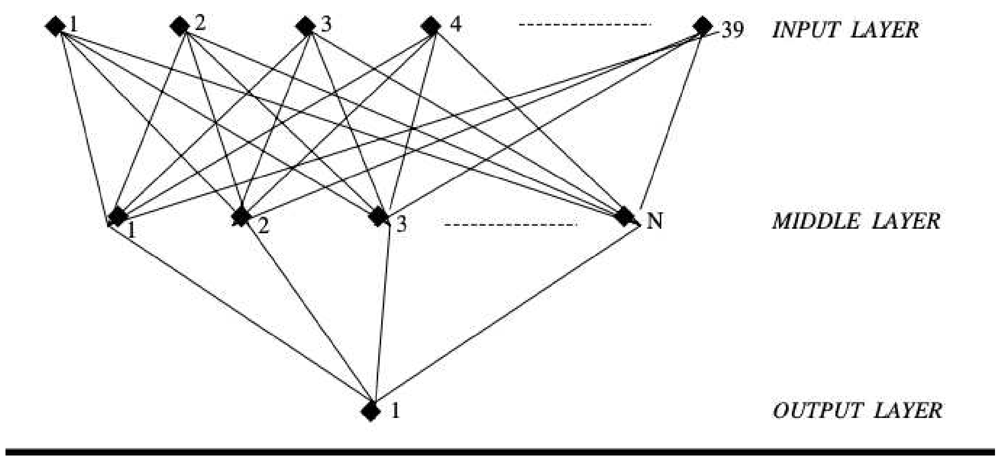
NN models is useful to map the relationships between variables which are really complex and between the inputs and outputs. Works to solve nonlinear problems due to its nonparametric nature.

Neural Networks don't make any assumptions regarding the distribution of the data and is useful even if the input data are noisy and/or incomplete. The possibility of adjusting the model once required by the environment which is changing.

This technique requires to define some elements to be implemented; the initial weights, the numbers of hidden layers, the learning rate (use to adjust the weights during the process which varies between 0 and 1), and the outlier factor coefficient have to be chosen.

A deeper description is needed for the learning rate; a small rate extends the time for the training of the model but allows to avoid the overreaction in the training process (Bell, 1997).

Figure 4



Source: (Bell, 1997)

The **Figure 4** shows the input layer that corresponds to each explanatory variable that has been chosen at the beginning (i.e. financial ratios selected for prediction purposes). The middle layers can be multiple (here only one hidden or middle layer is shown); the input layers is fully interconnected with the middle layer nodes. After the middle layers are created, a transfer

function is applied to show the output layer. Usually, the transfer (or activation) function for classification purposes is the sigmoid function but other functions that can be used are the hyperbolic tangent function or the multilayer perceptron.
The aim is to modify both the weights and the bias to getting closer to a more precise output. With backpropagation the algorithm will learn and weights are changed over and over to improve the output in particular the change in output is given by the equation:

$$\Delta output \approx \sum_j \frac{\partial output}{\partial w_j} \Delta w_j + \frac{\partial output}{\partial b} \Delta b$$

Changes in weights and bias will lead to small changes in output.

NN method also requires, as mentioned above, a training process. One of the most common training rules is the back-propagation algorithm which is based on the idea that the connection weights are responsible for the error of the output. With the backpropagation function the error is propagated back through the network the network and it is then used to adjust the connection weights to minimize the error, a gradient descent is commonly used; this allows to increase the efficiency. Another way to transfer the information is the Feed-forward that forwards the information from the input nodes to the output ones.

Swicegood and Clark (2001) applied to the same sample of regional banks both the MDA model and the NN in order to evaluate which model is more accurate. MDA correctly predicts 86.4% of the failed banks whilst the NN prediction power is 81.4%. Nevertheless, NN has a Type I error which is lower than the MDA. Indeed, NN shows a higher accuracy once a higher weight is given to Type I error.

Lee and Choi (2013) used the Back propagation Neural Network algorithm in order to indagate if it us a better predictor than the Multiple Discriminant Analysis. They consider as important features to predict the bank failures: profitability, growth, productivity, liquidity, and asset quality. The BNN outperforms the MDA since it can capture the nonlinear relationship between the independent variables.

Bell (1997) compared on the same sample a statistical and a machine learning technique: logit analysis and Artificial Neural Network have been used in order to understand which of the model have the highest prediction accuracy. The analysis revealed that none of the two methods dominates the other, leading the author to state that both ANN and Logit show a good predictive power in that specific context.

3.2.2 Support Vector Machines (SVM)

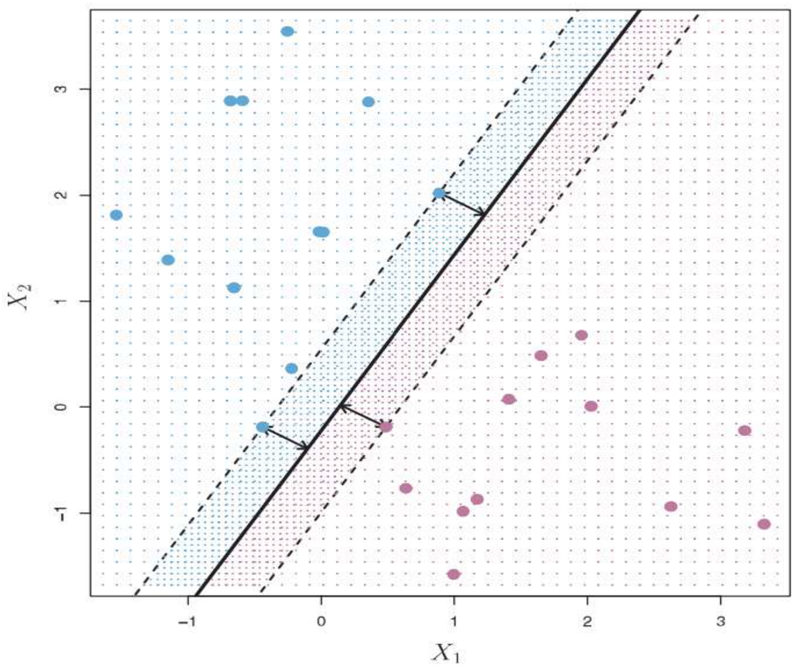
Support Vector Machines was developed by Vapnik in 1998 hence it is a relatively new technique if compared to others that gained popularity during the last years. It is based on the structural risk minimization principle leading to the possibility to reach the global optimum. Neural Network tends to have overfitting issues with respect to SVM.

Support Vector Machines is a supervised-learning technique that is used for classification purposes; it is helpful to divide the data into classes based on their position in the space. The space is divided by a hyperplane which is multidimensional (here 2 dimensions are considered). The distance between the hyperplane and the nearest point on each side of the plane is the margin.

The aim of SVM is to maximize the margins (also called support vectors), indeed bigger the margin, better the classification accuracy.

It is not required that the original space is linearly separable; kernel functions are helpful to transform the data into higher dimensional feature space.

Figure 5



Source: (James, 2013)

Figure 5 shows the hyperplane and the margins. The black thick line is the hyperplane which divides the plane into 2 (in our case failed and non-failed banks) and the dashed lines show the support vectors.
In this specific case in which the hyperplane is in 2 dimensions, it is defined by the simple equation:

$$f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0$$

Where $\beta_0, \beta_1, \beta_2$ are parameters.
One side of the plane is characterized by f(x)<0 and the other one f(x)>0. The magnitude indicates how far the observation is from the hyperplane; if f(x) is close to 0 the observation will be close to the plane causing more misclassification issues.

The figure also shows a grid composed by colored dotted points. This shows how the model is constructed; in particular in this case if an observation falls in the blue grid, the blue color will be assigned to it.
As mentioned above, the aim is to maximize the distance between the hyperplane and the support vectors. This results in a maximization problem shown by James et al. (2013):

$$\underset{\beta_0, \beta_1, \beta_2, \ldots, \beta_p, M}{\text{maximize}} \quad \text{M} \tag{1}$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 \tag{2}$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \geq M \tag{3}$$

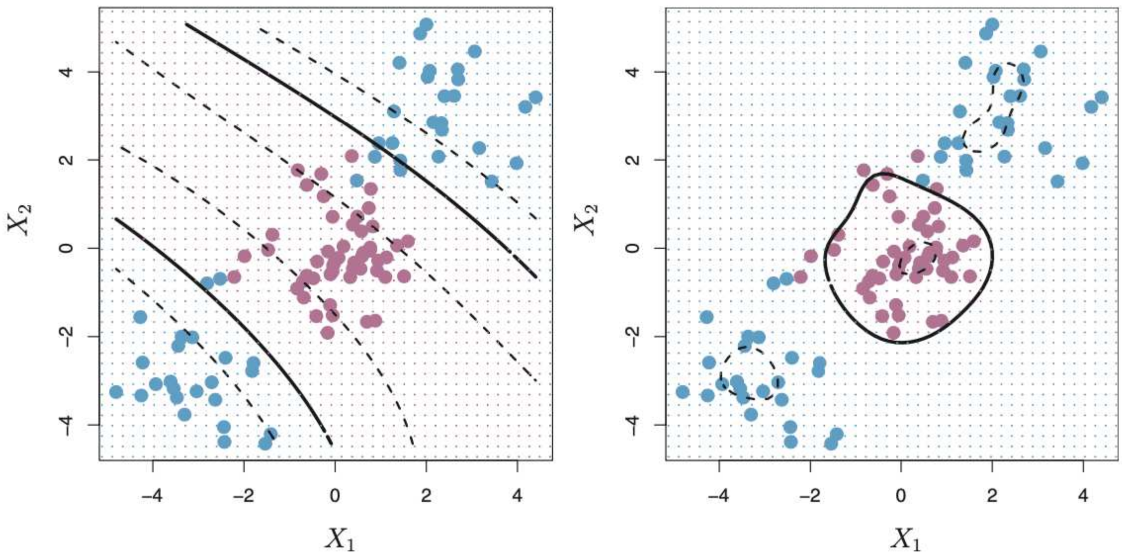
Equation (3) is the constraints which ensures that all the observations are classified on the correct side of the hyperplane given that M is positive. Also equation (2) ensures that the observations are at the correct distance (at least M) from the hyperplane.
The aim is to estimate the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ that maximize the margin of the hyperplane.
In cases in which the classes are not exactly separable the soft margins can be used and the maximal marginal classifier cannot be used; the support vector classifier has to be introduced which does not perform greatly when the data are non-linearly classified.

Support Vector Machines allows to convert a linear classifier into a classifier that allows to create non-linear decision boundaries automatically. One way to address for the non-linearity is the enlargement of the feature space introducing quadratic, cubic or higher order functions of the predictors.

As mentioned before, another way to take into account the nonlinearity is the enlargement of the feature space through the kernels. In particular, the classifier coming from the combination between a Support Vector Classifier and non-linear kernel, is called Support Vector Machine.

There exist different types of kernels; their primary goal is to understand how similar are two observations. Some of the most common kernels used for SVM purposes are the polynomial kernel (with different degrees), and the radial basis kernel.

Figure 6



Source: (James, 2013)

Figure 6 shows two examples of non-linearly separable data in which two different kernels have been applied. On the left example, a polynomial kernel with degree of 3 has been applied, whilst on the left the radial kernel has been chosen.

Min and Lee (2005) applied the SVM technique in order to predict the bank failures. They chose the radial basis function as kernel function selecting all the necessary parameters (kernel and penalty parameters). They also applied a fully connected BPNN to the same sample of data; the study showed that both the methods reveal to be accurate in order to predict the failing institutions.

Altinirmak, et al. (2016) defined the SVM as a good alternative for other techniques as the Neural Network, Random Forest, etc.

SVM shows a good performance when applied to real data; in particular, SVM outperforms NN showing a higher prediction accuracy and generating a reliable result.

Contrarily to the previous result, Ecer 2013 found that ANNs are more precise than SVM. In particular, the Type I and Type II error that result from the ANN method are smaller than the ones retrieved from SVM method.

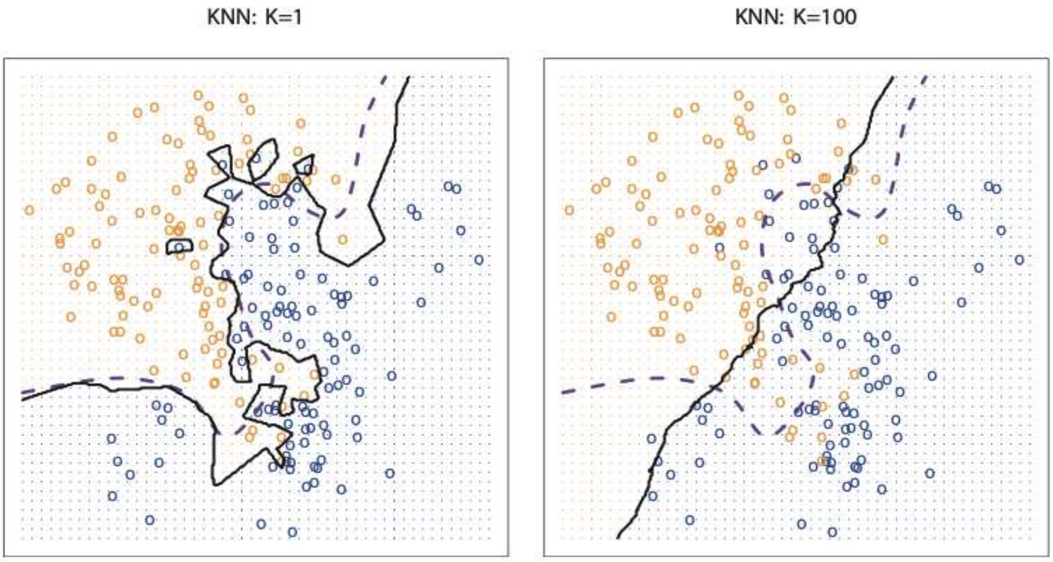
3.2.3 K-NN

K-NN (k-Nearest Neighbors) is a supervised non-parametric technique which focuses on the historical data and identifies a number (K, a positive integer) of similar features which are called nearest neighbors. The prediction process comes from averaging the historical outcomes.
It is instance-based meaning that it does not learn a model but memorizes the training elements which are then used for prediction purposes.
With K-NN the classified data are taken (represented by a vector of features); a new data needs to be classified given its position in the space. The K-NN techniques consists in the measurement of the distance (usually Euclidean distance is chosen) between the new data that needs to be classified and the data which are already classified. Classification process requires to assign a label to data on the basis of the "majority voting".
K-NN is a simple technique if compared to the ones previously mentioned but it is commonly used especially by the financial analyst since it is easy to implement and works even when the data are not linearly separable. K-NN is sensitive to outliers and it can cause over-fitting problems. This creates a low bias but a high variance in the model. The over-fitting problem can be addressed by choosing the "K":

- A small K leads to estimates which are more flexible but this opportunity will result in over-fitting issues and leading to an estimation with a high variance.
- A large K allows to have smoother decision boundaries and it creates estimates which are stable but not flexible (given the smoother boundary). The estimation will have less variance but higher bias (not accurate).

This trade-off is shown in **Figure 7**. The thick black line is the decision boundary; as mentioned above, the choice of a higher K leads to a smoother boundary but it is less flexible and involves a higher bias. Using flexible methods allows to decrease the training error but the test error could be high.

Figure 7



Source: (James, 2013)

K-NN has been used as a predictive tool for the bank failure field; it is an easy method that can show a high degree of accuracy.

Le and Viviani (2016) compared different methodologies. K-NN and ANN perform in a more effective way if compared to the statistical methods (Logit and LDA). The Early Warning Systems created using the Artificial Neural Network technique and the K-NN are able to detect and predict the most difficult cases.

3.2.4 Random Forest

Random Forest is a Machine Learning technique that can be used for either classification or regression purposes based on the so called "decision trees". Random Forest average the basic decision tree model allowing for a forecast which is more reliable.
A decision tree is composed by different elements which lend with a terminal node that contains the answer to the initial question on the basis of the characteristics of the variables that have been chosen.
Figure 8 shows the most important elements of a decision tree. The root node is the initial question that begins the analysis, it is usually a "Yes" or "No" question. The root node and the decision ones are called "internal nodes" that splits into some attributes, the terminal nodes are

the different answers that can be retrieved by following different paths and they are called "leaves".

Figure 8

Splitting

Source: author's elaboration

The decision trees are useful get some smaller set by splitting the subtrees. One of the most important decision that has to be taken while developing a decision tree consists in which feature to split; usually this choice is taken based on the entropy reduction (minimization of impure predictions) or on the information gain that come from the split.

One drawback of decision trees is the overfitting; in order to reduce this tendency, more decision trees can be aggregated. These trees are constructed in different ways and at the end the decisions are taking on the basis of the majority vote or on the weighted majority vote.

Random forest implicates the use of different initial samples in order to build different tree; as bagging technique, it uses bootstrapping (sampling with replacement) to create decision trees. Whenever a split in the tree is computed, a random sample of predictors (taken from the full set of predictors) is used as possible predictor. At each split, a new sub-sample is taken; the number of predictors that is taken into account is usually the square root of the total predictors. This decision is usually taken to avoid the effects of a strong predictors on the final result and also allows the trees to be uncorrelated to each other.

Random Forest technique have also been used as classification method to predict bank failures. Petropoulos et al. (2017) investigated the predictive power of Random Forest finding that this method has a higher predictive power if compared to Logit and Linear Discriminant Analysis but also if compared to Support Vector Machines and Neural Networks. The authors underline the positive aspects that characterize Random Forest: it is able to work with large datasets, it does not have correlation restrictions, efficiently handle the outliers.

The statistical and machine learning techniques that are useful for the prediction of bank and corporate failure. In the next chapter, some of these methodologies will be implemented in order to see which one has the highest prediction accuracy.

4. EMPIRICAL ANALYSIS

Finding a model that accurately predicts the failure of the banks is the aim of this analysis. The classification problem (default/no default) starts with the identification of the methodology that predicts in the most accurate way the banks' default based on the observation of different financial ratios, in particular comparing statistical and machine learning techniques.

The best model is then used to evaluate how the accuracy of the prediction varies with time; indeed, there will be analyzed data retrieved for 8 quarters prior to the failure.

Section 4.1 is focused on the description of the data: how the sample of banks has been chosen, what are the ratios utilized for the analysis, etc.

Section 4.2 is focused on the methodology: how multicollinearity has been addressed, how the most relevant variables have been effectively chosen, and how the models have been implemented.

The last part (Section 4.3) contains the results obtained from the empirical analysis.

4.1 Data

This section provides an overview of the data which have been used to implement the models. Paragraph 4.1.1 focuses on the selection of the sample of banks which has been selected to conduct the analysis. In paragraph 4.1.2 the list of the ratios initially chosen is described.

4.1.1 Selection of the Banks

The selection of the banks that are used to build the model is based on both the asset size and the geographic position (whenever possible).

A total of 50 large banks with the asset size bigger than \$ 500 million have been chosen for the empirical analysis; the time period analyzed goes from 2005 to 2015; the choice has been made to have a complete view about the situation of the banks pre and post financial crisis.

The list of the failed banks has been retrieved from the FDIC (Federal Deposit Insurance Corporation) database. FDIC is an independent agency founded by Congress in 1933, its main aim is to insure deposits (up to \$250,000 per depositor) and to supervise institutions in order to guarantee stability of the U.S. financial system.

Figure 9

Summary Statistics

Note: Dollars in millions; Adjusted for mergers

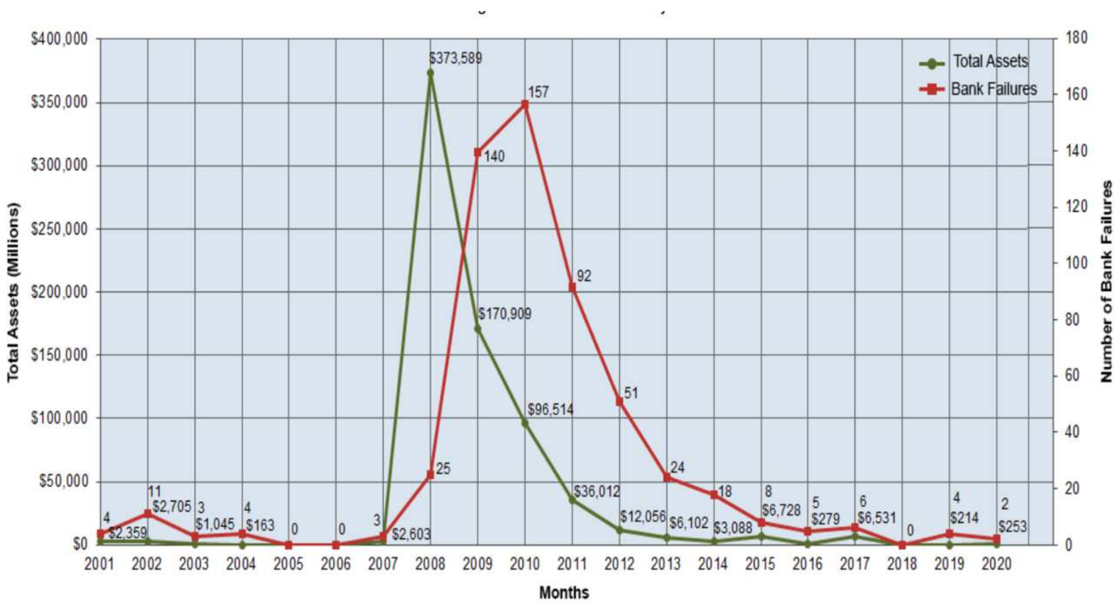
FDIC-Insured Institutions		
Number as of	7/16/2020	5,068
Assets as of	3/31/2020	$20,344,368
Deposits as of	3/31/2020	$15,824,395

Source: https://www.fdic.gov

Figure 9 shows the importance of the existence of FDIC; more than 5,000 institutions are insured (as of 7/16/2020) with value of deposits which is more than $15 million.

Between these 5,068 active institutions, 25 of them have been selected in order for them to match with the 25 failed banks previously chosen. The active banks have been selected on the basis of their asset size: looking at the asset size of the failed bank in the default quarter, the banks active have been matched accordingly. The data have been retrieved for 8 quarter (1st quarter in which the institutions defaulted and the previous 7 quarters) in order to obtain accurate data. This matching procedure has been improved (whenever possible) by selecting institutions with both the same asset size and the same location.

Figure 10



Source: https://www.fdic.gov

Figure 10 shows the number of FDIC insured banks that failed between 2000 and 2020 and the Total Assets in million dollars. One of the fundamental requirements that is needed for the analysis to be conducted is the regulatory framework: the institutions must be comply with the same requirements.

4.1.2 Selection of the ratios

After the selection of the banks that will be analyzed, it is important to choose the ratios on which the analysis will based. In the previous chapter the importance of the selection of the ratio has been explained: some financial ratios better identify the problematic banks than others. The different methodologies defined in Chapter 2 explain why the ratios are fundamental to detect banks that are not sound and which are the to be taken into account in order to get more accurate models.

For the purposes of this empirical analysis the studies conducted by Altman (1968), Le et al. (2018), Kumar Ravi et al. (2007), Beaver (1968), Sinkey (1975) have been analyzed to understand which ratios allows to predict the failure of the banks in a more accurate way.

The financial ratios can be grouped into 5 categories: Capital, Asset Quality, Management, Earnings, Liquidity (CAMEL) ratios. All the 19 ratios initially selected are listed hereafter:

- Equity/Total Assets (C): it indicates the quantity of the assets of the bank that has been generated through equity (hence owned by the shareholders). It can also be seen as the quantity of assets which is not held by the debtholders: the higher the ratio, the less the company is leveraged. Therefore, in case of liquidation of the bank, the ratio is useful to understand how much the shareholders are going to receive.

- Debt/Equity (C): it is useful in order to analyze the financial leverage of the banks. A high Debt/Equity ratio highlights that the bank is financing its operating side through debt. It is a useful tool to analyze the capital structure of the bank.

- Tier 1 Risk based capital ratio (C): Tier 1 Risk Based Capital Ratio is one fundamental indicator required by the Basel international Capital Standards after the 2008 financial crisis. This ratio is important to understand how the lack of capital was fundamental to absorb the losses and remain active and liquid. Many banks analyzed have a low Tier 1 Risk Based Capital Ratio, underlying the importance to have sufficient capital to avoid insolvency. The ratio is computed as Tier 1 Capital to Risk Weighted Assets. RWA

(Risk Weighted Assets) are the bank's assets to which a weight is assigned on the basis of their level of risk.

- Total Risk based capital ratio (C): it is computed as the sum of Tier 1 and Tier 2 Capital divided by RWA. Tier 1 Capital is composed by the shareholders' equity and the retained earnings whilst, the Tier 2 Capital is the "supplementary capital" (such as the reserves, etc.) and it is less secure than Tier 1 Capital. It is another instrument that allows to make a prediction with a higher rate of accuracy.

- Earning Assets/Total Assets (A): this ratio helps to understand which is the amount of assets which is helping the bank to generate earnings. The Earning Assets can be described as the ones which are able to generate wealth for the bank (including bonds, real estate properties, license, etc.). These earning assets are important because they allow the bank to get additional sources of income which does not come from the main operations.

- Loan and Lease Loss Allowance/Loans and Leases (A): the allowances for Loan and Lease Losses is a reserve that allows to take into account the possibility for the banks to face issues of bad debts or of collecting back the amount. The Allowance reduces the book values of the loans and leases in order for them to be consistent with the total amount that the bank is expecting to collect.

- Loan and Lease Loss Allowance/Noncurrent Loan and Leases (A): it is important to understand what is the percentage of allowance dedicated to the Noncurrent Loans and Leases: Loans and Leases 90 days past due or in nonaccrual status.

- Pretax ROA (A): is computed as the pretax net income divided by the average total assets. It allows to understand if the bank is able to allocate and manage the reserves resources.

- NOI/Total Assets (M): management ability is not always easy to quantify; some ratios can be used to evaluate it. One of them is the Net Operating Income to Total Assets ratio. NOI is useful to evaluate if the income generating investments are profitable.

- Efficiency Ratio (M): this ratio is computed as the noninterest expenses less amortization of intangible assets as percent of net interest income plus noninterest income. It allows to measure the amount of net operating revenues that are absorbed by the expenses. The lower the value of the ratio, the higher the ability of the bank to use its assets to generate more income.

- Cash Dividends/NI (M): is the amount of cash dividend paid to the shareholders as percentage of the net income. It shows how much the bank is returning to the shareholders as dividends compared to how much it could instead retain to add either cash or reserves or to reinvest. The ratio can also be useful to assess if the dividend policy is sustainable with respect to the financial situation of the bank.

- Retained Earnings/Average Equity (M): it is a measure of the quantity of retained earnings the bank is keeping if compared to the average equity. This ratio is important to understand how much the bank is keeping to reinvest and how much is instead paying out. A high ratio can be either a good or bad indicator: there are situations in which the bank retains more earnings than what is effectively needed.

- Noninterest income/Total Assets (E): Noninterest income can be defined as the income coming from activities that are not part of the core operations of the bank. It can be defined as "fee income" as the fees represent the majority of the noninterest income for a bank. When the interest rates are low, the banks may rely on noninterest income to stay active.

- NIM (E): it is a useful tool to analyze how much the bank is earning from the interests on the loans they made compared to the amount it has to pay as interest on deposits. It can be used as a reliable indicator on how much the bank is profitable and how much it is growing.

- ROA (E): it is one of the most common ratios used to evaluate the profitability of a bank. Indeed, it allows to evaluate how profitable a bank is compared to the total amount of assets owned. Thanks to this ratio, many evaluations on the bank can be conducted; it helps to understand if the institution is converting the money invested into income.

- ROE (E): a common measure of the bank's financial performance is the Return on Equity. It is computed by dividing the Net Income by the Equity; it can measure the ability of the management to use assets to create profits.

- Net Loans and Leases/Total Assets (L): Net Loans and Leases are computed as the Total Loans and Leases minus the Allowance for Loans and Leases Losses. This ratio shows what is the percentage of Loans and Leases issued by a bank with respect to the Total Assets. A high ratio might be a signal of scarce resources for the bank to survive to liquidity crises.

- Net Loans and Leases/Total Deposits (L): it is another ratio to check if the bank is currently liquid. A too high ratio might cause liquidity issues for the bank but a too low ratio might be a signal that the bank is not earning as much as it could effectively do.

- Domestic Deposits/Total Assets (L): it shows the percentage of the domestic deposits with respect to the Total Assets. Domestic Deposits are considered important because they are a cheap and reliable source of funds. A higher level of deposits helps the banks to be more stable and to be solvent.

All the data needed to compute the ratios have been retrieved from the Call Reports reported by the banks available on the FDIC database.

4.2 Methodology

In this section, the methodologies used are explained. As aforementioned, the aim is to implement different models (statistical and machine learning) in order to identify the most accurate. This model is subsequently used to evaluate how accuracy varies as time increases (8 quarters). The software used for all the computation is Python.

Significance of the Predictors

As seen in the previous section, the ratios which were considered at the beginning are 19. Once all the data have been retrieved, before the implementation of the models, it is important to identify which are the relevant variables for all the time periods taken into account.

In particular, t-tests have been conducted for this purpose for all the 8 quarters analyzed.

t-test compares the means of the two groups and allows to understand if the differences between them are statistically significant. A large t score indicates a wider difference between the groups. The hypotheses to be tested are:

H_0: There is no difference between the means of the two groups

H_1: The means between the two groups are significantly different

A 10% significance level has been used to conduct the test. The results for the first quarter before failure are reported in the **Table 2**:

Table 2

RATIO	t-test output
Debt/Equity ratio	Statistics=-12.155, p=0.000 The two groups are significantly different, reject H0
Equity/Total Assets ratio	Statistics=-10.951, p=0.000 The two groups are significantly different, reject H0
Tier 1 Risk Based Capital ratio	Statistics=-7.267, p=0.000 The two groups are significantly different, reject H0
Total Risk Based Capital ratio	Statistics=-7.287, p=0.000 The two groups are significantly different, reject H0
Earnings/Total Assets ratio	Statistics=-1.696, p=0.096 The two groups are significantly different, reject H0
Allowance for Loans and Leases Losses/Noncurrent Loans and Leases	Statistics=-3.967, p=0.000 The two groups are significantly different, reject H0
Allowance for Loans and Leases Losses/Loans and Leases	Statistics=5.388, p=0.000 The two groups are significantly different, reject H0

Pretax ROA	Statistics=-6.777, p=0.000 The two groups are significantly different, reject H0
NOI/Total Assets ratio	Statistics=-5.273, p=0.000 The two groups are significantly different, reject H0
Efficiency ratio	Statistics=2.378, p=0.021 The two groups are significantly different, reject H0
Cash Dividends/Net Income ratio	Statistics=0.991, p=0.327 The two groups are not statistically different, accept H0
Retained Earnings/Average Equity	Statistics=0.601, p=0.551 The two groups are not statistically different, accept H0
Noninterest Income/Total Assets	Statistics=-3.357, p=0.002 The two groups are significantly different, reject H0
Net Interest Margin	Statistics=-3.644, p=0.001 The two groups are significantly different, reject H0
ROA	Statistics=-6.931, p=0.000 The two groups are significantly different, reject H0
ROE	Statistics=0.634, p=0.529 The two groups are not statistically different, accept H0
Loans/Total Assets	Statistics=1.927, p=0.060 The two groups are significantly different, reject H0
Loans/Total Deposits	Statistics=0.263, p=0.793 The two groups are not statistically different, accept H0
Domestic Deposits/Total Assets	Statistics=2.117, p=0.039 The two groups are significantly different, reject H0

Source: author's elaboration

After analyzing the results of the t-test, some variables are dropped. These variables are: ROE, Retained Earnings/Average Equity, and Net Loans/Total Deposits; the other 15 ratios are used to implement the models.

Multicollinearity

Another element that may be an issue for the implementation of some methods is the multicollinearity which is a situation in which two or more predictors are highly (but not perfectly) correlated in a regression model. In fact, it can cause problems in the calculation of the predictors.

All the ratios that have been initially selected show, analyzing the literature, a great predictive power. It is important to understand if all of them can be used for the classification problem or if some deeper analysis has to be done.

Multicollinearity can cause big changes in the coefficients estimation when there are small changes in the data analyzed but it does not decrease the overall predictive power of the model implemented. In order to get coefficients' values which can be used for the analysis, multicollinearity has to be eliminated.

The correlation matrix (Appendix 4) computed for the statistically different variables shows that some variables are highly correlated between each other.

Some models (in particular Logit and LDA) are highly sensitive to multicollinearity.

Variance Inflation Factor measures the correlation of the explanatory variables in a regression model. In order to compute the VIF, a linear regression has been implemented since R^2 measure is needed. As previously mentioned, this is an analysis that will help to understand if multicollinearity exists for the set of predictors chosen but it is not going to impact the whole predictive power of the model.

The Variance Inflation Factor allows to detect the multicollinearity and it is an indicator of how much the variance of a predictor is inflated (in percentage). VIF is computed as:

$$VIF = \frac{1}{1 - R^2}$$

Where R^2 measures how close the data are to the fitted regression.

A high VIF value (more than 5) highlights a multicollinearity issue. **Figure 12** shows VIF values for the variables previously considered.

Figure 11

	variables	VIF
0	Equity/Total Assets	44.323275
1	Debt/Equity	46.643102
2	Tier 1 risk based capital ratio	832.911553
3	Total Risk based capital ratio	964.364921
4	Earnings/Total Assets	52.941739
5	loan and Lease Loss Allowance/Loans and Leases	7.701692
6	Loan and Lease Loss Allowance/Noncurrent Loss ...	2.938932
7	Pretax ROA	9.887941
8	NOI/Total Assets	5.359863
9	Efficiency Ratio	2.458779
10	Noninterest income/Total Assets	1.649260
11	NIM	14.472706
12	ROA	6.745480
13	Net Loans/Total Assets	45.664182
14	Domestic Deposits/Total Assets	59.474188

Source: author's elaboration

Principal Component Analysis (PCA)

Principal Component Analysis is a dimension reduction technique that helps to focus on the most important predictors.

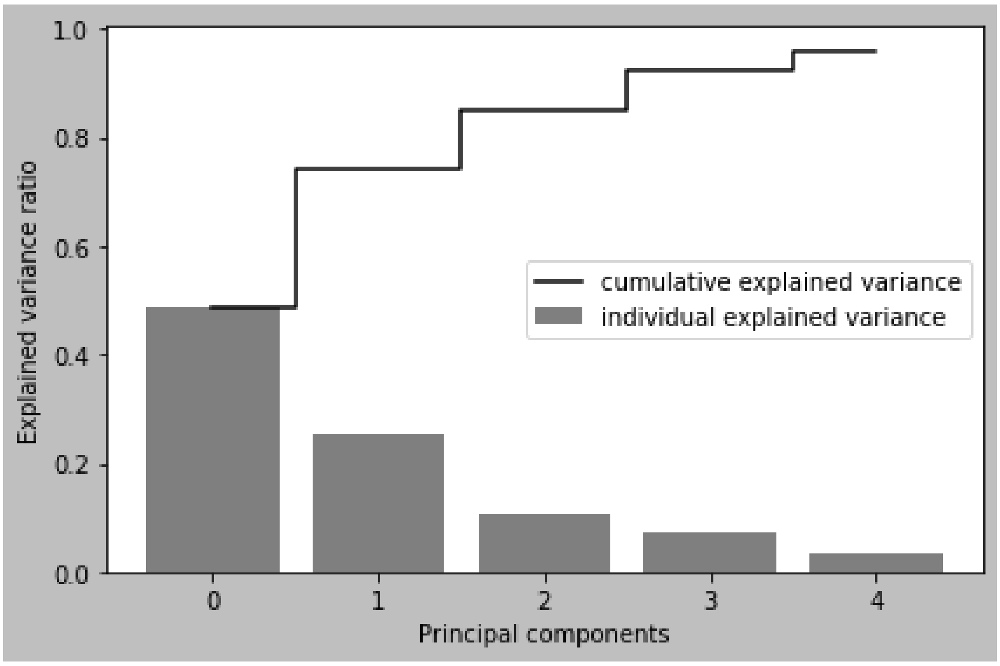
PCA starts with the computation of the correlation matrix and then the calculation of the eigenvalues and eigenvectors; the first principal component is given by the eigenvector associated with the largest eigenvalue (in general, the larger the eigenvalue, the higher is its contribution to the behavior of the data).

This analysis allows to perform a linear transformation of the original data: the first principal component (as a linear combination of original variables) explains the highest percentage of the variance. The important variables (as components) are extracted with the aim of explaining as much variance as possible, eliminating the problem of collinear variables. Indeed, all the correlated variables are included in the same Principal Component.

The following principal components are both uncorrelated between each other and from the first principal component and they capture the remaining variance of the model.

In addition to the creation of uncorrelated components, PCA is frequently used in order to improve the speed of the algorithm thanks to the dimensionality reduction.

Figure 12



Source: author's elaboration

For the purposes of the analysis, 5 Principal Components have been chosen. **Figure 13** graphically shows the individual and the cumulative percentage of variance explained by the components. In particular, the first Principal Component explains 48.68% of the variance, followed by the second PC (25.58%), the third (10.93%), the fourth (7.20%), and the fifth (3.36%). The total variance explained by the five Principal Components is 95.75%.

Analysis of the models

In order to analyze the different methods, the data for the defaulted banks and for the active banks have been randomly shuffled; the data are standardized and then the whole sample has been divided into training (25% of data) and test set (75% data). The training set is used to fit the model whilst the test set is used to evaluate how the model implemented fits on the training data. Python programming language has been used for the computations.

Different measures have been used to compare the prediction power of the models:

- Accuracy: Python's *accuracy_score* function allows to compute the accuracy of the model, an accuracy of 100% indicates that the model is able to fully predict the failures.

- Area Under the Receiver Operating Characteristic Curve (AUC): it is a common measure of prediction power of the models. ROC can be defined as a probability curve, whilst the AUC shows the performance of the classifier (quantifies the ability of the model of distinguish and separating between classes). A high AUC value indicates a better predictive power, in particular, an AUC value of 1 indicates a perfect prediction. The ROC curve is plotted as True Positive Rate (y-axis) against False Positive Rate (x-axis), it hence shows how the model is able to classify the input data.
- Gini coefficient: it can be a measure of the performance of a classifier. It gives an idea on the accuracy of the model. Its value ranges from 0 to 1: 0 is assigned to models which are not able to predict the outcome at all, whilst 1 is assigned to models which can perfectly predict the outcome.

$$Gini = 2(AUC) - 1$$

Additionally, a confusion matrix is computed to understand which are the Type I and Type II error.

4.3 Empirical Results

In this section, the results from the empirical analysis are presented. The confusion matrices and the accuracy levels will be shown for every implemented model; a comparison between the different methodologies will highlight which is the most accurate model (given the banks and ratio that have been chosen) that will be used to analyze how accuracy moves as the time from default increases (up to 8 quarters prior to the failure). All the methods that have been used are supervised. The problem presented (predict the defaulted and sound banks) is a classification issue; the output for all the methodologies is binary: 0 for banks that the model predicts to be defaulted and 1 for the sound banks.

4.3.1 Statistical Methods

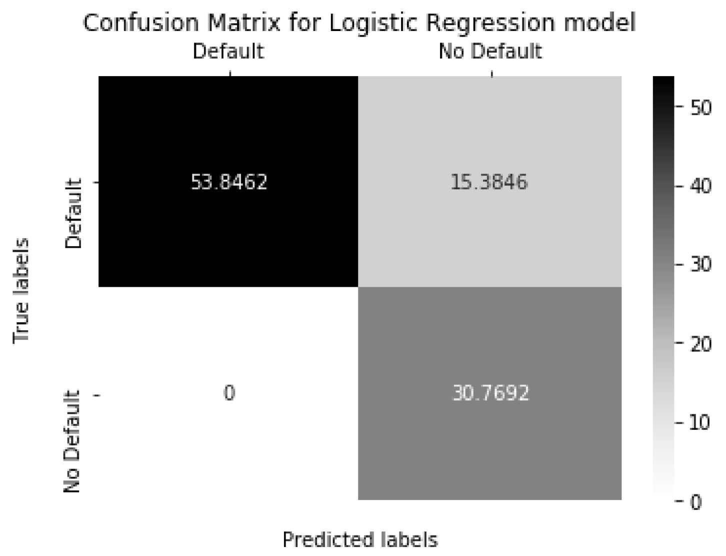
Logistic Regression

Logistic Regression (or logit) model can be used for classification purposes. It is a useful tool to forecast the probability of an event by looking at the data in the past. As previously mentioned, the output that this model gives is binary (0 or 1). The linear relationship between variables can be described as:

$$ln\frac{p_i}{1-p_i} = \sum_{j=1}^{n}\beta_j x_j + \beta_0$$

Where p_i is the probability of default and x_j defines the j[th] ratio.

Figure 13



Source: author's elaboration

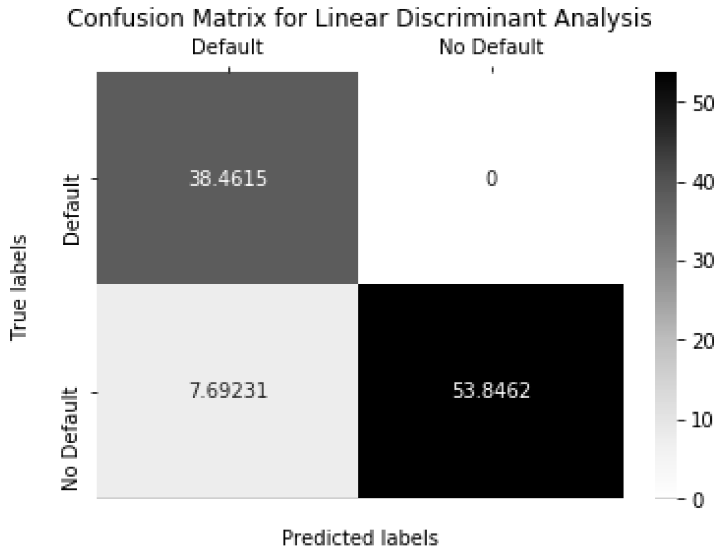
Figure 14 is the confusion matrix; the Type I error is 13.38% whilst the Type II error is 0%. The Type I error (predicting that the bank is sound when it is defaulting) is the most concerning one. The model shows an accuracy of 84.62% for the first quarter prior to failure, an AUC of 0.92 and a Gini Coefficient of 0.83.

Logit regression is a model with significant drawbacks: it requires linearity among predictors, normality and also independence. PCA has been used to address all these issues.

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) allows to derive a linear combination of the predictors that better help to predict the failed banks. As logistic regression, LDA has some requirements regarding the variables such as the linearity, the independence, and the normality. As for the previous analysis, PCA has been implemented before LDA to identify 5 independent components that allow to eliminate or reduce these problems.

Figure 14



Source: author's elaboration

LDA shows a Type I error of 0% whilst the Type II error is 7.7%. The accuracy of the model is 92.31%, AUC is 0.93 and Gini Coefficient is 0.85.

4.3.2 Machine Learning Methods

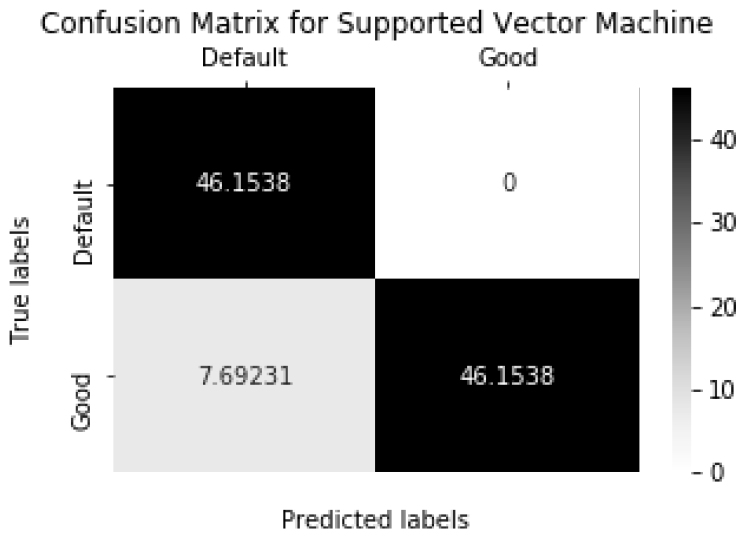
Supervised Machine learning models are popular techniques that are widely used for classification problems. These tools are able to learn from the experience in order to be able to improve the future performance in similar situations. The machine is able to learn if the predictive performance improves as the experience increases.

In this section, 4 Machine Learning Techniques will be analyzed in order to compare their performance with the statistical techniques previously implemented.

Support Vector Machines

Support Vector Machines (SVM) model is a powerful tool that allows to construct a decision surface to transform complex problems in simpler ones. A kernel has been applied (sigmoid function) to transform the nonlinear data.

Figure 15



Source: author's elaboration

Type I error is 0% whilst Type II error is 7.69%. The accuracy of the model is 92.31%, the AUC is 0.93 and Gini Coefficient is 0.86.

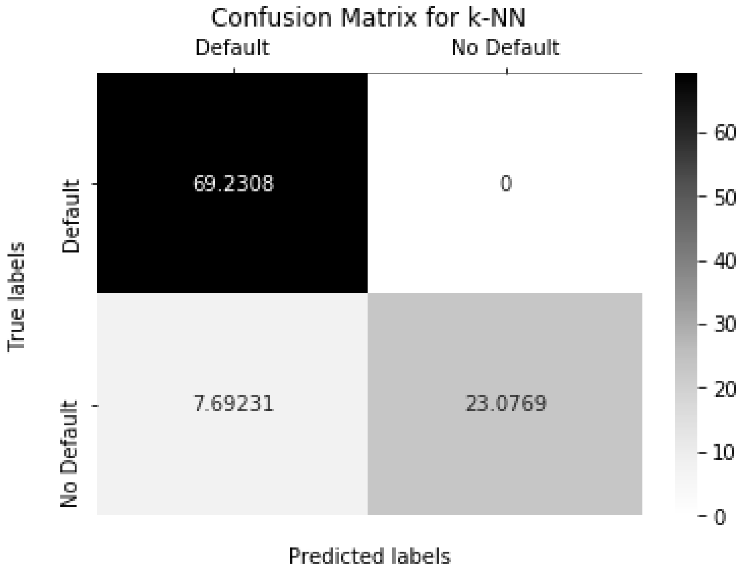
k-Nearest Neighbor

k-Nearest Neighbor (k-NN) assigns the new observation to the majority of the class among the k- closest data of the training set. For each new observation the data are updated. The distance between the observations is computed as Euclidean distance computed as follows:

$$dist(a,b) = \sqrt{(a_1 - b_1)^2 + \ldots + (a_n + b_n)^2}$$

Where a and b are two vectors of n features (which correspond to two banks).
In order to be implemented, the model requires to identify the k parameter. Considering the literature, the hypothesis behind the choice are numerous. For the purposes of this analysis a k of 4 was chosen given by the square root of the number of parameters taken into account.

Figure 16



Source: author's elaboration

k-NN shows Type I and Type II error equal to the SVM model, nevertheless the probability of predict correctly the defaulted institutions is higher for the k-NN model.
The accuracy of the model is 92%, the AUC is 0.87, and the Gini Coefficient is 0.75.

Random Forest

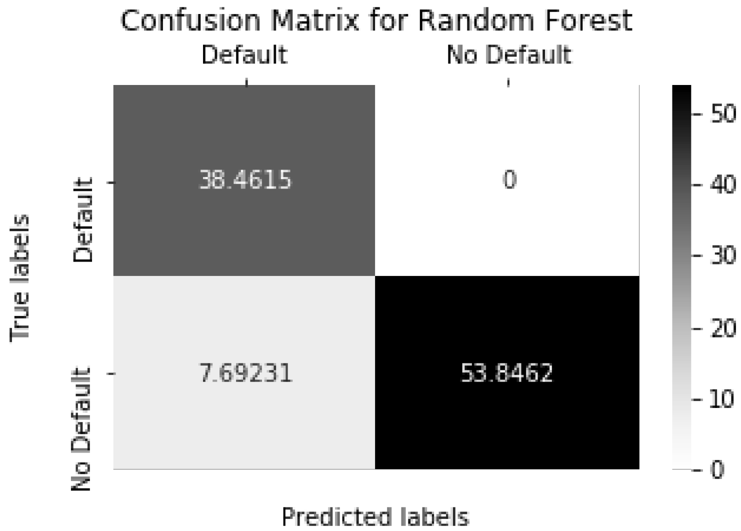
Random Forest (RF) is composed by uncorrelated decision trees. The lack of correlation is one of the strengthen point of the model. The wrong predictions of some trees are compensated by the good prediction of others. Random Forest is efficient while dealing with large databases and with a large number of predictors; it is also useful to reduce the overfitting and the variance in the prediction without increasing the bias.
Random Forest is one of the most common used algorithms since it is simple to implement and to visualize.
While implementing Random Forest some characteristics have to be defined: the number of trees in the forest has been set at 90 and the Gini index has been chosen to measure the impurities in the nodes. The Gini Impurity of the node is defined as the probability that the random sample in the node is misclassified and it is computed as:

$$Gini = 1 - \sum_{i=1}^{j} (p_i)^2$$

Where p_i is the probability of the outcome i.

The power of RF over other models (i.e. boosting and bagging) consists in the way in which every tree is created allowing the individual decision trees to be uncorrelated between each other. In particular, whenever a split has to be done in a tree, a subset of random predictors is chosen. The number of predictors chosen for each node is constant. For classification problems usually the random number of predictors m is usually equal to p/3 where p is the number of the initial predictors.
Additionally, to build the trees, a random sample of data is chosen. These samples are drawn with replacement (bootstrapping) meaning that the same sample can be used different time in the same tree. This allows to have multiple decision trees trained with different samples and also to both have a low bias and variance.
In this analysis the maximum depth of the trees has not been chosen, meaning that the nodes are expanded until the leaves are pure. The final predictions are mad by averaging the predictions retrieved from the 90 individual trees.

Figure 17



Source: author's elaboration

Confusion matrix shows low Type I and Type II error, the model better predicts the defaulted institutions. The accuracy of the model is 92.3%, the AUC is 0.94 and the Gini Coefficient is 0.88.

Backpropagation Neural Network

Backpropagation Neural Network (B-NN) is a supervised machine learning technique inspired to the human system of neurons. The B-NN in particular refines the error by sending it back to the previous layers and learning from this mechanism. In order to implementing a B-NN algorithm, the network has to be initialized and the weights and bias that are associated to the initial layer are decided. The neurons have to be activated through the input data, the weights and the bias previously selected. In order to get a first output from the algorithm, an activation function (sigmoid in this case) is used; the information is then forward propagated. The error for each output (0 or 1) neuron is then computed and backpropagated, the weights are updated and the network is trained. Once the algorithm is trained, the prediction can be done.

The prediction accuracy for the sample that has been chosen is 92%.

Random Forest shows to have the highest prediction accuracy if compared to the other model that have been implemented.

Logistic Regression is the model which shows the lowest Type II error (0%). This means that the model correctly classifies the banks which are not defaulting. Indeed, a positive Type I error highlights that the model classifies some banks which are defaulted as sound banks; this type of error is costly. The other models analyzed have a higher Type II error but a really low Type I error (the most concerning one). This is the reason why, even if logit shows the lowest Type II error, it is not used to conduct further analyses.

4.3.3 Changes of accuracy over time

As it has been previously shown, Random Forest is the technique which has the highest accuracy for the prediction of the bank failures. The aim of this section is to apply the RF technique to the 8 quarters prior to the failure. As previously mentioned, the t-test for all the quarter was computed to define the predictors that are going to be used; all the prediction variables are standardized, the correlation matrices have been computed. The accuracy measures previously mentioned have been used to evaluate the model.

For the **2^nd^ quarter prior to failure** the Type II error is higher if compared to the first quarter (15%) whilst the Type I error is 0%. The accuracy is 84.62%, the AUC is 0.89 and the Gini Coefficient is 0.78.

For the **3rd quarter prior to failure** the Type I error is 7.7% and Type II error is 15.39%. The accuracy is 76.91%, the AUC is 0.78 and the Gini Coefficient is 0.55.

For the **4th quarter prior to failure** the Type I error is 0% whilst the Type II error increased to 23.01%. The accuracy is 76.92% AUC is 0.75 and the Gini Coefficient is 0.50.

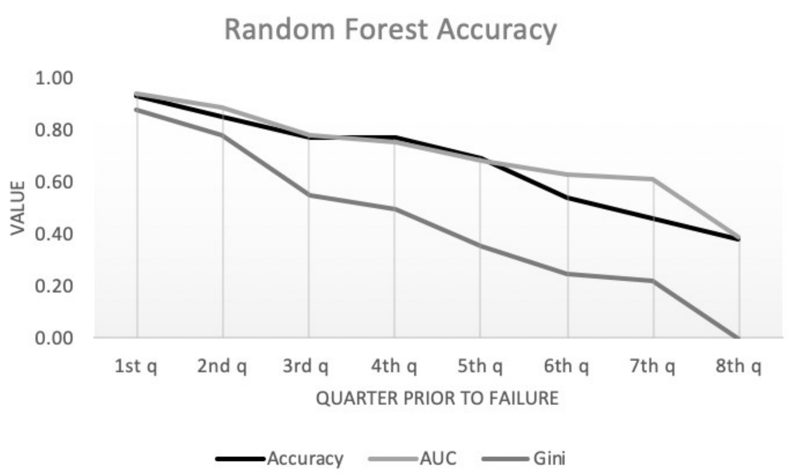
For the **5th quarter prior to failure** both the Type I and Type II error are 15.38%, the accuracy is 69.23%, the AUC is 0.68 and the Gini Coefficient is 0.35.

For the **6th quarter prior to failure** the Type I error is 46.15% and the Type II error is 0%. The accuracy is 53.85%, the AUC is 0.63, and the Gini Coefficient is 0.25.

For the **7th quarter prior to failure** the Type I error is 0% and the Type II error is 53.84%. The accuracy is 46.15%, the AUC is 0.61, and the Gini Coefficient is 0.22.

For the **8th quarter prior to failure** the Type I error is 23.08% whilst the Type II error is 38.46%. The Accuracy is 34.46%, the AUC is 0.39 and the Gini Coefficient is 0.

Figure 18



Source: author's elaboration

As it can be seen in **Figure 19**, all the accuracy and prediction power measures that have been used show a decreasing trend as the time from the default increases. The model is reliable up to 4-5 quarters prior to failure.

This analysis shows that, given an initial sample of 25 randomly chosen defaulted banks matched by asset size with banks still active (as March 2020), the different methodologies implemented have different degree of accuracy. In particular, confronting different accuracy measures, some models result to be less efficient in predicting the bank failure.

Overall, Random Forest is more accurate than the statistical methods implemented (logit and LDA) and of other machine learning techniques (k-NN, SVM, and B-NN) whilst, the least accurate model is LDA.

Concluding, Random Forest technique can be implemented as a reliable Early Warning System.

CONCLUSIONS

The failure of a bank creates concern since its effects reflects on the whole economy and on the other financial institutions through the so called "contagion effect". There exist many causes that lead a commercial bank to fail; they can both be internal to the bank or they can come from an external source.

The impossibility for the U.S. regulator to continuously supervise the institutions through the on-site examinations lead to the need to find an alternative way to verify the status of the institutions.

Many Early Warning Systems have been created by the government, mainly implemented through statistical techniques and taking the main information from the bank's financial statements.

The purpose of this thesis was to build off-site models using both statistical and machine learning techniques and to understand which one had the highest predictive power; the whole process was based on finding the best classification technique. In particular, 50 large (more than \$ 500 million assets) U.S. commercial bank have been analyzed: 25 failed banks between 2005 and 2015 have been matched with 25 still active banks based on the asset size at the time of the default and on the location whenever possible. All the institutions have been randomly selected from the FDIC database (hence only including insured banks).

Initially, 19 financial ratios have been retrieved for 8 quarters prior to the failure. Literature regarding the prediction of bank failure have been analyzed and all the model used have been explained in order to select the most useful ratios.

Student's t-tests for all the 8 quarters have been implemented to eliminate the predictors that didn't show statistically different means between groups. For the first quarter 15 ratios were used to conduct the analyses.

Logistic Regression and Linear Discriminant Analysis were the two statistical technique implemented, whilst K-Nearest Neighbors, Supported Vector Machines, Random Forest, and Backpropagation Neural Network have been implemented as machine learning techniques.

These 6 approaches are studied for the first quarter prior to failure of the institutions; in order to evaluate and compare them, accuracy, AUC, and Gini coefficient have been computed for each model. Confusion matrices have also been showed to compare Type I and Type II errors for each classification technique.

The implementation of each method had different requirements; statistical techniques need the predictors to be linear, normal, and independent. Multicollinearity revealed to be one of the biggest issues for the correct estimation of the parameters (as shown by the VIF). Therefore, a

Principal Component Analysis have been implemented before the Logit and LDA models. 5 uncorrelated Principal Components explaining 95.75% of the total variance have been used to implement the statistical models.

Machine learning techniques implemented don't have the same requirement for the predictors as the ones discussed for the statistical models; nevertheless, the main issue to address is the variance-bias tradeoff: models with low bias in the estimation of the parameter tend to have high variance.

For example, the k-NN model shows a high variance when the number of Nearest Neighbors (k) is small. Balancing the variance and bias is extremely important to achieve a good classification model that could accurately predict the default of the banks. Random Forest is a methodology that puts together different decision trees and which is able to decrease the variance without increasing the bias as described in Chapter 4.

The results for the first quarter prior to failure show that Random Forest shows higher value for accuracy, Gini Coefficient, and AUC resulting as the model with the highest predictive power. Logit shows the lowest Accuracy score but it presents the lowest Type II error (and the highest Type I error). LDA and SVM are better than Logit model (especially comparing the Type I errors) whilst k-NN shows the lowest Gini Coefficient.

Random Forest shows the best predictive power if compared to the other models implemented, meaning that for the first quarter prior to the failure RF is able to correctly classify 92.3% of the banks (with a Type II error of 7.3%).

The predictive power of the model has been analyzed for all the 8 quarters prior to failure: as expected, it decreases as the time from the failure increases. Type I and Type II errors increase too, leading to less reliable results.

This result highlights how difficult is to predict the failures when the analysis is done 2 years prior to the failure (8^{th} quarter); indeed, the accuracy is less than 35% and probability of misclassifying a default bank is 23.07% (Type I error).

Overall, all the models presented have a good predictive power for the first quarter prior to the failure without showing big differences between statistical and machine learning techniques (especially comparing the accuracy), highlighting that efficient Early Warning Systems can be created to predict the defaulting banks.

Nevertheless, the simple mechanism of the Random Forest showed a slightly higher accuracy than more complex models such as Back-Propagation Neural Network.

The creation of an accurate Early Warning System is a really important task for the regulators and the investors. For the former it can help to understand when to add requirements and reduce the overall cost associated to default; for the latter it allows to identify the sound banks and to be aware of the risk they are taking.

Deeper researches can be conducted using ensembled techniques; combining different machine learning techniques; this would allow to reduce the bias-variance tradeoff and create even more accurate models. Additionally, other information can be used to increase the predictive power of the presented models such as macroeconomic factors.

Moreover, it would also be useful to analyze the power of unsupervised techniques as Clustering (k-means and hierarchical clustering). Unsupervised machine learning algorithms are used for data which are not already labeled; they are complex to implement and can be less accurate but they can be useful to detect the hidden pattern of the data.

APPENDIX 1: List of Institutions

DEFAULTED BANKS	ACTIVE BANKS	YEAR OF FAILURE OF DEFAULTED BANKS
VANTUS BANK	CITIZENS TRI-COUNTY BANK	09/04/2009
SOLUTIONSBANK	SOUTHERN BANK	12/11/2009
THE FIRST STATE BANK	THE CORTLAND SAVINGS AND BANKING COMPANY	01/20/2012
THE PARK AVENUE BANK	NEWFIELD NATIONAL BANK	03/12/2010
WASHINGTON FIRST INTERNATIONAL BANK	OAK VALLEY COMMUNITY BANK	06/11/2010
SECURITY PACIFIC BANK	BANKERS' BANK OF THE WEST	11/07/2008
HOME NATIONAL BANK	UNITED BANKERS' BANK	07/09/2010
CAPITALSOUTH BANK	WAUCHULA STATE BANK	08/21/2009
PEOPLES COMMUNITY BANK (OH)	THE FARMERS & MERCHANTS BANK	07/31/2009
PENINSULA BANK	1ST SUMMIT BANK	06/25/2010
BANK OF FLORIDA - SOUTHWEST	THE FIRST NATIONAL BANK OF PALMERTON	05/28/2010
DARBY BANK & TRUST COMPANY	FIRST UNITED SECURITY BANK	11/12/2010
TEAMBANK, N.A.	SEATTLE BANK	03/20/2009
BARNES BANKING COMPANY	FALL RIVER FIVE CENTS SAVINGS BANK	01/15/2010

LIBERTYBANK	1ST CONSTITUTION BANK	07/30/2010
THE COLUMBIAN BANK AND TRUST COMPANY	HOME STATE BANK, NATIONAL ASSOCIATION	08/22/2008
ATLANTIC SOUTHERN BANK	CITIZENS 1ST BANK	05/20/2011
SAN DIEGO NATIONAL BANK	HANCOCK BANK	10/30/2009
DORAL BANK	FIRST BANK	02/27/2015
CALIFORNIA NATIONAL BANK	EVERBANK	10/30/2009
UNITED COMMERCIAL BANK	THIRD FEDERAL SAVINGS AND LOAN ASSOCIATION OF CLEVELAND	11/06/2009
IMPERIAL CAPITAL BANK	FARMERS AND MERCHANTS BANK OF LONG BEACH	12/18/2009
PARK NATIONAL BANK	PLAINSCAPITAL BANK	10/30/2009
AMTRUST BANK	BANKUNITED	12/04/2009
DOWNEY SAVINGS AND LOAN ASSOCIATION, F.A.	WASHINGTON FEDERAL SAVINGS AND LOAN ASSOCIATION	11/21/2008

APPENDIX 2: Student's t-tests

The results for the t-tests from the 2nd to the 8th quarters prior to the failure are hereafter presented as Python outputs. The variables which don't show a statistically different mean between the two groups are dropped to improve the quality of the predictive power.

Results for the t-test 2 quarters prior to the failure

```
t-test D/E
Statistics=-9.590, p=0.000
The two groups are significantly different, reject H0

t-test E/A
Statistics=-8.654, p=0.000
The two groups are significantly different, reject H0

t-test Tier 1 C ratio
Statistics=-6.919, p=0.000
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-6.898, p=0.000
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=-1.613, p=0.113
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=5.011, p=0.000
The two groups are significantly different, reject H0

t-test LLA/Noncurrent LL
Statistics=-3.207, p=0.002
The two groups are significantly different, reject H0

t-test Pretax ROA
Statistics=-1.237, p=0.222
The two groups are not statistically different, accept H0


t-test NOI/TA
Statistics=-6.166, p=0.000
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=3.296, p=0.002
The two groups are significantly different, reject H0

t-test Cash Dividends/NI
Statistics=-1.680, p=0.099
The two groups are significantly different, reject H0

t-test Retained Earnings/E
Statistics=-6.577, p=0.000
The two groups are significantly different, reject H0

t-test NII/A
Statistics=-1.150, p=0.256
The two groups are not statistically different, accept H0

t-test NIM
Statistics=-4.230, p=0.000
The two groups are significantly different, reject H0

t-test ROA
Statistics=-5.887, p=0.000
The two groups are significantly different, reject H0

t-test ROE
Statistics=-5.858, p=0.000
The two groups are significantly different, reject H0

t-test Loans/A
Statistics=2.436, p=0.019
The two groups are significantly different, reject H0

t-test Loans/Deposits
Statistics=-0.054, p=0.957
The two groups are not statistically different, accept H0

t-test Domestic Deposits/A
Statistics=2.197, p=0.033
The two groups are significantly different, reject H0
```

Results for the t-test 3 quarters prior to the failure

```
t-test D/E
Statistics=-7.782, p=0.000
The two groups are significantly different, reject H0

t-test E/A
Statistics=1.029, p=0.309
The two groups are not statistically different, accept H0

t-test Tier 1 C ratio
Statistics=-5.075, p=0.000
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-3.550, p=0.001
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=0.897, p=0.374
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=1.533, p=0.132
The two groups are not statistically different, accept H0

t-test LLA/Noncurrent LL
Statistics=-1.408, p=0.165
The two groups are not statistically different, accept H0

t-test Pretax ROA
Statistics=-6.901, p=0.000
The two groups are significantly different, reject H0

t-test NOI/TA
Statistics=-6.187, p=0.000
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=4.421, p=0.000
The two groups are significantly different, reject H0

t-test Cash Dividends/NI
Statistics=-1.951, p=0.057
The two groups are significantly different, reject H0

t-test Retained Earnings/E
Statistics=-6.845, p=0.000
The two groups are significantly different, reject H0

t-test NII/A
Statistics=-3.110, p=0.003
The two groups are significantly different, reject H0

t-test NIM
Statistics=-1.891, p=0.065
The two groups are significantly different, reject H0

t-test ROA
Statistics=-6.289, p=0.000
The two groups are significantly different, reject H0

t-test ROE
Statistics=-1.982, p=0.053
The two groups are significantly different, reject H0

t-test Loans/A
Statistics=1.044, p=0.302
The two groups are not statistically different, accept H0

t-test Loans/Deposits
Statistics=-0.418, p=0.678
The two groups are not statistically different, accept H0

t-test Domestic Deposits/A
Statistics=1.416, p=0.163
The two groups are not statistically different, accept H0
```

Results for the t-test 4 quarters prior to the failure

```
t-test D/E
Statistics=-6.039, p=0.000
The two groups are significantly different, reject H0

t-test E/A
Statistics=-4.470, p=0.000
The two groups are significantly different, reject H0

t-test Tier 1 C ratio
Statistics=-4.773, p=0.000
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-3.253, p=0.002
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=1.173, p=0.246
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=-0.719, p=0.476
The two groups are not statistically different, accept H0

t-test ROA
Statistics=-4.116, p=0.000
The two groups are significantly different, reject H0

t-test ROE
Statistics=-3.222, p=0.002
The two groups are significantly different, reject H0

t-test Loans/A
Statistics=3.329, p=0.002
The two groups are significantly different, reject H0

t-test Loans/Deposits
Statistics=-0.127, p=0.900
The two groups are not statistically different, accept H0

t-test Domestic Deposits/A
Statistics=1.432, p=0.159
The two groups are not statistically different, accept H0

t-test Pretax ROA
Statistics=-4.719, p=0.000
The two groups are significantly different, reject H0

t-test NOI/TA
Statistics=-2.562, p=0.014
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=1.142, p=0.259
The two groups are not statistically different, accept H0

t-test Cash Dividends/NI
Statistics=-2.284, p=0.027
The two groups are significantly different, reject H0

t-test Retained Earnings/E
Statistics=-3.830, p=0.000
The two groups are significantly different, reject H0

t-test NII/A
Statistics=-3.320, p=0.002
The two groups are significantly different, reject H0

t-test NIM
Statistics=-1.909, p=0.062
The two groups are significantly different, reject H0
```

Results for the t-test 5 quarters prior to the failure

```
t-test D/E
Statistics=-4.784, p=0.000
The two groups are significantly different, reject H0

t-test E/A
Statistics=-4.108, p=0.000
The two groups are significantly different, reject H0

t-test Tier 1 C ratio
Statistics=-4.333, p=0.000
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-4.172, p=0.000
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=0.000, p=1.000
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=3.332, p=0.002
The two groups are significantly different, reject H0

t-test LLA/Noncurrent LL
Statistics=-1.805, p=0.077
The two groups are significantly different, reject H0


t-test Pretax ROA
Statistics=-3.394, p=0.001
The two groups are significantly different, reject H0

t-test NOI/TA
Statistics=-3.606, p=0.001
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=1.101, p=0.276
The two groups are not statistically different, accept H0

t-test Cash Dividends/NI
Statistics=0.637, p=0.527
The two groups are not statistically different, accept H0

t-test Retained Earnings/E
Statistics=-4.854, p=0.000
The two groups are significantly different, reject H0

t-test NII/A
Statistics=-1.642, p=0.107
The two groups are not statistically different, accept H0
```


```
t-test NII/A
Statistics=-1.642, p=0.107
The two groups are not statistically different, accept H0
```


```
t-test NIM
Statistics=-2.074, p=0.043
The two groups are significantly different, reject H0

t-test ROA
Statistics=-4.276, p=0.000
The two groups are significantly different, reject H0

t-test ROE
Statistics=-5.501, p=0.000
The two groups are significantly different, reject H0

t-test Loans/A
Statistics=3.950, p=0.000
The two groups are significantly different, reject H0

t-test Loans/Deposits
Statistics=0.768, p=0.446
The two groups are not statistically different, accept H0

t-test Domestic Deposits/A
Statistics=1.440, p=0.156
The two groups are not statistically different, accept H0
```

Results for the t-test 6 quarters prior to the failure

```
t-test D/E
Statistics=-2.799, p=0.007
The two groups are significantly different, reject H0

t-test E/A
Statistics=-2.324, p=0.024
The two groups are significantly different, reject H0

t-test Tier 1 C ratio
Statistics=-3.613, p=0.001
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-3.589, p=0.001
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=-0.016, p=0.988
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=2.236, p=0.030
The two groups are significantly different, reject H0

t-test LLA/Noncurrent LL
Statistics=-1.040, p=0.304
The two groups are not statistically different, accept H0


t-test Pretax ROA
Statistics=-0.400, p=0.691
The two groups are not statistically different, accept H0

t-test NOI/TA
Statistics=-1.999, p=0.051
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=1.388, p=0.172
The two groups are not statistically different, accept H0

t-test Cash Dividends/NI
Statistics=-1.239, p=0.221
The two groups are not statistically different, accept H0

t-test Retained Earnings/E
Statistics=-1.124, p=0.267
The two groups are not statistically different, accept H0

t-test NII/A
Statistics=-0.845, p=0.402
The two groups are not statistically different, accept H0

t-test NIM
Statistics=-1.486, p=0.144
The two groups are not statistically different, accept H0


t-test ROA
Statistics=-2.184, p=0.034
The two groups are significantly different, reject H0

t-test ROE
Statistics=-3.251, p=0.002
The two groups are significantly different, reject H0

t-test Loans/A
Statistics=4.044, p=0.000
The two groups are significantly different, reject H0

t-test Loans/Deposits
Statistics=0.999, p=0.323
The two groups are not statistically different, accept H0

t-test Domestic Deposits/A
Statistics=0.906, p=0.369
The two groups are not statistically different, accept H0
```

Results for the t-test 7 quarters prior to the failure

```
t-test D/E
Statistics=-2.905, p=0.006
The two groups are significantly different, reject H0

t-test E/A
Statistics=-2.474, p=0.017
The two groups are significantly different, reject H0

t-test Tier 1 C ratio
Statistics=-4.106, p=0.000
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-3.840, p=0.000
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=-0.624, p=0.536
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=1.576, p=0.122
The two groups are not statistically different, accept H0

t-test LLA/Noncurrent LL
Statistics=-1.128, p=0.265
The two groups are not statistically different, accept H0

t-test Pretax ROA
Statistics=-1.821, p=0.075
The two groups are significantly different, reject H0

t-test NOI/TA
Statistics=-2.102, p=0.041
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=1.932, p=0.059
The two groups are significantly different, reject H0

t-test Cash Dividends/NI
Statistics=-0.950, p=0.347
The two groups are not statistically different, accept H0

t-test Retained Earnings/E
Statistics=-1.935, p=0.059
The two groups are significantly different, reject H0

t-test NII/A
Statistics=-1.370, p=0.177
The two groups are not statistically different, accept H0

t-test NIM
Statistics=-3.280, p=0.002
The two groups are significantly different, reject H0

t-test ROA
Statistics=-2.111, p=0.040
The two groups are significantly different, reject H0

t-test ROE
Statistics=-2.220, p=0.031
The two groups are significantly different, reject H0

t-test Loans/A
Statistics=-0.944, p=0.350
The two groups are not statistically different, accept H0

t-test Loans/Deposits
Statistics=2.719, p=0.009
The two groups are significantly different, reject H0

t-test Domestic Deposits/A
Statistics=0.533, p=0.597
The two groups are not statistically different, accept H0
```

Results for the t-test 8 quarters prior to the failure

```
t-test D/E
Statistics=-1.549, p=0.128
The two groups are not statistically different, accept H0

t-test E/A
Statistics=-0.885, p=0.380
The two groups are not statistically different, accept H0

t-test Tier 1 C ratio
Statistics=-3.019, p=0.004
The two groups are significantly different, reject H0

t-test Total Risk ratio
Statistics=-2.890, p=0.006
The two groups are significantly different, reject H0

t-test Earnings/A
Statistics=1.253, p=0.216
The two groups are not statistically different, accept H0

t-test LLA/LL
Statistics=1.251, p=0.217
The two groups are not statistically different, accept H0

t-test LLA/Noncurrent LL
Statistics=-0.410, p=0.684
The two groups are not statistically different, accept H0


t-test NIM
Statistics=-0.153, p=0.879
The two groups are not statistically different, accept H0

t-test ROA
Statistics=-2.797, p=0.007
The two groups are significantly different, reject H0

t-test ROE
Statistics=-1.442, p=0.156
The two groups are not statistically different, accept H0

t-test Loans/A
Statistics=4.544, p=0.000
The two groups are significantly different, reject H0

t-test Loans/Deposits
Statistics=2.988, p=0.004
The two groups are significantly different, reject H0

t-test Domestic Deposits/A
Statistics=1.251, p=0.217
The two groups are not statistically different, accept H0


t-test Pretax ROA
Statistics=-2.344, p=0.023
The two groups are significantly different, reject H0

t-test NOI/TA
Statistics=-2.536, p=0.015
The two groups are significantly different, reject H0

t-test Efficiency ratio
Statistics=1.358, p=0.181
The two groups are not statistically different, accept H0

t-test Cash Dividends/NI
Statistics=0.447, p=0.657
The two groups are not statistically different, accept H0

t-test Retained Earnings/E
Statistics=-1.161, p=0.251
The two groups are not statistically different, accept H0

t-test NII/A
Statistics=-2.196, p=0.033
The two groups are significantly different, reject H0
```

APPENDIX 3: ROC and AUC

Receiver Operating Characteristic (ROC) graph allows to efficiently visualize the classifier in order to compare it with others based on its performance. ROC is a two-dimensional graph: on the X-axis the False Positive Rate is plotted whilst on the Y-axis the True Positive Rate is showed. This graph is a useful tool for evaluating a model on the basis of the True Positives and False Positives it is able to detect. ROC can be also defined as a probability curve (Fawcett, 2006).

False Positive rate is given by:

$$FP\ rate = \frac{False\ Positives}{False\ Positives + True\ Negatives}$$

True Positive rate is given by:

$$TP\ rate = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Accuracy rate is given by:

$$Accuracy\ rate = \frac{True\ Positives + True\ Negatives}{Total\ Positives + Total\ Negatives}$$

Precision is given by:

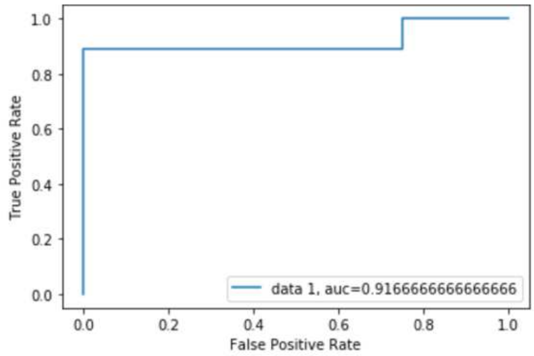
$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

AUC (Area Under ROC Curve) permits to recognize if the model is a good classifier. The correct prediction of the bank failures is a supervised classification problem; AUC defines if a technique is able to separate between defaulted and sound banks.

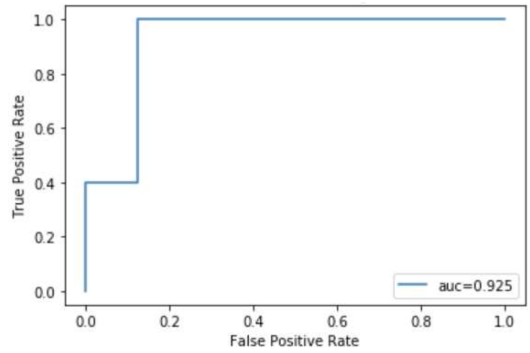
In particular, AUC value goes from 0 to 1. An AUC value of 1 indicates a classifier that perfectly discriminates between the two classes, meanwhile an AUC value of 0 indicates that the classifier isn't able to separate between classes. Therefore, the higher the AUC, the better is the predictive power of the model.

All the ROC graphs and AUC for the first quarter prior to failure are reported.

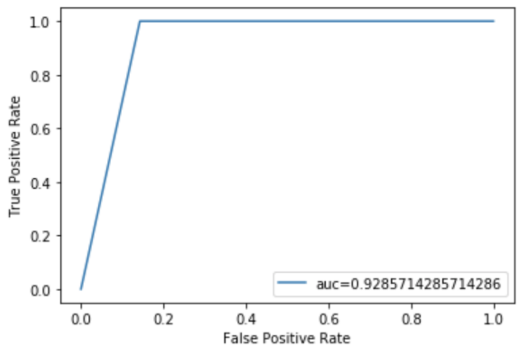
Logistic Regression ROC



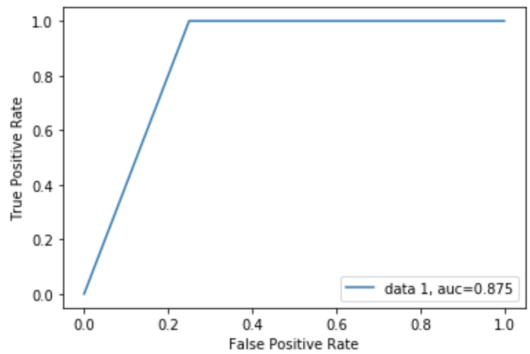
Linear Discriminant Analysis ROC



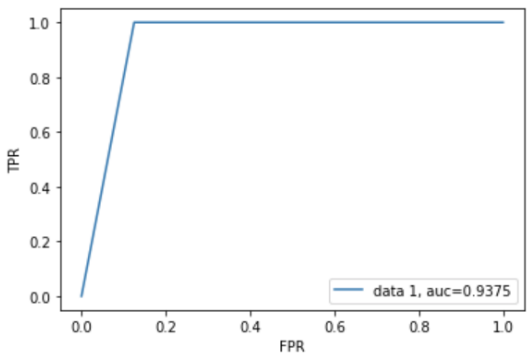
Support Vector Machines ROC



k-Nearest Neighbors ROC



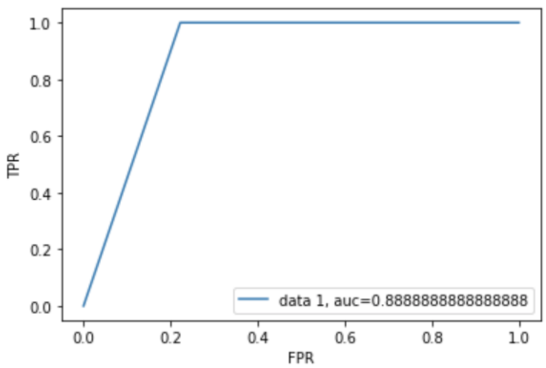
Random Forest ROC



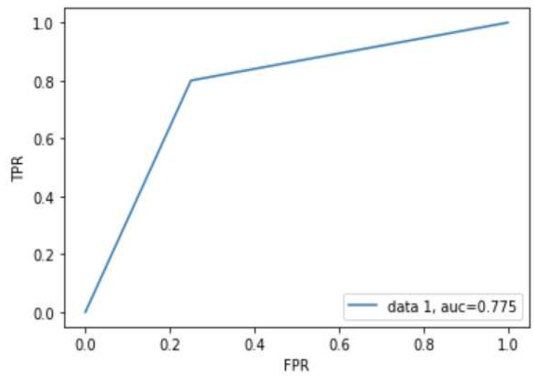
AUC values and the ROC curves shows that the model which better separates between the groups is Random Forest with an AUC of 0.94. The model that is good to discriminate between failed and non-failed institutions is k-Nearest Neighbors.

The ROC graphs and the AUC values for the Random Forest model implemented with the data retrieved for the other quarters analyzed are:

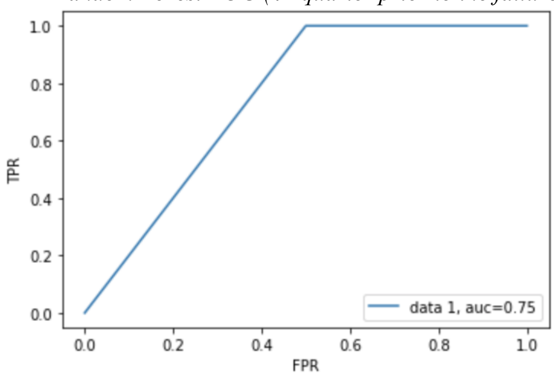
Random Forest ROC (2^{nd} quarter prior to the failure)



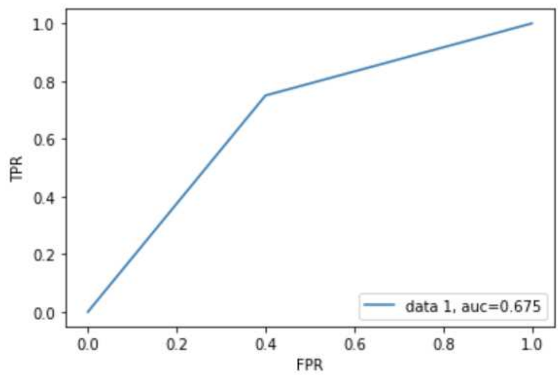
Random Forest ROC (3^{rd} quarter prior to the failure)



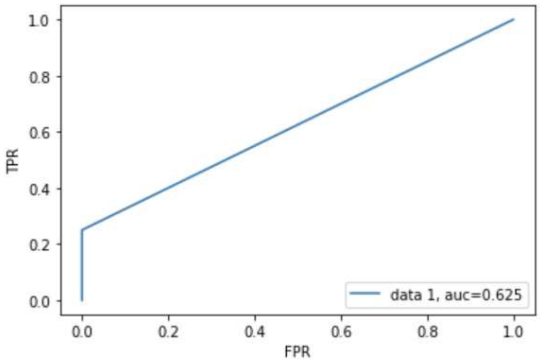
Random Forest ROC (4^{th} quarter prior to the failure)



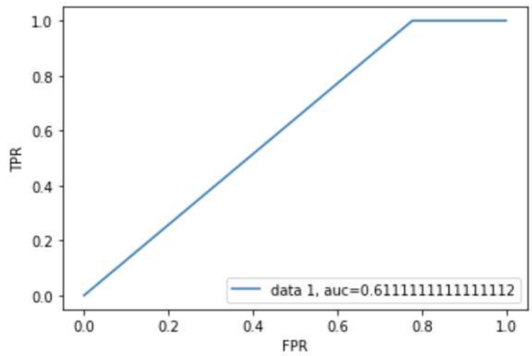
Random Forest ROC (5^{th} quarter prior to the failure)



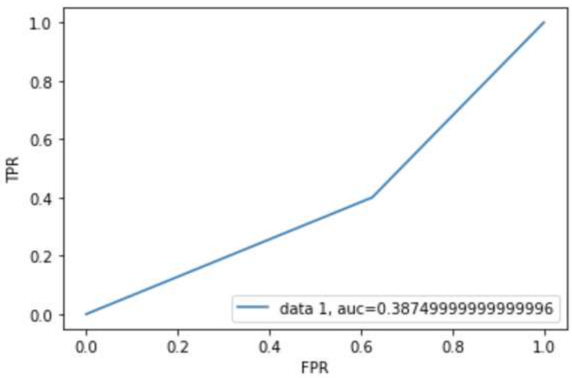
Random Forest ROC (6th quarter prior to the failure)



Random Forest ROC (7th quarter prior to the failure)



Random Forest ROC (8th quarter prior to the failure)



Source: author's elaboration

AUC values of Random Forest decrease when the oldest quarters are analyzed. In particular, the value goes from 0.89 for the 2nd quarter prior to failure to 0.39 for the 8th quarter prior to failure.

APPENDIX 4: CORRELATION MATRIX

	Equity/Total Assets	Debt/Equity	Tier 1 risk based capital ratio	Total Risk based capital ratio	Earnings/Total Assets	loan and Lease Loss Allowance/Loans and Leases	Loan and Lease Loss Allowance/Noncurrent Loss and Leases	Pretax ROA
Equity/Total Assets	1.000000	0.965515	0.857080	0.864595	0.189734	-0.595447	0.417074	0.695334
Debt/Equity	0.965515	1.000000	0.863376	0.871244	0.249206	-0.597598	0.403268	0.674254
Tier 1 risk based capital ratio	0.857080	0.863376	1.000000	0.998346	0.132702	-0.506662	0.332282	0.596025
Total Risk based capital ratio	0.864595	0.871244	0.998346	1.000000	0.131155	-0.510077	0.335426	0.600020
Earnings/Total Assets	0.189734	0.249206	0.132702	0.131155	1.000000	0.205883	0.094782	-0.065230
loan and Lease Loss Allowance/Loans and Leases	-0.595447	-0.597598	-0.506662	-0.510077	0.205883	1.000000	-0.308013	-0.698197
Loan and Lease Loss Allowance/Noncurrent Loss and Leases	0.417074	0.403268	0.332282	0.335426	0.094782	-0.308013	1.000000	0.363075
Pretax ROA	0.695334	0.674254	0.596025	0.600020	-0.065230	-0.698197	0.363075	1.000000
NOI/Total Assets	0.610349	0.607642	0.519250	0.533434	-0.001033	-0.598770	0.307347	0.834991
Efficiency Ratio	-0.241540	-0.252305	-0.225093	-0.216303	-0.050007	0.170934	-0.178784	-0.129528
Noninterest income/Total Assets	0.263758	0.295993	0.237451	0.238109	0.021640	-0.250270	0.071149	0.218955
NIM	0.529416	0.529292	0.549761	0.558761	-0.077788	-0.376706	0.074613	0.471992
ROA	0.694064	0.674454	0.581934	0.583159	0.007919	-0.637626	0.365687	0.881607
Net Loans/Total Assets	-0.289492	-0.247724	-0.489475	-0.482510	0.077040	0.085870	-0.303423	-0.383783
Domestic Deposits/Total Assets	-0.427653	-0.400419	-0.423873	-0.425549	-0.046925	0.358783	-0.172863	-0.200300

	NOI/Total Assets	Efficiency Ratio	Noninterest income/Total Assets	NIM	ROA	Net Loans/Total Assets	Domestic Deposits/Total Assets
Equity/Total Assets	0.610349	-0.241540	0.263758	0.529416	0.694064	-0.289492	-0.427653
Debt/Equity	0.607642	-0.252305	0.295993	0.529292	0.674454	-0.247724	-0.400419
Tier 1 risk based capital ratio	0.519250	-0.225093	0.237451	0.549761	0.581934	-0.489475	-0.423873
Total Risk based capital ratio	0.533434	-0.216303	0.238109	0.558761	0.583159	-0.482510	-0.425549
Earnings/Total Assets	-0.001033	-0.050007	0.021640	-0.077788	0.007919	0.077040	-0.046925
loan and Lease Loss Allowance/Loans and Leases	-0.598770	0.170934	-0.250270	-0.376706	-0.637626	0.085870	0.358783
Loan and Lease Loss Allowance/Noncurrent Loss and Leases	0.307347	-0.178784	0.071149	0.074613	0.365687	-0.303423	-0.172863
Pretax ROA	0.834991	-0.129528	0.218955	0.471992	0.881607	-0.383783	-0.200300
NOI/Total Assets	1.000000	-0.136459	0.228909	0.433987	0.768122	-0.390604	-0.283457
Efficiency Ratio	-0.136459	1.000000	-0.124078	-0.200751	-0.141652	0.006881	0.347086
Noninterest income/Total Assets	0.228909	-0.124078	1.000000	0.194254	0.222652	-0.006394	-0.079152
NIM	0.433987	-0.200751	0.194254	1.000000	0.394809	-0.283969	-0.088433
ROA	0.768122	-0.141652	0.222652	0.394809	1.000000	-0.330662	-0.205300
Net Loans/Total Assets	-0.390604	0.006881	-0.006394	-0.283969	-0.330662	1.000000	0.105292
Domestic Deposits/Total Assets	-0.283457	0.347086	-0.079152	-0.088433	-0.205300	0.105292	1.000000

REFERENCES

Adya, M. a. C. F., 1998. How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation. *Journal of Forecasting,* Volume 17, pp. 481-495.

Agur, I., 2013. Multiple Bank Regulators and Risk Taking. *Journal of Financial Stability,* Issue 9, pp. 259-268.

Altinirmak, S. e. a., 2016. Comparison of machine learning techniques for analyzing banks' financial distress. *The Journal of Social Sciences Institutes,* 19(36), pp. 291-303.

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance,* Volume 23, pp. 589-609.

Beaver, W. H., 1966. Financial Ratios As Predictors of Failure. *Journal of Accounting Research,* Volume 4, pp. 71-111.

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

Beck, T. e. a., 2003. Bank Concentration and Crises. *NBER Working Paper ,* Issue 9921, pp. 1-40.

Bell, T. B., 1997. Neural Nets or the Logit Model? A Comparison of Each MOdel's Ability to Predict Commercial Bank Failures. *Intelligent Systems in Accounting, Finance and Management ,* Volume 6, pp. 249-246.

Berger, A. D. S., 1998. The Information Content of Bank Examinations. *Journal of Financial Services Research,* 14(2), pp. 117-144.

Berger, A. e. a., 2000. Comparing Market and Supervisory Assessment of Bank Performance: Who Knows What and When?. *Journal of Money, Credit Banking,* 32(3), pp. 641-667.

Bjorksten, N. e. a., 2014. *U.S. Implementation of the Basel Capital Regulatory Framework,* s.l.: Congressional Research Service.

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

Boyd, J. H. a. D. N. G., 2005. The Theory of Bank Risk Taking and Competition Rivisited. *The Journal of Finance,* 60(3), pp. 1329-1343.

Branson, M. e. a., 2014. *Regulatory Consistency Assessment Programme. Assessment of Basel III Regulation.* United States, Bank for International Settlements.

Brinkmann, E. e. a., 1995. Risk-Based Capital Standards and the Credit Crunch. *Journal of Money, Credit and Banking ,* 27(3), pp. 848-863.

Canbas, S. e. a., 2005. Prediction of commercial bank failure via multivariate statistical analysis of financial structures: the Turkish case. *European Journal of Operational Research,* Volume 166, pp. 528-546.

Chen, W. a. D. Y., 2009. Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications,* Volume 36, pp. 4075-4086.

Cheong, C. e. a., 2019. Bank Failure: A New Approach to Prediction and Supervision. *Asian Journal of Finance & Accounting,* 11(1), pp. 111-140.

Clarke, R. L. e. a., 1988. *Bank Failure: An Evaluation of the Factors Contributing to the Failure of National Banks,* Washington, D.C.: Office of the Comptroller of the Currency.

Cole, A. R. e. a., 1995. FIMS: A New Financial Institutions Monitoring System for Banking Organizations. *Federal Reserve Bulletin,* Volume 81, pp. 1-41.

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

Collier, C. e. a., 2003. The SCOR System of Off-Site Monitoring: Its Objectives, Functioning, and Performance. *FDIC Banking Review,* 15(3), pp. 1-40.

Cox, R. A. K. a. W. G. W., 2014. Predicting the US bank failure: A discriminant analysis. *Economic Analysis and Policy ,* Volume 44, pp. 202-211.

Curry, T. e. a., 2003. Using Market Information to Help Identify Distressed Institutions: A Regulatory Perspective. *FDIC Banking Review,* 15(3), pp. 1-16.

Detragiache, E. e. a., 1998. Financial Liberalization and Financial Fragility. *International Monetary Fund,* pp. 1-36.

Detragiache, E. e. a., 1998. The Determinants of Banking Crises in Developing and Developed Countries. *International Monetary Fund,* pp. 81-109.

Ecer, F., 2013. Comparing the bank failure prediction performance of neural networks and support vector machines: the turkish case. *Economic Research,* Volume 26, pp. 81-98.

Estrella, A. e. a., 2000. Capital Ratios as Predictors of Bank Failure. *FRBNY Economic Policy Review,* pp. 33-52.

FDIC, 2019. *Basic Examination Concepts and Guidelines.* s.l., RMS Manual of Examination Policies.

Flannery, M. J. e. a., 1999. The Value of a Government Monitor for U.S. Banking Firms. *Journal of Money, Credit Banking,* 31(1), pp. 14-34.

Gilbert, R. A. e. a., 2002. Could a CAMELS Downgrade Model Improve Off-Site Surveillance?. *The Federal Reserve Bank of St. Louis,* pp. 47-64.

Hasman, A., 2013. A Critical Review of Contagion Risk in Banking. *Journal of. Economic Surveys,* 27(5), pp. 978-995.

Heyliger, W. E. a. H. D. P., 1991. Predicting Small Bank Failure. *The Journal of Entrepreneurial Finance,* 1(2), pp. 125-140.

Jagtiani, J. e. a., 2003. Early warning models for bank supervision: Simple could do better. *Federal Reserve Bank of Chicago ,* pp. 50-60.

James, G. e. a., 2013. *An Introduction to Statistical Learning with Applications in R.* 1st Edition ed. New York: Springer.

Kaufman, G. G., 1994. Bank Contagion: A Review of the Theory and Evidence. *Journal of Financial Services Research,* pp. 123-150.

Kaufman, G. G., 1996. Bank Failures, Systemic Risk , and Bank Regulation. *Working Paper Series Federal Reserve Bank of Chicago,* pp. 1-38.

Kumar Ravi, P. e. a., 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques- A review. *European Journal of Operational Research,* Volume 180, pp. 1-28.

Le, H. H. a. V. J., 2016. Predicting Bank Failure: An Improvement By Implementing a Machine Learning Appproach to Classical Financial Ratios. *Research in Internaational Business and Finance ,* Volume 44, pp. 16-25.

Lee, S. a. C. W. S., 2013. A multi-industry bankruptcy prediction model using. back-propagation neural network ad multivariate discriminant analysis. *Expert Systems with Applications,* Volume 40, pp. 2941-2946.

Leo, M. e. a., 2019. Machine Learning in Banking Risk Management: A literature Review. *Risks,* Volume 7, pp. 1-22.

Lucca, D. e. a., 2014. Inconsistent Regulators: Evidence from Banking. *The Quarterly Journal of Economics,* pp. 889-938.

Lucca, D. e. a., 2014. Inconsistent Regulators: Evidence From Banking. *The Quarterly Journal of Economics,* pp. 889-938.

Martin, D., 1977. Early Warning of Bank Failure A Logit Regression Approach. *Journal of Banking and Finance,* Volume 1, pp. 249-276.

Meyer, P. a. P. H. W., 1970. Prediction of Bank Failures. *The Journal of Finance,* 25(3), pp. 853-868.

Min, J. H. a. L. Y., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel parameters. *Expert Systems with Applications,* Volume 28, pp. 603-614.

Molinero, C. e. a., 2001. Bank Failure: a multidimensional scaling approach. *The European Journal of Finance,* Volume 7, pp. 165-183.

Ohlson, J. A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research,* Volume 8, pp. 109-131.

Petropoulos, A. e. a., 2017. Predicting bank insolvencies using machine learning techniques. *Bank of Greece,* pp. 1-42.

Santonu, B., 2003. Why do Banks Fail?. *International Review of Applied Economics,* 17(3), pp. 231-248.

Sinkey, J. F., 1975. A Multivariate Statistical Analysis of the Characteristics of Problem Banks. *The Journal of Finance,* 30(1), pp. 21-36.

Swicegood, P. a. C. J. A., 2001. Off-site Monitoring System for PredictingBank Underperformance: A Comparison of Neural Networks, Discriminant Analysis, and Professional Human Judgement. *International Journal of Intelligent Systems in Accounting, Finance & Management ,* Volume 10, pp. 169-186.

Taha, Z., 2013. Bank Failure Prediction with Logistic Regression. *International Journal of Economics and Financial Issues,* Volume 3, pp. 1-7.

Tai, B. Y. K. e. a., 1986. A Multivariate Statistical Analysis of the Characteristics of Problem Firms in Hong Kong. *Asia Pacific Journal of Management,* 3(2), pp. 121-127.

Thomson, J. B., 1991. Predicting Bank Failures in the 1980s. *Federal Reserve Bank of Cleveland,* Volume 27, pp. 9-20.

Todd, K. J., 1996. Text of Uniform Financial Institution Rating System. *Federal Register,* 61(245), pp. 67021-67029.

Wang, G. e. a., 2014. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Application,* Volume 41, pp. 2353-2361.

Wheelock, D. C., 2011. Banking Industry Consolidation and Market Structure: Impact of the Financial Crisis and Recession. *Federal Reserve Bank of St. Louis,* Issue 93, pp. 419-438.

Wu, C. a. W. X., 2000. A Neural Network Approach for Analyzing Small Business Lending Decisions. *Review of Qualitative and Accounting,* Volume 15, pp. 259-276.

Zavgren, C. V., 1985. Assessing the vulnerability to failure of american industrial firms: a logistic analysis. *Journal of Business Finance& Accounting,* Volume 12, pp. 19-45.

Zhao, H. e. a., 2009. Effects of feature construction on classification performance: An empirical study in bank failure prediction. *Expert Systems with Applications,* Volume 36, pp. 2633-2644.