Banking failure prediction: a boosting classification tree approach

Alexandre Momparler, Pedro Carmona & Francisco Climent

Published online: 11 Jan 2016.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

Banking failure prediction: a boosting classification tree approach
La Predicción Del Fracaso Bancario Con La Metodología "Boosting Classification Tree"

Alexandre Momparler[a], Pedro Carmona[b] and Francisco Climent[c]

[a]Department of Corporate Finance, Facultat d'Economia, Universitat de València, València, Spain;
[b]Department of Accounting, Facultat d'Economia, Universitat de València, València, Spain; [c]Department of Financial Economics, Facultat d'Economia, Universitat de València, València, Spain

ABSTRACT

The recent financial crisis shows that failure of some financial institutions can cause other banks to fail and ultimately cause damage to the financial system worldwide. Eurozone banks that experienced either liquidity or solvency problems during the financial markets turmoil were bailed out by their national governments with the financial support and supervision of the European Union. This paper applies the boosted classification tree methodology to predict failure in the banking sector and identifies four key scorecard variables that are worth tracking closely in order to anticipate and prevent bank financial distress. The data used in this study comprises 2006–2012 annual series of 25 financial ratios of 155 banks in the Eurozone. The findings indicate that the greater the size and the higher the income from non-operating items and net loans to deposits, the more likely is bank failure; conversely, the higher the Interbank ratio the lower the chances of bank financial distress. For the sake of their own financial soundness, banks should fund lending activities through clients' deposits and should avoid relying excessively on non-recurring sources of income.

RESUMEN

La reciente crisis financiera muestra que el fracaso de algunas instituciones financieras puede producir la quiebra en cadena de otras entidades financieras y, en última instancia, originar graves problemas al sistema financiero mundial. Los bancos de la zona euro que experimentaron problemas de liquidez o solvencia durante las turbulencias de los mercados financieros fueron rescatados por sus gobiernos nacionales con el apoyo y la supervisión financiera de la Unión Europea. En este trabajo se aplica la metodología Boosting Classification Tree con el objeto de predecir el fracaso en el sector bancario identificando cuatro indicadores clave, cuyo seguimiento es primordial para anticipar y prevenir problemas financieros en dicho sector. La muestra utilizada en este estudio se compone de series anuales de 25 ratios financieras de 155 bancos de la Zona Euro para el período 2006–2012. Los resultados indican que a mayor tamaño, mayores ingresos extraordinarios y mayor ratio préstamos/depósitos, más probable es el fracaso de un banco. En cambio, cuanto mayor sea

CONTACT Pedro Carmona ✉ pedro.carmona@uv.es

la ratio préstamos al interbancario/prestamos del interbancario, la probabilidad de que la entidad tenga problemas financieros es menor. Por lo tanto, con el objetivo de mejorar su estabilidad financiera, la banca debería financiar su actividad crediticia principalmente a través de los depósitos de los clientes, evitando una dependencia excesiva de fuentes de ingresos extraordinarios no recurrentes.

1. Introduction

The recent financial crisis shows that financial institutions cannot perform much-needed banking functions while experiencing either liquidity or solvency problems. Additionally, in the case of major bank failures, those functions could not simply be shut down or substituted without significant systemic damage.

The banking sector plays a special role in the economy and has critical functions which are essential for economic activity. Banks collect funds (deposits and other forms of debt) from private persons and businesses (financial and non-financial). They carry out maturity transformation and provide loans for households and businesses allowing savings to be allocated mainly to investments. They also manage payment transactions that are crucial for all sectors of the economy and society. The banking business is based on the trust of stakeholders. Banks' most important capital is their reputation. If confidence is lost, depositors and other short-term debt holders (time deposits or certificates of deposit) would immediately try to withdraw their funds. This would make the bank unavoidably bankrupt since no bank holds sufficient liquid assets to cover all short-term liabilities. Bank failures are capable of undermining financial stability, especially if they lead to a loss of depositor confidence in other banks. During the last financial crisis, these issues led governments to, for the most part, recapitalise and save failing banks (or banks in financial distress).

The financial crisis has illustrated that the failure of some financial institutions would cause other financial institutions to fail and ultimately, cause wider damage to the financial system. The turmoil created after the failure of Lehman Brothers in September 2008, which the US Government decided not to save, demonstrated the materialisation of this risk. If a financial institution fails, other banks that provide funds to it would not get access to those funds. This would cause liquidity problems for them that would make these banks vulnerable too. If their short-term debt holders and depositors consider that it is better to withdraw funds from these vulnerable banks, then a domino effect is likely to take place. This could result in liquidity and ultimately solvency problems for a significant part of the financial sector. Capital markets may also experience shocks and payment systems be disrupted.

Rescuing banks with public funds (bail-out) helped to avoid what could have been economic depression on a scale not seen before, but it has also created a number of medium- to long-term problems that are becoming increasingly apparent (European Commission, 2012):

- The distortion of competition: institutions that are perceived by the market as being systemic are often perceived to benefit from an implicit state guarantee.

- The realisation of moral hazard: as the state guarantee risks encouraging excessive risk-taking within systemic institutions.
- Growth regardless of synergy gains: as non-systemic institutions have an incentive to reach systemic importance, thus leading banks to expand beyond their ideal economic size.
- Increased burden on public finances: public intervention cost taxpayers substantial sums of money and even put some Member States' public finances at risk. Between October 2008 and October 2011, the EU Commission approved EUR 4.5 trillion (equivalent to 37% of EU GDP) in state aid measures to financial institutions.

Well-managed banks should always try to find borrowers who will pay high interest rates and are unlikely to default on their loans. The causes of failure of risk management controls at banks should be of great concern to both bank shareholders and supervisors. According to the European Commission (2012), the underlying causes of risk management failures are:

- Growth-oriented compensation schemes may become an incentive for excessive risk-taking behavior of executives and traders within the financial institution.
- Focus on short-term performance that may cause damage to the financial institution when risks materialise.
- Income diversification: overextension of non-traditional banking activities.
- Excessive concentration of the loan portfolio in a particular industry (e.g. real estate).

Many central banks have employed various early warning systems to monitor bank risks for years. However, the number of banking crises experienced over recent decades shows that upholding the bank system is no easy task. The literature on bank financial distress draws heavily on the CAMELS[1] ratings system, introduced by US regulators in 1979.

This paper aims to identify a number of leading indicators (or scorecard variables) that may help anticipate bank failure. The accurate and permanent tracking of these scorecard variables may facilitate timely recognition of risk management control failures and ultimately prevent banks' financial distress. To this end, a machine learning method (ML), boosted regression trees (BRT), is used to predict bank failure and the results of our research suggest that the use of this methodology in the banking sector is appropriate. ML methodology assumes that the generation of data is an intricate process, and it attempts to obtain a response by observing inputs and responses and finding prevailing patterns. This enhances a model's predictive ability, and focuses on what is being predicted and how predicting accuracy should be assessed.

One advantage of single decision trees is their simplicity, but boosting produces a model with hundreds to thousands of trees, presenting a challenge in regard to understanding the final model. However, BRT does not have to be treated like a black box, and we show how the models can be summarised, evaluated and interpreted similarly to conventional regression models. Although BRT models are complex, they can be

summarised in ways that give powerful insight, and their predictive power is superior to most traditional modelling methods.

Our paper makes the following important contributions. To the best of our knowledge it is the first time that BRT methodology is applied to predict failure in the Eurozone banking sector. Second, our paper identifies four key scorecard variables that, considered jointly, may be helpful to anticipate and prevent banks' financial distress. In addition, the results obtained are consistent with generally accepted financial principles and the robustness tests carried out confirm the remarkable prediction accuracy of the model.

The remainder of this paper proceeds as follows. The relevant literature is reviewed and data is described in the subsequent sections. Then, the statistical study and methodology are introduced in the next section. In the following section the results of the study are presented. Next, the one-year-before-failure BRT model is compared against two- and three-year-before-failure models. BRT is also compared to alternative methodologies. Finally, conclusions and managerial implications are put forward in the last section.

2. Literature review

In this section we discuss prior research on financial sector turmoil and its effects on the economy. Thus, a report by the European Commission Competition (2011) provides a comprehensive account of how the Commission's state aid policy responded to the financial and economic crisis, and examines the extent to which the objectives pursued by this policy can be considered as having been met. In so doing, it contributes to the wide policy debate that has been opened by the unprecedented use of state aid during the crisis, and provides a comprehensive factual background and insights for the new rules that are in the making as regards rescue and restructuring aid (for both financial and non-financial firms) and bank resolution and regulation. The main conclusion that can be drawn is that state aid to the financial sector and to the real economy under state aid control by the Commission has been effective in reducing financial instability, improving the functioning of financial markets and cushioning the effects of the crisis on the real economy.

As for research that investigates the European case, Savona, De Lisa, Dorodnykh and Vallascas (2013) examined whether pre-crisis bank characteristics explain state support to European banks during the global financial crisis. The authors show that, before the crisis, supported and non-supported banks differed in many aspects and these differences reflect bank characteristics at the core of the regulatory agenda. In particular, bank size and the related too-big-to-fail concerns play a dominant role in explaining state support in Europe. The results suggest that income diversification produces a decline in the expected bail-out costs for public finance in large banks and an increase of these costs in small and medium banks, and highlight the importance of ad hoc prudential requirements and cross-country resolution regimes for large European banks and justify the introduction of regulatory restrictions on income diversification if applied to banks of small and medium size or if they lead to a significant decrease in the size of large banks.

Additionally, Betz, Oprică, Peltonen and Sarlin (2014) developed an early-warning model for predicting vulnerabilities leading to distress in European banks using both bank and country-level data. The sample is based on the availability of balance sheet and income-statement data in Bloomberg. The observation period starts in 2000Q1 and ends in 2013Q2. The key findings of the paper are that complementing bank-specific vulnerabilities with indicators for macro-financial imbalances and banking sector vulnerabilities improves model performance and yields useful out-of-sample predictions of bank distress during the current financial crisis. Furthermore, the results of the evaluation framework show that a policymaker has to be substantially more concerned with missing bank distress than issuing false alarms for the model to be useful. The evaluations also imply that it is important to give more emphasis to systemically important and large banks for a policymaker concerned with systemic risk.

With regard to the research works on the US case, Ellul and Yerramilli (2013) constructed a risk management index (RMI) to measure the strength and independence of the risk management function at 72 bank holding companies for the period 1994–2009. Over the period 1995–2010, bank holding companies with a higher lagged RMI had lower tail risk and higher return on assets, all else being equal, and these banks had higher stock returns during crisis years, but there is no association between RMI and stock returns during normal non-crisis years. Overall, these results suggest that a strong and independent risk management function can curtail tail risk exposures at banks and possibly enhance value, particularly during crisis years.

In their analysis of banks' sources of income, DeYoung and Torna (2013) tested whether income from non-traditional banking activities contributed to the failure of hundreds of US commercial banks during the financial crisis. The authors observed bank characteristics each quarter from 2008:Q1 to 2010:Q2 to estimate the probability of bank failure from 2008:Q3 to 2010:Q4 and used a multi-period logit model. The results indicate that the probability of distressed bank failure declined with pure fee-based non-traditional activities such as securities brokerage and insurance sales, but increased with asset-based non-traditional activities such as venture capital, investment banking and asset securitisation. Furthermore, banks that engaged in risky non-traditional activities also tended to take risk in their traditional lines of business, suggesting that deregulation was neither a necessary nor a sufficient condition for bank failure during the crisis.

Other research efforts have been directed to study the relationships between the information provided by audit reports and financial crisis, and the contribution here of Jin, Kanagaretnam and Lobo (2011) is noteworthy. Their paper examines the ability of selected accounting and audit quality variables measured over a period prior to the financial crisis (the four quarters of 2006) to predict banks that subsequently failed during the financial crisis, and employs two sets of samples from the US: a troubled banks sample that includes those that failed in or after 2007 as well as banks classified as being troubled based on profitability, loan quality and balance sheet position in 2007, and a full sample that includes all banks with available required data. The authors used the 2006 quarterly data to measure the accounting and auditing predictor variables, and the 2007 annual data to identify troubled banks in that year. The paper used logistic regressions with clustered robust errors to account for potential serial and cross-sectional correlation. The results indicate that banks audited by reputable auditors

have a lower probability of failure and confirm the general belief that the recent banking crisis in the US was primarily driven by credit problems.

In terms of methodology we must highlight the thorough review by Demyanyk and Hasan (2010). This article provides a summary of methodologies and empirical results obtained in several economics and operations research papers that attempt to explain, predict or suggest remedies for financial crises or banking defaults. The paper provides an analysis of financial and economic circumstances associated with the sub-prime mortgage crisis in the US, along with an extensive review of intelligence techniques used in the operations research literature to predict bank failures.

At this point, we review the papers which used boosting for financial distress prediction. For instance, Alfaro, García, Gámez and Elizondo (2008) compared the prediction accuracy of artificial neural networks and AdaBoost on a set of European firms, considering the usual predicting variables (e.g. financial ratios), as well as qualitative variables, such as firm size, activity and legal structure. They concluded that the AdaBoost approach decreases the generalisation error by about 30% with respect to the error produced with a neural network. Comparing discriminant analysis, Cortés, Martínez and Rubio (2008) presented AdaBoost as a classification technique that can be used successfully in business failure forecasting. Accordingly, it was applied to a sample of 1.180 Spanish firms and has proved to be more accurate than discriminant analysis.

In a recent study Sun, Li, Huang and He (2014) summarised, analysed and assessed the current literature on financial distress prediction (FDP). Current FDP methods were categorized and reviewed through a combination of qualitative and quantitative selection. This review paper is valuable to guide research and applications in the field.

In a research work by Wang, Ma and Yang (2014), the authors proposed a new and improved FS-Boosting to predict corporate bankruptcy. By means of a feature selection strategy, FS-Boosting can improve boosting performance as base learners in FS-Boosting can achieve higher accuracy and diversity. For testing and illustration purposes, two real-world bankruptcy datasets were selected to demonstrate the effectiveness and feasibility of FS-Boosting. Experimental results reveal that FS-Boosting can be used as an alternative method for corporate bankruptcy prediction.

Finally, Kim, Kang, and Kim (2015) applied a geometric mean-based boosting algorithm (GMBoost) to handle data imbalance problems. To assess GMBoost performance, they applied it to bankruptcy prediction. The results and the comparative analysis with AdaBoost and cost-sensitive boosting indicate that GMBoost has advantages in terms of higher predictive power and robust learning capability when dealing with either imbalanced or balanced data.

3. Data

The data used in this study includes 38 annual financial ratios and total assets figures of 168 publicly traded financial institutions in the euro area for the period 2006–2012, comprising a total of 1069 observations. Firm size, as measured by total assets, is included because it has proved to be a good predictor or financial failure (Assadian & Ford, 1997).

Table 1. List of bailed-out banks.

Name	Country
Oesterreichische Volksbanken AG	Austria
Erste Group Bank AG	Austria
KBC Groep NV/ KBC Groupe SA-KBC Group	Belgium
Dexia	Belgium
IKB Deutsche Industriebank AG	Germany
Deutsche Bank AG	Germany
Commerzbank AG	Germany
Aareal Bank AG	Germany
Caja de Ahorros del Mediterraneo CAM	Spain
Soci,t, G,n,rale	France
Cr,dit Agricole S.A.	France
BNP Paribas	France
TT Hellenic Postbank S.A	Greece
T Bank S.A	Greece
Proton Bank S.A.	Greece
Piraeus Bank SA	Greece
National Bank of Greece SA	Greece
General Bank of Greece SA	Greece
Attica Bank SA-Bank of Attica SA	Greece
Alpha Bank AE	Greece
Agricultural Bank of Greece	Greece
Bank of Ireland-Governor and Company of the Bank of Ireland	Ireland
Allied Irish Banks plc	Ireland
Unione di Banche Italiane Scpa-UBI Banca	Italy
UniCredit SpA	Italy
Mediobanca SpA	Italy
Intesa Sanpaolo	Italy
Credito Valtellinese Soc Coop	Italy
Credito Emiliano SpA-CREDEM	Italy
Banco Popolare – Societ… Cooperativa-Banco Popolare	Italy
Banca Popolare di Spoleto SpA	Italy
Banca Popolare di Sondrio Societa Cooperativa per Azioni	Italy
Banca Popolare di Milano SCaRL	Italy
Banca popolare dell'Etruria e del Lazio Soc. coop.	Italy
Banca popolare dell'Emilia Romagna	Italy
Banca Monte dei Paschi di Siena SpA-Gruppo Monte dei Paschi di Siena	Italy
Banca Ifis SpA	Italy
Banca Carige SpA	Italy
ING Groep NV	Holland
Banco Espirito Santo SA	Portugal
Banco BPI SA	Portugal

A clear distinction is made between banks who received government financial aid (43) and those that did not get financial support (125). The source of information is Orbis database, which is widely used by other authors to analyse the banking sector through financial ratios (i.e. Gambetta, Zorio-Grima & García-Benau, 2015; García-Posada & Mora-Sanguinetti, 2014).

In order to operate with the data that best anticipates financial distress, we eliminate observations corresponding to the year in which distressed banks receive aid in order to avoid the presence of distorted financial ratios in the model (Table 1). Besides, observations for all years following receipt of aid are eliminated. Finally, the observation prior to the year of aid receipt is considered as that in which the credit institution is in financial distress. After this selection process, the number of valid observations is down to 1026.

Finally, in order to minimise the effect of lack of information available for certain institutions from the sample, additional data filtering is required. First, ratios unavailable for more than 30% of sampled banks have been removed. Accordingly, we are left

Table 2. Explanatory variables.

KEY	Variable
1. Asset Quality	
R1	LOAN LOSS RESERVE/GROSS LOANS
R2	LOAN LOSS PROVISIONS/NET INTEREST REVENUE
2. Operating Ratios	
R3	NET INTEREST MARGIN
R4	NET INT REVENUE/AVERAGE ASSETS
R5	OTHER OPERATING INCOME/AVERAGE ASSETS
R6	NON-INTEREST EXPENSES/AVERAGE ASSETS
R7	PRE-TAX OPERATING INCOME/AVERAGE ASSETS
R8	NON-OPERATING ITEMS & TAXES/AVERAGE ASSETS
R9	RETURN ON AVERAGE ASSETS
R10	RETURN ON AVERAGE EQUITY
R11	NON-OPERATING ITEMS/NET INCOME
R12	COST TO INCOME RATIO
R13	RECURRING EARNING POWER
3. Capital Ratios	
R14	EQUITY/TOT ASSETS
R15	EQUITY/NET LOANS
R16	EQUITY/DEPOSITS & SHORT-TERM FUNDING
R17	EQUITY/LIABILITIES
R18	CAPITAL FUNDS/TOTAL ASSETS
R19	CAPITAL FUNDS/LIABILITIES
4. Liquidity Ratios	
R20	INTERBANK RATIO
R21	NET LOANS/TOTAL ASSETS
R22	NET LOANS/DEPOSITS & SHORT-TERM FUNDING
R23	NET LOANS/TOTAL DEPOSITS & BORROWINGS
R24	LIQUID ASSETS/DEPOSITS & SHORT-TERM FUNDING
R25	LIQUID ASSETS/TOTAL DEPOSITS & BORROWINGS
5. Balance Sheet	
TA	TOTAL ASSETS

Table 3. Bailed-out banks.

Stage	Number of banks	Bailed out	Non-distressed	Ratios	Remaining observations
Initial observations	168	43	125	38	1069
Discard bail-out and subsequent years' observations	168	43	125	38	1026
Discard unavailable observations	155	33	122	25	749

with 25 ratios (see Table 2 and Annex). Second, observations for those entities which are not available for at least 50% of the years under consideration have been eliminated. As a result, the number of institutions in the sample is down to 155 banks, 33 of which received state aid and 122 did not. Ultimately, the number of available observations is reduced to 749 (Table 3).

4. Statistical study and methodology

In our study we use accounting and financial information to achieve a good prediction of banks that failed during the financial crisis in Europe in or after 2007. The objective in this classification question is to determine which variables are involved in the failure or success of a bank. To this end we use a relatively new procedure, BRT, applying the

stochastic gradient boosting algorithm. Boosting is one of the approaches to build ensemble learning models as noted below, but it is a forward, stagewise procedure which makes it an exclusive sequential method. In addition, BRT allows for extreme results, non-linearity, missing data and can handle different types of predictors such as categorical variables.

The aim of this study is to assess whether Friedman's stochastic gradient is appropriate to predict bank failure. Comparing results among alternative ensemble methods or other ML models is not the primary goal of this paper, although two alternative models are considered for comparative purposes.

Boosting is a class of iterative technique that seeks to minimise overall error by introducing additional models based on error from previous iterations. It is particularly useful and frequently used in the context of classification problems. Among the many different boosting methods, the most common are AdaBoost, gradient boosting, and stochastic gradient boosting. These methods use multiple models to produce an aggregate model whose predictive power is better than individual models used alone (Chambers & Dinsmore, 2015). Bagging is another ensemble method whose algorithm selects multiple subsamples from an original training data set, and random forest is a version of bagging that adds a random selection of features or predictors.

According to Kuhn and Johnson (2013), boosting algorithms were developed in the early 1990s when a number of weak classifiers were combined to produce an ensemble classifier with a superior generalised misclassification error rate. AdaBoost provided a practical implementation of boosting a weak learner into a strong learner. Later, the AdaBoost algorithm was connected to statistical concepts such as loss functions, facilitating several algorithmic generalisations to classification problems, which gave rise to the gradient boosting machine of Friedman (2001). Unlike random forest, the trees in boosting are dependent on past trees, have minimum depth and contribute unequally to the final model. Friedman recognised that his gradient boosting machine could be susceptible to over-fitting and he proposed a remedy to constrain the learning process by employing regularisation or shrinkage, which means only a fraction of the current predicted value is added to the previous iteration's predicted value. After Friedman published his gradient boosting machine, he evaluated some of the features of bagging techniques and updated the boosting machine algorithm with a random sampling scheme, and the new procedure was termed as "stochastic gradient boosting".

Boosting-based models are a useful practical tool for different predictive tasks, as they can consistently provide higher accuracy results compared with conventional models (Natekin & Knoll, 2013). Boosting is a method for improving model accuracy, based on the idea that it is easier to find and average many rough rules of thumb than to find a single, highly accurate prediction rule (Schapire, 2003). The boosting machine algorithm we apply in our study uses Friedman's stochastic gradient, which according to Kuhn and Johnson (2013) is now widely accepted as the boosting algorithm of choice among practitioners. The stochasticity improves predictive performance, reducing the variance of the final model by using only a random subset of data to fit each new tree (Friedman 2002a).

BRT is a method that upgrades weak learning algorithms – trees – into strong learning algorithms. The gradient boosting method takes both a loss function and a

weak learner, and the algorithm seeks to find an additive model that minimises the loss function (Kuhn & Johnson, 2013). The idea behind a decision tree is to partition the space of input variables into homogenous rectangular areas by a tree-based rule system (Natekin & Knoll, 2013). In a classification tree, leaves represent the class labels (bank failure or not bank failure) and the results are simple and easy to visualise and are highly interpretable, but suffer from model instability and may not produce optimal predictive performance. Boosting involves producing multiple trees which are then combined to yield a single consensus prediction; and combining a large number of trees can often result in significant improvements in model accuracy (James, Witten, Hastie & Tibshirani, 2013).

To put it another way, BRT involves creating multiple copies of the original training data set using the bootstrap, fitting a separate decision tree to each copy, and then combining all of the trees in order to create a single predictive model. The progress of the trees is sequential, and each iteration adjusts the case weights based on the accuracy of a sample's prediction. In stochastic boosting the construction of each tree depends strongly on the trees that have already been grown (Hastie, Tibshirani & Friedman, 2009). The boosting procedure consecutively fits new models to provide a more accurate estimate of the response variable (Natekin & Knoll, 2013).

Therefore, fitting multiple trees in BRT overcomes the biggest drawback of single-tree models (weak learner): their relatively poor predictive performance (Elith, Leathwick & Hastie, 2008). Unlike fitting a single large decision tree to the data, which amounts to fitting the data hard and potentially over-fitting or overemphasising patterns that are not reproducible and have poor accuracy when predicting a new sample, the boosting approach instead fits a series of small decision trees. The BRT algorithm fits a decision tree to the residuals from the model and adds this decision tree to the fitted function in order to update the residuals. Each of these trees can be rather small, with just a few terminal nodes. By fitting small trees to the residuals, this slowly improves the model (James et al., 2013). The final BRT model is a linear combination of many trees (usually hundreds to thousands) that can be thought of as a regression model where each term is a tree (Elith et al., 2008). Figure 1 summarises the algorithm presented in the stochastic gradient tree-boosting algorithm employed in our study.

1 Select tree depth, D, and number of iterations or trees, K

2 Compute the average response for a baseline model and use this as the initial predicted value for each sample

3 Randomly select a fraction of the training data

4 **for** $k = 1 \ to \ K$ **do**

5 Compute the residual, the difference between the observed value and the *current* predicted value, for each sample

6 Fit a regression tree of depth, D, using the residuals as the response

7 Predict each sample using the regression tree fit in the previous step

8 Update the predicted value of each sample by adding a fraction λ (shrinkage parameter) of previous iteration's predicted value to the predicted value generated in the previous step

9 **end**

Figure 1. Stochastic gradient boosting algorithm. Source: Taken and adapted from Kuhn and Johnson (2013).

As mentioned above, boosting requires a weak learner. We chose trees because they make an excellent learner base for boosting, for several reasons (Kuhn & Johnson, 2013):

- First, they have the flexibility to be weak learners by simply restricting their depth.
- Second, separate trees can easily added together, much like individual predictors can be added together in a regression model, to generate a prediction.
- And third, training time for generated trees is very rapid. Hence, results from individual trees can be directly aggregated thus making them inherently suitable for an additive modelling process.

BRT process requires the specification of three important parameters. First, the learning rate or shrinkage parameter (λ), which is generally a low positive number. This determines how quickly the algorithm adapts or the contribution of each tree to the growing model, and can take values between 0 and 1. Ridgeway (2009) suggests that small values of the learning parameter (<0.01) work best. This reduces the size of incremental steps and thus penalises the importance of each consecutive iteration. The intuition behind this technique is that it is better to improve a mode by taking many small steps than by taking fewer large steps. (Natekin & Knoll, 2013). To achieve good performance, a very small value of this parameter can require a very large number of trees, increasing computation time in finding an optimal model (the value of the parameter is inversely proportional to the computation time required). Tree depth (D), also known as interaction depth, is the second parameter, and is the number of splits in each tree controlling the complexity of the boosted structure; the third parameter is the number of trees in the model (K). The optimal number of trees in a BRT model can be determined by the first two parameters or by means of cross-validation methods.

It is very important to decide on the number of trees. As James et al. (2013) emphasise, combining a large number of trees can often result in dramatic improvements in training accuracy, at the expense of some loss in prediction accuracy. In other words, using a very large tree value in a BRT model might lead to over-fitting.

In our study, to predict bank failure we build a classification model to predict a qualitative response variable consisting of banks that failed during the financial crisis in Europe in or after 2007. It is based on the BRT technique explained above and computes the probability that a bank failure occurs. In successive boosting steps, the algorithm seeks to find an additive model that minimises the loss function (deviance), learning from the relationship between the response and its predictors and then improving the banking classification. The explanatory variables are a set of accounting and financial variables, most in the form of ratios (see Table 2).

All models were fitted in *R* (R Core Team, 2015) version 3.1.3, and for BRT using *gbm* package version 2.1 (Ridgeway, 2015), *dismo* package version 1.0-12 (Hijmans, Phillips, Leathwick & Elith, 2015) and *caret* package version 6.0-41 (Kuhn, 2015).

5. Results

Before proceeding with the analysis we checked for the presence of both multicollinearity and high correlations between variables through the matrix values of Pearson's correlation coefficient. We observed high correlations between some of these variables, which made us drop certain ratios.[2] Therefore, the analysis does not take into account the following ratios of Table 2: *R4, R9, R16, R17, R18, R19, R21* and *R25*.

Regarding the descriptive statistics provided in Table 4, we first focus on median values, which are not unduly affected by outliers. Note that there are no major differences in most median values for the two subsets of banks, failed and non-failed, indicating that we are dealing with a challenging classification problem. There are only two variables that show noteworthy differences, "Total Assets" and "Non-operating items/Net income", which will later be identified as relevant variables by the BRT model.

Except for "Total Assets" and "Non-operating items/Net income", the standard deviation is higher in the case of banks that did not receive government financial aid, which is indicative of their greater dispersion due, to a large extent, to the larger number of banks that make up the non-failed banks subset. The same explanation applies to the minimum and maximum values displayed in Table 4.

5.1. *Tuning parameter selection*

Almost all predictive modelling techniques have tuning parameters that enable the model to flex to find the structure in the data. An important concern about building a model is the its resulting generalisation capabilities and, if the model is not built properly, it can easily over-fit the data. Hence we must use the existing data to identify settings for the model's parameters that yield the best and most realistic predictive

Table 4. Summary of descriptive statistics of data by bank failure.

Variable	Mean		SD		Median		Min		Max	
	NF	F	NF	F	NF	F	NF	F	NF	F
R1	3.47	2.66	3.93	1.60	2.55	2.31	0.10	0.36	43.97	5.79
R2	22.40	23.78	46.24	15.93	16.05	21.24	−400.00	1.44	373.08	76.87
R3	1.98	1.87	2.77	0.88	1.83	1.84	−25.68	0.31	23.53	4.31
R5	6.82	0.97	17.38	0.50	1.40	0.92	−3.24	−0.03	142.42	2.36
R6	7.22	2.11	16.38	0.83	2.62	2.09	−1.12	0.68	128.44	4.02
R7	1.42	0.59	4.78	0.57	0.95	0.56	−35.71	−0.85	37.13	2.02
R8	−0.57	−0.19	2.22	0.32	−0.27	−0.13	−19.90	−1.28	21.24	0.52
R10	6.96	7.18	16.64	10.54	8.03	7.38	−166.00	−17.10	51.47	25.34
R11	2.47	7.43	92.78	97.44	0.16	5.61	−583.15	−227.93	702.47	346.42
R12	65.32	65.27	26.28	19.06	60.92	62.75	10.68	40.10	352.31	148.70
R13	1.88	1.02	4.38	0.60	1.32	1.01	−35.71	−0.28	37.13	2.56
R14	15.22	5.70	18.57	2.27	8.21	6.04	−0.97	1.45	91.38	9.64
R15	53.63	10.12	119.95	3.24	13.50	9.81	−1.52	2.90	950.00	17.91
R20	130.04	56.09	187.75	42.82	55.00	40.63	0.87	12.15	989.36	191.37
R22	89.05	104.32	66.10	34.50	88.68	98.70	0.00	47.08	861.17	186.48
R23	67.98	67.69	42.63	20.49	73.41	74.74	0.00	13.57	600.00	94.33
R24	58.00	34.07	119.98	32.76	23.18	21.27	0.82	8.64	971.26	149.63
TA	101.04	463.63	309.12	671.95	11.70	141.73	0.01	4.41	3065.09	2494.41

F: Failure: banks that received government financial aid.
NF: Non-failure: banks that did not receive government financial aid.
Variables are defined in Table 2 (TA in millions of euros).
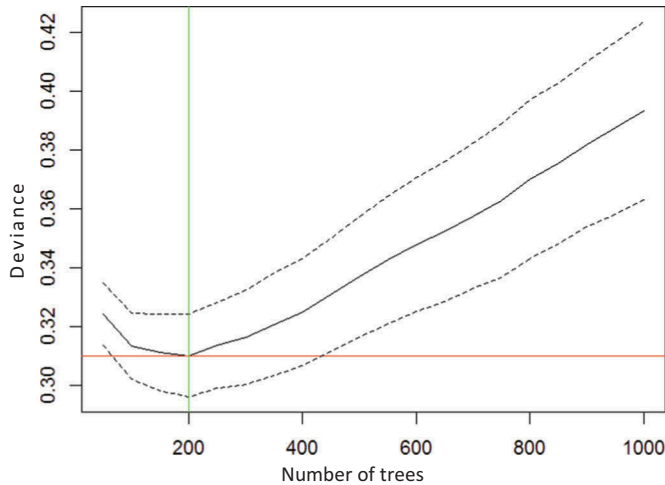
Figure 2. Number of trees for Model 1 of bank failure. Learning rate = 0.01. Number of split in each tree = 5. Estimated tenfold cross-validation deviance = 0.310. Optimal number of trees = 200.

performance (Kuhn & Johnson, 2013). We used all the data to identify settings for the model parameters and we start by fitting Model 1, a bank failure model with a learning rate of 0.01 and 5 splits in each tree as a first estimation, using the rules of thumb discussed in Elith et al. (2008). Our objective is to find the combination of the three important parameters mentioned above (learning rate, number of splits and number of trees) that achieves minimum error for predictions to test samples (minimum predictive error).

Controlling these parameters is particularly important for BRT because its sequential model fitting allows trees to be added continuously and then over-fitting of the training data. It made a model with 200 trees (Figure 2), which was performed with tenfold cross-validation in order to determine the optimal number of trees. The cross-validation method for identifying the optimal number of trees in our boosted classification tree model works as follows, using deviance reduction as the measure of success:

(1) Split the data randomly into 10 disjoint non-overlapping subsets of approximately equal size.
(2) Make 10 different training sets each comprising a unique combination of nine groups. Therefore, for each training set there is a held-out group that acts as a validation set.
(3) Fit 10 BRT models simultaneously on each training set, starting with a number of 50 trees. Obtain validation predictive performance on their respective held-out groups (hold-out deviance). The validation performance is estimated by averaging the 10 resulting performances estimates (average hold-out deviance).
(4) Increase the number of trees in each model by 50 and repeat step (3).
(5) The process identifies the optimal number of trees as that at which the hold-out deviance is minimised.

The solid black curve in Figure 2 denotes the deviance mean for the excluded or hold-out folds of the cross-validation and for each number of possible trees, and the dotted curves denote the interval of approximately one standard error for the changes in predictive deviance (loss function). The horizontal red line denotes the minimum of the deviance mean (0.310), and the vertical green line the optimal number of trees at which that arises. Above the optimal number of trees (200), predictive deviance increases indicating that over-fitting is occurring. The deviance is a measure that plays the role of residual sum of squares for a broader category of models such as tree-based classification methods; the lower the deviance, the better the fit; that is, a small deviance indicates a model that provides a good fit to the data.[3]

We will not modify the chosen number of five splits in each tree for the rest of the tuning parameter selection process because, as Natekin and Knoll (2013) argue, there is much evidence that even complex models with rich tree structure (interaction depth >20) provide almost no benefit over compact trees (interaction depth ≈ 5). Therefore, we only consider a trade-off between the number of trees and learning rate; and to deal with this trade-off we rely on the cross-validation procedure.

The foregoing is a first estimate of settings, and the next step is to obtain a better model by means of reducing the learning rate, as suggested by Elith et al. (2008). Decreasing the learning rate increases the number of trees required, and in general a smaller learning parameter and larger number of trees are preferable because this decreases the contribution of each tree and the final model better predicts the response. In fact, the best strategy appears to be to set the learning rate very low (<0.1). This yields important improvements for regression and for probability estimation; the corresponding improvements in misclassification are less, but still considerable. The price paid for these improvements is computational: lower values of learning rate give rise to higher values of boosting iterations, and computation is proportional to the latter (Hastie et al., 2009).

Reducing learning rate, setting a value of 0.005 and keeping tree complexity with five splits in each tree provides a model with 400 trees (Model 2, see Figure 3). This was again performed with tenfold cross-validation in order to determine the number of trees. According to the horizontal red and vertical green lines (see above), the optimal number of trees arises with a hold-out deviance of 0.304, which is lower than that of Model 1.

Models with too many variables might over-fit sample data and thus out-of-sample predictions might be worse than indicated by in-sample performance and be more susceptible to invalid predictions, reducing their generality. Controlling the number of variables implies that a parsimonious model provides better accuracy in predictions. Therefore, we dropped different combinations of variables of Model 2 and evaluated the reduction in deviance, with the aim of obtaining a simplified and better model in terms of prediction. Particularly, and as Hijmans et al. (2015) mention, the R function *gbm. simplify* takes the cross-validated model and then assesses the potential to remove predictors using k-fold cross-validation. This is done for each fold, removing the lowest contributing predictor and repeating this process for a set number of steps. After the removal of each predictor, the change in predictive deviance is computed relative to that obtained when using all predictors. The results show the ideal number of variables
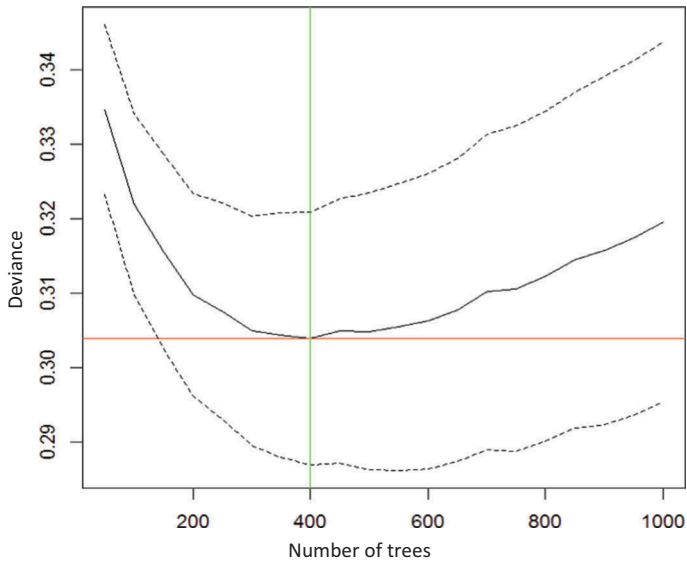
Figure 3. Number of trees for Model 2 of bank failure. Learning rate = 0.005. Number of split in each tree = 5. Estimated tenfold cross-validation deviance = 0.304. Optimal number of trees = 400.
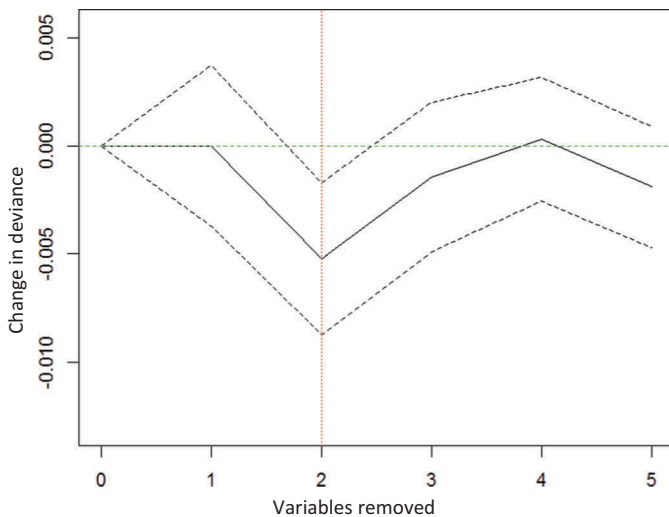
Figure 4. Number of variables to remove. Assessing the potential to remove predictors using tenfold cross-validation change in predictive deviance. R8 and R13 are dropped.

to be dropped is two: *R8* and *R13* (see Figure 4). After dropping these two variables Model 3 is fitted, resulting in 400 trees and a lower deviance than Model 2.

We did not test learning parameters lower than 0.005 because very low values of the regularisation of this parameter will increase awareness of over-fitting (Natekin & Knoll, 2013)

5.2. *Variable importance*

Variable importance can be computed for BRT by assessing the overall improvement in the optimisation criteria for each predictor. This measure is based on the number of times a variable is selected for splitting. To obtain the overall influence of a variable in the model, this influence should be averaged over all trees. The influences are further standardised so that they add up to 100% (Natekin & Knoll, 2013).

As illustrated in Figure 5 the most relative important variables for Model 3 with a higher relative influence on the response variable are *logA, R20, R11* and *R22*, which are expressed relative to the maximum (the contribution of each variable is scaled so that the sum of all variables adds to 100 and higher numbers indicate a stronger influence on bank failure). Although the collection of boosted trees is much more difficult to interpret than a single tree, it is possible to obtain an overall summary of the importance of each predictor using the Gini index, and a high value indicates an important predictor. Variable importance is computed adding up the total amount Gini index decreases by splits over a given predictor, averaged over the total number of bootstrapped sets of trees used.

Additionally, Figure 6 shows the effect of the four most important predictors identified on the response variable, bank failure. These are partial dependence plots which depict the marginal effect of a variable on bank failure probability. Or, to put it another way, each figure illustrates the partial dependence of the model on a predictor variable, after accounting for the average joint effect of the other predictors; and while this may not provide a comprehensive description, it can show general trends and provide a useful basis for interpretation (Friedman, 2002b). Natekin and Knoll (2013) state these graphs might not be a perfect representation of the captured effects; however, partial dependence plots can provide a useful basis for interpretation that has been found practical in different applications.
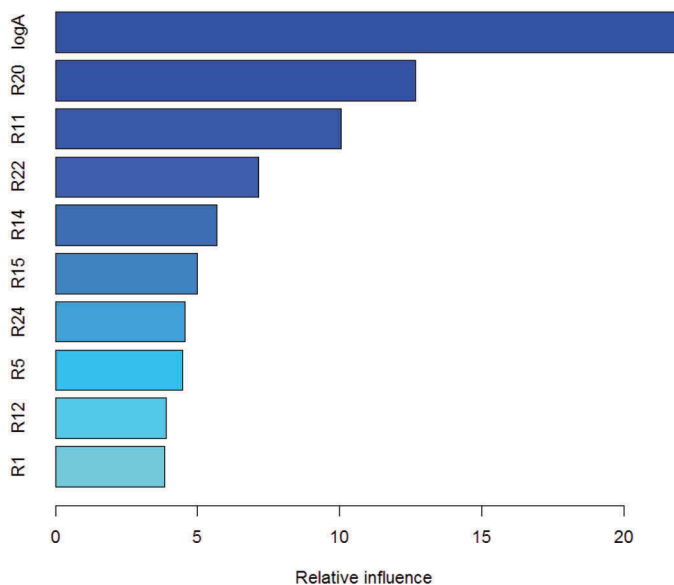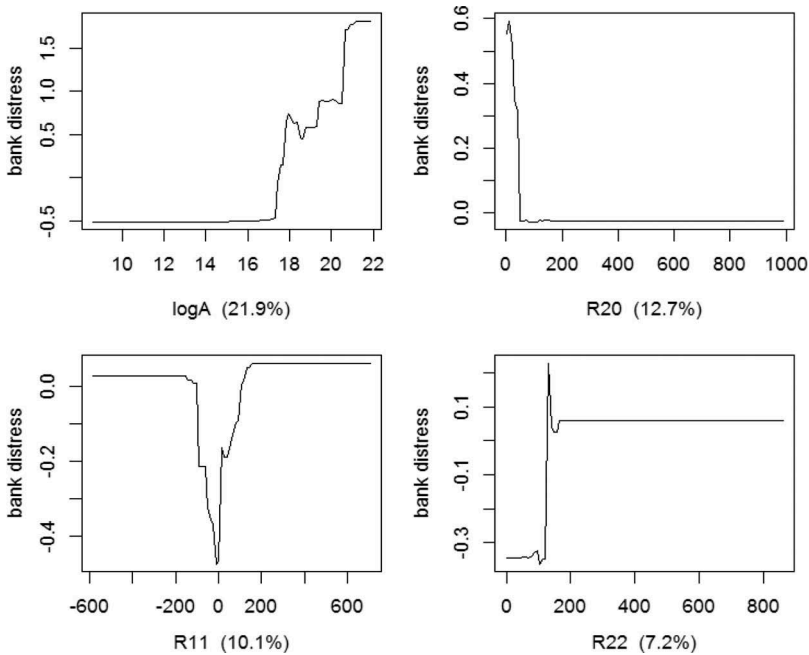

Figure 5. BRT. Variable importance plot for bank failure.

Figure 6. Partial dependence plots for the four most influential variables. logA, natural logarithm of total assets; R20, interbank ratio; R11, non-operating items divided by net income; R22, net loans/ deposits & short-term funding.

Bank failure is best explained by four influential variables, indicating that there is a positive relationship between bank failure and the natural logarithm of total assets, ratios *R11* and *R22*. Accordingly, the higher the values of "Total Assets", "Non-operating items to net income" and "Net loans to deposits & short-term funding", the higher the chance of bank failure. Conversely, there is a negative relationship for ratio *R20*, meaning that the lower the "Interbank ratio" the higher the likelihood of bank failure. Note that these results are consistent with the economic interpretation of the ratios.

5.3. Model accuracy

The boxplot of the predicted values of the BRT model for bank distress shows that these values are higher in cases of bank distress and lower in cases of non-distress, although some parts at the end of the boxplots overlap each other (Figure 7, left panel).

Therefore, there is no perfect probability cut-off between the groups and it is not possible to draw a horizontal line that completely separates the predicted probability of bank failure, that is, the estimated model does not produce a perfect prediction. Nevertheless, and as displayed in Figure 7 (right panel), the lowest number of mis-classifications is found for a threshold value of 0.20.[4]

Table 5 shows the confusion matrix of the prediction results when choosing a threshold of 0.20. The table layout allows visualisation of the performance of the BRT technique used to fit the model, comparing the actual outcomes to the predicted ones.
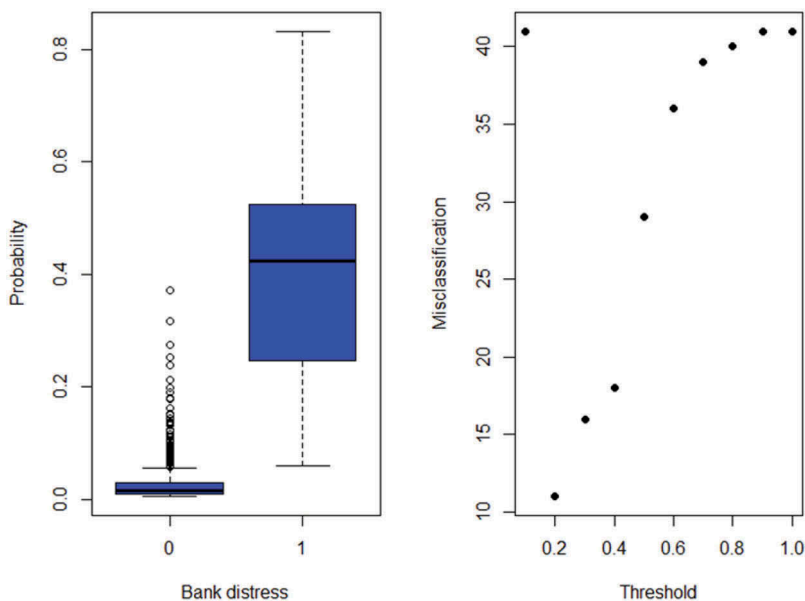
Figure 7. Best threshold for predicting bank distress. Bank distress is a dummy variable taking a value of 1 if a bank has received state support, and 0 otherwise.

**Table 5.** BRT confusion matrix for bank distress.

		Predicted		
		Distress		
Actual		No	Yes	Accuracy
Distress	No	**706**	10	98.60%
	Yes	4	**29**	87.87%
Total				98.13%

Results for a threshold value of 0.20.
N = 749.

The probability obtained by the BRT model is transformed into a binary prediction that is 1 if it exceeds the threshold of 0.20, and zero otherwise. The global percentage of errors or the actual cases incorrectly classified rose to 1.87% (100–98.13). which is a very low rate. The model has a sensitivity rate of 87.87% (true positive rate) and a specificity rate of 98.60% (true negative rate).

Additionally, in order to measure the overall performance of our model, Figure 8 shows the ROC[5] curve for the BRT classifier on all banking data, a graphic for simultaneously displaying the two types of error for all possible thresholds. That is, given the probability of bank failure, there are many alternative thresholds that can be evaluated (not just a 50% threshold); and for each threshold we can calculate the sensitivity and specificity. The overall performance of a classifier, summarised over all possible thresholds, is given by the area under the ROC curve (AUC). An ideal ROC curve will hug the top left corner, so the larger area under the ROC curve the better the classifier. The AUC statistic is used for a quantitative assessment of the model, and for
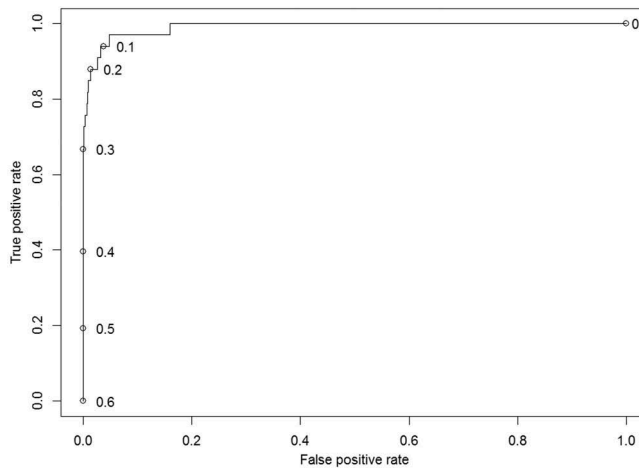
Figure 8. ROC Curve for BRT classifier. Numbers alongside the ROC Curve show the different thresholds. AUC is 0.99.

the data analysed the AUC is 0.99, very close to the maximum of 1, indicating that this is a high-performing classification model.

5.4. *Model prediction assessment*

Model assessment is the process of evaluating a model's performance. Many modern classification and regression models are highly adaptable: they are capable of modelling complex relationships. However, they can very easily overemphasise patterns that are not reproducible (Kuhn & Johnson, 2013). In our BRT model all the results shown are computed on the training data set, that is, all available banking data and, as shown above, the model fits very well but predictive performance should not be estimated on training data. The following step is to evaluate the performance of the model on the test data set computing the test error rate. The test error is the average error that results from using a statistical learning method to predict the response to a new observation, that is, a measurement that was not used in the training data set. Given a data set, the use of a particular statistical learning method is appropriate if it results in a low test error (Hastie et al., 2009). The test error can be calculated if a designated test set is available, but in our case all banking data is used to fit or train the model and there is no hold-out data set or testing set. Therefore, we estimate the power or accuracy of the BRT model by means of an alternative technique such as cross-validation (already used as discussed above for model development or parameter tuning). This method is utilised for estimating the accuracy of data test using the training data, that is, cross-validation can be used to estimate the test error associated with a given statistical learning method in order to evaluate its performance.

Cross-validation provides a means for testing the model on withheld portions of data, while still using all data at some stage to fit the model (Elith et al., 2008). Table 6 displays the results of a tenfold cross-validation on the training data where the global final misclassification error is 5.88% (accuracy 94.12%). Although this percentage is not

**Table 6.** Tenfold cross-validation to estimate model accuracy.

| Fold	Error
1	4.000%
2	4.000%
3	8.000%
4	4.000%
5	6.667%
6	8.000%
7	2.667%
8	6.667%
9	5.333%
10	9.459%
Mean	5.879%

as low as the 1.87% obtained above using all the training data, we can clearly state that the estimated validation accuracy is quite good.

We also performed a tenfold cross-validation for AUC on the training data, and the score was 0.84. Though this result is quite high it is lower, as expected, than the 0.99 obtained for all training data as shown in Figure 8. Leathwick, Elith, Francis, Hastie and Taylor (2006) point out that while over-fitting is often seen as a problem in statistical modelling, prediction with BRT to independent data is not compromised and is generally superior to other methods.

6. BRT comparison of two and three years before bank failure

In order to test the accuracy of our final model, we fitted two more BRT, in the two-years-before-failure and the three-years-before-failure settings. Figure 9 displays variable importance relative to the maximum contributor. While the resulting combination of relevant variables is different from the one-year-before-failure model, note that the size variable is present in all three models. However, the size variable becomes less relevant in the three-years-before-failure model.
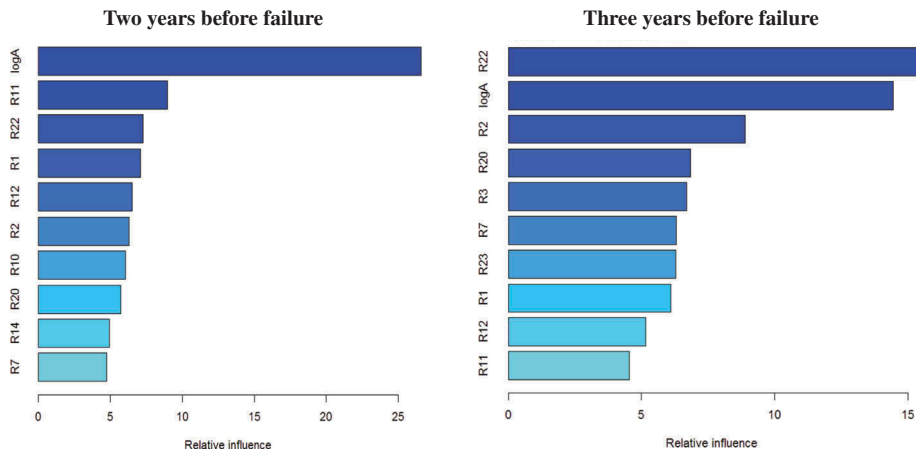
Figure 9. BRT. Variable importance two and three years before bank failure.
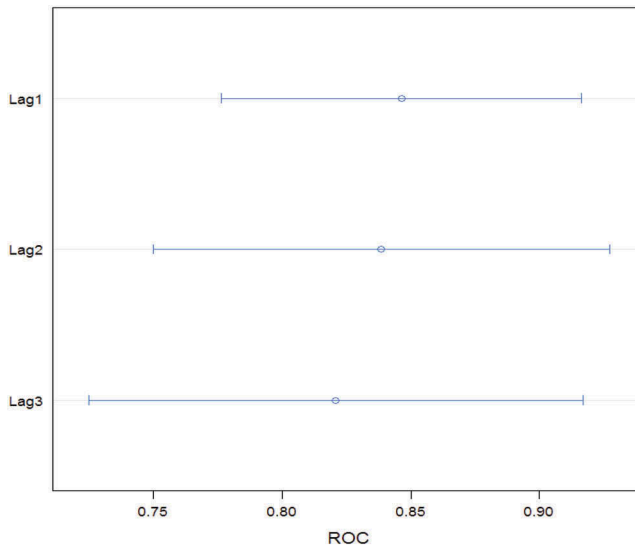
Figure 10. Confidence intervals for the AUC (ROC) performance distribution. Lag1, one year before bank failure; Lag2, two years before bank failure; Lag3, three year before bank failure.

The ROC curve is used for a quantitative assessment of these two new models. For the former the cross-validation AUC was 0.83 and for the latter 0.82, which are lower than for one-year-before-failure BRT (0.85). However p-values assessing possible differences among the three models are too high, suggesting none has a significantly better AUC performance metric and that the new models fail to indicate any difference in performance. The 95% confidence interval for the AUC performance distributions in Figure 10 indicates they are very similar indeed.

7. Comparison to other methodologies

In order to explain and assess the contribution of our study, in this section the results discussed above are compared to other approaches to business failure prediction. More specifically, we focus on two different methodologies: logistic regression and lasso regression. We compared the results to two other accepted ML algorithms and the results show that BRT achieves a better model performance. Comparing results to other ML methods is beyond the scope of this paper.

7.1. *Logistic regression*

Logistic regression is widely applied to bankruptcy prediction. In this section we follow a conventional approach in the literature, which is about observing distress conditions in banks employing a logit design. Logistic regression is a linear model that has important advantages in terms of inference and is often very competitive in relation to non-linear methods, particularly in terms of model interpretability.

Initially there are 26 explanatory variables (see Table 2) that can predict bank failure, but as mentioned above we dropped eight of them because we observed high correlations.

Table 7. Logistic regression for bank distress.

	Estimate	Std. Error	z-value	Pr(>|z|)	
Constant	−7.643	2.754	−2.775	0.006	**
R3	1.333	0.442	3.019	0.003	**
R13	−1.377	0.545	−2.526	0.012	*
R15	−0.185	0.072	−2.57	0.010	*
R20	−0.011	0.005	−2.308	0.021	*
R24	0.026	0.012	2.172	0.030	*
logA	0.347	0.141	2.462	0.013	*

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1.
AIC: 181.51.
Variables are defined in Table 2.

It is likely that some of these 18 remaining variables used in a logistic regression model are in fact not associated with bank failure. Including inappropriate variables could produce an unnecessarily complex model and introduce noise to the results. By eliminating these unnecessary variables, it is possible to achieve a simpler model that is easier to interpret. In this section we automatically perform a variable selection, excluding irrelevant predictors, in order to obtain the best parsimonious logistic regression model. As stated by Calcagno and Mazancourt (2010), it is often the case that you are bound to assess what terms in the model formulation are important to describe the dependent variable; that is, which one should remain, and which one could be dropped.

Following a subset selection approach, the objective consists of fitting a model using least squares on a minimised set of variables. It requires identifying the subset of the explanatory variables that is best related to the prediction of bankruptcy. For this purpose, *bestglm* package version 0.34 for R (McLeod & Xu, 2014) provides a convenient algorithm that selects the best logistic regression model according to a specified metric. A summary of the resulting model is reported in Table 7. The best fit was found using the Akaike information criterion (AIC). Unlike BRT, logistic regression does not handle missing values and thus after dropping these incomplete observations we have a total of 348 available data for the analysis.

The variables *logA, R3, R13, R15, R20* and *R24* are statistically significant at the 5% level. *logA* (total assets) is one of BRT four main predictors (see Figure 5), and its positive sign as well as the sign of the rest of significant variables are consistent with their economic interpretation.

Figure 11 displays the overall performance of the logistic regression on training data; it shows an AUC of 0.83 (left panel). For the optimal threshold of 0.131 the global accuracy of model goes up to 77.87%, as illustrated by the confusion matrix of the logistic regression prediction of banking failure (right panel). We also completed a tenfold cross-validation procedure to estimate the out of sampling model's performance, and it is 0.77 for AUC and 76.16% for accuracy, which signal lower predictive power results.

7.2. *Lasso regression*

Lasso regression introduces a shrinkage penalty (λ) that decreases the coefficient estimates of a linear regression towards zero, performing as a result a variable selection. The penalty has the effect of making some coefficient estimates zero when it is very
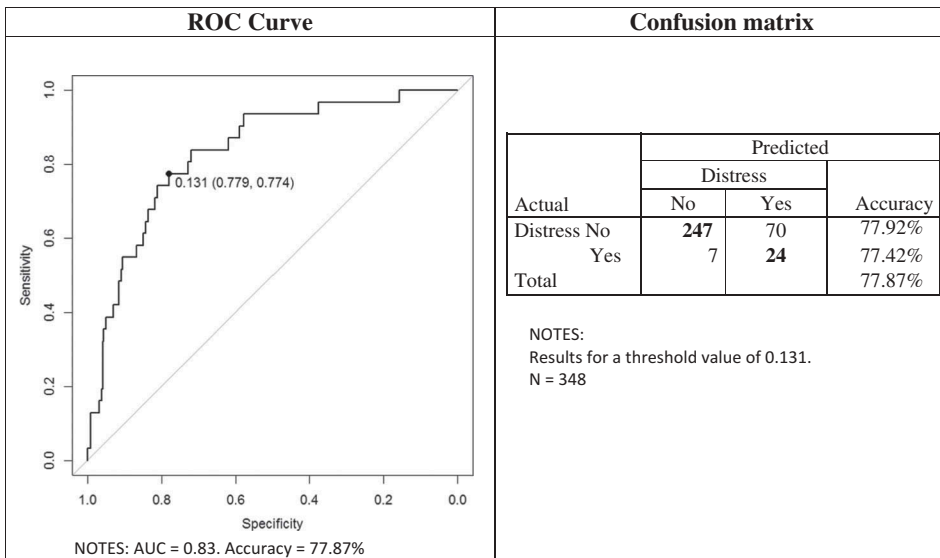
ROC Curve	Confusion matrix

Confusion matrix:

		Predicted		
		Distress		
Actual	No	Yes	Accuracy
Distress No	**247**	70	77.92%
Yes	7	**24**	77.42%
Total			77.87%

NOTES:
Results for a threshold value of 0.131.
N = 348

NOTES: AUC = 0.83. Accuracy = 77.87%

Figure 11. Logistic regression performance on training data.

high. As with least squares, lasso regression seeks coefficient estimates that fit the data well by making the residuals small but penalising the number of non-zero coefficients (James et al., 2013). Therefore, models generated from lasso are generally much easier to interpret than those created by linear regression, producing parsimonious models that involve only a subset of the variables (Hastie et al., 2009). In order to obtain a good set of bank failure predictors it is very important to determine correctly the tuning parameter λ, and for this purpose we utilise cross-validation. A simple model is more interpretable because involves only a subset of the explanatory variables that can predict bank failure, and a complex model always performs worse for generalisation purposes.

Lasso is carried out using *R caret* package version 6.0-41 (Kuhn, 2015). Performing tenfold cross-validation reveals $\lambda = 0.013$ as the optimal value to obtaining a good model fit. The results in Figure 12 indicate that *R3, logA, R14* and *R13* are the most important variables identified through lasso regression. Many of these predictors of bank failure and their signs are consistent with those detected using BRT and logistic regression.

The overall performance of lasso regression on training data is shown in Figure 13. The left panel shows an AUC of 0.81 and, with a 0.115 model threshold, the confusion matrix indicates a global accuracy of 72.13% for banking failure prediction (right panel). In order to estimate the out-of-sample performance of lasso regression we also carry out tenfold cross-validation, obtaining 0.77 for AUC and 70.98% for accuracy.

In summary, according to the results – summarised and displayed in Table 8 – in regard to bank FDP, BRT outperforms other established classifier techniques such as logistic regression or lasso regression.
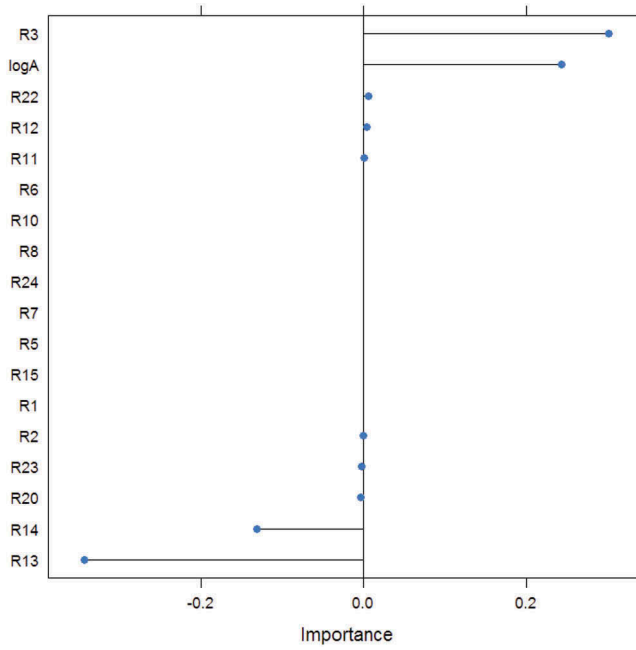
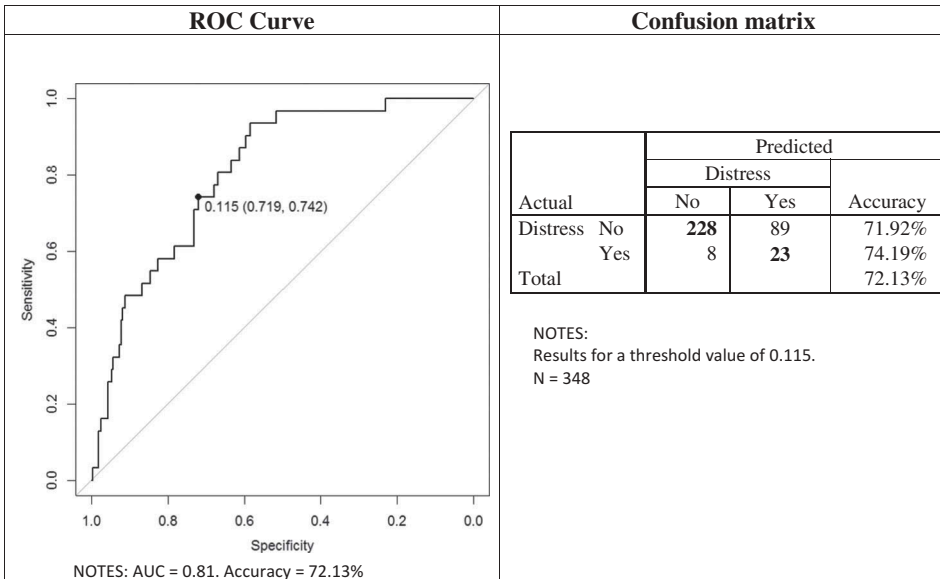**Figure 12.** Lasso. Variable importance plot for bank failure.

ROC Curve	Confusion matrix

ROC Curve plot with point labeled 0.115 (0.719, 0.742). NOTES: AUC = 0.81. Accuracy = 72.13%

		Predicted		
		Distress		
Actual		No	Yes	Accuracy
Distress	No	**228**	89	71.92%
	Yes	8	**23**	74.19%
Total				72.13%

NOTES:
Results for a threshold value of 0.115.
N = 348

Figure 13. Lasso regression performance on training data.

Table 8. Summary of model performance for bank distress.

Model	Performance on training data		Out of sample performance	
	AUC	Accuracy	AUC	Accuracy
Boosted Regression Trees	0.99	98.13%	0.84	94.12%
Logistic Regression	0.83	77.87%	0.77	76.16%
Lasso Regression	0.81	72.13%	0.77	70.98%

Out-of-sample performance estimated by tenfold cross-validation.

8. Conclusions

The main objective of this paper is to help predict and prevent failure in the banking sector. To this end, the empirical analysis performed with the boosted regression trees (BRT) methodology (a machine learning approach) identifies a set of key leading indicators that, considered jointly, may be helpful in anticipating and averting bank financial distress.

The BRT method combines the strengths of two algorithms: regression trees models that relate a response to their predictors by recursive binary splits, and boosting, an adaptive method for combining many simple models to give improved predictive performance. BRT has rarely been used in bank failure analysis, even though there is clear evidence of strong predictive performance and consistent identification of relevant variables and interactions.

Here, we focus on the discussion of the outcomes of the optimal performing model: the one-year-before-failure BRT model. It is remarkable that out of the 18 explanatory variables analysed, only four showed a joint relative importance of greater than 50% on the response variable, that is, bank financial distress. Accordingly, the interbank ratio has an inverse relationship with the possibility of being bailed out, indicating that relying heavily on costly interbank market loans to finance operations may be a strong indicator of potential financial difficulties. Also, the greater the size of the bank in terms of total assets, the greater the likelihood of potential financial distress. Likewise, the higher the net loans to deposits and income from non-traditional banking activities, the higher the chance of future financial distress.

9. Managerial implications

With regard to the applied implications of our study, we focus on the one-year-before-failure BRT model. The four most influential variables in predicting bank failure were identified in this study, and these are all directly related to key aspects of the risk management control function in financial institutions. Constant follow-up of these key scorecard variables may facilitate prompt recognition of risk management control failures and eventually avoid bank failure.

Banks that finance their operations by taking loans from other banks (low interbank ratio values) are using an expensive source of funds, very sensitive to financial crises, that can freeze interbank markets and cut off funding. In addition, constant borrowing of funds from other banks may signal a bank's incompetence to finance through ordinary clients' deposits, which is traditionally the primary source of funds.

Obtaining a significant percentage of net income from unusual (non-operating items) sources of income might improve the bottom line in a given year, but it is often the result of a one-off event that does not represent a steady and sustainable form of income. Relying too much on these erratic sources of income does not ensure long-term financial soundness.

Non-systemic institutions have an incentive to reach systemic importance, thus leading banks to expand beyond their ideal economic size. In addition, large banks are more willing to take higher risks than small financial institutions. As large banks have access to a wider variety of sources of funding than small ones, they are more prone to lending money without appropriate risk control and assessment.

Banks with net loans exceeding client deposits tend to resort to more expensive sources of funds to finance their lending activities. In turn, this expensive funding will have a direct impact on the banks' bottom line.

Finally, the model presented should allow managers in financial institutions to take advantage of the high predicting power of the four key scorecard variables identified and keep a permanent watch on these relevant predictors of bank failure. Managers may be able to avert financial distress before it happens by taking actions on their own, rather than waiting for government intervention. Regulators, on their part, may find the model useful in identifying and giving early warning to potentially distressed banks.

10. Future research

This study opens some future research prospects. The application of the boosted classification tree methodology to a broader Eurozone sample comprising both listed and non-listed financial institutions may enhance the reliability of results. Additionally, extending this study to other non-Eurozone countries may reveal failed financial institution-specific traits across countries.

Regarding the methodology applied in this study, note that "Boosted Classification Tree" is just one machine learning model, and there are a number of alternative machine learning models that may be appropriate to approach banking failure prediction: "Extreme Gradient Boosting", "Neural Network Models", "Support Vector Machines", "Random Forest Models" or "Discriminant Analysis Models." The suitability of these machine learning models should be explored in future research works.

11. Limitations

The inclusion of some macroeconomic variables in the model might have improved the results. However, given the high predicting power of the model, the inclusion of macroeconomic variables may cause an over-fitting problem with no significant improvement.

Another limitation is the nature of the sample: as we included only listed banks in the sample there is a large number that are left out. However, note that most non-listed banks are either mid-sized or small financial institutions.

There are different ways to classify distressed banks, but we deem the reception of government financial aid as the most reliable factor to identify failed banks. This is especially true in the Eurozone case, where virtually all distressed banks have been bailed out by governments.

Notes

1. An acronym for Capital adequacy, Asset quality, Management quality, Earnings, Liquidity and Sensitivity to market risk.
2. The correlation table is not reported here to save space but is available upon request.
3. A saturated model is a perfect model with as many parameters as observations that explains all variance in the response. Deviance measures the difference in "fit" of a candidate model and that of the saturated model and represents the loss in predictive performance due to a suboptimal model.
4. We could have used a 0.50 cut-off value, which is very commonly used, but it is not necessarily the best depending on the classification model.
5. Acronym for receiver operating characteristics.

Acknowledgement

Disclosure statement

Funding

References

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, *45*, 110–122. doi:10.1016/j.dss.2007.12.002

Assadian, A., & Ford, J. M. (1997). Determinants of business failure: The role of firm size. *Journal of Economics and Finance*, *21*, 15–23. doi:10.1007/BF02929018

Betz, F., Oprică, S., Peltonen, T. A., & Sarlin, P. (2014). Predicting distress in European banks. *Journal of Banking & Finance*, *45*, 225–241. doi:10.1016/j.jbankfin.2013.11.041

Calcagno, V., & Mazancourt, C. (2010). Glmulti: An R package for easy automated modelselection with (generalized) linear models. *Journal of Statistical Software*, *34*(12), 1–29. doi:10.18637/jss.v034.i12

Chambers, M., & Dinsmore, T. W. (2015). *Advanced analytics methodologies: Driving business value with analytics*. Upper Saddle River, NJ: Pearson Education.

Cortés, E. A., Martínez, M. G., & Rubio, N. G. (2008). Linear discriminant analysis versus adaboost for failure forecasting. *Revista Española De Financiación Y Contabilidad*, *XXXVII* (137, Enero–Marzo), 13–32. doi:10.1080/02102412.2008.10779637

Demyanyk, Y., & Hasan, I. (2010). Financial crises and bank failures: A review of prediction methods. *Omega*, *38*, 315–324. doi:10.1016/j.omega.2009.09.007

DeYoung, R., & Torna, G. (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation*, *22*, 397–421. doi:10.1016/j.jfi.2013.01.001

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*, 802–813. doi:10.1111/j.1365-2656.2008.01390.x

Ellul, A., & Yerramilli, V. (2013). Stronger risk controls, lower risk: Evidence from U.S. Bank Holding Companies. *The Journal of Finance*, LXVIII(5), 1757–1803. doi:10.1111/jofi.12057

European Commission. (2012). *Commission staff working document*. Brussels: Summary of the Impact Assessment. 6.6.2012; SWD (2012) 167.

European Commission Competition. (2011, October). *The effects of temporary state aid rules adopted in the context of the financial and economic crisis* (Commission Staff Working Paper). Brussels: Author.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. doi:10.1214/aos/1013203451

Friedman, J. H. (2002a). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. doi:10.1016/S0167-9473(01)00065-2

Friedman, J. H. (2002b). Tutorial: Getting Started with MART in R Retrieved from http://statweb.stanford.edu/~jhf/r-mart/tutorial/tutorial.pdf

Gambetta, N., Zorio-Grima, A., & García-Benau, M. A. (2015). Complaints management and bank risk profile. *Journal of Business Research*, 68, 1599–1601. doi:10.1016/j.jbusres.2015.02.002

García-Posada, M., & Mora-Sanguinetti, J. S. (2014). Are there alternatives to bankruptcy? A study of small business distress in Spain. *Journal of the Spanish Economic Association*, 5, 287–332.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. New York, NY: Sprinter.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2015). Dismo: Species distribution modeling. R package version 1.0-12. Retrieved from http://CRAN.R-project.org/package=dismo

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (2nd ed.). New York, NY: Springer.

Jin, J. Y., Kanagaretnam, K., & Lobo, G. J. (2011). Ability of accounting and audit quality variables to predict bank failure during the financial crisis. *Journal of Banking & Finance*, 35, 2811–2819. doi:10.1016/j.jbankfin.2011.03.005

Kim, M. J., Kang, D. K., & Kim, H. B. (2015). Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications*, 42(3), 1074–1082.

Kuhn, M. (2015). Caret: Classification and Regression Training. R package version 6.0-41. Retrieved from http://CRAN.R-project.org/package=caret

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Marine Ecology Progress Series*, 321, 267–281. doi:10.3354/meps321267

McLeod, A. I., & Xu, C. (2014). Bestglm: Best Subset GLM. R package version 0.34. Retrieved from http://CRAN.R-project.org/package=bestglm

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7(21). doi:10.3389/fnbot.2013.00021

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ridgeway, G. (2009). Generalized Boosted Models: A Guide to the gbm Package. Retrieved from https://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/inst/doc/gbm.pdf?revision=18&root=gbm&pathrev=22

Ridgeway, G. (2015). Gbm: Generalized Boosted Regression Models. R package version 2.1.1. Retrieved from http://CRAN.R-project.org/package=gbm

Savona, P., De Lisa, R., Dorodnykh, E., & Vallascas, F. (2013, June). *Explaining state support to European banks during the crisis: Which implications for regulation? Fondo interbancario di tutela dei depositi. Interbank deposit protection fund* (Working Paper No. 12). Interbank deposit protection fund [Fondo interbancario di tutela dei depositi].

Schapire, R. (2003). The boosting approach to machine learning - an overview. In D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, & B. Yu (eds), *MSRI workshop on nonlinear estimation and classification, 2002*. New York, NY: Springer.

Sun, J., Li, H., Huang, Q.-H., & He, K.-Y. (2014). Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, *57*, 41–56. doi:10.1016/j.knosys.2013.12.006

Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, *41*(5), 2353–2361. doi:10.1016/j.eswa.2013.09.033

Annex. Explanatory variables

All banks in the sample filed complete financial statements. The financial ratios obtained can be grouped into the following categories:

- Ratios of asset quality
- Operating ratios
- Capital ratios
- Liquidity ratios
- Balance sheet

(1) Ratios of asset quality

Assess quality rating includes an examination of credit risk associated with a particular asset. These assets usually are loans and investment portfolios. Asset quality, also referred to as portfolio quality, is a key aspect of financial performance for financial institutions. The loan portfolio is typically the predominant component of financial institutions' asset base. The asset quality ratios used in this study are important financial performance metrics for the financial institutions.

(2) Operating ratios

Bank's operating ratios are essentially equivalent to a regular company's operating margin, measuring the overall earning performance of a bank and its efficiency in utilising assets, liabilities and equity.

(3) Capital ratios

The financial structure ratios measure the riskiness of business in terms of leverage (debt gearing). They provide information about the combination of financial resources used to carry out the required investments to support the bank's business operations. These indicators measure the degree of the bank's dependence on each type of financial resources it may use. Likewise they are used to assess the adequacy of the liquidity of the banks and ensure they have adequate cash flow to meet all obligations in a timely and cost-effective manner.

(4) Liquidity ratios

Liquidity or short-term solvency ratios are closely linked to the financial equilibrium, indicating the bank's ability to meet its short-term obligations. That is, liquidity refers to the bank's ability to convert assets to cash and its ability to pay its financial obligations by their due date. These ratios relate items or assets and liabilities of the current structure of the business, both assets and liabilities. They are used to determine entity liquidity.

The different ratio categories used in the empirical test are shown in Table 2. The number of 25 ratios selected is the result of a filtering process of an initial amount of 38 ratios. as mentioned previously and displayed in Table 3.

(5) Balance sheet

Total Assets is the final amount of all gross investments, cash and equivalents, receivables and other assets as presented on the balance sheet.