EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Michael Bräuning, Despo Malikkidou,
Stefano Scalone, Giorgio Scricco

A new approach to Early Warning Systems for small European banks

No 2348 / December 2019

Abstract

This paper describes a machine learning technique to timely identify cases of individual bank financial distress. Our work represents the first attempt in the literature to develop an early warning system specifically for small European banks.

We employ a machine learning technique, and build a decision tree model using a dataset of official supervisory reporting, complemented with qualitative banking sector and macroeconomic variables.

We propose a new and wider definition of financial distress, in order to capture bank distress cases at an earlier stage with respect to the existing literature on bank failures; by doing so, given the rarity of bank defaults in Europe we significantly increase the number of events on which to estimate the model, thus increasing the model precision; in this way we identify bank crises at an earlier stage with respect to the usual default definition, therefore leaving a time window for supervisory intervention.

The Quinlan C5.0 algorithm we use to estimate the model also allows us to adopt a conservative approach to misclassification: as we deal with bank distress cases, we consider missing a distress event twice as costly as raising a false flag.

Our final model comprises 12 variables in 19 nodes, and outperforms a logit model estimation, which we use to benchmark our analysis; validation and back testing also suggest that the good performance of our model is relatively stable and robust.

JEL Classification: E58, C01, C50.

Keywords: Machine Learning, Bank Distress, Decision Tree, Quinlan.

1 Non-technical summary

In this paper we build an early warning system to identify cases of financial distress at the level of individual institutions in the European banking sector.

The motivation for this work stems from the desire to support the National Competent Authorities in their daily supervisory work, as well as Divisions within the European Central Bank conducting oversight functions, by providing them with a purely quantitative tool to complement the expert knowledge component of the supervisory risk assessment work.

This model represents only one of the many tools in the hands of the supervisory, to contribute to help with her work, and its role is solely that of contributing to shape and prioritise efforts towards certain institutions, whose situation might need to be followed more closely. In this respect, it should in no way be thought to be the trigger of any direct or indirect supervisory action.

There is a tendency in the literature to build Early Warning Systems to predict bank defaults. This is an undoubtedly useful theoretical exercise, but the identification happens too late for the supervisor to intervene. The goal of our model is to identify an institution when in distress, with a broader definition based on previous literature and regulation (Bank Recovery and Resolution Directive, SSM Framework Regulation, Directive 2014/49/EU). With this change of definition, we obtain a sample of 350 distress cases over a time span of only six quarters, allowing us to estimate the model in a more precise manner.

The dataset we use is based on bank specific variables coming from quarterly supervisory data (mainly COREP and FINREP), complemented with banking sector specific variables (e.g. whether a bank is a member of an Institution Protecting Scheme) and macro-economic indicators. Our panel is particularly wide, as we collect data for around 3,000 institutions.

We estimate a decision tree model using a machine learning technique of supervised learning and using Quinlan C5.0 algorithm. This methodology allows us to adopt a conservative approach to misclassification: we in fact consider missing a distress event twice as costly as raising a false flag, given the supervisory framework in which we operate.

We build our model in a forward-looking perspective, in order to make the identification of distress cases timely enough - and therefore leaving a time window for the supervisor to intervene: we map distress events at time t to explanatory variables at t-h, where h is the prediction horizon of one quarter.

Several challenges were identified in the data pre-processing phase, given the ample dataset of more than 3,000 variables we started with; first of all, not all banks are requested to report every data point in supervisory reporting, this making our dataset full of blanks and missing values. We cleaned the data both horizontally - eliminating institutions with too many missing values - and vertically - deleting variables where the majority of values is missing, correlation too high or variance nearly zero. Second, not all institutions are subject to the same reporting requirements; we therefore built a mapping between institutions reporting NGAAP (75% of the sample) and IFRS (the remaining 25%).

The final result is a decision tree model with 12 variables and 19 nodes. Banks are first split based on their level of profitability (adjusted for reporting standards), after which the top predicting variable for a low-profitability institution is the geographical location of the institution (particularly important for small institutions, often strongly linked to the territory in which they operate and with a not much diversified business model); on the other hand, if a bank has a relatively solid profit making capacity, the model looks at credit risk. In general, from the supervisory point of view, the model seems to recognise what the main risk dimensions are.

This early warning system represents a useful tool in the hands of the supervisor, as it complements her work in a purely quantitative manner to help prioritise and understand which institutions should be followed more closely. It in fact represents a widely used tool in the SSM daily oversight work on Less Significant Institutions.[1]

[1]As defined by the European supervisor, according to a precise set of rules contained in the SSM Guide to Banking Supervision:
https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssmguidebankingsupervision201411.en.pdf

2 Introduction

Models for the early identification of bank financial distress represent a useful tool both for the theoretical work of the researcher and the practical daily use of the supervisor. They in fact help the researcher understand what is that drives a bank into distress and tailor its investigation on bank crises, and allow for the timely intervention of the supervisor and in most cases the triggering of policy actions before the financial situation of an institution further deteriorates.

With this work, we propose an early warning model for the timely identification of distress of financial institutions, based on a large sample of small European banks.

Existing and comparable models are usually based on conventional modeling techniques such as multivariate logit models, and are calibrated using only a very small number of default (and not just distress) events. We here propose to innovate the current theoretical framework along two lines: first, we create a new definition for distress event inspired by literature and regulation, and obtain a sample of distress events in our dataset of small European banks that is significantly larger than most other works in the literature which focus on bank defaults. Second, we propose a machine learning methodology to build a decision three, which notably improves the predictive performance with respect to the most usual modeling techniques (we benchmark our decision tree with a logit estimation).

As said, our proposed approach in defining bank distress enlarges the usual sample size of distress events and therefore improves the learning of the model. We propose to classify banks as distressed leveraging on the BRRD's early interventions measures and on its criteria for categorizing banks as failing or likely to fail. Since this definition does not constitute the final stage of a bank's failure, the system will predict the distress event[2] stage early enough to allow the supervisor to intervene and adopt preemptive measures to tackle the financial deterioration case.

This paper makes use of a decision tree model, a technique often applied in machine learning for classification problems, with the goal to construct a flexible and interpretable signalling tool for banking supervision. The proposed early warning system (EWS) has the ability to predict individual cases of bank distress and identify which the variables driving the financial deterioration are.

This theoretical framework is applied to a unique dataset of more than 3,000 small European

[2]Whose definition is outlined in section 5.

banks, the so called Less Significant Institutions (LSIs[3]).

The remainder of this paper is organised as follows: section 3 introduces the framework by analysing the European banking sector, section 4 summarises the past relevant literature on the topic, section 5 describes the dataset used for the estimation, section 6 presents the model and the main results and section 7 concludes.

3 Background

This paper describes an early warning system for the European banking sector. This is an explicit choice that is both practical - as granular and frequent data is available for SSM institutions - and theoretical - since wider samples relative to wider geographical areas might have even stronger heterogeneity problems, and it might therefore be difficult to build a one-fits-all model for a global sample (Davis, 2011). The European banking system is extremely assorted and comprises institutions of different size, scope and business model. Banks range from big globally significant institutions to small local savings and cooperative banks, together with retail banks of different sizes, investment banks, custodians, asset managers, among the others.

In November 2014 the European Central Bank assumed responsibility for the supervision of euro area banks[4] and centralised the direct supervision of around 120 *significant* banking groups,[5] and set up supervisory standards for the remaining 3,000 smaller institutions (classified as *less significant* - LSIs from here on), the direct supervision of which is left to the National Competent Authorities (but still conducted in close collaboration with the ECB).

In this framework, the strong fragmentation of the European banking system called for a rigorous quantitative approach like the model we present in this paper to complement and facilitate the daily work of the analyst. Moreover, only recently there have been attempts in the literature to develop a model for the early detection of distress cases (see e.g. Rosa and Gartner (2018)), as researchers tend to focus on bank failures or on systemic bank crises[6] rather than single-bank distress: despite representing an undoubtedly useful theoretical exercise, the practical usefulness of such models is limited as once a bank is defaulting the situation is often irreversible. The reason for this is also technical, as it is often the lack of data that pushes

[3]As defined by the Single Supervisory Mechanism of the European Central Bank.
[4]https://www.ecb.europa.eu/press/pr/date/2014/html/pr141104.en.html
[5]Data source: EU Banking Supervision Website.
[6]On this see the extensive literature reviews by Kumar Ravi and Ravi (2006) and Davis and Karim (2010) for banking crises and Gramich et al. (2010) for systemic banking crises.

researchers to analyse the problem from the systemic point of view or to focus on bank distress; we instead make extensive use of supervisory data to estimate our model, which is significantly more granular and frequent. The change of perspective from bank default to bank distress that we propose in this work helps to detect cases of financial difficulties early enough to allow for a time frame for supervisory intervention.

Looking more precisely at the composition of the European banking sector, some of its key features contributed to shape the model we present in the following sections. While the business model of significant institutions is very heterogeneous, the vast majority of LSIs is of two types, *savings* and *cooperative banks* (Bülbül et al., 2013).[7] Even though these types of bank business models are extremely diffuse throughout Europe, the composition of the system of *less significant institutions* is unevenly distributed across countries and notably concentrated in certain jurisdictions, namely Germany, Austria and Italy. The main reason behind this is historical, as even though most European countries have very deeply-rooted local systems of savings and cooperative banks, in some jurisdictions these smaller institutions are unified in a single consortium or network (e.g. Credit Agricole in France or Rabobank in the Netherlands), and therefore are considered as *significant*.

As a consequence, the sample of European *less significant institutions* is strongly polarised towards a few countries, with Italy, Austria and Germany alone accounting for more than 80% of European LSIs.

This strong polarisation and the high absolute number of institutions raise some questions on how to timely identify cases of financial distress even before or as soon as they materialise, in order to give the supervisor enough time to intervene. All these peculiar data features are reflected in our model, and will be explained more in details in the next sections.

4 Literature Review

Economics is only one of the disciplines that makes use of early warning systems, which are used in the most various subjects from disaster management for natural events, to medicine for timely identification of diseases, to the world of social media. Researchers in these subjects often borrow calculation and modelling techniques from other disciplines like physics and engineering,

[7] These types of banks originated in Europe between the late 18$^{\text{th}}$ and early 19$^{\text{th}}$ Centuries, with the goal of offering banking services to farmers, workers and small entrepreneurs, which at the time were facing extreme difficulties to access credit.

the same approach we use here with boosting decision trees and machine learning.

Our model is developed at the level of the single bank, an approach that differs with the general strand of literature: in economics, in fact, much more diffused are early warning systems to timely recognise signs of potential systemic risks, both in the banking sector and at the country level. For the former, it is the case for example of Drehmann and Juselius (2013) and Aldasoro et al. (2018), who analyse the area under the ROC curve to identify early warning indicators (EWIs) that might represent potential source of vulnerabilities; for the latter, Kaminsky (2000)'s seminal paper, developing an early warning model of financial vulnerabilities by creating some composite indicators of financial distress, based on the commonly shared hypothesis that economic distress and fragility of an economy are good predictors of financial crises.

The literature on single indicators of financial distress is vast, and often exceeds (or consciously diverges from) the scope of canonical early warning systems. A fundamental reference on the topic is Kaminsky and Reinhard (1999), who focus on currency crises. They adapt the methodology of Stock and Watson (1989) for leading indicators to develop a first step in the design of an early warning system that would help detecting a domestic financial crisis. The main signals are indeed currency and exchange rate expectations, but the authors underline how micro data on banking would be the natural complement of their analysis: they find that currency crises deepen banking crises, and analyse the deep interlinkages between these two sectors.

Focusing on models for banking crises, there is a tendency in the literature to identify indicators signalling banking problems, rather than building a fully specified model to predict bank distress. This is the case for example of Honohan (1997), with the author proposing a set of variables, each of which with a threshold that if exceeded can successfully predict a situation of bank distress. This apparently dogmatic approach however leaves room for the discretionary intervention of the supervisor, the *expert judgement* procedure that is extremely common in this strand of work.

One of the first milestones in the literature on bank distress is represented by the pioneering work by Sinkey (1975), who adapted Altman (1968) to predict bank crises in a framework of multiple discriminant analysis. As outcome of the model, the variables indicated as good predictors for bank distress are asset composition, loan characteristics, capital adequacy, sources of revenues, efficiency and profitability.

Many of these variables correspond to the widely used CAMELS indicators, initially published by the Federal Financial Institutions Examination Council in 1979 and further developed by the Federal Reserve in 1995. This set of indicators represents a supervisory rating system to evaluate a bank's financial conditions and operations, and is widely used in the literature on banking distress identification. The indicators are:

- Capital adequacy,
- Asset quality,
- Management,
- Earnings,
- Liquidity,
- Sensitivity to market risk,

and represent a useful starting point for the variable selection of any banking model.

Not surprisingly, the interest in the literature on predicting bank distress events peaked after the 2007 financial crisis: Jin et al. (2011 and 2013) further developed the CAMELS approach by complementing the six indicators with data on banks' internal controls on risk-taking and audit quality variables, to find an improved predictive rate. Cole and White (2012) found that measures of commercial real estate investments are also relevant for predicting bank distress. Betz et al. (2013) also use the CAMELS indicators as a starting point to build their early warning system on European banks.

A good recap of the variables most widely used in the literature is in Oat et al. (2013) who, despite focusing on explanatory variables for systemic risk and financial distress, depict a list that contains many of the variables that compose also our model.

We partly rely on this literature to construct the initial dataset for our decision tree. We in fact build the model using three sets of variabels: bank specific indicators, banking-sector and country level macro-financial variables using the CAMELS approach as one of our starting points.

As the occurrence of a crisis - no matter how it is defined - can be easily described by a dummy variable, a common approach in the literature is to use logit/probit models. Thomson (1992) and Cole and Gunther (1998) estimate logit and probit models to show that vulnerability indicators covering the CAMELS dimensions are good predictors of banking failures. This is indeed a valid methodology, that we use to benchmark our estimations via the decision tree.

The attention on early warning systems for banking is on the rise also at the institutional level. In this context, our paper fits the new wave of works on the topic from authorities and central banks (see for example Lang et al. 2018 and Ferriani et al. 2019).

5 Data and Sample of Distress Events

Our model uses supervisory reporting, micro- and macro-economic data covering around 3,000 European less significant institutions over a period of four years. The sources of bank-level data are the Common Reporting (COREP, containing information on capital adequacy and risk-specific information), available on quarterly basis since December 2014, and the Financial Reporting (FINREP, which includes balance sheet items, the statement of profit and loss and detailed breakdowns of assets and liabilities by product and counterparty), available with different frequency since December 2014 and on quarterly basis since March 2017. COREP covers the entire LSI universe since its inception, FINREP instead presents different reporting frequency and levels of granularity on the basis of the complexity and size of the reporting entity; therefore for our purposes it has been integrated with an ad-hoc data collection carried out by the Single Supervisory Mechanism on bi-annual basis. The macro-economic data are mainly obtained from the ECB Statistical Data Warehouse, complemented by the Eurostat and the OECD for the regional data.

In order for an early warning system on bank distress to be useful in practice for the the supervisor and the policy maker, the recognition of the distress event must be timely enough to allow a buffer of time for intervention. If we consider the failure or liquidation of a bank as triggering event, as often defined in the literature (see e.g. the literature review by Gissel et al. (2007), or the more recent works by Jin et al. (2011) or Cole and White (2012)), we lose the practical validity of the model, which would in turn be helpful only for ex post calibrations. Moreover, bank failures in Europe are relatively rare, this making the estimation of such an early warning system even more challenging. We therefore relax the traditional hypothesis of considering only bank crises or defaults as positive events in the sample (as e.g. done by Kaminsky and Reinhard 1999, who mark the beginning of a banking crisis by a bank run leading to closure, merging or take-over by the public sector of a bank, or large-scale government assistance), and instead consider all financial distress cases. By doing so, the sample of distress events significantly grows in size, allowing us to obtain precise estimations despite the short time horizon on which we span our model.

The large number of small institutions in Europe, together with the high quality of our supervisory data allows us to build a large dataset of distress events; we start by describing our relaxed hypothesis to identify bank distress events.

We use a mixed approach, and base our definition of distress on the Banking Recovery and Resolution Directive complemented by one of the four conventional types of financial distress in Betz et al. (2013), and end up with a database of more than 350 distress events throughout a sample of only six quarters, starting from 2014 Q4 for six subsequent quarters.

Institutions which enter in financial distress are kept in the sample in the following quarters, as the model is run on a quarterly basis and the interest is also on the dynamics of banks entering and exiting the distress status. Institutions are only excluded when they default, so that our sample at each point in time represents the universe of Less Significant Institutions operating in Europe.

In general, we consider a bank to be in financial distress if:

- It is deemed to be failing or likely to fail within the meaning of Article 32 of the BRRD. For categorizing a bank as failing or likely to fail, indicators assessing whether a bank has breached the minimum capital requirements or capital buffers are constructed;
- It meets the conditions for early intervention pursuant to Article 27 of the BRRD. The triggers used to meet the conditions of early interventions consist of indicators for assessing if a bank is close to breaching minimum capital requirements;
- In case of the removal of the senior management and management body of the institution, in its entirely or with regard to individuals, in line with article 28 of the BRRD;
- It is placed under special administration and/or is appointed of a temporary administrator pursuant to Article 29 of the BRRD;
- There is a rapid and significant deterioration of its financial situation according to Article 96 of the Framework Regulation. This is based on expert judgement by national central banks and in-house qualitative and data;

- When there is an indication that the supervised LSI can no longer be relied upon to fulfil their obligations towards their creditors or where there is an indication of circumstances that could lead to a determination that the LSI concerned is unable to repay the deposits as referred to in Article 2 (1) (8) of Directive 2014/49/EU;
- One of the four types of conventional bank distress events proposed by Betz et al (2013) (i.e. bankruptcies, liquidations, state interventions and forced mergers) is met.

The distribution of distress events is not homogeneous across countries, and the frequency of cases varies significantly among jurisdictions. This is mainly due to two different factors, first the different levels of fragmentation of the banking system in Europe - with our sample of institutions strongly polarised towards certain geographies, and second the economic situation of the countries - for which distress cases are much more frequent in countries with a relatively weak economic situation.

The current prediction horizon of only one quarter was chosen due to the availability of data. A longer prediction horizon would in general be desirable, despite however resulting in a smaller sample of distress events - which in turn would negatively affect the predictive performance of the model.[8]

6 Methodology

6.1 Data pre-processing

Data pre-processing steps are required to ensure that unreliable and noisy data as well as irrelevant and redundant information is eliminated prior to the modelling phase. As such, the final training dataset used for the analysis is of high quality, thus increasing the efficiency and performance of the final model. This represents a key step in our process, as we start from the extremely vast dataset of supervisory reporting consisting of more than 3,000 variables, and end up with a final sample of only 12.

As explained above, the extensive and detailed dataset we use to build our model relies mainly on SSM supervisory reporting data. The available data is extensive and variables tend to be highly correlated - if not linear combinations of others - therefore data pre-processing represents a crucial step for the construction of our early warning system.

[8]The extension of the prediction horizon will however be the subject of future developments of this work.

We start by cleaning the data, in order to eliminate incomplete or uninformative data; many banks do not report some data points, following the proportionality criteria that inspires the supervisory reporting framework. We therefore eliminate variables for which the majority of values is missing, or where the variance across the sample is close to zero. This first step already reduces the number of total indicators to less than 500, this giving a hint of the importance of the data pre-processing procedure.

The second step is a simple transformation of the data, with the goal of increasing consistency and comparability across institutions. More than one accounting standard coexists in Europe, this complicating the job of the supervisor; we apply a transformation technique, based on a mapping of these heterogeneous data points to make different data sources somehow coherent. In a following stage, we normalise our variables through the creation of ratios in order to increase comparability.

We remove explanatory variables which are too highly correlated, using a simple threshold of 0.9 and select the final set of indicators based on their ability to predict distress. Variables are ranked according to their importance, captured by the individual Area under the Receiver Operating Characteristic curve (AUC) for each indicator; using this technique, we select the top 100 variables in terms of predictive performance which we use as starting point for the decision tree.

6.2 Decision trees

Our early warning system is based on a decision tree methodology. The predictive performance of this technique is very high, both in and out of sample, as demonstrated by the data on accuracy presented in section 6; moreover, this methodology well handles missing values, a common issue in the early warning literature (Mitchell 1997), and one of the main flaws in our dataset. Finally, the decision tree methodology is relatively transparent, and allows supervisors to interpret the output tree and understand which indicators affect bank distress, thus minimizing the risk of creating a *black box* model.

A decision tree is a classification technique commonly used in machine learning. The tree recursively identifies the significant indicators and their respective thresholds which best split the sample into the pre-determined classes (in our case distress and no distress).

More in details, a decision tree is a hierarchical model that identifies local regions (leaf nodes) via a sequence of recursive splits. Starting from the root of the tree, at each non-leaf node a test is run and a decision (one of the branches) is taken, depending on the outcome of the test. This process is repeated recursively, until a leaf node is reached; at that point the splitting stops, and the observation si labelled with the value of the leaf node.

The splitting procedure simply defines a region of the n-dimensional input space, where n is the number of distinct variables present in the tree; all the observations falling in the same localised region are given the same label by the model.

Among all possible splits, the best one is taken. The goodness of fit is measured in terms of *purity* and *entropy* (Quinlan, 1986). The purity of a node is a measure of the quality of the split: a node is pure if after the split, for all branches, all observations assigned to one branch belong to the same class. One possible way to measure the (im)purity of a node is *Entropy*:

$$H = -\sum_{i=1}^{n}(P_i log_2 P_i) \tag{1}$$

which is linked to the underlying probability of occurrence of value i: high entropy indicates that all classes are (nearly) equally likely (high impurity), while low entropy indicates that few classes are likely, and others are rarely observed (high purity).

The concept that guides the choice of the split is the maximization of the *information gain*, based on *conditional entropy*:

$$IG = H - (H_L * p_L + H_R * p_R) \tag{2}$$

where H is the *entropy* of the parent node, H_L is the entropy of the left node, and p_L is the probability that a random input is sent to the left node.

The final output of this recursive classification technique is a tree, illustrating a set of if-then rules (decision nodes) to reach a final decision on the classification (leaf nodes). In our case, for each bank, the classification starts from the root decision node, and based on predictors values create a path along the tree untila leaf node is reached, classifying if the bank is in distress or not.

6.3 Model and results

We employ Quinlan's C5.0 algoritm to build the classification tree model. The C5.0 algoritm is one of the most commonly used, as it is relatively fast and accurate, as well as efficient in handling missing data and removing unhelpful attributes.[9]

In training the model, we select a number of specification options:

- We impose a relatively short prediction horizon (1-3 months ahead of distress as starting point), given the short term (<1 year) scope of this EWS. By considering pre-default events as target variable, we ensure that the system has a forward looking perspective;

- We impose asymmetric misclassification costs when assessing the performance of explanatory variables: we consider Type I errors (missing a distress events) to be twice as costly as Type II errors (issuing a false alarm). In principle, this assumes that when faced with a tradeoff of issuing more false alarms or missing a distress event, the policymaker would take a conservative stance and choose the former;

- To increase the robustness of simple decision trees (which is relatively low, as underlined by Alessi and Detken (2014), who use a Random Forest method to overcome the problem), we employ a boosting technique á la Freund et al. (1999) to identify which variables to include in the final version of the tree.[10]

- As there is no univocal rule to choose one particular tree among the estimated ones, we use the boosting technique to simulate the creation of a large number of trees, and select the 20 variables that rank highest as of importance (measured in terms of how often they appear in the trials).

- We complement the variable selection with both expert judgement and quantitative measures: in particular, for evaluating the performance of the model, we rely on the area under the Receiver-Operating-Characteristic curve (AUC) and Cohen's kappa statistic, both standard measures of accuracy in the early warning system literature (e.g. Peltonen at al. (2015)).

The final tree is composed of 19 nodes, covering 12 different explanatory variables, and is represented in Figure 1.[11]

[9]For a literature review of Data Mining Algoritms see Wu et al. (2008). The relative R environment used in this paper refers to Kuhn et al. (2015).

[10]Boosting is a technique for generating and combining multiple classifiers to improve the predictive accuracy of the model. Instead of using a single tree, n separate decision trees (trials) are grown and combined to make predictions. The error rate of the boosted classifier is often substantially lower than that of single trees.

[11]Please note that the trees represented in this version of the paper are somehow anonymised, i.e. without the precise splitting thresholds of each node.

The variables included in the model are:

- Adjusted profitability;
- Non-performing loans (NPL) ratio;
- Non-performing loans coverage ratio;
- Deficit-to-GDP ratio;
- GDP growth;
- Liquidity coverage ratio (LCR);
- Leverage ratio;
- Equity exposures;
- Exposures in default;
- Two proxies for market risk;[12]
- Membership in an institutional protection scheme (IPS),

The indicator of the parent node is profitability, adjusted for the different accounting standards of the banks in the sample. The node splits the banks between profit (right branch) and loss (left branch) making. The remaining variables are a mix of macro-economic indicators (deficit-to-GDP ratio, and real GDP growth) and banking indicators covering the most important risks: credit (non-performing loans ratio, non-performing loans coverage ratio, exposure in default), liquidity (liquidity coverage ratio), market (captured by the sum of trading financial assets and financial liabilities held for trading over total asset and net gains on financial assets and liabilities held for trading over total operating income), capital (leverage ratio and equity exposure), together with the qualitative information of whether a bank is member of an institutional protection scheme (IPS).

In the framework of supervised learning, the role of our tree is not only to find an efficient method to identify which banks are in financial distress, but also to suggest which variables are significant and how they model the distress. In this perspective, it is interesting to analyse the

[12]The sum of trading financial assets and financial liabilities held for trading over total assets, and net gains on financial assets and liabilities held for trading over total operating income.

main paths through the tree: if a bank is making profits (parent node to the right), profitability is likely to not be an issue, and the model suggests to investigate credit risk (first node to the right is the NPL ratio). If credit risk is deemed as material, i.e. if the bank has a high level of non-performing loans, the model moves to analyse whether these NPLs are covered by sufficient allowances. On the other hand, for banks with low non-performing loans (NPL ratio node to the left), market risk becomes relevant in capturing distress: this is intuitive, as banks strongly relying on income from market activities are subject to higher volatility and potential distress, especially in countries with fragile economic fundamentals (high level of deficit over GDP ratio).

On the opposite side of the tree, if a bank is deeply unprofitable (parent node to the left, first node to the left), if it is operating in a country with a low growth of GDP it is automatically labelled as being in distress. When instead the bank has relatively high equity exposures combined with a weak LCR ratio, then the distress will be determined by the eventual membership in an IPS; Institution Protecting Schemes in fact protect banks from financial distress, therefore making member institutions less vulnerable. Finally, for moderately unprofitable banks, the economic conditions of the country in which they operate (proxied by deficit over GDP) and the leverage ratio indicate whether a bank is in distress or not.

The predictive performance of the model is very high, both in and out of sample (which we conduct via a 10-fold cross validation testing). As depicted in Table 1, in the training data the true positive rate is 0.89, while the false positive rate (Type II error) and false negative rate (Type I error) equal 0.01 and 0.11, respectively. The AUC is 0.95, much closer to the unity value of a perfect classifier, than to the 0.5 value of a purely random one. The Cohen's kappa statistic is also high, at a value of 0.89. The out of sample performance (25% of the initial observations), based on a random split is also satisfactory, with an AUC of 0.92 and a Cohen's kappa statistics of 0.80. Type I and II errors are therefore comparable to the in-sample error rates, and remain at adequately low levels.

Table 1: Validation Results Full Sample

Measures	In Sample (train)	Out of Sample (test)
Type I error rate	0.01	0.03
Type II error rate	0.11	0.10
AUC	0.95	0.92
Cohen's Kappa	0.89	0.80

It is useful to remark how the model is estimated by introducing an unbalance between level-1 and level-2 errors: false positives are in fact considered to be less problematic than missed distress events, and therefore weight half.

We benchmark the results of our decision tree model with a Logit regression, a typical approach in the literature when predicting a binary categorical target variables (in our case the event of a case of financial distress).

For the data pre-processing we instead follow the same exact steps of the decision tree: we clean the dataset by eliiminating variables with too many missing values, with almost zero variance and with too high correlation[13]. We select the model variables with a LASSO regression (á la Tibshirani, 1996), in order to prevent overfitting; this methodology minimises the sum of the squared errors just like a normal linear estimation process, but bounds on the sum of the (absolute) value of the coefficients.

Data is then randomly split into training (75%) and test (25%) sample, in order to allow for independent performance validation of the model. We follow Lang et al. (2018) and choose the shrinkage parameter lambda through cross-validation, in order to obtain the model which optimises the out-of-sample forecasting performance.

The results are still relatively accurate, but the Logit model misses many more distress events than the decision tree (see table below). The Logit is in fact significantly more sensitive to missing values, and therefore fails to detect distress events relative to institutions reporting a sufficient number of NAs, unlike the decision tree.

Table 2: Validation Results Logit Model

Measures	In Sample (train)	Out of Sample (test)
Type I error rate	0.01	0.02
Type II error rate	0.18	0.19
AUC	0.95	0.90
Accuracy	0.97	0.97

The supervised learning estimation framework of the decision tree allows us to perform a precise *ex post* back-testing analysis. In the course of 2017 the model correctly identified 79% of distress events. The missing financial deterioration cases were mainly triggered by qualitative

[13] We build the correlation matrix and proceed to reduce pairwise correlation by selecting the couples of too highly correlated variables, and eliminating the one with the largest mean absolute correlation.

features (e.g. governance issues), which our purely quantitative model fails to capture.

7 Conclusions

This paper develops an innovative model to identify cases of bank financial distress, using a subsample of 3,000 small European institutions[14] for a time period of six quarters (between 2014 and 2016).

We build a sample of distress cases based on European regulation, to early detect future cases of financial deterioration rather than simply referring to banks that are already in or close to default. With a broad definition of financial deterioration, our sample of distress events is significantly larger than any other work in the literature, despite a relatively short time series of data.

On this sample we construct a decision tree model, which accurately classifies banks into distressed or non-distressed; the prediction horizon of the model is one-three months, a time span that would in our view give the supervisor enough time to trigger supervisory action.

We find that the predictive power of our model is extremely high, and the decision tree steadily outperforms the Logit approach, the most widely used methodology to predict binary classifications which we use as benchmark.

As a final remark, further extensions of this work should go in the direction of increasing the prediction horizon (currently 1-3 months), and could include the development of an *ad hoc* tree for each business model. Unfortunately, the data does not allow us to do this, yet, as given the polarisation of the dataset towards retail lenders, the sample of remaining business models is simply not numerous enough to allow for a proper estimation.

A clear limitation of this work is the length of the time series, which we try to partly overcome with the size of the panel and the granularity of the data. However, we plan to re-estimate the model with a longer time series when it will be available, to continue to back test the predicting performance of the model and to analyse the changes in the business models of the banks; the latter in particular could be achieved in two ways: first, by re-estimating the model to understand how the environment changed, and what new variables contribute to describing the business of an institution; second, by conducting a case-by-case analysis for institutions for which the model changes classification: in this way, it would be possible to get some insights on

[14] Excluding the so called Significant Institutions, as defined by the SSM.

how the changes in monetary policy and the economic environment influence the behaviour of banks (think for example of the current prolonged period of low interest rates, which might be pushing retail banks that see their interest margins eroded into finding new sources of income).

Finally, further research on the topic could focus on the severity of bank distress events. In this respect, a multi-class target model could help classify different levels of distress like as Mild, Moderate and Severe. Such classification would be important for supervisors in order to efficiently allocate resources and improve financial surveillance.

References

[1] Aldasoro I., Borio C. and Drehmann M., 2018, *Early Warning Indicators of Banking Crises: Expanding the Family*, BIS Quarterly Review, March 18, 29-45.

[2] Alessi L. and Detken C., 2014, *Identifying Excessive Credit Growth and Leverage*, ECB Working Paper n.1723.

[3] Alessi L., Antunes A., Babecky J., Baltussen S., Behn M., Bonfim D., Bush O., Detken C., Frost J., Guimaraes R., Havranek T., Joy M., Kauko K., Mateju J., Monteiro N., Neudorfer B., Peltonen T., Rodrigues P., Rusnak M., Schudel W., Sigmund M., Stremmel H., Smidkova K., van Tilburg R., Vasicek B. and Zigraiova D., *Comparing different early warning systems: Results from a horse race competition among members of the Macro-prudential Research Network*, MPRA Paper 62194, University Library of Munich, Germany.

[4] Alpaydin E. I., 2010, *Introduction to Machine Learning*, 2^{nd} Edition, The MIT Press.

[5] Altman E. I., 1968, *Financial Ratios, Discrimimant Analysis and the Prediction of Corporate Bankruptcy*, The Journal of Finance, 23, 589-609.

[6] Altman E. I., Marco G., and Varetto F., 1994 , *Corporate Distress Diagnosis: Comparisons Using Linear Discriminate Analysis and Neural Networks*, Journal of Banking and Finance, 18, 505-529.

[7] Atiya A., 2001, *Bankruptcy Prediction for Credit Risk using Neural Networks: A Survey and new Results*, Transactions on Neural Networks, 12(4), 929-935.

[8] Betz F., Opric S., Peltonen T. A., and Sarlin P., 2014, *Predicting Distress in European Banks*, Journal of Banking and Finance, 45, 225-241.

[9] Boritz, J. E. and D. B. Kennedy, 1995, *Effectiveness of Neural Network Types for Prediction of Business Failure*, Expert Systems with Applications, 9(4), 95-112.

[10] Bülbül D., Schmidt R. H. and Schüwer U., 2013, *Savings Banks and Cooperative Banks in Europe*, SAFE White Paper Series n.5.

[11] Cole, R. A., and J. Gunther, 1998, *Predicting Bank Failures: A Comparison of on- and off-site Monitoring Systems*, Journal of Financial Services Research 13(2), 103-117.

[12] Cole, R. A. and L. J. White, 2012, *Deja Vu all over again: The Causes of U.S. Commercial Bank Failures this Time Around*, Journal of Financial Services Research 42(1-2), 5-29.

[13] Davis P. E. and Karim D., 2007, *Comparing Early Warning Systems for Banking Crises*, Journal of Financial Stability, 4(2), 89-120.

[14] Davis P. E., Karim D. and Liadze I., *Should multivariate early warning systems for banking crises pool across regions?*, Review of World Economics, 147(4), 693-716.

[15] Drehmann M. and Juselius M., 2013, *Evaluating Early Warning Indicators of Banking Crises: Satisfying Policy Requirements*, BIS Working Papers n.421.

[16] Ferriani F., Cornacchia W., Farroni P., Ferrara E., Guarino F. and Pisanti F., 2019, *An Early Warning System for Less Significant Italian Banks*, Banca d'Italia Occasional Papers n.480.

[17] Flannery M. J., 1998, *Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence*, Journal of Money, Credit and Banking, 30(3), 273-305.

[18] Freund, Y., Schapire, R., and Abe, N., 1999, *A Short Introduction to Boosting*, Journal-Japanese Society For Artificial Intelligence, 14, 771-780, n.1612.

[19] Frydman H., Altman E. I. and Kao D. L., 1985, *Introduction Recursive Partitioning for Financial Classification: The Case of Financial Distress*, Journal of Finance, 40(1), 269-291.

[20] Gaytn A. and Johnson C. A., 2002, *A Review of the Literature on Early Warning Systems for Banking Crises*, Central Bank of Chile Working Papers n.183.

[21] Gepp A., Kumar K. and Bhattacharya S., 2010, *Business Failure Prediction using Decision Trees*, Journal of Forecasting, 29(6), 536-555.

[22] Gepp A. and Kumar K., 2015, *Predicting Financial Distress: a Comparison of Survival Analysis and Decision Tree Techniques*, 11^{th} International Conference on Data Mining and Warehousing.

[23] Gissel J. L., Giacomino D. and Akers M. D., 2008, *A Review of Bankruptcy Prediction Studies: 1930-Present*, Journal of Financial Education, 33, 1-42.

[24] Gonzlez-Hermosillo M. B., 1999, *Determinants of ex-ante Banking System Distress: a Macro-micro Empirical Exploration of some Recent Episodes*, International Monetary Fund.

[25] Gramith D., Miller G. L., Oet M. V. and Ong S. J., 2018, *Early Warning Systems for Systemic Banking Risk: Critical Review and Modeling Implications*, Banks and Bank Systems, 5(2), 199-211.

[26] Honohan P., 1997, *Banking System Failures in Developing and Transition Countries: Diagnosis and Predictions*, BIS Working Papers n.39.

[27] Jin, J., K. Kanagaretnam, and G. Lobo, 2011, *Ability of Accounting and Audit Quality Variables to Predict Bank Failure During the Financial Crisis*, Journal of Banking & Finance 35(11), 2811-2819.

[28] Jin, J., K. Kanagaretnam, G. Lobo, and R. Mathieu, 2013, *Impact of FDICIA Internal Controls on Bank Risk Taking*, Journal of Banking & Finance 37(2), 614-624.

[29] Joos P., Vanhoof K., Ooghe H. and Sierens N., 1998, *Credit Classification: a Comparison of LOGIT Models and Decision Trees*, Proceeding Notes of the Workshop on APplication of Machine Learning and Data Mining in Finance, 59-72.

[30] Kaminsky G. L., 1999, *Currency and Banking Crises: the Early Warning of Distress*, IMF Working Paper, WP/99/178.

[31] Kaminsky G. L. and Reinhart C. M., 1999, *The Twin Crises: the Causes of Banking and Balance-of-Payments Problems*, The American Economic Review, v.89 n.3, 473-500.

[32] Kohavi R. and Quinlan R., *Decision Tree Discovery*, in Handbook of Data Mining and Knowledge Discovery, 267-276, Klosgen & Zytkow, Oxford University Press, Oxford.

[33] Kuhn M., Weston S., Coulter N., and Culp M., 2015, *C5.0 Decision Trees and Rule-Based Models*.

[34] Kumar Ravi P. and Ravi V., 2006, *Bankruptcy prediction in banks and firms via statistical and intelligent techniques A review*, European Journal of Operational Research, 180(1), 1-28.

[35] Lang J. H., Peltonen T. A. and Sarlin P., 2018, *A Framework for Early-warning Modeling with an Application to Banks*, ECB Working Paper n.2182.

[36] Meyer P. A. and Pifer H. W., 1970, *Prediction of Bank Failures*, Journal of Finance, 25(4), 853-868.

[37] Mitchell T., 1997, *Machine Learning*, 1st edition, McGraw-Hill, New York.

[38] Ng G. S., Quek C. and Jiang H., 2008, *FCMAC-EWS: A Bank Failure Early Warning System based on a Novel Localized pattern Learning and Semantically Associative Fuzzy Neural Network*, Expert Systems with Applications, 34(2), 989-1003.

[39] Oet M. V., Bianco T., Gramlich D. and Ong S. J., *SAFE: An Early Warning System for Systemic Banking Risk*, Bank Finance, 37(11), 4510-4533.

[40] Peltonen, T. A., Piloiu, A., and Sarlin, P., 2015, *Network Linkages to Predict Bank Distress*, ECB Working Paper n.1828.

[41] Quinlan J.R., 1986, *Induction of Decision Trees*, Machine learning 1, 81-106.

[42] Roengpitya R., Tarashev N. and Tsatsaronis K., 2014, *Bank Business Models*, BIS Quarterly Review, December 2014.

[43] Rosa P. S. and Gartner R. I., 2018, *Financial Distress in Brazilian Banks: an Early Warning Model*, Revista Contabilidade & Finanças, 29(77), 312-331.

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

[45] SSM, 2014, *Guide to Banking Supervision*.

[46] Stock J. H. and Watson M., 1989, *New Indices of Coincident and Leading Economic Indicators*, National Bureau of Economic Research Macroeconomics Annual 1989, NBER.

[47] Tibshirani R., 1996, *Regression Shrinkage and Selection via the LASSO*, Journal of the Royal Statistical Society Series B, 58(1), 267-288.

[48] Thomson, J. B., 1991, *Predicting Bank Failures in the 1980s*, Economic Review-Federal Reserve Bank of Cleveland, 27(1), 9.

[49] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., *et alii*, 2008, *Top 10 Algorithms in Data Mining*, Knowledge and Information Systems, 14(1), 1-37.

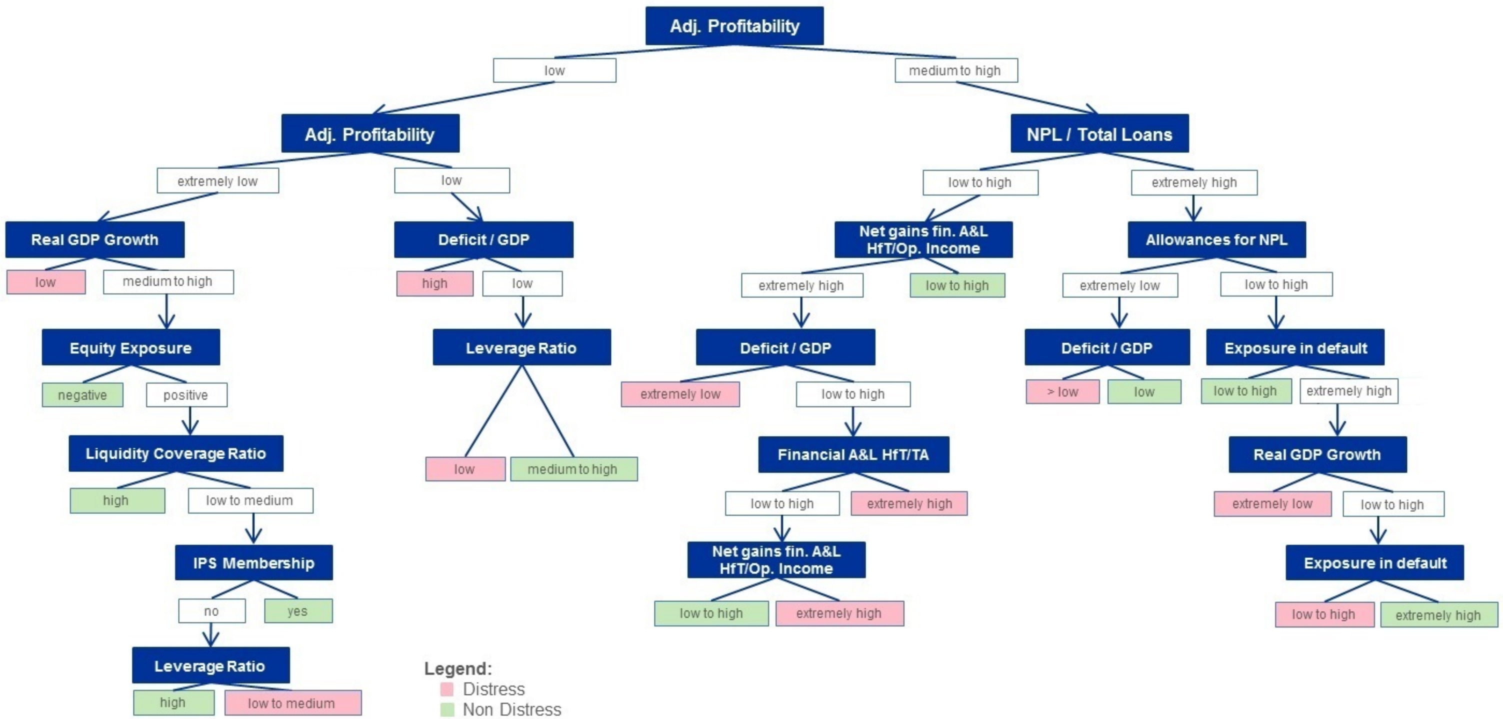

Figure 1: Decision tree for full sample

Acknowledgements

We are extremely grateful to Caroline Liesegang, Gianluca Trevisan and Giuseppe Vulpes for their support and the valuable comments received.

Michael Bräuning
European Central Bank, Frankfurt am Main, Germany; email: michael.brauning@ecb.europa.eu

Despo Malikkidou
European Banking Authority, Paris, France; email: despo.malikkidou@eba.europa.eu

Stefano Scalone
European Central Bank, Frankfurt am Main, Germany; email: stefano.scalone@ecb.europa.eu

Giorgio Scricco
European Central Bank, Frankfurt am Main, Germany; email: giorgio.scricco@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

This paper can be downloaded without charge from www.ecb.europa.eu, from the Social Science Research Network electronic library or from RePEc: Research Papers in Economics. Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website.

PDF ISBN 978-92-899-3979-9 ISSN 1725-2806 doi:10.2866/128724 QB-AR-19-129-EN-N