Εκπαίδευση, Δια Βίου Μάθηση, Έρευνα και Τεχνολογική Ανάπτυξη, Καινοτομία και Οικονομία

Τόμ. 2 (2019)

Πρακτικά του 2ου Πανελληνίου Επιστημονικού Συνεδρίου με Διεθνή Συμμετοχή «Ελλάδα-Ευρώπη 2020: Εκπαίδευση, Δια Βίου Μάθηση, Έρευνα, Νέες Τεχνολογίες, Καινοτομία και Οικονομία», Λαμία 28, 29, 30 Σεπτεμβρίου 2018

ΕΛΛΗΝΙΚΟ ΙΝΣΤΙΤΟΥΤΟ ΟΙΚΟΝΟΜΙΚΩΝ
ΤΗΣ ΕΚΠΑΙΔΕΥΣΗΣ & ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ,
ΤΗΣ ΕΡΕΥΝΑΣ & ΚΑΙΝΟΤΟΜΙΑΣ

Πρακτικά

2ου Πανελλήνιου Συνεδρίου με Διεθνή Συμμετοχή
"Ελλάδα - Ευρώπη 2020:
Εκπαίδευση, Διά Βίου Μάθηση, Έρευνα,
Νέες Τεχνολογίες, Καινοτομία και Οικονομία"

Υπό την Αιγίδα της
Α.Ε. του Προέδρου της Δημοκρατίας
κυρίου Προκοπίου Παυλόπουλου

28-30 Σεπτεμβρίου 2018, Λαμία

Οργάνωση:
- Ελληνικό Ινστιτούτο Οικονομικών της Εκπαίδευσης & Διά Βίου Μάθησης, της Έρευνας & Καινοτομίας
- Πανεπιστήμιο Θεσσαλίας

Συνεργασία:
- Περιφέρεια Στερεάς Ελλάδας
- Δήμος Λαμιέων

Επιμέλεια Πρακτικών
Ε. Καραΐσκου & Γ. Κουτρομάνος

Banking supervision with the use of innovative statistical techniques

Vasileios Siakoulis, Anastasios Petropoulos, Evangelos Stavroulakis, Nikolaos Vlachogiannakis, Ioannis Tsikripis

doi: 10.12681/elrie.1528

Banking supervision with the use of innovative statistical techniques

Petropoulos Anastasios[1], Siakoulis Vasilis[2], Stavroulakis Evangelos[3], Vlachogiannakis Nikolaos[4], Tsikripis Ioannis[5]

apetropoulos@bankofgreece.gr,vsiakoulis@bankofgreece.gr, estavroulakis@bankofgreece.gr, nvlachogiannakis@bankofgreece.gr, itsikripis@bankofgreece.gr

[1]PhD, Bank of Greece,[2]PhD, Bank of Greece,[3]PhD, Bank of Greece,[4]PhD, Bank of Greece,[5]PhD, Bank of Greece

Abstract

Proactively monitoring and assessing the economic health of financial system has always been the cornerstone of supervisory authorities for supporting informed and timely decision making. Bank of Greece as the competent supervisory authority for the Greek banking system evaluates both the riskiness of banks on an individual level and the health of the financial system in total, from a macro prudential perspective. In accomplishing those targets, the Bank of Greece could make use –inter alia- of various statistical methods along with expert judgment. In this work, we employ a series of innovative modeling techniques in the prediction of individual bank insolvencies and generalized financial crises. Our empirical results indicate that innovative statistical techniques, i.e. Deep Learning and Machine Learning methodologies, have superior out of sample and out of time predictive performance in comparison to traditionally employed methods in finance, such as Logistic Regression, Classification Tress, and Linear Discriminant Analysis. In essence, we build an Early Warning System for bank insolvencies and another one for stock market crises, which could complement the assessments performed by micro-prudential and macroprudential authorities. In short, the holistic monitoring of the resilience of the financial system would steer decision making, via triggering the imposition of any necessary targeted corrective actions, leading vulnerable institutions back to viable business performance and the financial system back to balanced operation.

Keywords: Bank’s insolvencies, Financial Crises, Machine Learning, Deep Learning, Big Data, Banking Supervision

JEL: G01, G21, C53

1. Introduction – Banking Supervision and Innovation

The recent global financial crisis disrupted significantly the economic growth and had severe socio-economic and fiscal effects in most parts of the world. In several countries the sovereign had to step in and provide support in order to avoid the full collapse of the banking system. In Greece, the global financial crisis unveiled the large economic imbalances that were built up after euro accession, which ultimately triggered an unprecedented sovereign and banking crisis. More specifically, the Greek economy suffered from the “twin” deficits, namely, the fiscal and the current account deficits, which were the result of a strong fiscal expansion financed mostly by external borrowing. At the same time, private indebtedness had increased on the back of a sizeable domestic credit growth. As the cross-border flows dropped dramatically after the eruption of the global financial crisis in 2008, Greece remained exposed to these economic imbalances and avoid complete collapse only due to the economic support receive by the IMF and its EU partners. The crisis also weighed negatively on the Greek banking system that lost access to the capital and liquidity markets and had to resort to the Emergency Liquidity Assistance of the Central Bank to address the massive deposits outflows. The asset quality deteriorated significantly due to worsened balance sheet of both corporate and households. The elevated loan loss provisions, the impact from the participation in the sovereign debt restructuring, and the subdued profitability resulted in three rounds of recapitalizations for the four core banks and the resolution of several non-systemic players.

The supervisory response to the global financial crisis was immense. The Basel Committee on Banking Supervision updated the rulebook incorporating all lessons learned from the crisis, in the so-called Basel III accord that was transposed in the EU framework via the CRR/CRD IV. The supervisory requirements increased significantly both in quantitative and qualitative terms. The macro prudential perspective in supervision gained momentum, whereas several requirements regarding the recovery and resolution of financial institutions were introduced. The effort made by the supervisory authorities and the official sector targeted to the buildup of a resilience financial sector that will be capable to absorb the impact from a future crisis. However, future financial crisis cannot be precluded, whereas we can never be sure about the shape and length of an imminent crisis. Therefore, considering the importance of early warning systems in order to mitigate or even preempt a financial crisis, we used innovative statistical techniques to build up tools for predicting crisis both at the level of individual banks and at the whole financial system level.

The working paper is structured as follows: In Chapter 2 an overview of the innovative methodologies employed by Bank of Greece is presented whereas in Chapter 3 the relevant evaluation measures are shown. The respective applications of the innovative methodologies are described analytically in Chapter 4 (Bank insolvencies prediction) and Chapter 5 (Stock Market Crisis prediction) whereas some conclusions and regulatory implications are discussed in Chapter 6.

2. Innovative Methodologies

Random Forests (RF) is a popular method for modeling classification problems. Since its inception (Breiman 2000) RFs has gained significant ground and is frequently used in many machine learning applications across various fields of the academic community. To build the considered Random Forests, we employed the “randomForest” package in R. The basic philosophy of Random forests is based on combining three concepts: i) classification or regression decision trees; ii) bootstrap aggregation or bagging; and iii) random subspaces. It adopts a divide-and-conquer approach to capture non-linearities in the data and perform pattern recognition. Its core principle is that a group of “weak learners” combined, can form a “strong predictor” model.

Support Vector Machines (henceforth SVMs) are a family of non-linear, large-margin binary classifiers. SVMs estimate a separating hyperplane that achieves maximum separability between the data of the two modeled classes (Vapnik, 1998). The main drawbacks of SVMs stem from the fact that they constitute black-box models, thus limiting their potential of offering deeper intuition and visualization of the obtained results and inference procedure.

Neural Networks constitute a well-known machine learning technique that is broadly used in credit rating classification problems. Classification problems are characterized by the availability of big datasets, many explanatory variables, and the possibility of noise existence in the data. Experimental results offer evidence that neural networks are able to capture complex non-linear patterns in the analyzed data. As such, it is no coincidence that the current literature offers numerous structural variations of Neural Networks depending on the number of layers, the flow of information and the algorithms used to train them.

XGBoost (eXtreme Gradient Boosting) is an advanced implementation of gradient boosting algorithm, offering increased efficiency, accuracy and scalability over RFs and NNs. It supports fitting various kinds of objective functions, including regression, classification and ranking. XGBoost offers increased flexibility, since optimization is performed on an extended set of hyperparameters, while it fully supports online training, without the danger of catastrophic forgetting.

Deep learning has been an active field of research in the recent years, as it has achieved significant breakthroughs in the fields of computer vision and language understanding. However, their application in the field of finance is rather limited. Our approach consists in building a multi-layer perceptron using the MXNET package of R. We postulated modern deep models that are up to five hidden layers deep and comprise various numbers of neurons. Model selection using cross-validation was performed by maximizing the area under the curve metric.

3. Evaluation Measures

Classification accuracy is the main criterion to assess the efficacy of each method and to select the most robust one. In this section, we present a series of metrics that are broadly used by the Bank of Greece for quantitatively estimating the discriminatory power of each scoring model. In evaluating the classification accuracy, we focus on the following measures

- G-mean: The geometric mean G-mean is the product of sensitivity and specificity. This metric indicates the balance between classification performances on the majority and minority class. A poor performance in prediction of the positive cases will lead to a low G-mean value, even if the negative cases are correctly classified from the algorithm.
- LR-: The negative likelihood ratio is the ratio between the probability of predicting a case as negative when it is actually positive, and the probability to predict a case as negative when it is truly negative. A lower negative likelihood ratio means better performance on the negative cases, which is the main point of interest in this study as we model bank failures.
- DP: Discriminant power is a measure that summarizes sensitivity and specificity.

For DP values higher than 3 then the algorithm distinguishes well between positive and negative cases.

- BA: The balanced accuracy is the average of Sensitivity and Specificity. If the classifier performs equally well on either class, this term reduces to the conventional accuracy measure. In contrast, if the conventional accuracy is high merely because the classifier takes advantage of good prediction on the majority class, the balanced accuracy will drop thus signaling any performance issues. That is, BA doesn’t disregard the accuracy of the model in the minority class.
- Youden’s γ: Youden’s index is a linear transformation of the mean sensitivity and specificity therefore it is difficult to interpret. As a general rule, a higher value of Youden’s γ indicates better ability of the algorithm to avoid misclassifying banks.
- AUC: The area under the ROC curve (Area Under Curve, AUC) is a summary indicator of the performance of a classifier into a single metric. The AUC can be estimated through various techniques, the most commonly used being the trapezoidal method. The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. In practice, the value of AUC varies between 0.5 and 1 with a value above 0.8 to denote a very good performance of the algorithm.

4. Predicting bank insolvencies using machine learning techniques

Supervisory authorities are mandated to protect depositors’ interests, via ensuring that financial institutions are able to survive under business as usual conditions and are capable to absorb adverse market shocks. Hence, the comprehensive assessment of the current financial conditions of a bank as well as the evaluation of its future sustainability is the cornerstone of proactive banking supervision. To distinguish between strong and weak banks, supervisory authorities make use of early warning expert systems or/and statistical modeling techniques.

The outcome of this analysis can drive the imposition of targeted regulatory measures. These measures can take the form of preemptive corrective actions addressing vulnerabilities of weaker banks and as a result can increase their resilience on a going concern basis. On the other hand, in specific cases of likely to fail banks, whose return to viability is considered rather improbable, it will provide the necessary evidence to the supervisory in order to take actions from a gone concern perspective. Essentially, supervisory actions serve in retaining depositors' confidence to the financial system by ensuring soundness of individual banks or resolution of failing banks in an orderly manner, should this be necessary in order to avoid any domino effect that can even trigger a systemic financial crisis.

In the last decades various statistical methodologies have been exploited to aggregate bank specific information into a single figure in order to distinguish between solvent and insolvent financial institutions. These classification methods range from simple Discriminant analysis (Altman 1968 and Cox 2014) and Logit/Probit regressions (Ohlson 1980, Cole and Wu 2014), to advanced machine learning techniques, conditional inference trees and Neural Networks (Messai & Gallali 2015). At the same time, other novel modeling approaches such as Random Forests (RF) (Breiman 2000) have not been employed up to now in the problem of assessing bank failures, regardless of these models being really popular for modeling classification problems in recent years.

In this work we employ a series of innovative techniques in predicting bank insolvencies, such as Random Forests, Support Vector Machines, Neural Networks and Random Forests of Conditional Inference Trees whereas we benchmark their results based on widely employed techniques such as Logistic Regression and Linear Discriminant analysis.

4.1. Literature Review

There is an extensive literature on the various methods and analyses performed, regarding the prediction of bank default. Messai and Gallali (2015) by applying discriminant analysis, logistic regression and artificial intelligence methods along with Cole and Wu (2014) who focused on time-varying hazard models and probit models, supported the view that CAMELS risk ratios are the most relevant and significant factors in predicting a bank default. The former pointed also that the neural network method performed better compared to the other models.

Cole and White (2010) examined the defaults of US commercial banks that occurred in 2009 by examining supervisory indicators as well as additional portfolio variables, such as real-estate loans and mortgages, which proved to be important as early warning indicators. Cox and Wang (2014) also focused on supervisory indicators, while they also incorporated risk factors that were overlooked by the literature prior the US financial crisis in 2007-2009.

Mayes and Stremmel (2014) incorporated supervisory indicators and macroeconomic variables in the framework of Logistic Regression and discrete survival time analysis methods. Betz et al. (2013) combined supervisory indicators with country-level data in order to improve the performance of the model in terms of Type I error and out-of-sample validation over different forecast horizons. Poghosyan and Čihák (2009) used supervisory indicators together with other factors related to depositor discipline, contagion effect among banks, macroeconomic environment, banking market concentration and the financial market. The results show that indicators related to capitalization, asset quality and profitability can effectively identify weak banks.

4.2. Data Collection and Variable Selection

We have collected information on non-failed entities, failed entities, and entities that received state assistance, from the database of the Federal Deposit Insurance Corporation (FDIC), an independent agency created by the US Congress in order to maintain the stability and the public confidence in the financial system. The collected information is related to all US banks, while the adopted definition of a default event in this dataset includes all bank failures and assistance transactions of all FDIC-insured institutions. Under the proposed framework, each entity is categorized either as solvent or as insolvent based on the indicators provided by FDIC.

The dataset covers the 2008-2014 period; a 7 years' period with quarterly information resulting in dataset with more than 175,000 records. The selected time period, seems to approximate a full economic cycle, in terms of the Default Rate evolution. Figure 1, shows the number of records included in each observation quarter and the corresponding default rate. The default rates significantly increased in the first half of sample, compared to the second half. Specifically, the Default Rates follow an increasing trend in the 2008-2009 period, where they peak at 2.5% in the third quarter of 2009. Thereafter, they follow a decreasing trend. The default rates seem to have flattened out in 2013, further decreasing during 2014, reaching 0.1% in the fourth quarter of 2014.

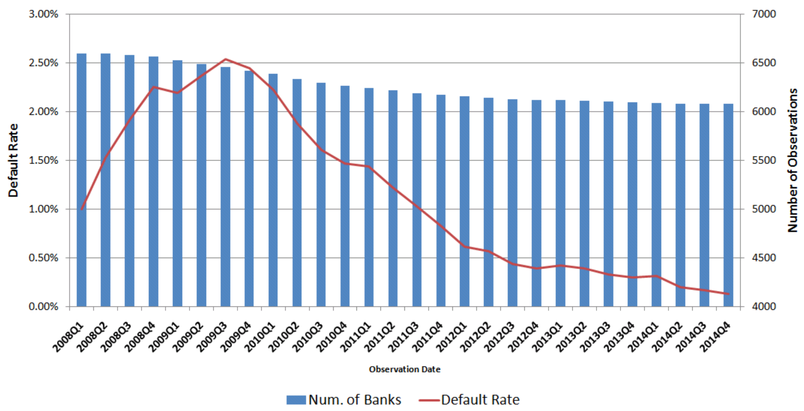


Figure 1: Bank defaults in the USA for the period 2008-2014 (source: FDIC)

The dataset was split into three parts (Figure 2). Our original development sample contains 101.641 observations that can be divided into 100.068 solvent and 1573 insolvent cases, and we call it “Full in-sample”. The overbalanced nature of our dataset, which presents a preponderance of solvent banks (i.e. good cases), does not facilitate the training of complex techniques. To this end, we created a new training sample (called “Short in-sample”), including randomly chosen 10% of the good cases and all the bad cases. So, the final training sample used to develop our models contains 10.001 good cases and 1.572 bad cases, reaching 11.573 observations in total. For the purpose of fine tuning the parameters of the random forests and neural networks specifications, we further equally divide the short in-sample dataset into training and validation sub-samples (50% each). In short, the term “Short in-sample” refers to the more balanced dataset, while the term “Full in-sample” refers to the sample that includes all the good cases. As already mentioned, the “Out-of-sample” dataset refers to the 20% randomly selected observations covering the years 2008-2012. Finally, the “Out-of-time sample” refers to the data for the years 2013-2014.

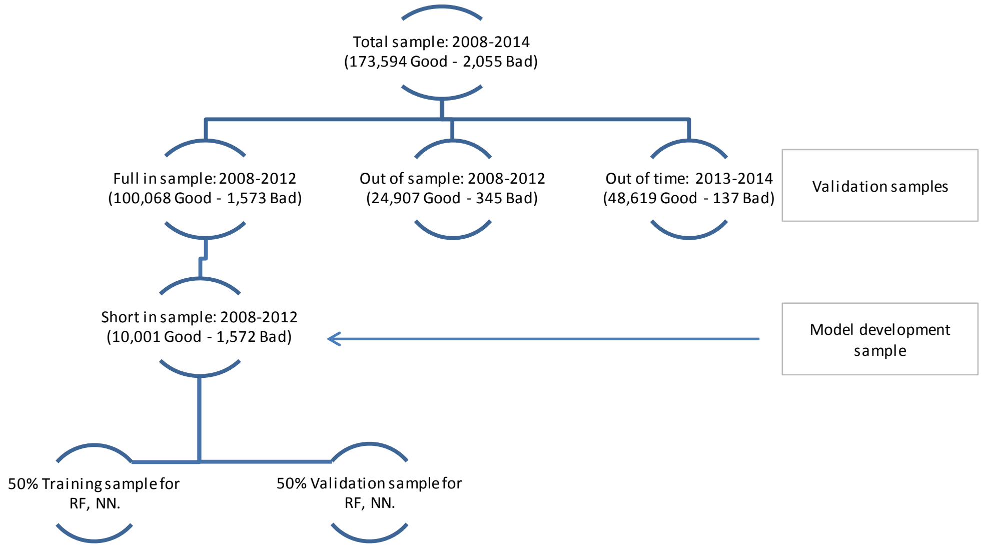
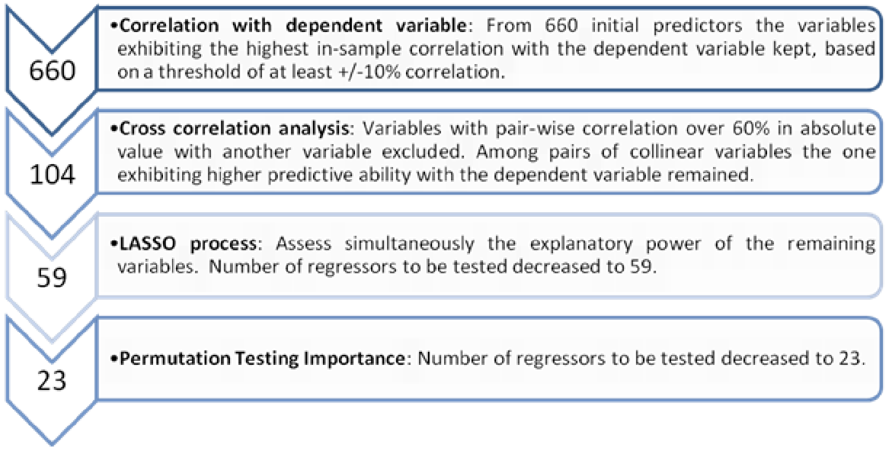

Total sample: 2008-2014
(173,594 Good - 2,055 Bad)
Full in sample: 2008-2012
(100,068 Good - 1,573 Bad)
Out of sample: 2008-2012
(24,907 Good - 345 Bad)
Out of time: 2013-2014
(48,619 Good - 137 Bad)
Validation samples
Short in sample: 2008-2012
(10,001 Good - 1,572 Bad)
Model development sample
50% Training sample for RF, NN.
50% Validation sample for RF, NN.

Figure 2: Model development and Validation samples

In developing our model specifications, we examine an extended set of variables that follow under the classification categories of CAMELS (i.e. Capital, Asset Quality, Management, Earnings, Liquidity, and Sensitivity to market risk). The variables employed and the relative transformations are shown analytically in Part I of the Appendix. The variable generation process led to a set of 660 predictors as potential candidates for our modeling procedures. The so-obtained set of time-series was narrowed down in four consecutive stages (Figure 3):



- **Correlation with dependent variable**: From 660 initial predictors the variables exhibiting the highest in-sample correlation with the dependent variable kept, based on a threshold of at least +/-10% correlation.
- **Cross correlation analysis**: Variables with pair-wise correlation over 60% in absolute value with another variable excluded. Among pairs of collinear variables the one exhibiting higher predictive ability with the dependent variable remained.
- **LASSO process**: Assess simultaneously the explanatory power of the remaining variables. Number of regressors to be tested decreased to 59.
- **Permutation Testing Importance**: Number of regressors to be tested decreased to 23.

Figure 3: Variable selection process (number of variables enclosed in arrows)

4.3. Results

In terms of performance accuracy, we focus on the out of sample and out of time accuracy of the employed specifications. When examining the out-of-sample (Table 1) performance, RFs are the best across almost all performance measures, while logistic regression seems also to be an adequate tool for assessing bank failure probability as it is ranked second. Regarding out-of-time performance, presented in Table 1, Random Forests and Neural Networks provide again the best fit, with the former method exhibiting marginally better performance in 5 criteria and better performance in 1 criterion relative to the latter. Logistic regression performs poorly in the out-of-time period, as it shows the worst performance in 6 out of 8 criteria.

Summarizing it is evident that the proposed RF rating system exhibits higher discriminatory power compared to the considered benchmark models when taking into account the skewness of the data. More importantly, the performance of RF is more stable and more consistent across all test samples, resulting in lower performance variability.

Table 1: Validation Measures – Dependent Variable concerns a bank failure event. AUROC (Area Under the Curve), G-mean (Geometric Mean), LR (negative likelihood ratio), DP (Discriminant power), Youden's index, BA (Balanced Accuracy) and WBA (Weighted Balanced Accuracy)

Out-of sample performance metrics

	Logit	**LDA**	**RF**	**SVM**	**NN**	**CRF**
AUROC	0.990	0.983	0.990	0.992	0.980	0.989
G-mean	0.919	0.905	0.934	0.916	0.922	0.907
LR-	0.144	0.169	0.113	0.150	0.130	0.165
DP	3.239	3.099	3.352	3.268	3.051	3.147
BA	0.921	0.908	0.935	0.919	0.923	0.910
Youden	0.842	0.816	0.871	0.837	0.847	0.821

Out-of time performance metrics

	Logit	**LDA**	**RF**	**SVM**	**NN**	**CRF**
AUROC	0.990	0.974	0.976	0.993	0.990	0.965
G-mean	0.741	0.824	0.862	0.819	0.862	0.838
LR-	0.452	0.321	0.255	0.329	0.255	0.296
DP	3.684	3.590	3.793	3.804	3.722	3.668
BA	0.774	0.839	0.871	0.835	0.871	0.851
Youden	0.548	0.677	0.743	0.670	0.742	0.702

5. An innovative forecasting framework for stock market crisis events

The analysis of interdependence and contagion in financial markets presents a challenging analysis topic for supervisory authorities. This is especially true in times of financial turmoil, as investors and policy makers have strong interests in knowing whether and how the crisis propagates between markets and countries. Our approach comprises a solid forecasting mechanism concerning the probability of a stock market crash event in various time frames. The developed approach combines different machine learning algorithms in modelling data from 39 countries that cover a large spectrum of economies. More precisely, we leverage the merits of a series of techniques including Classification Trees, Support Vector Machines, Random Forests, Neural Networks, Extreme Gradient Boosting, and Deep Neural Networks.

5.1. Literature Review

The use of Machine Learning Techniques in the development of early warning systems for financial crisis is rather limited in the existing literature. Cuneyt et al (2014) developed three different early warning systems, based on artificial neural networks (ANN), decision trees, and logistic regression, and tested them on the Turkish economy; artificial neural networks yielded the best performance in their analyses. Atsalakis et al. (2016) focused on 1-day stock market forecasting, specifically during stressed periods, and employed a neuro-fuzzy modeling methodology. Oztekin et al. (2016) also focused on prediction of daily stock price. Their work deployed and integrated adaptive neuro-fuzzy inference systems, artificial neural networks, and support vector machines. Dopke et al. (2017) implemented boosted regression Trees for predicting recessions. Finally, Dabrowski et al. (2016) investigated dynamic Bayesian network models and showed that they can provide significantly more precise early-warnings compared to logistic regression.

5.2. Data collection and processing

A crisis "event" for each country was identified when the daily return of the Stock Index was below the first percentile of the empirical distribution of returns. The initial empirical distribution of returns was calculated based on the stock index returns of the first 200 observations, covering the period 10/01/1996 - 15/10/1996. For each subsequent record, the empirical distribution of returns was re-calculated in order to incorporate the new observation, and an event was identified if the return was below the first percentile of the new empirical distribution. Thus, for the latest observation in the sample (i.e. 15/12/2017), the empirical distribution of returns was based on the 10/01/1996 – 14/12/2017 period.

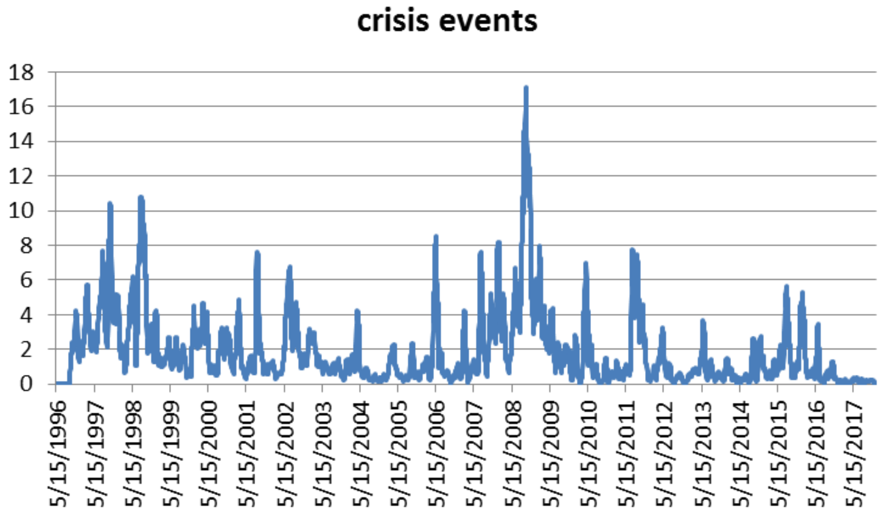


Figure 4: Number of countries exhibiting an extreme stock market fall (exceedance less than 1% percentile of the empirical distribution) movement (co-exceedances)

Figure 4 illustrates the number of countries exhibiting an extreme stock market fall during the selected 22 years period. We observe at least 10 events with global impact, with the most severe being the global economic crisis of 2008. The events are identified based on the number of stock market exceedances (less than 1% percentile of the empirical distribution) across the 39 countries in the sample (co-exceedances). More recently, i.e. from 2011 to 2017, it seems that the global market is more stable as less variability is observed, even though some events are also present.

Having identified the "events" occurring in each of the 39 countries in our sample on the basis of the daily movement of the corresponding stock indices, we proceeded to event aggregation at a region level, i.e. America, Asia, Europe, and Global. The essence of these binary variables is to capture the "significant events" within a region, i.e. events that had a collective impact on many stock markets in the region. The selected thresholds were determined in relation to the number of stock markets inof each region. Specifically, we postulated the following thresholds:

- America: At least 3 country specific events per day.
- Asia: At least 6 country specific events per day.
- Europe: At least country specific 8 events per day.
- Global: At least 2 regions are in crisis mode on a daily basis.

Based on the above outcomes, we created two classes of predictive binary variables. The first one measures whether there is a significant event in the next working day (Glob 1), both on regional basis (America, Asia, Europe) and globally; the second one measures whether there is a significant event during the next 20 working days (Glob 20). To fit the developed machine learning models, we use the binary variables for a global crash as our dependent variables, whereas the created binary target variables pertaining to each region (America, Asia and Europe) serve as independent lagged predictors.

The explanatory covariates employed in the study encompass information from stock, bond, currency, and commodity markets, along withcredit spreads and volatility indices. The covariates and their relative transformations are shown analytically in Part II of the Appendix. This variable generation process led to a set of almost 2700 potential predictors to be tested in the developed machine learning models.

5.3. Variable selection

The initially constructed dataset comprises an enormous number of independent variables, which is clearly disproportional to size of the dataset as we are dealing with around 2700 variables over around 5400 days. Fitting a machine learning model to such a huge number of independent variables (relative to the size of the dataset) is doomed to suffer from the so-called curse of dimensionality problem. That is, the fitted classifier may seem to yield very good performance in the training dataset, but it turns out to generalize very poorly, yielding a catastrophically low performance outcome in the test data. Thus, to ensure a good performance outcome for our model, we need to implement a robust independent variable (feature) selection stage, so as to limit the number of used features to the absolutely necessary. Besides, apart from increasing the generalization capabilities of the fitted models, such a reduction is also important for increasing the computational efficiency of the explored machine learning algorithms.

Figure 5 provides an overview of the adopted feature selection procedure. It comprises three phases: In the first phase, we employ three methodologies that independently assign importance to the available features: Boruta, LASSO, and a qualitative criteria-driven filter method. In the second phase, a balanced score is produced for each variable. In the third phase, we impose a heuristically determined cut-off score, and discard all features that do not reach this score. This way, a total of 131 explanatory variables are eventually selected to be retained.

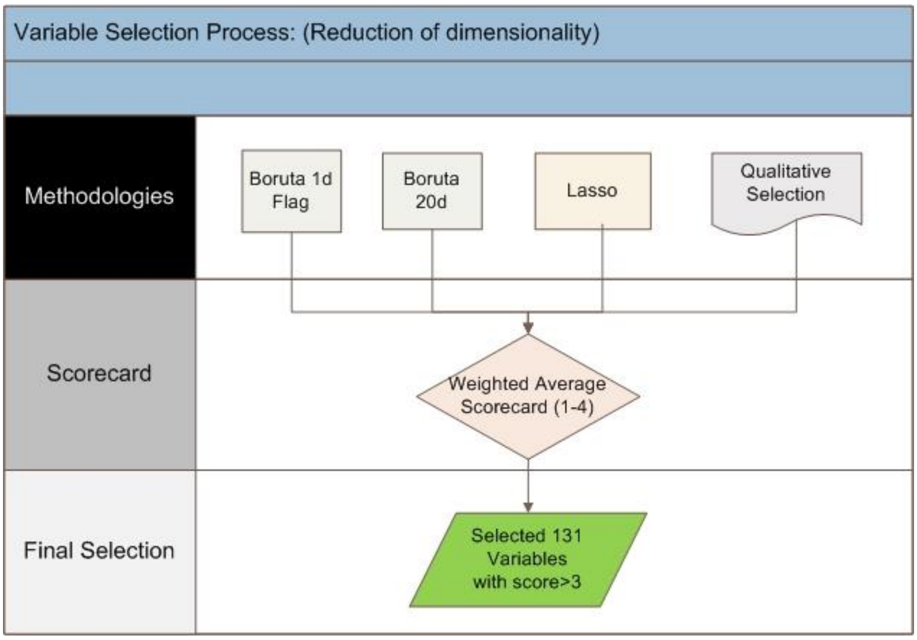

Variable Selection Process: (Reduction of dimensionality)
Methodologies
Boruta 1d Flag
Boruta 20d
Lasso
Qualitative Selection
Scorecard
Weighted Average Scorecard (1-4)
Final Selection
Selected 131 Variables with score>3

Figure 5: Process for variable scoring and selection for reducing dimensionality

The Boruta algorithm is based on a postulated Random Forest model. Based on the inferences of this Random Forest, features are removed from the training set, and model training is performed. Boruta infers the importance of each independent variable (feature) in the obtained predictive outcomes by creating shadow features. On the other hand, LASSO is a regression model that penalizes the number of model parameters in its objective function as a means of excluding irrelevant variables from the model. One of the most important features of LASSO is its ability to cope with high numbers of independent variables (features) relative to the available training observations, which is pertinent in the context of our study. We performed LASSO analysis by using the GLMNET package in R, which offers a very fast way to select best model using both cross-validation and the Bayesian Information Criterion (BIC).

Finally, the employed qualitative criteria-driven method consists of evaluating the individual correlation of each feature against the dependent variables. The rankings of variables produced by each of the above methods are combined by applying weighted average scoring. Specifically, each selection method assigns each variable 1 or 0, depending on whether they are selected by the corresponding method or not. Then, each score is multiplied by the weight assigned to each selection method. In our study, Boruta outputs are assigned a slightly higher weight due to its extensive analysis of the features in the dataset. Eventually, the final score obtained for each explanatory variable ranges in the interval from 0 to 4. To obtain the final selection, a cut-off score of 3 is applied, yielding a narrowed group of 131 candidate variables.

5.4. Results

We evaluate the predictive performance of the developed methods in the dataset covering the years 2011-2017; in the following, we refer to this part of the dataset as the “Out-of-time” sample.

As we observe in Table 3, in both horizons (1 day and 20 days) the MXNET algorithm provides the best empirical performance. This is followed by the XGBoost methodology in the case of the 20-day horizon, and the Neural Network in the case of the 1-day horizon. Hence, MXNET deep neural networks offer significantly superior predictive accuracy both in the 1-day and 20-day forecasting setup on the test sample. Another remark is that, by moving from simple neural networks to deep networks, we are able to infer richer and subtler dynamics from the data, thus increasing our capacity in modeling nonlinearities and cross-correlations among financial market variables.

Summarizing the results across all metrics in the test sample, it is evident that the MXNET system exhibits higher discriminatory power compared to all the considered benchmark models when taking into account the skewness of the data. At this point, it is important to stress that a non-anticipated crash in the global stock markets may come at a much higher cost for the economy compared to generating to false-alarm. Hence, it is crucial for supervisory purposes to achieve the maximum possible accuracy in predicting imminent crises via a developed Early Warning System for economic and financial crisis.

Table 2: Validation Measures – Dependent Variables concern a stock crisis occurring on 20-days horizon (Glob 20) and 1-day horizon (Glob 1). AUROC (Area Under the Curve), G-mean (Geometric Mean), LP (positive likelihood ratio), LR (negative likelihood ratio), DP (Discriminant power), Youden's index, BA (Balanced Accuracy)

Glob20	**Logit**	**CART**	**RF**	**SVM**	**NN**	**XGBOOST**	**MXNET**
AUROC	0.630	0.654	0.739	0.708	0.677	0.743	0.783
G-mean	0.549	0.594	0.616	0.591	0.596	0.635	0.638
LP	3.116	3.474	3.858	2.947	3.485	4.153	4.083
LR	0.743	0.680	0.645	0.690	0.677	0.615	0.610
DP	0.790	0.899	0.987	0.800	0.904	1.053	1.048
Youden	0.229	0.284	0.316	0.267	0.286	0.343	0.346
BA	0.615	0.642	0.658	0.634	0.643	0.672	0.673
Glob1	**Logit**	**CART**	**RF**	**SVM**	**NN**	**XGBOOST**	**MXNET**
AUROC	0.698	0.640	0.741	0.708	0.776	0.737	0.807
G-mean	0.610	0.606	0.583	0.610	0.557	0.611	0.682
LP	4.114	3.669	3.664	4.114	3.428	4.237	5.142
LR	0.652	0.661	0.692	0.652	0.728	0.650	0.537
DP	1.016	0.945	0.919	1.016	0.854	1.034	1.246
Youden	0.313	0.301	0.276	0.313	0.244	0.316	0.417
BA	0.657	0.651	0.638	0.622	0.657	0.658	0.708

Further, we present in Figures 6 and 7 the ROC curves corresponding to the models analyzed. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the modeling approach. The corresponding ROC curve of deep neural networks is higher over all the considered competitors regarding both the explored dependent variables (pertaining to the one-day and 20-day horizons). Hence, we obtain yet another strong evidence supporting the high degree of efficacy and generalization capacity of the proposed deep learning system.

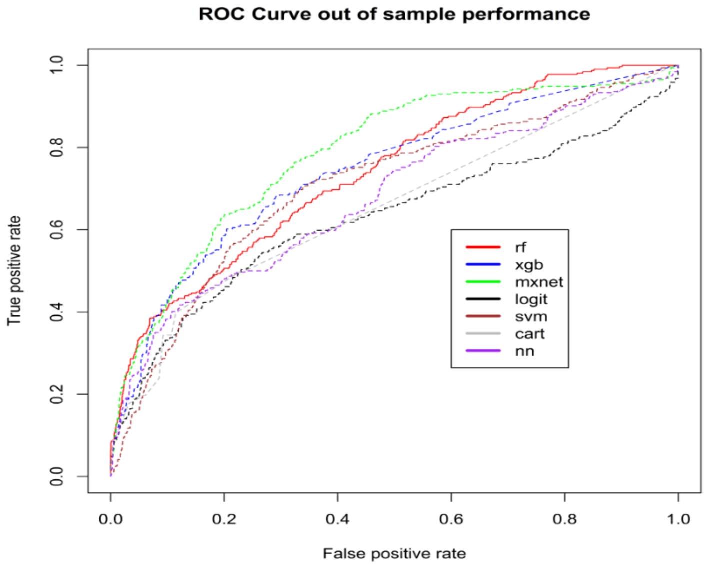


Figure 6: ROC curve for forecasting a stock crisis occurring on the 20-day horizon (Glob 20)

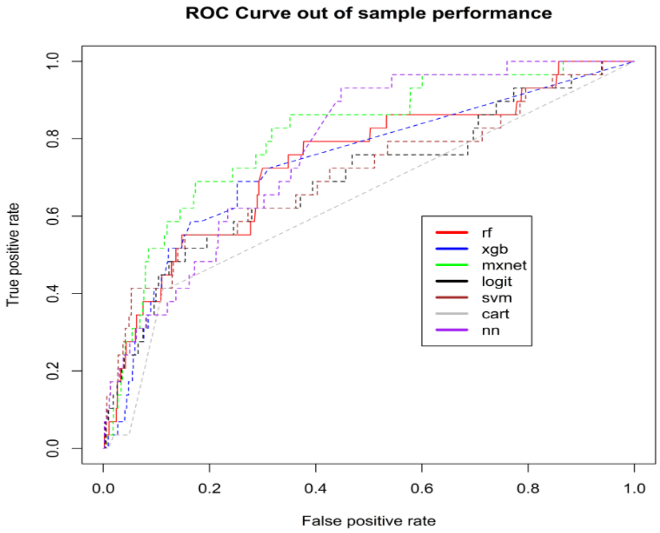


Figure 7: ROC curve for forecasting a stock crisis occurring on the one-day horizon (Glob 1)

6. Conclusions – Regulatory Implications

Our empirical results indicate that innovative statistical techniques, i.e. Deep Learning and Machine Learning methodologies, have significant predictive power. The use of these models can be used by the micro-prudential supervisors as a complement to their existing tools, notably to the Supervisory Review and Evaluation Process (SREP). Macro-prudential supervisors could also benefit by the use of these models predicting stock market crisis but also taking into consideration that systemic crisis may be created by the collapse of individual banks.

References

Altunbas, Y., Manganelli, S., & Marques-Ibanez, D. (2011). *Bank Risk during the Great Recession: Do business models matter?* ECB Working Paper No. 1394.

Atsalakis, G., Protopapadakis, E., & Valavanis, K. (2016). Stock trend forecasting in turbulent market periods using neuro-fuzzy systems. *Operational Research*, *16*(2), 245–269.

Babecký, J, Havranek, T, Mateju, J., Rusnak, M., Smidkova, K., & Vacisek, B. (2014). Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability, 15,* 1–17.

Bae, K., Karolyi, A., & Stulz, R. (2003). A new approach to measuring financial contagion. *Review of Financial Studies, 16,* 717–763.

Bekkar, M., Kheliouane, H., & Taklit, A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications, 3*(10), 2224–5782.

Berger, A., & Bouwman, H. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, *109*(1), 146–176.

Betz, F., Oprica, S., Peltonen, T., & Sarlin, P. (2014). Predicting distress in European banks. *Journal of Banking & Finance, 45,* 225–241.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5–32.

Breiman, L., Friedman, J., Stone, C., & Ohlsen, R. (1984). *Classification and regression trees*. CRC press.

Bussière, M. (2013). In Defense of Early Warning Signals. *Banque de France* Working Paper No. 420.

Chiaramonte, L., Croci, E., & Poli, F. (2015). Should we trust the Z-score? Evidence from the European Banking Industry. *Global Finance Journal, 28,* 111–131.

Christiansen, C., & Ranaldo, A. (2009). Extreme coexceedances in new EU member states' stock markets. *Journal of banking & finance, 33*(6), 1048-1057.

Cole, R. A., & Qiongbing, W. (2009). Hazard versus probit in predicting US bank failures: a regulatory perspective over two crises. *22nd Australasian Finance and Banking Conference*.

Cole, R. A., and Lawrence J. W. (2012). Déjà vu all over again: The causes of US commercial bank failures this time around. *Journal of Financial Services Research, 42*(1-2), 5–29.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273–297.

Cox, R., & Wang, G. (2014). Predicting the US bank failure: A discriminant analysis. *Economic Analysis and Policy, 44*(2), 202–211.

Cuneyt, S., Oztekin, A., Ozkan, B., Serkan, G. & Erkam, G. (2014). Developing an early warning system to predict currency crises. *European Journal of Operational Research*, *237,* 1095–1104. 10.1016.

Dabrowski, J., Beyers, C., & De Villiers, DP. (2016). Systemic Banking Crisis Early Warning Systems Using Dynamic Bayesian Networks. *Expert Systems with Applications.*

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. *International Journal of Forecasting, 33*(4), 745–759.

Estrella, A., & Mishkin, F. (1996). The yield curve as a predictor of US recessions. Current issues in economic and finance. *Federal Reserve Bank of New York, 2*(7).

Faranda, D., Flavio, M., Giachino, E., Vaienti, S., & Dubrulle, B. (2015). Early warnings indicators of financial crises via auto regressive moving average models. *Communications in Nonlinear Science and Numerical Simulation, 29*(1-3), 233–239.

Forbes, K., & Rigobon, R. (2002). No contagion, only interdependence: measuring stock market comovements. *The journal of Finance*, *57*(5), 2223–2261.

Halling, M., & Hayden, E. (2006). Bank failure prediction: a two-step survival time approach. *IFC Bulletin, 28*.

Kolari, J., Glennon, D., Shin, H., & Caputo, M. (2002). Predicting large US commercial bank failures. *Journal of Economics and Business, 54*(4), 361–387.

Lall, P. (2014). Factors affecting US Banking Performance: Evidence From the 2007-2013 Financial Crisis. *International Journal, 3*(6), 282–295.

LeCun, Y., Bengio Y., & Hinton G., *Deep learning.* Nature.

Markwat, T., Kole, E., & Van Dijk, D. (2009). Contagion as a domino effect in global stock markets. *Journal of Banking & Finance, 33*(11), 1996–2012.

Mayes, D., & Stremmel, H. (2012). *The effectiveness of capital adequacy measures in predicting bank distress.* SUERF studies.

Messai, AS., & Gallali, MI. (2015). Financial Leading Indicators of Banking Distress: A Micro Prudential Approach-Evidence from Europe. *Asian Social Science, 11*(21), 78–90.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, *18*(1), 109–131.

Oztekin, A., Kizilaslan, R., Freund, S., & Iseri, A (2016). A data analytic approach to forecasting daily stock returns in an emerging market. *European Journal of Operational Research*, *253*(3), 697-710.

Poghosyan, T., & Čihák, M. (2009). Distress in European Banks: An Analysis Based on a New Dataset. *IMF* Working Papers, 1–37.

Srivastava, N, Hinton, J., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.

Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory*, *1*. New York: Wiley.

Vašíček, B., Zingraoiva, D., Hoeberichts, M., Vermuelen, R., Smidkova, K., & De Haan, J. (2017) Leading indicators of financial stress: New evidence. *Journal of Financial Stability, 28*, 240–257.

Wanke, P., Azad, AK., Barros, CP., & Hadi-Vencheh, A. (2015). *Predicting performance in ASEAN banks: an integrated fuzzy MCDM–neural network approach.* Expert Systems.