BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

A liquidity risk early warning indicator for Italian banks: a machine learning approach

by Maria Ludovica Drudi and Stefano Nobili

June 2021

Number 1337

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

A liquidity risk early warning indicator for Italian banks: a machine learning approach

by Maria Ludovica Drudi and Stefano Nobili

Number 1337 - June 2021

The papers published in the Temi di discussione *series describe preliminary results and are made available to the public to encourage discussion and elicit comments.*

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

ISSN 1594-7939 (print)
ISSN 2281-3950 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

A LIQUIDITY RISK EARLY WARNING INDICATOR FOR ITALIAN BANKS: A MACHINE LEARNING APPROACH

by Maria Ludovica Drudi[†] and Stefano Nobili[‡]

Abstract

The paper develops an early warning system to identify banks that could face liquidity crises. To obtain a robust system for measuring banks' liquidity vulnerabilities, we compare the predictive performance of three models – logistic LASSO, random forest and Extreme Gradient Boosting – and of their combination. Using a comprehensive dataset of liquidity crisis events between December 2014 and January 2020, our early warning models' signals are calibrated according to the policymaker's preferences between type I and II errors. Unlike most of the literature, which focuses on default risk and typically proposes a forecast horizon ranging from 4 to 6 quarters, we analyse liquidity risk and we consider a 3-month forecast horizon. The key finding is that combining different estimation procedures improves model performance and yields accurate out-of-sample predictions. The results show that the combined models achieve an extremely low percentage of false negatives, lower than the values usually reported in the literature, while at the same time limiting the number of false positives.

JEL Classification: C52, C53, G21, E58.
Keywords: banking crisis, early warning models, liquidity risk, lender of last resort, machine learning.
DOI: 10.32057/0.TD.2021.1337

Contents

[†] Bank of Italy, Directorate General for Economics, Statistics and Research.
[‡] Bank of Italy, Directorate General for Markets and Payment Systems.

1. Introduction[1]

The main economic function performed by a bank is intermediation, i.e. the transfer of financial resources from those who have them to those who instead lack them. This activity takes place through the transformation of maturities: banks collect short-term or demand liabilities (deposits) from the public and transform them into less liquid assets by financing projects over longer time horizons and exposing themselves to liquidity risk. This liquidity and maturity transformation can create an incentive for investors to withdraw quickly funds in adverse situations. Creditors' loss of confidence towards a bank can trigger a liquidity crisis (Diamond and Dybvig, 1983), which could also cause difficulties to other financial institutions, threatening the overall stability of the system and adversely affecting the economy. Historians and economists often refer to widespread creditor and investor runs as "financial panics".

To deal with the effects of these structural weaknesses and prevent contagion risks in a particularly strategic sector such as the credit sector, central banks can intervene as a lender of last resort (LOLR). Providing Emergency Liquidity Assistance (ELA) represents the most traditional tool to contrast financial instability resulting from a liquidity crisis (BIS CGFS Papers, 2017; Dobler et al., 2016). ELA to financial institutions is a core responsibility of central banks because of their unique ability to create liquid assets in the form of central bank reserves, their central position within the payment system and their macroeconomic stabilization objective. The global financial crisis served as a reminder of the critical importance of the LOLR function in restoring financial stability. The provision of ELA by the central bank should only be considered when other funding solutions have already been fully explored, but it represents a process that may have to happen very fast. To be able to intervene promptly or to adopt pre-emptive actions, central banks must have a set of methodological tools useful to anticipate the occurrence of these situations of instability. These tools are useful to attempt anticipating the need for liquidity support in advance before formal action is required, so

[1] Any views expressed in this paper are the authors' and do not necessarily represent those of the Bank of Italy. We thank Luigi Cannari, Patrizia Ceccacci, Sara Cecchetti, Gioia Cellai, Paolo Del Giovane, Taneli Makinen, Franco Panfili, Rosario Romeo, Stefano Siviero, Marco Taboga and two anonymous referees for valuable comments and suggestions.

as to enable better information gathering and preparation. After the global financial crisis, central banks paid particular attention to developing or strengthening their early warning (EW) models. An EW system pursues the early detection of difficulties in order to start addressing the problems immediately. According to the literature, it is possible to distinguish two main strands of EW models. The first regards studies and models at the micro level designed to signal at an early stage the potential distress of individual institutions (e.g. Bräuning et al., 2019, Ferriani et al., 2019). The second is based on macro models attempting to identify the build-up of macro-financial vulnerabilities that threat the soundness of the banking and financial system as a whole (e.g. Aldasoro et al., 2018, Alessi and Detken, 2018, Beutel et al, 2018). The latter models are useful to guide decisions on when to activate macroprudential tools targeting excessive leverage or credit growth. The present paper focuses on the first strand, applying to Italian banks an EW system that concentrates on liquidity risk.

The models applied to predicting distress at bank level are usually multivariate systems that aim to estimate the probability of an institution's distress within a given time horizon based on a set of input indicators. While not normally providing a fully reliable signal of any impending weaknesses, these EW systems are particularly useful to highlight those intermediaries that present greater vulnerabilities, being therefore more exposed to the risk of entering a crisis. Historically, they have focused on estimating banks' probabilities of incurring a crisis starting from the analysis of historical default cases. In these models, the dependent variable typically assumes the value 1 if a bank is in default in a given period and 0 otherwise. Input variables are often related to the main intermediaries' risk profiles such as capital adequacy, credit, profitability, governance and control systems and liquidity.

The present paper contributes to the literature on EW models applied to the individual financial institutions in several ways. First, using the data available to the Bank of Italy, we shift the focus from the analysis of insolvency risk to liquidity risk introducing a novel dataset of bank liquidity crisis events, starting from banks that resorted to the central bank's ELA.

Unlike most of the literature, which focuses on the analysis of the insolvency risk and typically proposes a forecast horizon ranging from 4 to 6 quarters, we consider as the main forecast horizon the 3-month one. This choice depends on the fact that liquidity crises by their nature tend to arise quickly and require prompt intervention. Indeed, the choice of the forecast horizon has to satisfy the trade-off between being as short as possible in order to obtain very accurate estimates and being large enough to allow central banks to take corrective actions that could possibly avoid the onset of the crisis.

We compare the predictive capabilities of three different models, and then proceed to combine them. The three models are the logistic LASSO, the random forest and the Extreme Gradient Boosting. Our results notably improve the predictive performance with respect to the most usual modeling techniques. To identify the model specification that optimizes its out-of-sample predictive capacity (Tibshirani, 1996), the first approach uses a LASSO (Least Absolute Shrinkage and Selection Operator) cross-validation technique, based on maximizing a binomial logarithmic likelihood function subject to a penalty term. The second classification method, introduced by Breiman (2001), is a generalization of decision tree models and computes the estimates by aggregating the forecasts derived from a large number of different trees. By combining the trees predictions, the variance of the estimates decreases, making them more efficient[2]. This phenomenon occurs because random forest models generate heterogeneity between the different decision paths by randomly choosing a subset of observations on which the estimate is made (bagging) and considering only a random subset of indicators at each step (attribute bagging). Finally, the third method, exactly as the random forest, is a classification model that generalizes the decision trees models. This time, the trees are combined sequentially by working on the residuals of the previous tree, thereby reducing estimates bias. The Extreme Gradient Boosting (XGBoost) algorithm, proposed by Chen and Guestrin (2016), has become a popular supervised learning algorithm, in particular for financial time series data. To avoid the overfitting problem, this algorithm uses features subsampling as the random forest and a regularization parameter to reduce the final prediction sensitivity to single observations.

[2] Decision trees are extremely sensitive to the specific dataset they are based on. Sometimes they show the problem of overfitting: they have a high in sample performance, but are not very versatile if applied outside the sample.

The literature on insolvency or liquidity risk of banks usually presents individual models' results. In contrast, to obtain a more robust system for measuring banks' liquidity vulnerabilities, this work follows the approach of comparing and aggregating the information obtained by the three methods. Indeed, it is likely that different approaches capture different types of fragility, and therefore can complement each other, giving value to their simultaneous use (Holopainen and Sarlin, 2017).

Following Sarlin (2013), the models' signals are calibrated according to the policymaker's preferences between type I (missing crises) and type II (false alarm) errors modeled by a loss function. The signals are then aggregated using both a simple average and a weighted average with a measure of relative usefulness as weight. To take into account central bank's preference for models that minimize type I error, in line with bank early warning literature (e.g. Holopainen and Sarlin, 2017), for each method the loss function has been optimized attributing a weight to missing a crisis far greater than the one of a false alarm. The rationale behind this choice follows the fact that an EW signal should trigger actions by the authority to investigate and possibly restore the liquidity situation of the bank. Should the analysis reveal that the signal is false, there is no loss of credibility for the policymaker as model results are not published.

The key finding of the paper is that complementing different estimation procedures improves model performance and yields accurate out-of-sample predictions of banks' liquidity crises. Results show that the combined model manages to have an extremely low percentage of false negatives (missing a crisis) in out-of-sample estimates, equal to 10%, while at the same time limiting the number of false alarms too. These results are lower than the values usually reported in the literature. This could depend on our choice to test the model on a 3-month forecast horizon, which is perfectly suitable for liquidity risk models, but shorter than the horizon usually considered in insolvency risk models.

This work fits two strands of literature: the first relating to forecasting individual bank crisis and the second one to optimizing early warning signals through models. An initial contribution to the first strand is the work of Martin (1977), which uses a logistic regression based on balance sheet variables (CAMEL rating system: Capital adequacy, Asset quality, Management quality, Earnings, Liquidity) to predict the default of a set of US banks between 1970 and 1976. Subsequent literature focused on identifying the explanatory variables to be included in the model to improve its predictive performance (Thomson, 1992; Cole and Gunther, 1998; González-Hermosillo, 1999; Jagtiani et al., 2003; Kraft and Galac, 2007). Other EW models included market indicators, as they are more responsive in identifying bank's health status (Flannery 1998; Krainer and Lopez 2003; Campbell et al., 2008). Finally, Poghosyan and Čihák (2011) built an EW system in which a bank's crisis is identified through a massive search for some keywords in the press reports.

Works like Kaminsky et al. (1998) have contributed to the second strand of literature. They introduced the so-called "signal approach" based on the identification of thresholds beyond which the individual indicators signal a crisis. Demirguc-Kunt and Detragiache (2000) introduced the use of a loss function that allows the authority to weigh its cost of missing a crisis (type I error) differently from a false alarm (type II error). Subsequent works modified the structure of the loss function to adapt it to the specific characteristics they need (Alessi and Detken, 2011; Lo Duca and Peltonen, 2013; Sarlin, 2013). Lang et al. (2018) propose a system to derive EW models with an excellent out-of-sample forecasting performance. The use of a loss function that evaluates the model in terms of preferences between type I and type II errors is associated with a regularized logistic regression (logistic LASSO regression) that allows selecting the model specification with the best forecasting properties.

More recently, the EW literature has started using a wide range of nonparametric techniques that can be generically traced back to machine learning. Tam and Kiang (1992), Alam et al. (2000), Boyacioglu et al. (2009), among others, present EW models based on neural networks, a non-linear approach in which the identification of institutions facing difficulties is based on some interconnection measures between explanatory variables. The recent works of Tanaka et al. (2016), Holopainen and Sarlin (2017), Alessi and Detken (2018) use a random forest to improve the forecasting performance of EW models. Bräuning et al. (2019) identify cases of individual bank financial distress by estimating a decision tree model based on the Quinlan C5.0 algorithm.

The remaining part of the paper is organized as follows. Section 2 focuses on data and variables definition; in Section 3, we present the loss function and a measure of usefulness. Section 4 describes the classification models. In Section 5, we compare the results and characteristics of the individual models and we propose two combination methods. Section 6 describes the results from two robustness checks and Section 7 finally concludes.

2. Data and variables

The EW system developed in this paper uses Bank of Italy's confidential data on recourse to ELA as well as supervisory reporting and market data, covering all banks that operate in Italy over a period of more than 5 years, from December 2014 to January 2020. The sources of bank-level data are the Common Reporting (COREP, containing capital adequacy and risk-specific information), the Financial Reporting (FINREP, which includes balance sheet items and detailed breakdowns of assets and liabilities by product and counterparty), both available since December 2014, and other datasets from national supervisory reporting[3].

The EW model estimation is based on the interaction between a set of attributes, selected from those that can better signal the onset of potential crises, and a binary target variable that takes value 1 if a bank is going through a period of crisis and 0 otherwise. To construct this binary variable it is necessary to identify which banks are facing a crisis. To the best of our knowledge, we adopt an innovative perspective with respect to the prevailing literature. Indeed, we focus on the analysis of liquidity risk identifying the period in which the liquidity distress starts, builds up and possibly outbreaks. This choice helps to detect cases of financial difficulties early enough to allow for a timely intervention of the authorities (see e.g. Rosa and Gartner, 2018 and Bräuning et al., 2019).

In order to have an EW system on bank liquidity crises, we defined our target variable in an extensive way. First, we considered banks that resorted to the central bank's ELA. Central banks may provide ELA when solvent financial institutions face temporary liquidity shortages and after other market and/or private funding sources have been exhausted. ELA aims at mitigating sudden funding pressures that could, if not addressed, jeopardize financial stability. By providing liquidity support on a temporary basis, central

[3] These variables come from banks' balance sheet data at the highest degree of consolidation. To get the monthly equivalent for variables available only on a quarterly basis, we kept the last value constant.

banks seek to prevent a systemic loss of confidence that can spread to otherwise healthy institutions. The Bank of Italy (BoI) provides ELA under the European System of Central Banks (ESCB) ELA Agreement, published by the European Central Bank (ECB) in May 2017, which places the ELA function within the prerogatives of National Central Banks. In BoI, the function to provide ELA is allocated to the competent organizational unit within the Markets and Payment Systems Directorate.

Considering only banks' recourse to ELA, however, would be too narrow as ELA is only one of the possible outcome of a liquidity crisis. Moreover, banks' recourse to ELA in Italy has not been frequent. For this reason, we extended the definition of liquidity crises to include other possible manifestations of banks' liquidity deterioration. Among other situations that arguably indicate a state of liquidity distress, we considered banks subject to restrictive measures on participation to monetary policy operations. Indeed, according to the General Documentation (Guideline 2015/510 of the ECB and following amendments)[4], the Eurosystem allows the participation to monetary policy operations only to financially sound institutions. If this condition is not respected at a certain point in time or there are instances that are deemed to affect negatively the financial soundness of the bank, the Eurosystem can limit, suspend or exclude it from monetary policy operations. These measures typically follow a reduction of capital ratios under the minimum regulatory requirements; but even if they are conceptually closer to solvency than illiquidity risk (though, of course, telling the two apart is difficult) they usually have immediate negative consequences on the bank's liquidity position.

In addition, to promptly identify liquidity distress events, we took into account banks subject to enhanced monitoring by the BoI's Supervisory Directorate and brought to the attention of the BoI's Markets and Payment Systems Directorate, regardless of whether or not the bank was eventually given access to ELA. Indeed, the Supervisory Department puts banks under enhanced monitoring when liquidity is starting to deteriorate and the frequency of regular supervisory liquidity reporting is deemed insufficient to adequately monitor emerging liquidity strains.

---

[4] https://www.ecb.europa.eu/ecb/legal/pdf/oj_jol_2015_091_r_0002_en_txt.pdf

We also included banks that received State support either in the form of Government Guaranteed Banks Bonds (GGBBs) or in the form of a precautionary recapitalization[5].

Moreover, we considered banks under early intervention measures as defined by the BRRD[6], e.g. banks placed under special administration and/or subject to the appointment of a temporary administrator, regardless of the fact that special or temporary administration succeed in rehabilitating the bank. Although not all banks in the sample were placed under early intervention measures due to liquidity problems, all these events are connected directly or indirectly to a liquidity deterioration. Indeed, even if the aim of early intervention measures is restoring ailing institutions in the medium time horizon, in the short time it could produce a negative effect on the liquidity situation, due to deposit withdrawals by the customers. This assumption is confirmed by performing a panel fixed effect regression where the independent variable is a dummy that takes value 1 if the bank in that month was placed under early intervention measures and 0 otherwise and the dependent variable is the monthly change of the sight deposit variation (Table A1). In this way, we are able to capture changes in the deposit trend due to early intervention. In line with our expectations, the coefficient is negative and statistically significant.

---

5 Art. 32, point 4, letter d of the BRRD: "[…] the extraordinary public financial support takes any of the following forms: i) a State guarantee to back liquidity facilities provided by central banks according to the central banks' conditions; ii) a State guarantee of newly issued liabilities; iii) an injection of own funds or purchase of capital instruments at prices and on terms that do not confer an advantage upon the institution […].

6 The early intervention measures are set in Artt. 27 - 29 of the BRRD. Art. 27: "Where an institution infringes or, due, inter alia, to a rapidly deteriorating financial condition, including deteriorating liquidity situation, […] is likely in the near future to infringe the requirements of Regulation (EU) No 575/2013, Directive 2013/36/EU, Title II of Directive 2014/65/EU or any of Articles 3 to 7, 14 to 17, and 24, 25 and 26 of Regulation (EU) No 600/2014, Member States shall ensure that competent authorities have at their disposal […] at least the following measures: a) require the management body of the institution to implement one or more of the arrangements or measures set out in the recovery plan […]; b) require the management body of the institution to examine the situation, identify measures to overcome any problems identified and draw up an action program to overcome those problems […]; c) require the management body of the institution to convene […]; d) require one or more members of the management body or senior management to be removed or replaced […]; e) require the management body of the institution to draw up a plan for negotiation on restructuring of debt […]; f) require changes to the institution's business strategy; g) require changes to the legal or operational structures of the institution; h) acquire, including through on-site inspections and provide to the resolution authority, all the information necessary in order to update the resolution plan and prepare for the possible resolution of the institution […]." Art. 28 "Removal of senior management and management body", Art. 29 "Temporary administrator".

Finally, if not already detected, we included banks deemed to be failing or likely to fail (FOLTF) for liquidity reasons[7].

Usually a bank in liquidity crisis sequentially experiences various events of liquidity distress. Given the importance of the time dimension for a model aiming to predict a crisis, we considered the beginning of the crisis from the starting date, if available, of the first liquidity distress event. Indeed, the full-blown crisis event is often preceded by a period in which the bank's liquidity position gradually deteriorates, although each situation is different from the others. In some cases, it is possible to identify precisely the triggering event, other times the cause is more remote and it is not easy to define an unequivocal starting point. To improve the accuracy in identifying crisis periods, we carried out an analysis on the CDS and bonds (senior and subordinated) market trends, as their dynamics usually anticipate the evolution of the issuer's health[8]. The model is run on a monthly basis. In line with the majority of the EW literature, institutions that enter in liquidity distress are recorded as stressed banks as long as the crisis is ongoing (the target variable takes value 1), while at the end of that period they are changed back to their original value of 0. This choice depends on the fact that we are interested in understanding the dynamics of banks entering and exiting the distress status.

Following these criteria, during the analysed period (December 2014 - January 2020), we identified 31 banks that have gone through periods of liquidity crisis of different lengths, for a total of 527 out of 12,761 observations. The crisis events show a peak during 2016, reaching a maximum of 6.2% of the total number of observations in the year (Figure A1).

To check the robustness of our results, we considered two further specifications of the target variable. The first one is based on a more restrictive definition of liquidity events. We considered only the events more closely linked to a liquidity deterioration such as resorting to central bank's ELA, benefiting from State support through Government

[7] The FOLTF for liquidity reasons is set in Art. 32 of the BRRD. Art.32, point 4 "[…] an institution shall be deemed to be failing or likely to fail in one or more of the following circumstances: […] (c) the institution is or there are objective elements to support a determination that the institution will, in the near future, be unable to pay its debts or other liabilities as they fall due. […]". So far, no Italian bank has been declared FOLTF for liquidity reasons.

[8] For intermediaries for which such data are available, the onset of the crisis has been made to coincide with increases in premiums and / or returns higher than the 90th percentile of the historical distribution.

Guaranteed Banks Bonds (GGBBs) or, finally, being subject to enhanced liquidity monitoring by the BoI's Supervisory Directorate and brought to the attention of the BoI's Markets and Payment Systems Directorate. Following these criteria, we identified 15 banks that have gone through periods of liquidity crisis of different lengths, for a total of 376 out of 12,761 observations.

In the second robustness check, we considered as liquidity crisis event only the first month in which one of the events previously described occurred and excluded from the dataset the following observations related to that event for that bank. However, we allow for the possibility that during the period a bank incurs in different liquidity crises. We identified 39 observations associated to a liquidity crisis event. In this setting, given the limited occurrence of crisis events (0.3%), we oversampled the minority class by simply duplicating examples from that class in the training dataset prior to fitting the model (e.g. Chawla et al., 2002).

Early-warning models aim at signalling crisis events in advance, but the specific forecast horizon will depend on the application at hand. We considered as the main forecast horizon for our analyses the 3-month one, a time span that would give the authorities enough time to adopt pre-emptive actions. This choice depends on the fact that liquidity crises by their nature tend to quickly build-up and require prompt intervention. Indeed, the choice of the forecast horizon has to satisfy the trade-off between being as short as possible in order to obtain very accurate estimates and being large enough to allow central banks to take corrective actions that could possibly avoid the onset of the crisis.

As attributes, we considered 20 indicators selected in order to capture the peculiarities of the financing structure of Italian banks.

The indicator that best provides a first measure of banks' liquidity is the Liquidity Coverage Ratio (LCR). The LCR is a coverage index that aims at ensuring that an intermediary maintains an adequate stock of freely available high quality liquid assets (HQLA) to meet any liquidity needs over 30 days in a stressful scenario. It is computed according to the following formula:

$$LCR = \frac{HQLA}{Outflows - \min(Inflows;\ 75\%\ Outflows)}$$

where outflows are calculated by multiplying the outstanding balances of various categories or types of liabilities and off-balance sheet commitments by the rates at which they are expected to run off or be drawn down in a specified stress scenario. Inflows are calculated by multiplying the outstanding balances of various categories of contractual receivables by the rates at which they are expected to flow in under the scenario up to an aggregate cap of 75% of outflows. The LCR has been a regulatory requirement since October 2015 and starting from 1st January 2018 the lowest allowed value is 100%.

The asset encumbrance ratio (AE) and the asset encumbrance eligible ratio (AEE) both provide a measure of the share of assets tied up in financing transactions, but the second one considers only the subset of assets eligible for monetary policy operations. The eligibility requirement represents an approximation of the marketability of an asset; therefore, the AEE index measures the encumbered portion of the most easily tradable assets. They can be computed according to the following formulas:

$$AE = \frac{Encumbered\ assets + Collateral\ received\ and\ reused}{Total\ assets + Total\ collateral\ received}$$

$$AEE = \frac{Encumbered\ eligible\ assets + Eligile\ collateral\ received\ and\ reused}{Total\ eligible\ assets + Total\ eligible\ collateral\ received}$$

From the liquidity risk point of view, the analysis of the encumbrance is important since the scarce availability of unencumbered collateral means that the bank has lower financing capacity in stressful situations. In addition, high and increasing AE levels, to the extent that they represent an element of fragility of the intermediary, could lead to greater difficulties in placing unsecured financing instruments on the market.

Then, we consider two variables related to sight deposits: their variation and a coverage index. The first one is a weighted average of the deposits monthly changes during the previous 12 months, with a greater weight (w_s) for the most recent data.

$$Sight\ deposits\ change = \sum_{t=1}^{T} \sum_{s=0}^{11} \Delta\, Sight\ deposits_{t-s} * w_s$$

$$w_s = q^s, \qquad q = 0.9$$

This allows to capture their temporal evolution and, therefore, to have information about the depositors' confidence level toward a bank.

On the other hand, the second indicator is computed as the difference between cash and / or easily disposable assets[9] and bonds maturing within the year over total sight deposits.

$$Sight\ deposits\ and\ bond\ coverage\ ratio = \frac{Liquid\ assets\ - Maturing\ bonds}{Sight\ deposits}$$

$$Liquid\ assets = Excess\ liquidity + Pool\ overcollateralization + Uncommitted\ eligible\ securities$$

The indicator provides a measure of the bank's ability to cope with a reduction in sight deposits through its cash and cash equivalents, under the assumption that it is unable to renew the bonds falling due within the year.

Starting from the components of this latter indicator, we built two other attributes. The first one is the ratio between our measure of liquid assets and bank’s total assets, while the second one is the ratio between liquid assets and sight deposits, without considering the maturing bonds.

The Eurosystem dependency index is the ratio between the share of Eurosystem refinancing to total funding and a level of capitalization, measured by the CET1 ratio:

$$Eurosystem\ dependency\ index = \frac{Eurosystem\ refinancing}{Total\ funding * CET1}$$

The inclusion of the capital ratio allows discriminating banks that resort to monetary policy operations to optimize their financing cost from those that, instead, resort to it because they have difficulties in accessing the market. To be thorough, we also included in the models the ratio of Eurosystem refinancing over total funding and the ratio of Eurosystem refinancing over total assets.

[9] The easily disposable assets are composed of liquidity deposited in the current accounts with the Bank of Italy in excess with respect to the minimum reserve requirement (excess liquidity), overcollateralization in the monetary policy collateral pool and uncommitted eligible securities outside the collateral pool.

The fragility index on the bond market is computed as the difference between net issues[10] and bonds falling due in the following 12 months, where both measures are reported on a monthly base, over total outstanding bonds:

$$Fragility\ index\ on\ the\ bond\ market\ = \frac{Net\ issues - Maturing\ bonds}{Total\ outstanding\ bonds}$$

This indicator is included as it allows comparing the ability to renew bond financing with the degree of reliance on this source of funding.

With reference to the bond market, we also computed two other attributes: the ratio between net issues and total outstanding bonds and the ratio between bonds falling due in the following 12 months and total outstanding bonds.

The Net Cash Outflows (NCO) over total assets indicator, comparing the LCR denominator with total assets, allows to analyze the liquidity risk from a complementary perspective with respect to the one of the LCR and to obtain a relative measure of the net outflows in a 30-day horizon.

$$Net\ Cash\ Outflows\ over\ total\ assets\ = \frac{Net\ Cash\ Outflows}{Total\ assets}$$

Then, we included the capital ratios (CET1, Tier1 and Total Capital Ratio) and the NPL ratio as their deterioration usually involves a worsening of an institution liquidity situation. Furthermore, based on the rules set out in the Eurosystem's General Documentation, a failure to comply with capital requirements, undermining banks' solvency, could result in the limitation of the existing refinancing or in the suspension from monetary policy operations with potential consequences on banks liquidity position.

Finally, we considered two variables related to the overall system: the indicator of systemic liquidity risk in the Italian financial markets proposed by Iachini and Nobili (2014), and the 10-year BTP yield. The former is a coincident indicator of systemic liquidity risk that exploit standard portfolio theory to take account of the systemic dimension of liquidity stress in the financial market. Three sub-indices reflecting liquidity

[10] Net issues are monthly equivalent values of quarterly data moving averages.

stress in specific market segments (the money market, the secondary market for government securities, and the stock and corporate bond markets) are aggregated in the systemic liquidity risk indicator in the same way as individual risks are combined in order to quantify the overall portfolio risk. The aggregation takes into account the time-varying cross-correlations between the sub-indices, using a multivariate GARCH approach. This is able to capture abrupt changes in the correlations and makes it possible for the indicator to identify systemic liquidity events precisely. The inclusion of the 10-year BTP yield aims at controlling for the interconnection between financial health of banks and of sovereigns, the so-called sovereign-bank nexus. Indeed, a souring of the sovereign funding conditions might cause an increase in banks' risk. These two variables are useful to classify the effect on single bank's liquidity vulnerability stemming from the overall Italian financial market.

Tables A2, A3 and A4 show respectively some descriptive statistics for each of these indicators and their correlation matrices, both non-standardized and standardized (through a transformation based on their empirical cumulative distribution function, CDF, involving the computation of order statistics[11]). From the analysis of the correlation matrices, it appears evident that our indicators are constructed to focus on specific and different aspects of funding liquidity risk thereby showing low correlation. In line with expectations, the LCR is negatively correlated with all the variables based on Eurosystem refinancing. Besides, AER and AEER are correlated among each other and with the Eurosystem dependency index. Finally, the correlations between capital ratios are strong.

3. A policymaker's loss function and the usefulness measure

Early warning models typically forecast crisis probabilities; by combining them with thresholds based on policymaker's preferences, they produce binary signals (in this framework liquidity crisis/no liquidity crisis). Converting probabilities into clear signals

[11] We prefer the CDF approach to "classic" standardization (i.e. by subtracting the sample mean from the raw indicator and dividing this difference by the sample standard deviation). Classic standardization, in fact, implicitly assumes variables to be normally distributed; but since many indicators violate this assumption, the risk that the results obtained from the use of standardized variables are sensitive to outlier observations is enhanced. The CDF transformation projects the indicators into variables which are unit-free and measured on an ordinal scale with range (0,1].

helps to inform policymakers' ultimate decision on whether and when to take action. Moreover, it allows for a straightforward evaluation of predictions in terms of right or wrong signals.

For this reason, early-warning models require a signal evaluation framework that resembles the decision problem faced by the policymaker. The signal evaluation framework focuses on a policymaker with a relative preference between type I (false negative) and II (false positive) errors and the utility obtained by using a model versus not using it. Policymaker's preferences between type I and II errors should internalize the expected costs of a banking crisis and of a false alarm.

The occurrence of a crisis can be represented with a binary state variable $I_j(0) \in \{0,1\}$ (with observation $j=1,2,\ldots,N$), where $I_j(0)=0$ indicates a tranquil period and $I_j(0) = 1$ a distress period. However, to enable policy actions to avoid the build-up of single financial institutions' liquidity crises, the early-warning model should identify a crisis period in advance $I_j(h) \in \{0,1\}$ with a specified forecast horizon h, meaning that $I_j(h) = 1$ if $I_{j+h}(0) = 1$.

To identify events $I_j(h)$ using the information coming from the explanatory indicators, econometric models can be used. Indeed, models estimate for each bank a crisis probability $p_j \in \{0,1\}$, which is then converted into a binary prediction P_j that assumes value 1 if p_j exceeds a specified threshold (ϑ) and 0 otherwise. Given model probabilities p_j, the policymaker should focus on choosing a threshold such that her loss is minimized. Indeed, lower threshold values will lower the probability of failing to signal a crisis at the cost of increasing the number of false alarms. The choice of the threshold should be conservative since for a central bank, which is typically more worried by the negative consequences of an unexpected crisis, the cost of a false alarm is lower than the cost of missing a crisis (Betz et al., 2013). The correspondence between the prediction P_j and the empirical event I_j can be summarized by a so-called confusion (or contingency) matrix, which returns a representation of the statistical classification accuracy. Each column of the matrix represents the empirical crisis/ no crisis events, while each row represents crisis/ no crisis events predicted by the model (Table 1).

Table 1 – Confusion matrix

	No crisis	Crisis
No signal	True Negative (TN)	False Negative (FN)
Signal	False Positive (FP)	True Positive (TP)

The evaluation framework in this paper follows Sarlin (2013), where the optimal value of the threshold ϑ is obtained by turning policymakers' preferences between type I and II errors into a loss function. Type I errors represent the proportion of missed crises relative to the number of crises in the sample, $T_1 = \frac{FN}{(TP+FN)}$, and type II errors the proportion of false alarms relative to the number of no crisis periods in the sample, $T_2 = \frac{FP}{(FP+TN)}$.

The loss of a policymaker depends on T_1 and T_2, weighted according to her relative preferences between missing crises (μ) and issuing false alarms ($1-\mu$). By accounting for unconditional probabilities of crisis events ($P_1 = \frac{(TP+FN)}{(FP+TP+FN+TN)}$) and of no crisis periods ($P_2 = \frac{(FP+TN)}{(FP+TP+FN+TN)}$)[12], the loss function can be written as follows:

$$L(\mu) = \mu T_1 P_1 + (1-\mu) T_2 P_2$$

Using the loss function $L(\mu)$, it is possible to define the usefulness of a model. Following Sarlin (2013), we derive the absolute usefulness (U_a) of a model by subtracting the loss generated by the model from the loss of ignoring it:

$$U_a(\mu) = min(\mu P_1, (1-\mu)P_2) - L(\mu)$$

The absolute usefulness (U_a) gives a measure of the utility a policymaker can obtain from using a model with respect to not using it. A policymaker that does not have a model could achieve a loss equal to $min(\mu P_1, (1-\mu)P_2)$, depending on her relative preferences, μ. More specifically, as the unconditional probabilities are commonly imbalanced and the policymaker, based on her preference μ, may be more concerned about one class of

[12] Since crisis events are rare, errors must be scaled by the observations.

events, she could achieve a loss of μP_1 by never signaling or a loss of $(1-\mu)P_2$ by always signaling. In order to compare two different models, it is necessary to introduce the concept of relative usefulness, which is computed by comparing the absolute usefulness with the maximum possible usefulness of the model. That is, U_r reports U_a as a percentage of the usefulness that a policymaker would gain with a perfectly performing model (which corresponds to a model with a loss $L(\mu)=0$):

$$U_r(\mu) = \frac{U_a(\mu)}{min(\mu P_1, (1-\mu)P_2)}$$

In this paper, the value of μ has been set on values that allow to attribute a greater weight to missing a crisis compared to a false alarm. Using different values of μ in a range feasible to a conservative policy maker (0.75, 0.98)[13], the relative usefulness of the models appears to be very high (Figure A2).

4. The classification models

The aim of each classification algorithm is to identify which state of the world (crisis or no-crisis) a new observation belongs to, based on one or more predictive attributes. Classification is an exercise of supervised learning, which starts from a set of correctly classified observations (training set). In this paper, we applied three different classification models (logistic LASSO, random forest and Extreme Gradient Boosting) and we compared their results with the ones of a classic logistic regression.[14] They allow us to identify the state of the world to which each observation belongs through the optimization of the utility function presented in Section 3.

[13] These values of μ have been selected according to the relevant literature and correspond to weighting a missing crisis between 3 and 49 times more than a false alarm.

[14] Among all possible available methods, we selected the logistic LASSO because it is a flexible, analytical and intuitive method for early warning purposes. On the other hand, the random forest and XGBoost are well-known and established methods for forecasting.

Logistic LASSO

Logit models are widely used in the early warning literature (Frankel and Rose, 1996; Bussiere and Fratzscher, 2006; Lo Duca and Peltonen, 2013). They depend on two main assumptions. The first one is that the target binary variable is driven by a latent process (y^*), linked to the binary variable y by a logistic transformation:

$$\Lambda(y) = \frac{e^{y^*}}{1 + e^{y^*}}$$

Second, this latent process can be written as a linear combination of the explanatory variables:

$$y^* = X\beta + \varepsilon$$

Compared to machine learning methods, the main advantage of a logit model is that it is easy to interpret because it is based on a clear statistical model, which explicitly models uncertainty. At the same time, however, it is restricted to a specific functional form.

Since we are interested in forecasting the liquidity crisis probability of each bank, our first approach applies a classic logistic regression in combination with a LASSO (Least Absolute Shrinkage and Selection Operator; Tibshirani, 1996) cross-validation technique to set the penalty parameter that determines the complexity of the model. The aim of regularized methods[15] is to avoid overfitting. For this reason, the LASSO method maximizes the binomial logarithmic likelihood function $l(\beta|y)$ subject to a penalty term, which depends on the estimation of the coefficients:

$$max_\beta \; l(\beta|y) - \lambda \sum_i ||\beta_i||_1$$

$$l(\beta|y) = \frac{1}{N * T} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t} * \beta' x_{i,t} - log\,(1 + e^{\beta' x_{i,t}})$$

[15] Regularization is used to obtain a better out-of-sample performance by preventing the algorithm from overfitting the training dataset. We use as regularized term the L1 norm i.e. the sum of the absolute values of the coefficients.

where λ is the LASSO penalty parameter that determines the complexity of the model and $||\beta_i||_1$ is the L1 norm. Following Lang et al. (2018) we make use of L1 regularization in order to shrink variables towards zero, producing sparse models. In this way, the modeling technique is also a simultaneous variable selection device. Since different values of λ yield different model specifications, its selection becomes a key step. This decision highly depends on the purpose of the model. Since in this case we are building an early warning model for predictive purposes, we choose the parameter in order to maximize the model's forecasting performance. In line with Lang et al. (2018), we apply the cross-validation technique in combination with the relative usefulness presented in Section 4. The aim of cross-validation is to estimate the ability of a model to make predictions on previously unseen data. The typical approach when using cross-validation is to randomly shuffle the data and split them into k folds or blocks of the same size. Each fold is a subset of the data comprising t/k randomly assigned observations, where t is the number of observations. After this split, it is possible to estimate k times the penalty coefficient, each time using only k-1 subsets and leaving the last one for validation. However, some variants of t/k-fold cross-validation have been proposed especially for dependent data. One procedure largely used in literature is the blocked cross-validation[16]. The main difference with the standard procedure is that there is no initial random shuffling of observations. In time series, this generates k blocks of contiguous observations: in this way, the natural order of observations is kept within each block, but broken across them. Cross-validation allows improving estimates by computing the value of the penalty coefficient that maximizes each time the performance over the sample left out for validation. The optimal λ is the one associated with the highest cross-validated relative usefulness. To apply the blocked cross-validation to our training data, we split the sample by year into 5 folds of similar dimension. In this way, we select the optimal λ

[16] See e.g. Bergmeir and Benìtez (2012), Oliveira et al. (2018), Arlot and Celisse (2018), Cerquiera et al. (2019). Bergmeir and Benìtez (2012) present a comparative study of estimation procedures using stationary time series. Their empirical results show evidence that in such conditions blocked cross-validation procedures yield more accurate estimates than a traditional out-of-sample approach. Oliveira et al. (2018) perform an extensive empirical study of performance estimation for forecasting problems with spatio-temporal time series. In these cases, they suggest the use of blocking when using a cross-validation estimation procedure. Arlot and Celisse (2018) and Cerquiera et al. (2019) present surveys on the model selection performances of cross-validation procedures and on the application of these methods to time series forecasting tasks respectively, and affirm that, when data are stationary, blocked cross-validation and similar approaches can be used to deal with (short range) dependencies.

as the largest value for which the mean-squared prediction error is within one standard error from the minimum loss (Figure A3)[17].

The Logistic LASSO has determined the optimal model specification, which in our case is based on 10 out of the 20 explanatory variables (Figure A4) [18]. The first three columns of Table A5 show the estimates of the Logistic Lasso whereas the last three columns report the corresponding panel logit model estimates for the LASSO specification including coefficient standard errors and significance levels, which are not available for LASSO models. The results appear in line with expectations. Increases in the LCR, sight deposits change, capital ratios, the ratio between liquid assets and total assets and improvements in the fragility index on the bond market decrease the probability of a liquidity crisis in the subsequent 3 months. On the contrary, increases in the AE, the AEE, the NPL ratio and the net cash outflows over total assets increase this probability. In Table A5, we also report the estimates for 1- and 2-month forecast horizons as a robustness check.

Random forest

The random forest method, introduced by Breiman (1996, 2001), is a sophisticated classification model based on machine learning techniques. To the best of our knowledge, it has been the most frequently employed machine learning method in the early warning literature (Tanaka et al., 2016; Holopainen and Sarlin, 2017; Alessi and Detken, 2018). It is obtained from the aggregation of different classification trees, which represent its constituent elements.

A decision tree is a very simple and effective method to obtain a classifier and in machine learning it is used as a predictive model. The aim is to classify accurately the observations of the dataset based on some input variables. It consists of a root, branches (interior nodes) and leafs (final nodes). Each node of the tree is associated with a particular decision rule based on the values of a specific variable x and on a threshold τ. The decision rule assigns the observations to the right subtree if $x_i < \tau_i$ and to the left one otherwise. Observations

[17] The lambda that minimizes the mean-squared prediction error is equal to 4.2.

[18] The LASSO model selects the following variables: LCR, Change in sight deposits, AE, AEE, CET1, Eurosystem dependency index, NPL Ratio, Fragility index on the bond market, Net cash outflows over total assets and the ratio between liquid assets and total assets.

pass through innumerable nodes and reach up to the leaves, which indicate the category associated with the decision.

Normally a decision tree is built using learning techniques starting from the initial data set, which can be divided into two subsets: the training set, on which the structure of the tree is created, and the test set that is used to test the accuracy of the predictive model created. To estimate a tree, the algorithm chooses simultaneously the variables and thresholds to split the data, based on a measure of gain from each possible split and some stopping criteria to limit the complexity of the tree.

When decision trees have a complex structure, they can present overfitting problems on the training-set and show high variance. For this reason, when building a random forest it is important to generate a certain heterogeneity between different decision paths by randomly choosing a subset of observations on which the estimate is carried out (bagging) and considering only a random subset of attributes at each step (attribute bagging). Both components are needed in order to sufficiently reduce individual trees correlation and achieve the desired variance reduction, while maintaining a high degree of prediction accuracy.

More precisely, the bootstrap creates each time different sub-samples used as training sets for the construction of a complex decision tree (called weak learner). The data left outside the training set (out of bag data) are used for tree validation. Randomness becomes part of the classifiers construction and has the aim of increasing their diversity and thus diminishing the correlation. This process is repeated randomly *N* times and the results of each tree are aggregated together to obtain a single estimate. The aggregated result of the random forest is nothing but the class returned by the largest number of trees. In this way, although introducing some distortion, it is possible to reduce drastically the variance of the estimates thereby having predictions that are more accurate and solving the overfitting problem.

Analytically, given a set of B identically distributed random variables $\{Z_1,\dots,Z_B\}$ with $corr\left(Z_i, Z_j\right) = \varrho$, $E\left(Z_j\right) = \mu$ and $Var\left(Z_j\right) = \sigma^2$, if we assume that the empirical average is equal to $\bar{Z} = \frac{1}{B}\sum_b Z_b$ then we obtain:

$$Var\ (\bar{Z}) = \varrho\sigma^2 + (1 - \varrho)\frac{\sigma^2}{B}.$$

Since the random forest model assumes the weak learners are extremely complex trees, which have high variance but low bias, the main idea is to reduce the variance by decreasing the correlation between the different trees (ϱ).

In line with the literature, our random forest aggregates the results of 200 decision trees and assumes a 3-month forecast horizon. As we have done for the logistic LASSO, to calculate the threshold value that allows discriminating the crisis events, we applied the loss function of Sarlin assuming a value of $\mu \in$ (75%, 98%), equivalent to putting a weight on the false negative outcome (missing a liquidity crisis) between 3 and 49 times greater than the one of a false alarm.

Extreme Gradient Boosting

The idea of Boosting is similar to the one of the random forest, i.e. improving model performance by aggregating different weak learners (in our case different decision trees). This time, however, the selected weak learners have high bias, but low variance. The boosting method combines them sequentially by working on the residuals of the previous weak learner. To be more precise, the method starts by adapting a very simple model (in our case a tree with only two nodes) to the training dataset. Then by looking at the residuals, it fits another model by giving a higher weight to the observations erroneously classified and repeats this process until a stopping rule is satisfied. The final output of the model is obtained by taking a weighted majority vote on the sequence of trees generated.

In contrast with the random forest, the parameters tuning in the Boosting can be extremely complex. First, it is important to set the number of trees (B) the model is going to use: if this number is too high, then the model will overfit the data. This parameter must be selected via cross validation. Second, we need to choose the number of splits (*d*), which controls the complexity of the weak learners. By assumption, *d* must be a small number since we are interested in weak learners with small variance and high bias. If *d* = 1 we get a stump i.e. a tree with only the root. Since by having *d* splits we will use *d* attributes at most, it is useful to consider it as the parameter controlling the level of interaction between variables that we want to consider. In our case, we set *d* = 2. Third, the learning rate, which controls the learning speed of the model, is usually a small positive number, in our case is equal to 0.1. The smaller is the value of the learning rate, the higher is the number of trees needed to reach good predictive performance.

The first version of Boosting in the literature is known as Adaboost (Freund and Shapire, 1997) and was introduced to improve decision trees performance in binary classification models. Now, there exist different versions of Boosting, the one we consider in this paper is the Gradient Boosting (Friedman et al., 2000). With respect to Adaboost, this version generalizes the concept of residuals. Indeed, at each step, the model computes the error as the opposite of the gradient of the loss function, approximates this value with a tree of depth *d* and then updates the fit. This process is repeated B times.

In this paper, we implement Gradient Boosting through the Extreme Gradient Boosting (XGBoost) algorithm, an optimized implementation proposed by Chen and Guestrin (2016) that has become a popular supervised learning algorithm, in particular for financial time series data. To avoid the overfitting problem the algorithm uses features subsampling as the random forest and a regularization parameter to reduce the final prediction sensitivity to single observations.

5. Results and model combination

Following the approach of Holopainen and Sarlin (2017), this paper aims at comparing the signals generated individually by the logistic LASSO, the random forest and XGBoost and combining them in order to take into account their complementarity.

Although all models show excellent in sample performance in absolute terms, the XGBoost performs slightly better (Table A6). With forecasting models, however, it is essential to evaluate the models' out-of-sample performance. In order to do so, we implemented a series of steps:

- First, we divided the sample into two subgroups: one for estimation (training data) and the other for test. Then, an expanding window is constructed so that the estimation is repeated by adding each time the subsequent observation (in our case, a month) and testing the results on the left out part of the dataset. This approach allows taking into account the serial and cross-sectional correlation of the data (Holopainen and Sarlin, 2017). The training dataset considers the data of the Italian banks between December 2014 and September 2018, while the data between October 2018 and January 2020 are used from time to time for the out-of-sample test. Thus, the model is tested 14 times.

- Focusing on the training set, we applied the three estimation methods. In order to get estimates that are more robust we divided the train set into k subgroups and we repeated the estimation k times, so that each time the estimate is made on k-1 subgroups, leaving the last one for validation (see Section 4 for more details about the single estimation methods).
- Once obtained the probabilities of incurring in a liquidity crisis in the subsequent 3 months for each estimation method, we optimized the loss function in order to get the optimal threshold (ϑ^*) for each model. To take into account central bank's preference for models that minimize type I error, for each method the utility function has been optimized attributing a greater weight to missing a crisis compared to a false alarm: $\mu \in (0.75, 0.98)$.
- By comparing the probabilities of each estimation method with the different optimal thresholds, we defined the signals (0, 1) of the models.

Table 2 shows the models out-of-sample evaluation measures in correspondence of the value of $\mu = 0.97$, which maximizes models' relative usefulness. In addition to the usefulness and the relative usefulness, Table 2 also reports the Accuracy, which computes the ratio of correctly classified observations over its total, and two measures of precision: precision positives and precision negatives. These latter ratios compute the percentage of model outcomes (signal/ no signal) that are correct. Finally, we add the false positive and false negative rates, which give a measure of how many times the model incorrectly classifies crisis cases or no crisis cases. As a comparison, we add in Table 2 the evaluation measures of the classical logistic regression. The results confirm that machine learning methods outperform the classical logistic regression, showing similar and very high out-of-sample performance.

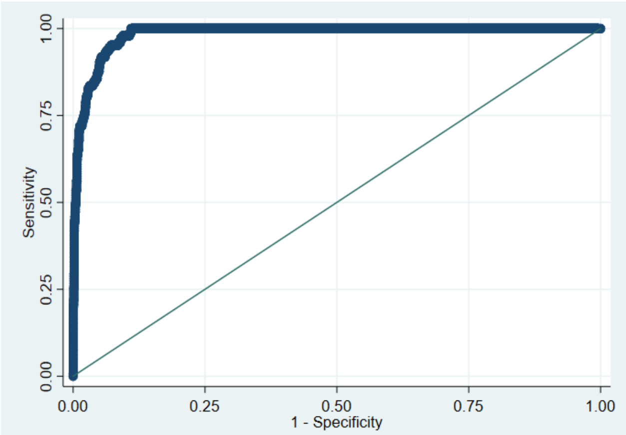
Table 2 – Models out-of-sample evaluation measures

	Logit	Logistic LASSO	Random forest	XGBoost
Recall positives	67%	90%	90%	84%
Recall negatives	97%	95%	94%	98%
FP rate	3%	5%	6%	2%
FN rate	33%	9%	10%	16%
Precision positives	82%	95%	93%	98%
Precision negatives	79%	92%	85%	88%
Accuracy	82%	93%	87%	91%
Usefulness	0.00	0.00	0.02	0.02
Relative usefulness	84.6%	83.2%	92.8%	97.2%
μ	**0.97**			

Note: The Table reports the following measures to evaluate the performance of the models: Recall positives (o TP rate) = TP/(TP+FN), Recall negatives (o TN rate) = TN/(TN+FP), Precision positives = TP/(TP+FP), Precision negatives = TN/(TN+FN), Accuracy = (TP+TN)/ (TP+TN+FP+FN), FP rate = FP/(FP+TN) e FN rate = FN/(FN+TP).

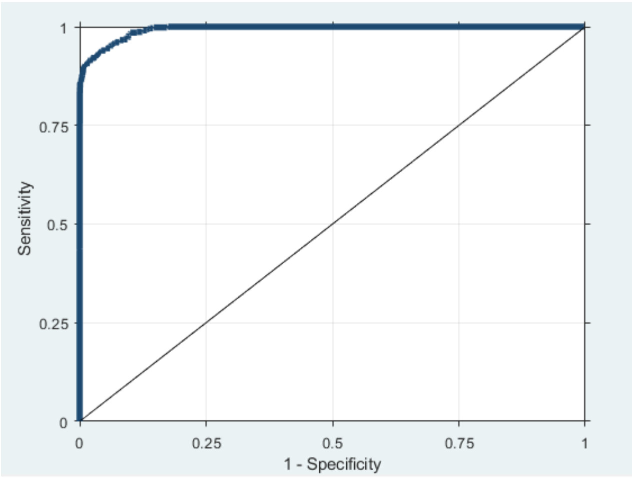
Another important evaluation metric for checking any classification model's performance is the AUROC (Area under the Receiver Operating Characteristic, ROC; Figure 1). The AUROC is a normalized index between 0 and 1, where a value close to 1 indicates a model that can perfectly discriminate the banks in crisis from the healthy ones, while the value 0.5 indicates a model that divides the banks randomly. The ROC curve plots the relationship between two measures, Sensitivity and 1– Specificity, as the threshold varies. Sensitivity is equal to the recall positive ratio or $TP\ rate\ =\ \frac{TP}{(TP+FN)}$. This ratio represents the percentage of crisis events correctly classified by the model. Regarding the second measure, since the Specificity is equal to the so-called recall negative ratio or $TN\ rate\ =\ \frac{TN}{(TN+FP)}$, which computes the ability of the model to correctly classify the no crisis observations, it follows that 1 – Specificity is equal to the $FP\ rate\ =\ \frac{FP}{(FP+TN)}$, thus providing a measure of how many times the model incorrectly classifies the no crisis cases. The Figure shows that the logistic LASSO has an excellent performance and, for the 3-month forecast horizon, the AUROC is equal to 0.97.

Figure 1 – AUROC logistic LASSO



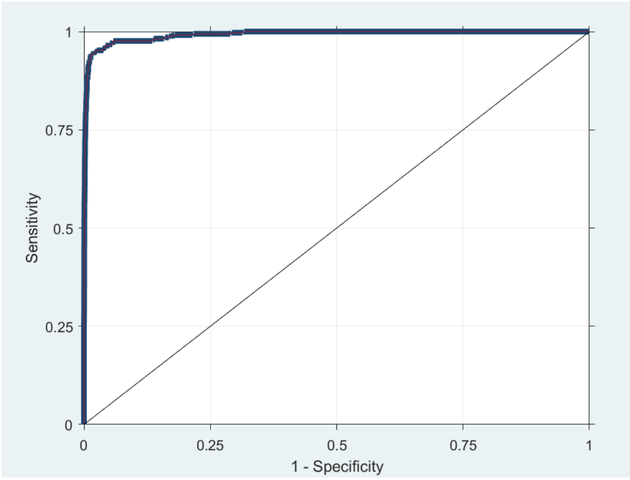
The AUROC of the random forest appears to be slightly lower (Figure 2), although it is extremely high in absolute value (0.96).

Figure 2 – AUROC random forest



Finally, the AUROC of XGBoost appears to be the best among the three methods and is equal to 0.98 (Figure 3).

Figure 3 – AUROC XGBoost



Changing the parameter of policymakers' preferences between type I and II errors in the feasible range $\mu \in (0.75, 0.98)$, which corresponds to weighting a missing crisis between 3 and 49 times more compared to a false alarm, leads to optimal results in line with the ones just presented (Table A7). The relationship between the value of μ and the model relative usefulness is non-monotonic (Figure A2). As the value of μ increases, the model relative usefulness increases until reaching its maximum value and then it decreases again. For all the models the maximum relative usefulness is reached when $\mu = 0.97$.

To obtain a robust early warning indicator that combines the strengths of the individual models, we decided to aggregate the probabilities obtained with the three methods. The literature on forecast combination (e.g. Stock and Watson, 2004) tends to prefer the use of simple aggregation methods such as the average, in particular when there is high multicollinearity between the forecasts (Timmermann, 2006).

Holopainen and Sarlin (2017) proposed two different combination methods. The first one computes the combined probabilities as the simple average of the individual estimates. In this case, however, it is not possible to give greater weight to the model that has a better performance. To consider this, the second method weights the aggregation using the relative usefulness value of each model. In this way, the model with the best performance will provide a higher contribution to the output of the combined model.

After having obtained the probabilities of incurring in a liquidity crisis for each estimation model, we followed the approach of Holopainen and Sarlin (2017). For the weighted average we used as weights the relative usefulness obtained by each estimation model when $\mu = 0.97$, which is the value that maximizes the relative usefulness for the three models.

Once combined the probabilities using these two different methods of aggregation, we computed new optimal thresholds by maximizing the utility function for different values of the parameter $\mu \in (0.75, 0.98)$, exactly as we did for the individual models. Based on the values of the optimal thresholds, we defined the signals 0 and 1. As it was the case for the individual models, also for the combined models the relationship between the value of μ and the model relative usefulness is non-monotonic (Figure A2). As the value of μ increases, model relative usefulness measures increase until reaching their maximum value and then they decrease again. Both the weighted average and the simple average model relative usefulness is maximized when $\mu = 0.95$ (Table A8).

The optimal value of μ is used to compute the evaluation measures. The analysis shows that the aggregate models manage to have an extremely low percentage of false negatives (10%; Table 3), lower than the values usually reported in the literature. However, the weighted average model achieves slightly better results than the simple average one, in line with the results of Holopainen and Sarlin (2017).

Table 3 – Combined models evaluation measures

	Weighted average	**Simple average**
Recall positives	90%	90%
Recall negatives	99%	99%
FP rate	1%	1%
FN rate	10%	10%
Precision positives	99%	98%
Precision negatives	92%	92%
Accuracy	95%	94%
Usefulness	0.02	0.02
Relative usefulness	90.6%	90.0%
μ	**0.95**	

As a further test of the goodness of our estimates, we analyzed the trend of the probability of incurring in a liquidity crisis or “probability of illiquidity”, derived from the weighted average model. For all the banks that have incurred in a liquidity crisis, we calculated the average probability of illiquidity up to 9 months before the liquidity distress event. Figure A5 shows that in the sample the average probability of illiquidity increases dramatically three months before the liquidity distress event. Since the model’s probability of illiquidity is estimated as a forecast with a 3-month horizon, the analysis of the trend of liquidity deterioration confirms the goodness of estimates.

Based on the probability of illiquidity, obtained from the weighted average model for each bank in each month, it is also possible to construct the historical time series of the percentage of banks whose probability is above the optimal threshold level. In January 2020, the percentage of these banks with respect to the Italian system as a whole appears to be contained compared to the historical distribution, reaching the minimum level over the last 5 years (Figure A6).

Another way to use the probability of illiquidity is to compute the difference between the model optimal threshold and this variable i.e. the “distance from illiquidity”. This distance allows us to rank the banks from the healthier to the more risky ones and to compute the banks’ distribution for each month. Figure A7 shows the historical evolution of this distribution between December 2018 and January 2020. The graph highlights that during that period the median distance from illiquidity has remained stable, whereas the average has decreased, particularly in the last few months.

6. Robustness checks

To check the robustness of our results we considered two different specifications of the model that differ one another for the definition of the target variable, as described in the Section 2.

In the first specification, the target variable considers as crisis events only the events more closely linked to a liquidity deterioration such as: resorting to central bank’s ELA, benefiting from State support through GGBBs or, finally, being subject to enhanced liquidity monitoring by the BoI’s Supervisory Directorate and brought to the attention of the BoI’s Markets and Payment Systems Directorate.

Exactly as we have done for the baseline model, we estimated the probability of incurring in a liquidity crisis in the following three months with the three methods (logistic LASSO, random forest and XGBoost) and we compare the results with the traditional logistic regression. Table 4 shows the out-of-sample evaluation measures. To take into account central bank’s preference for models that minimize type I error, for each method the optimal threshold (ϑ^{*}), based on which we compute all the performance measures, is obtained by optimizing the loss function attributing a greater weight to missing a crisis compared to a false alarm. Table 4 shows the evaluation measures in correspondence of the value of $\mu = 0.97$ which maximizes models’ relative usefulness.

The results confirm that machine learning methods outperform the classical logistic regression, with the random forest performing slightly better than the others. By comparing Table 4 with Table 2, it appears that the results of this second specification are marginally below the ones of the baseline model. Table A9 shows that changing the parameter of policymakers' preferences between type I and II errors in the feasible range $\mu \in (0.75, 0.98)$ leads to optimal results, which are in line with the ones just presented.

Table 4 – Out-of-sample evaluation measures

	Logit	**Logistic LASSO**	**Random forest**	**XGBoost**
Recall positives	70%	74%	88%	83%
Recall negatives	97%	96%	99%	99%
FP rate	3%	4%	1%	1%
FN rate	30%	26%	12%	17%
Precision positives	88%	82%	98%	99%
Precision negatives	79%	84%	91%	87%
Accuracy	83%	85%	93%	91%
Usefulness	0.00	0.00	0.02	0.02
Relative usefulness	88.4%	85.7%	92.7%	98.1%

μ	**0.97**

As for the baseline model, we combined the probabilities obtained from the three estimation methods through a simple average and a weighted average using as weights the relative usefulness of each method obtained when $\mu = 0.97$. Table 5 shows that also with this model specification combining model probabilities improves the performance and yields to accurate out-of-sample predictions of bank liquidity crises, although slightly lower that the ones of the baseline specification. Table A10 reports the evaluation measures of the two combined models changing the parameter of policymakers' preferences between type I and II errors in the feasible range $\mu \in (0.75, 0.98)$.

Table 5 – Combined models evaluation measures

	Weighted average	Simple average
Recall positives	90%	87%
Recall negatives	99%	99%
FP rate	1%	1%
FN rate	10%	13%
Precision positives	98%	99%
Precision negatives	92%	91%
Accuracy	94%	93%
Usefulness	0.02	0.02
Relative usefulness	96%	97%
μ	**0.97**	

Finally, we built a third specification, in which we considered as liquidity crisis event only the first month in which one of the events described in Section 2 occurred and excluded from the dataset the following observations related to that event for that bank. However, we allow for the possibility that during the sample period a bank incurs in different liquidity crises. In this setting, given the limited occurrence of crisis events, we oversampled the minority class by simply duplicating examples from that class in the training dataset prior to fitting the model.

Table 6 shows the results of the different estimation methods when $\mu = 0.95$, value that maximizes models' relative usefulness. In line with the other specifications, machine learning algorithms outperform the traditional logistic regression, with the random forest performing better than the others. The comparison between Tables 6 and 2 highlights that the results of the baseline model are better than the ones of this third specification. Table A11 shows the results of the model changing the parameter of policymakers' preferences.

Table 6 – Out-of-sample evaluation measures

	Logit	Logistic LASSO	Random forest	XGBoost
Recall positives	11%	49%	86%	75%
Recall negatives	98%	58%	94%	61%
FP rate	2%	42%	6%	39%
FN rate	89%	51%	14%	25%
Precision positives	36%	58%	93%	74%
Precision negatives	53%	54%	89%	82%
Accuracy	54%	53%	90%	68%
Usefulness	0.00	0.01	0.00	0.00
Relative usefulness	48.7%	63.7%	70.7%	73.0%
μ	**0.95**			

As before, we combined the probabilities obtained from the three estimation methods through a simple average and a weighted average using as weights the relative usefulness of each method. Table 7 shows that also with this model specification combining model probabilities improves the performance and yields to accurate out-of-sample predictions of bank liquidity crises, although slightly lower that the ones of the baseline specification. Table A12 reports the evaluation measures of the two combined models changing the parameter of policymakers' preferences.

Table 7 – Combined models evaluation measures

	Weighted average	Simple average
Recall positives	90%	71%
Recall negatives	92%	94%
FP rate	8%	6%
FN rate	10%	29%
Precision positives	92%	87%
Precision negatives	93%	82%
Accuracy	91%	83%
Usefulness	0.00	0.00
Relative usefulness	57.4%	61.3%
μ	**0.97**	

7. Conclusion

The paper presents an early-warning model to predict individual banks that could face a liquidity crisis. We started by identifying the liquidity crisis events affecting Italian banks in the period December 2014 - January 2020. To do this, we introduced an innovative dataset that complements banks that resorted to the central bank's ELA with those which have been subject to restrictive measures on participation to monetary policy operations and to enhanced liquidity monitoring by the Supervisory Authority. Finally, we also considered banks placed under special or temporary administration, subject to State aid and deemed to be failing or likely to fail (FOLTF) for liquidity reasons.

Subsequently, we applied three different models to estimate the probability for each bank to incur in a liquidity crisis in the following 3 months: the logistic LASSO with blocked cross-validation, the random forest and the Extreme Gradient Boosting.

The early warning literature usually presents the results coming from individual models. On the contrary, this paper aims at comparing and combining the information produced by the three approaches in order to obtain a more robust method for measuring banks vulnerabilities from a liquidity point of view. Indeed, it is likely that different approaches capture different linkages among variables that focus on specific types of fragility, and therefore can complement each other, giving value to their simultaneous use.

Moreover, the signals of the early warning models are calibrated according to a policymaker's preference between type I and II errors. The evaluation framework assumes that a central bank has to be substantially more concerned with missing a bank liquidity crisis than issuing false alarms. Indeed, an EW signal should trigger actions by the authority to investigate and possibly restore the liquidity situation of the bank. Should the analysis reveal that the signal is false, there is no loss of credibility for the policymaker as model results are not published. This is the reason why, for each method, the value of the threshold that allows discriminating the crisis cases has been computed by optimizing the utility function weighting the error of missing a crisis far more than a false alarm.

Finally, we compared the goodness of the out-of-sample fit of the models through some performance measures based on the confusion matrix and the AUROC. Results show that machine learning algorithms perform better than the traditional logistic regression.

The key finding of the paper is that complementing the models improves the performance and yields to accurate out-of-sample predictions of bank liquidity crises. In line with Holopainen and Sarlin (2017), we propose two different combination methods. The first one computes the combined probabilities as the simple average of the individual estimates. In this case, however, it is not possible to give greater weight to the model that has a better performance. To consider this, the second method weights the aggregation using the relative usefulness value of each model. In this way, the model with the best performance will provide a higher contribution to the output of the combined model.

The results show that the combined models manage to have an extremely low percentage of false negatives (missing crises), lower than the values usually reported in the literature, while at the same time limiting the number of false positives. In addition, in line with the results of Holopainen and Sarlin (2017), the weighted average model performs slightly better than the simple average one.

To check the robustness of our results, we considered two further specifications of the target variable. The first one is based on a more restrictive definition of liquidity events. We considered only the events more closely linked to a liquidity deterioration such as resorting to central bank's ELA, benefiting from State support through GGBBs or, finally, being subject to enhanced liquidity monitoring by the BoI's Supervisory Directorate and brought to the attention of the BoI's Markets and Payment Systems Directorate. In the second robustness check, we considered as liquidity crisis event only the first month in which one of the events previously described occurred and excluded from the dataset the following observations related to that event for that bank. Both specifications confirm that combining model probabilities allows improving the predictive performance, although the baseline specification results appear to be better.

Ensemble classifiers that aggregate the prediction of multiple base models have been dominating the scene for several years (e.g. Lessmann et al, 2015). From this experience, it emerges that future extensions of this work should go in the direction of further increasing the number of considered models (e.g. adding support vector machines and neural networks), since adding more models helps to capture different aspects of fragility and to improve the forecasting performance.

Another improvement of the model could be including other attributes tied to the liquidity risk analysis, which could contribute to better describing the banks' business. In addition, it could be also useful to include other macro-financial variables, which could help to better capture the effect that macroeconomic imbalances may have on individual banks. A limitation of this work could be the length of the time series, which we try to overcome with the size of the panel and the granularity of the data. With longer time series, when they will be available, we will continue to back-test the predicting performance of the model.

Anyway, the signal obtained from our model could be very useful for regular monitoring exercises, as they can help central banks to detect in advance banks that could be in need of central bank's Emergency Liquidity Assistance (ELA).

References

Alam P., D. Booth, K. Lee and T. Thordarson (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study”, Expert Systems with Applications, 18, 185-199.

Aldasoro I., C. Borio and M. Drehmann (2018). Early warning indicators of banking crises: expanding the family. BIS Quarterly Review, March 2018.

Alessi L. and C. Detken (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. European Journal of Political Economy, 27 (3), 520 – 533.

Alessi L. and C. Detken (2018). Identifying Excessive Credit Growth and Leverage. Journal of Financial Stability 35, 215-225.

Arlot S. and A. Celisse (2018). A survey of Cross-validation procedures for model selection. https://arxiv.org/abs/0907.4728v1

Bank for International Settlement (2017). Designing frameworks for central bank liquidity assistance: addressing new challenges. CGFS Papers number 58.

Bergmeir C. and J. M. Benìtez (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences 191, 192 - 213.

Betz F., S. Oprică, T. A. Peltonen and P. Sarlin (2013). Predicting Distress in European Banks. ECB Working Paper Series number 1597.

Beutel J., S. List and G. von Schweinitz, (2018). An evaluation of early warning models for systemic banking crises: Does machine learning improve predictions? Deutsche Bundesbank Discussion Paper number 48/2018.

Boyacioglu M. A., Y. Kara and O. K. Baykan, (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. Expert Systems with Applications, 36(2), 3355–3366.

Bräuning M., D. Malikkidou, S. Scalone and G. Scricco (2019). A new approach to Early Warning Systems for small European banks. ECB Working Paper Series number 2348.

Breiman L. (1996). Bagging Predictors. Machine Learning 24 (2), 123-140.

Breiman L. (2001). Random forests. Machine Learning, 45, 5–32, 2001.

Bussiere M. and M. Fratzscher, (2006). Towards a new early warning system of financial crises. Journal of International Money and Finance, 25 (6), 953-973.

Campbell J., J. Hilscher and J. Szilagyi (2008). In search of distress risk. The Journal of Finance, 58, 2899 – 2939.

Cerquiera V., L. Torgo and I. Mozetic (2019). Evaluating time series forecasting models. https://arxiv.org/abs/1905.11744v1.

Chawla N., K. Bowyer, L. Hall and W. Kegelmeyer (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16 (2002), 321–357.

Chen T. and C. Guestrin (2016). XGBoost: a scalable tree boosting system. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

Cole R. A. and J. Gunther (1998). Predicting bank failures: A comparison of on- and on-site monitoring systems. Journal of Financial Services Research 13, 103 – 117.

Demirguc-Kunt A. and E. Detragiache (2000). Monitoring banking sector fragility. A multivariate logit approach. World Bank Economic Review, 2000, 14 (2), 287 – 307.

Diamond D. W. and Dybvig P.H. (1983). Bank Runs, Deposit Insurance, and Liquidity. The Journal of Political Economy, Vol. 91, No. 3, pp. 401-419.

Dobler M., S. Gray, D. Murphy and B. Radzewicz-Bak (2016). The Lender of Last Resort Function after the Global Financial Crisis. IMF Working Paper WP/16/10.

Ferriani F., W. Cornacchia, P. Farroni, E. Ferrara, F. Guarino and F. Pisanti (2019). An early warning system for less significant Italian banks. Questioni di Economia e Finanza (Occasional papers) 480, Bank of Italy.

Flannery M. (1998). Using market information in prudential bank supervision: A review of the US empirical evidence. Journal of Money, Credit and Banking 30, 273 – 305.

Frankel J. A. and A. K. Rose (1996). Currency Crashes in Emerging Markets: An Empirical Treatment. Journal of International Economics 41 (3), 351-366.

Freund Y and R. E. Shapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences 55, 119-39.

Friedman J., T. Hastie, and R. Tibshirani (2000). Additive logistic regression: a statistical view of boosting. Special invited paper. The Annals of Statistics 2000, Vol. 28, No. 2, 337–407.

Gonzalez-Hermosillo B. (1999). Determinants of ex-ante banking system distress: A macro-micro empirical exploration of some recent episodes. IMF Working Paper 99/33.

Holopainen M. and P. Sarlin (2017). Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty. Quantitative Finance 17 (12), 1-31.

Iachini E. and S. Nobili (2014). An indicator of systemic liquidity risk in the Italian financial markets. Questioni di Economia e Finanza (Occasional Papers) 217, Bank of Italy.

Jagtiani J., J. Kolari, C. Lemieux and H. Shin (2003). Early warning models for bank supervision: simpler could be better. Working paper.

Kaminsky G., S. Lizondo, and C. M. Reinhart (1998). Leading Indicators of Currency Crises. IMF Staff Papers, 45 (1), 1 – 48.

Kraft E. and T. Galac (2007). Deposit interest rates, asset risk and bank failure in Croatia. Journal of Financial Stability, 2, 312 – 336.

Krainer J. and J.A. Lopez (2003). Using equity market information to monitor banking institutions. Working paper.

Lang J. H., T. A. Peltonen and P. Sarlin (2018). A framework for early-warning modeling with an application to banks. ECB Working Paper Series number 2182.

Lessmann S., B. Baesens, H. V. Seow and L. C. Thomas, (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124- 136.

Lo Duca M. and T. A. Peltonen, (2013). Assessing systemic risks and predicting systemic events. Journal of Banking & Finance, 37 (7), 2183 – 2195.

Martin D. (1977). Early warning of bank failure: A logit regression approach. Journal of Banking and Finance, 1 (3), 249 – 276.

Oliveira M., L. Torgo and V. S. Costa (2018). Evaluation procedures for forecasting with spatio-temporal data. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 703 – 718. Springer.

Poghosyan T. and M. Čihák (2011). Distress in European banks: An analysis based on a new dataset. Journal of Financial Services Research, 2011, 40, 163 – 184.

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

Sarlin P. (2013). On biologically inspired predictions of the global financial crisis. Neural Computing and Applications, 2013, 24 (3 – 4), 663 – 673.

Stock J. and M. W. Watson (2004). Combination Forecasts of Output Growth in a Seven-Country Data Set. Journal of Forecasting 23, 405–430.

Tam K. Y. and M. Kiang (1992). Predicting bank failures: a neural network approach. Decision Sciences, 23, 926-947.

Tanaka K., T. Kinkyo and S. Hamori (2016). Random Forests-Based Early Warning. System for Bank Failures. Economics Letters 148, 118-121.

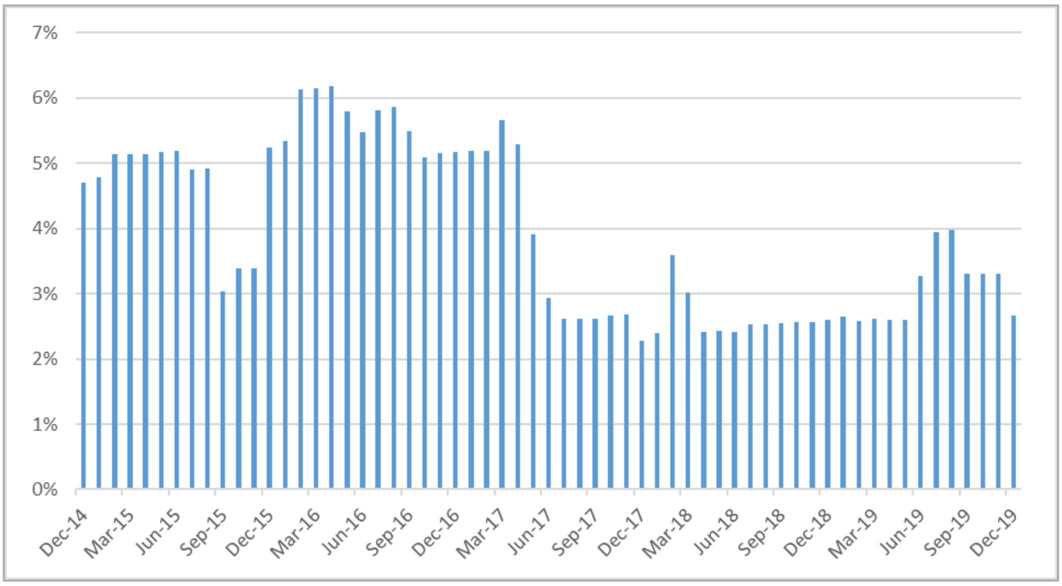
Thomson J. (1992). Modeling the regulator's closure option: a two-step logit regression approach. Journal of Financial Services Research 6, 5 – 23.

Tibshirani R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society (Series B), 58(1):267 – 288.

Timmermann A. (2006). Forecast Combinations. Handbook of Economic Forecasting, in: G. Elliott & C. Granger & A. Timmermann (ed.), Handbook of Economic Forecasting, edition 1, volume 1, chapter 4, pages 135-196, Elsevier.

Figures and tables

Figure A1 – Historical distribution of the target variable



Note: percentage of liquidity crisis events over total number of observation per month.

Figure A2 – Models' relative usefulness changing μ

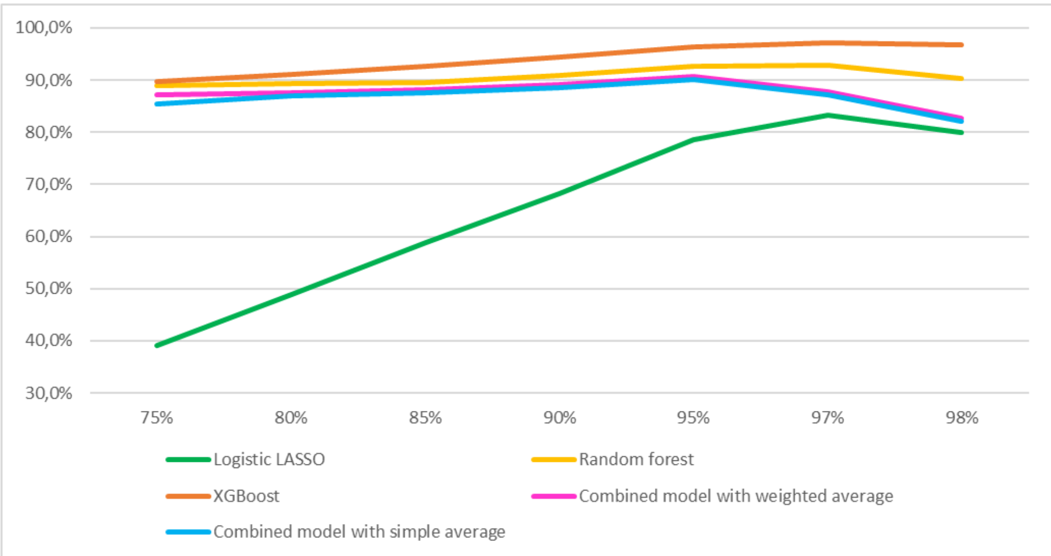
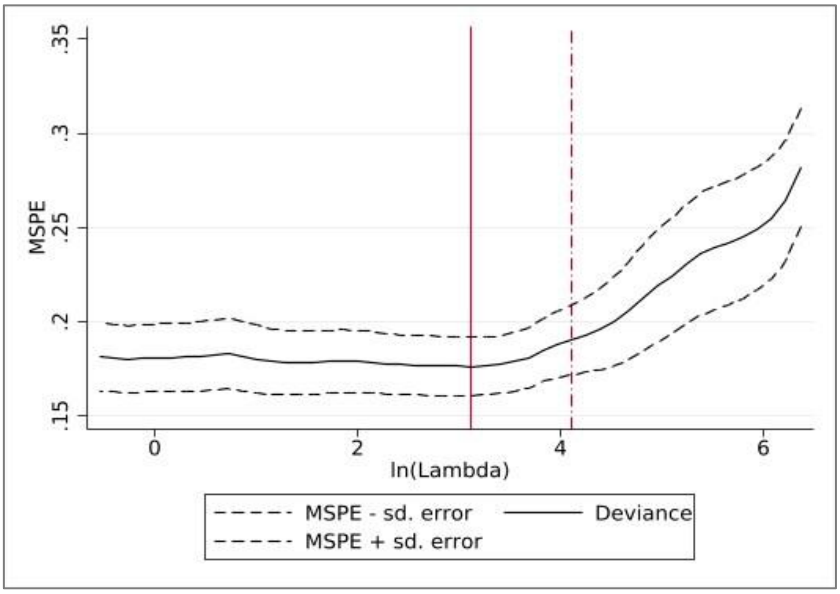


Figure A3 – Cross validation function of the logistic LASSO



Note: the red line represents the value of ln(λ) that minimizes the loss function and the dotted red line represents the largest lambda for which the mean-squared prediction error is within one standard error from the minimum loss.

Figure A4 – Logistic LASSO coefficients paths

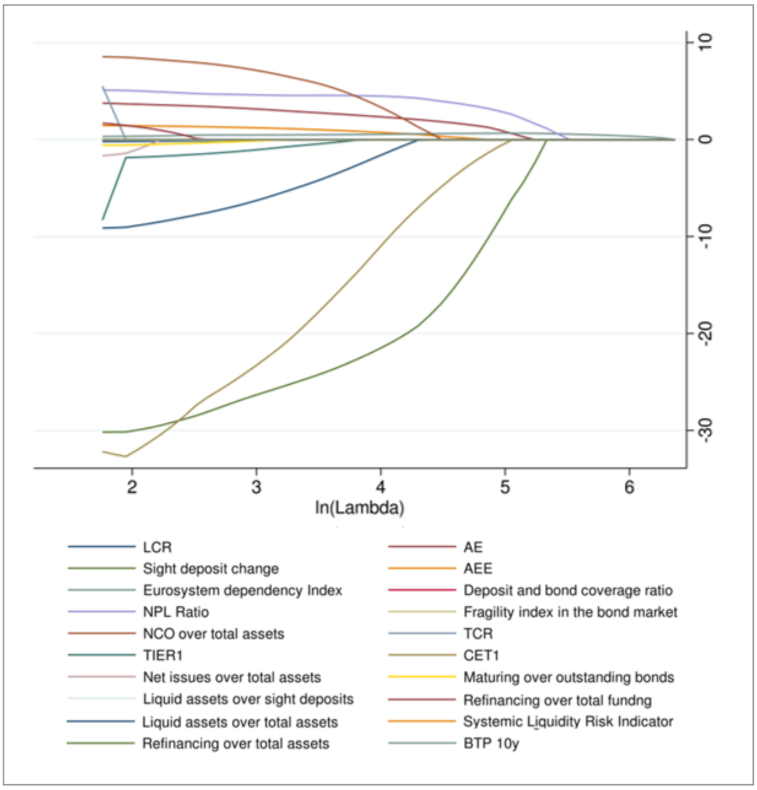
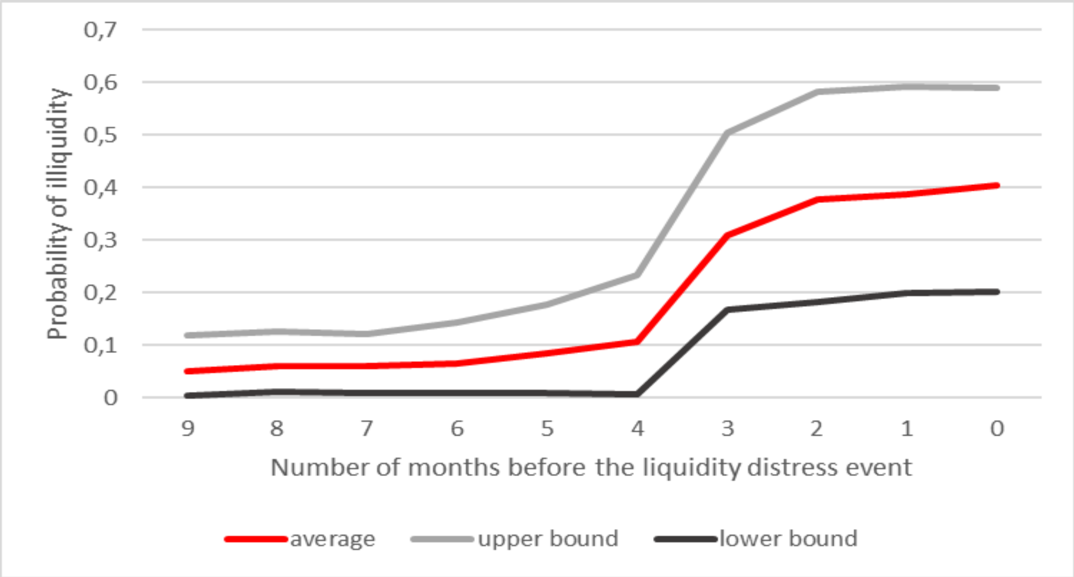
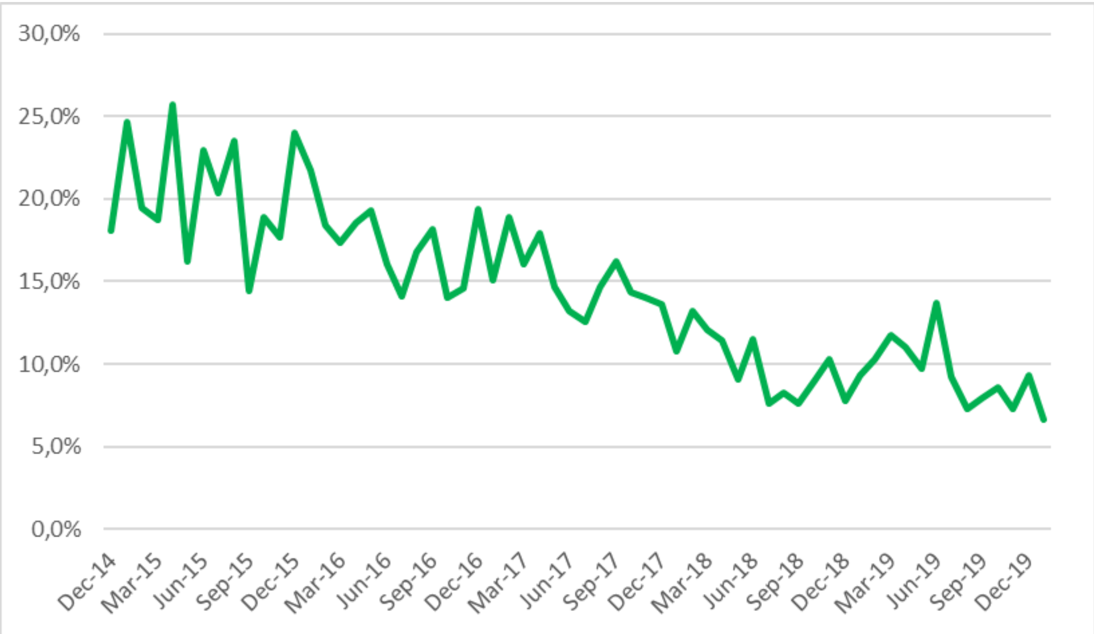


Figure A5 – Probability of illiquidity during the 9 months before the liquidity crisis



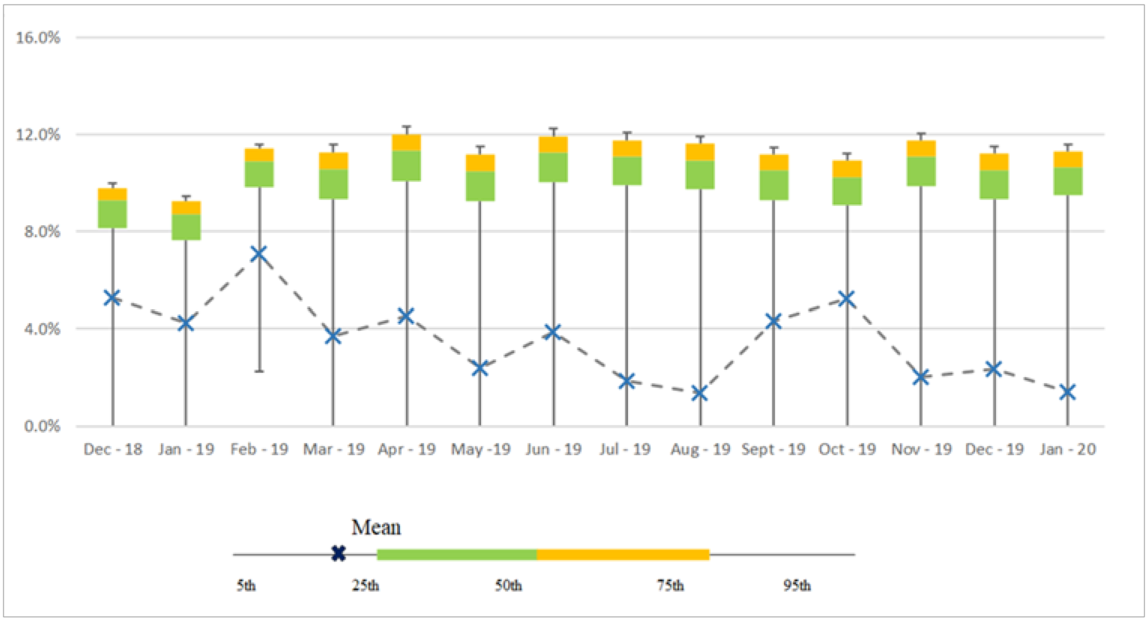
Note: The Figure displays the average probability of illiquidity for liquidity-distressed banks computed by the weighted average model with a 3-month forecast horizon; the Figure also shows the lower and upper bounds of the 95 per cent confidence interval.

Figure A6 – Historical evolution of the estimated percentage of vulnerable banks



Note: Vulnerable banks are those whose probability is above the optimal threshold level.

Figure A7 – Historical evolution of the distribution of the distance from the threshold



Note: The distance from illiquidity is computed considering for each bank the probability and the threshold obtained from the weighted average model.

Table A1 – Effects of early interventions on sight deposit change

	Monthly change in sight deposit
Special/Temporary Administration	-0.0278***
	(0.00739)
Constant	-0.000256
	(0.000173)
Observations	12,339
Number of banks	286
R-squared	0.001

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2 – Attributes descriptive statistics

	Observations	Mean	Standard Deviation	10 percentile	Median	90 percentile
LCR	11,802	306.3%	260.6%	104.2%	264.8%	714.3%
AE	11,068	18.3%	13.9%	3.0%	15.1%	36.8%
AEE	11,995	46.0%	26.7%	7.6%	46.2%	82.0%
Fragility index on the bond market	8,758	-0.03	4.88	-0.13	-0.04	0.00
Sight deposits and bond coverage ratio	12,550	29.0%	40.7%	5.6%	25.0%	77.0%
Eurosystem dependency index	5,158	110.4%	100.7%	32.2%	89.7%	199.6%
Change in sight deposits	12,617	2.3%	29.0%	-0.5%	0.8%	2.6%
CET1 ratio	12,495	19.5%	16.3%	10.5%	15.7%	30.4%
TIER 1	12,366	20.0%	16.4%	10.8%	16.2%	31.3%
Total Capital ratio	12,382	20.9%	16.9%	12.2%	16.9%	31.4%
Net cash outflows over total assets	11,899	7.3%	35.3%	1.2%	4.1%	11.1%
NPL Ratio	12,318	13.6%	11.4%	1.5%	10.6%	29.5%
Net bond issues over total asets	12,755	-0.2%	0.8%	-0.5%	0.0%	0.0%
Maturing bonds over outstanding bonds	12,755	21.0%	26.1%	0.0%	14.1%	55.4%
Liquid assets over sight deposits	12,659	53.0%	118.9%	13.1%	30.9%	82.5%
Liquid assets over total assets	12,622	14.1%	10.8%	4.5%	11.3%	26.4%
Eurosystem refinancing over total funding	12,754	6.2%	11.2%	0.0%	0.0%	16.7%
Eurosystem refinancing over total assets	12,754	4.4%	7.5%	0.0%	0.0%	12.9%
Systemic Liquidity Risk Indicator	12,755	9.4%	6.1%	3.0%	8.8%	18.3%
BTP 10y	12,755	1.9%	0.6%	1.2%	1.8%	2.7%

Table A3 – Attributes' correlation matrix

	LCR	AE	AEE	Fragility index on the bond market	Sight deposits and bond coverage ratio	Eurosystem dependency index	Change in sight deposits	CET1 ratio	TIER 1	Total Capital ratio	Net cash outflows over total assets	NPL Ratio	Net bond issues over total asets	Maturing bonds over outstanding bonds	Liquid assets over sight deposits	Liquid assets over total assets	Eurosystem refinancing over total funding	Eurosystem refinancing over total assets	Systemic Liquidity Risk Indicator	BTP 10y
LCR	1																			
AE	-0.1588	1																		
AEE	-0.1314	0.7235	1																	
Fragility index on the bond market	0.0081	0.0019	-0.0395	1																
Sight deposits and bond coverage ratio	0.0048	-0.0035	-0.0062	0.0005	1															
Eurosystem dependency index	-0.1556	0.3702	0.266	-0.0281	0.011	1														
Change in sight deposits	-0.0013	0.0123	-0.0095	0.0012	-0.0004	0.0249	1													
CET1 ratio	0.1754	-0.0169	-0.0646	0.0203	0.0276	-0.2899	0.07	1												
TIER 1	0.1469	0.0254	-0.0269	0.0191	0.0254	-0.2572	0.0644	0.9022	1											
Total Capital ratio	0.1266	0.0503	0.0002	0.0145	0.0243	-0.2304	0.0615	0.846	0.9737	1										
Net cash outflows over total assets	-0.066	0.0204	-0.0128	0.002	-0.0012	0.0298	-0.0031	-0.0228	-0.0218	-0.0129	1									
NPL Ratio	-0.0816	-0.0732	0.0502	0.0273	-0.0158	0.0481	-0.0388	-0.302	-0.3004	-0.2746	-0.0535	1								
Net bond issues over total asets	0.026	0.0085	-0.0056	0.0116	0.0143	-0.0242	-0.0154	0.0499	0.0512	0.0499	-0.0138	-0.0538	1							
Maturing bonds over total outstanding bonds	-0.0869	0.0304	0.0828	-0.0031	-0.0298	0.0527	0.0177	-0.1719	-0.1779	-0.1908	0.0051	0.1877	-0.0741	1						
Liquid assets over sight deposits	0.0196	0.1142	0.0121	0.0019	0.9926	0.0072	-0.0174	0.0189	0.0183	0.0203	0.0849	-0.0593	0.0143	-0.0192	1					
Liquid assets over total assets	0.1775	-0.225	-0.4158	0.0185	0.0092	-0.1615	0.0398	0.2489	0.2296	0.2177	0.2566	-0.2848	0.0327	-0.1563	0.0024	1				
Eurosystem refinancing over total funding	-0.1298	0.4165	0.3019	-0.0128	0.0054	0.8142	0.0057	0.0646	0.0706	0.0829	0.1178	0.0019	-0.0005	0.0207	0.0143	-0.1381	1			
Eurosystem refinancing over total assets	-0.134	0.4353	0.3117	-0.0122	-0.0012	0.7799	0.0069	-0.0049	0.0069	0.0171	0.0989	0.002	-0.0079	0.0511	0.0067	-0.1253	0.9446	1		
Systemic Liquidity Risk Indicator	-0.0391	0.0101	-0.0326	-0.017	0.0096	0.0151	0.0003	-0.0029	0.0041	0.004	0.0561	-0.1349	0.028	-0.0469	0.0086	0.0351	0.0166	0.0191	1	
BTP 10y	-0.0981	0.0103	-0.0024	-0.0159	0.0019	0.0393	-0.0156	-0.0114	-0.0066	-0.0085	0.032	-0.1423	0.0581	-0.0376	-0.0013	-0.0193	0.0485	0.0559	0.4623	1

Table A4 – Standardized attributes' correlation matrix

	LCR	AE	AEE	Fragility index on the bond market	Sight deposits and bond coverage ratio	Eurosystem dependency index	Change in sight deposits	CET1 ratio	TIER 1	Total Capital ratio	Net cash outflows over total assets	NPL Ratio	Net bond issues over total asets	Maturing bonds over outstanding bonds	Liquid assets over sight deposits	Liquid assets over total assets	Eurosystem refinancing over total funding	Eurosystem refinancing over total assets	Systemic Liquidity Risk Indicator	BTP 10y
LCR	1																			
AE	0.0046	1																		
AEE	0.1039	0.7402	1																	
Fragility index on the bond market	0.1381	0.0912	0.1634	1																
Sight deposits and bond coverage ratio	0.3155	0.0093	0.1895	0.3333	1															
Eurosystem dependency index	0.2090	0.4199	0.3549	0.2349	0.1120	1														
Change in sight deposits	0.1314	-0.0001	0.0095	0.0087	0.1414	0.0230	1													
CET1 ratio	0.3920	0.0907	0.0985	0.2245	0.3001	0.3406	0.1570	1												
TIER 1	0.3583	0.0677	0.0885	0.2368	0.2988	0.3124	0.1513	0.8945	1											
Total Capital ratio	0.3370	0.0368	0.0706	0.2561	0.3127	0.2854	0.1413	0.8317	0.9564	1										
Net cash outflows over total assets	0.3197	0.1039	-0.0847	-0.0039	-0.0911	0.1237	-0.0023	0.1218	0.0876	0.0694	1									
NPL Ratio	0.1199	0.0674	0.1950	0.2918	0.3229	0.2150	0.1260	0.2539	0.2678	0.2544	-0.1264	1								
Net bond issues over total asets	0.0782	-0.0024	0.0957	0.7077	0.3197	0.1879	0.0154	0.228	0.233	0.2406	-0.0316	0.3119	1							
Maturing bonds over total outstanding bonds	0.1053	0.0721	0.1545	0.764	0.3931	0.2528	0.0328	0.2535	0.2618	0.28	-0.015	0.3553	0.5543	1						
Liquid assets over sight deposits	0.2703	0.0213	0.194	0.2193	0.837	0.0523	0.1158	0.2347	0.2396	0.2624	-0.0377	0.2236	0.1846	0.163	1					
Liquid assets over total assets	0.2999	0.2065	0.392	0.1651	0.6463	0.1498	0.1496	0.2702	0.2556	0.2486	-0.1919	0.2495	0.1457	0.1713	0.6958	1				
Eurosystem refinancing over total funding	-0.3404	-0.4005	-0.2157	0.0286	0.0139	-0.5461	-0.0428	-0.2204	-0.1813	-0.1493	-0.2761	0.0029	0.1084	0.109	-0.0198	-0.062	1			
Eurosystem refinancing over total assets	-0.3412	-0.4044	-0.2158	0.0239	0.0098	-0.5451	-0.045	-0.2303	-0.1914	-0.163	-0.2814	0.0045	0.1052	0.1051	-0.0241	-0.0558	0.9958	1		
Systemic Liquidity Risk Indicator	0.0219	0.0095	-0.0314	-0.0775	-0.0534	0.0015	0.007	-0.0008	-0.009	-0.0117	0.0633	-0.1125	-0.1152	-0.1126	-0.0401	-0.0326	-0.0416	-0.0422	1	
BTP 10y	0.0952	0.0252	-0.0071	-0.0636	-0.0346	-0.0782	0.0093	-0.0148	-0.0124	-0.0031	0.1015	-0.1093	-0.1044	-0.1055	0.0455	0.0237	-0.1793	-0.1794	0.3713	1

Note: To obtain standardized variables we applied a two-step procedure. First, we transformed variables in a way such that the higher is the indicator the better is the bank's situation, then we standardized them by dividing each observation of each variable for the total amount of observations. In this way, variables are made unit-free and are measured on an ordinal scale.

Table A5 –Logistic LASSO and Logit regressions with 1-, 2-, 3-month forecast horizon

	(1)	(2)	(3)	(4)	(5)	(6)
	LASSO	LASSO	LASSO	Logit	Logit	Logit
	1m ahead	2m ahead	3m ahead	1m ahead	2m ahead	3m ahead
LCR	-0.001	-0.002	-0.004	-0.2611***	-0.324***	-0.306***
				(0.083)	(0.058)	(0.064)
AE	2.484	2.103	2.227	4.637***	4.044***	3.861***
				(0.731)	(0.722)	(0.725)
Change in sight deposits	-20.078	-19.969	-20.782	-24.924***	-22.462***	-20.229***
				(5.769)	(5.641)	(5.360)
AEE	0.593	0.679	0.721	1.119*	1.181**	1.161**
				(0.575)	(0.545)	(0.539)
CET1	-5.089	-6.242	-9.379	-41.575***	-40.994***	-41.006***
				(3.566)	(3.499)	(3.597)
Eurosystem dependency	0.636	0.557	0.557	0.543***	0.532***	0.451***
				(0.083)	(0.088)	(0.096)
NPL Ratio	4.300	4.387	4.387	4.565***	4.248***	3.984***
				(1.075)	(1.032)	(1.009)
Bond market fragility index	-0.061	-0.060	-0.060	-0.613***	-0.610***	-0.538***
				(0.142)	(0.145)	(0.160)
NCO over total assets	0.851	2.611	2.611	9.640***	8.466***	8.022***
				(2.140)	(1.685)	(1.747)
Liquid assets over total assets		-0.018	-0.923	-13.205***	-12.493***	-11.876***
				(3.252)	(2.944)	(2.928)
Constant	-4.644	-4.523	-4.152	-1.088*	-0.825	-0.6089
				(0.638)	(0.628)	(0.6121)
Time dummies	NO	NO	NO	YES	YES	YES
Observations	10,117	9,898	9,674	10,117	9,898	9,674
Pseudo R2				0.5329	0.5166	0.4865

Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
Note: coefficients in columns from (4) to (6) are obtained from a traditional panel logit model estimated on the variables selected by the LASSO.

Table A6 – Models in sample evaluation measures changing μ

Logistic LASSO

	$\mu = 0.75$	$\mu = 0.8$	$\mu = 0.85$	$\mu = 0.9$	$\mu = 0.95$	$\mu = 0.97$	$\mu = 0.98$
Recall positives	69%	70%	79%	79%	89%	89%	89%
Recall negatives	97%	97%	95%	95%	92%	91%	91%
FP rate	3%	3%	5%	5%	8%	9%	9%
FN rate	31%	30%	21%	21%	11%	11%	11%
Precision positives	96%	95%	94%	94%	91%	91%	91%
Precision negatives	76%	76%	82%	82%	89%	89%	89%
Accuracy	83%	83%	87%	87%	90%	90%	90%
Usefulness	0.01	0.01	0.01	0.01	0.00	0.00	0.00
Relative usefulness	38%	46%	53%	63%	76%	79%	73%
Lambda star	10.7%	10.5%	7.8%	7.7%	5.8%	5.6%	5.6%

Random forest

	$\mu = 0.75$	$\mu = 0.8$	$\mu = 0.85$	$\mu = 0.9$	$\mu = 0.95$	$\mu = 0.97$	$\mu = 0.98$
Recall positives	76%	76%	77%	80%	80%	99%	99%
Recall negatives	100%	100%	100%	99%	99%	78%	78%
FP rate	0%	0%	0%	1%	1%	22%	22%
FN rate	24%	24%	23%	20%	20%	1%	1%
Precision positives	100%	100%	100%	99%	98%	82%	82%
Precision negatives	81%	81%	81%	83%	83%	99%	99%
Accuracy	88%	88%	88%	89%	89%	88%	88%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.01	0.01
Relative usefulness	89%	90%	91%	93%	94%	95%	94%
Lambda star	22.1%	22.1%	20.8%	12.9%	11.6%	4.1%	4.1%

XGBoost

	$\mu = 0.75$	$\mu = 0.8$	$\mu = 0.85$	$\mu = 0.9$	$\mu = 0.95$	$\mu = 0.97$	$\mu = 0.98$
Recall positives	92%	92%	94%	94%	94%	94%	98%
Recall negatives	99%	99%	99%	99%	99%	99%	94%
FP rate	1%	1%	1%	1%	1%	1%	6%
FN rate	8%	8%	6%	6%	6%	6%	2%
Precision positives	99%	99%	99%	99%	98%	98%	94%
Precision negatives	92%	92%	94%	94%	95%	95%	98%
Accuracy	95%	95%	96%	96%	96%	96%	96%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.01
Relative usefulness	82%	85%	87%	89%	92%	93%	90%
Lambda star	22.3%	22.3%	14.5%	14.5%	13.3%	13.3%	3.2%

Note: The table reports the following measures to evaluate the performance of the models: Recall positives (o TP rate) = TP/(TP+FN), Recall negatives (o TN rate) = TN/(TN+FP), Precision positives = TP/(TP+FP), Precision negatives = TN/(TN+FN), Accuracy = (TP+TN)/ (TP+TN+FP+FN), FP rate = FP/(FP+TN) e FN rate = FN/(FN+TP).

Table A7 – Models' out-of-sample evaluation measures changing μ

Logit

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	51%	53%	59%	61%	61%	67%	68%
Recall negatives	99%	99%	98%	98%	98%	97%	97%
FP rate	1%	1%	2%	2%	2%	3%	3%
FN rate	49%	47%	41%	39%	39%	33%	32%
Precision positives	83%	83%	83%	83%	82%	82%	81%
Precision negatives	71%	72%	75%	76%	76%	79%	80%
Accuracy	75%	76%	79%	79%	79%	82%	83%
Usefulness	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Relative usefulness	57.3%	61.6%	67.1%	73.9%	81.6%	84.6%	79.9%

Logistic LASSO

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	66%	87%	87%	89%	89%	90%	90%
Recall negatives	99%	98%	98%	98%	97%	95%	95%
FP rate	1%	2%	2%	2%	3%	5%	5%
FN rate	29%	11%	11%	9%	9%	9%	9%
Precision positives	98%	98%	98%	97%	96%	95%	94%
Precision negatives	77%	90%	90%	91%	91%	92%	92%
Accuracy	82%	92%	92%	93%	93%	93%	92%
Usefulness	0.01	0.01	0.01	0.01	0.00	0.00	-0.01
Relative usefulness	39.0%	48.9%	58.9%	68.3%	78.5%	83.2%	79.9%

Random forest

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	84%	84%	84%	89%	89%	90%	93%
Recall negatives	99%	99%	99%	99%	95%	94%	78%
FP rate	1%	1%	1%	1%	5%	6%	22%
FN rate	16%	16%	16%	11%	11%	10%	7%
Precision positives	92%	92%	92%	98%	94%	93%	82%
Precision negatives	83%	83%	83%	85%	85%	85%	89%
Accuracy	87%	87%	87%	89%	87%	87%	81%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.01
Relative usefulness	88.9%	89.3%	89.6%	90.9%	92.7%	92.8%	90.2%

XGBoost

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	81%	81%	81%	81%	84%	84%	86%
Recall negatives	99%	99%	99%	99%	99%	98%	98%
FP rate	1%	1%	1%	1%	1%	2%	2%
FN rate	19%	19%	19%	19%	16%	16%	14%
Precision positives	99%	99%	99%	99%	98%	98%	97%
Precision negatives	86%	86%	86%	86%	88%	88%	90%
Accuracy	90%	90%	90%	90%	91%	91%	92%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Relative usefulness	89.8%	91.0%	92.7%	94.4%	96.4%	97.2%	96.8%

Table A8 – Combined model evaluation measures changing μ

Combined model with weighted average

	$\mu = 0.75$	$\mu = 0.8$	$\mu = 0.85$	$\mu = 0.9$	$\mu = 0.95$	$\mu = 0.97$	$\mu = 0.98$
Recall positives	85%	92%	92%	96%	96%	96%	96%
Recall negatives	100%	100%	100%	99%	99%	98%	98%
Precision positives	93%	100%	100%	99%	99%	98%	98%
Precision negatives	90%	94%	94%	96%	96%	96%	96%
Accuracy	92%	96%	96%	97%	97%	97%	97%
FP rate	0%	0%	0%	1%	1%	2%	2%
FN rate	15%	8%	8%	4%	4%	4%	4%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Relative usefulness	87%	87%	88%	89%	92%	94%	92%

Combined model with simple average

	$\mu = 0.75$	$\mu = 0.8$	$\mu = 0.85$	$\mu = 0.9$	$\mu = 0.95$	$\mu = 0.97$	$\mu = 0.98$
Recall positives	92%	96%	96%	96%	96%	96%	96%
Recall negatives	100%	100%	99%	99%	98%	97%	97%
Precision positives	100%	100%	99%	99%	98%	97%	97%
Precision negatives	94%	96%	96%	96%	96%	96%	96%
Accuracy	96%	98%	98%	97%	97%	96%	96%
FP rate	0%	0%	1%	1%	2%	3%	3%
FN rate	8%	4%	4%	4%	4%	4%	4%
Usefulness	0.03	0.03	0.03	0.03	0.02	0.02	0.02
Relative usefulness	86%	87%	88%	90%	93%	91%	88%

Note: The Table reports the following measures to evaluate the performance of the models: Recall positives (o TP rate) = TP/(TP+FN), Recall negatives (o TN rate) = TN/(TN+FP), Precision positives = TP/(TP+FP), Precision negatives = TN/(TN+FN), Accuracy = (TP+TN)/ (TP+TN+FP+FN), FP rate = FP/(FP+TN) e FN rate = FN/(FN+TP).

Table A9 – Second specification out-of-sample evaluation measures changing μ

Logit

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	59%	62%	62%	66%	70%	70%	70%
Recall negatives	98%	98%	98%	98%	97%	97%	96%
FP rate	2%	2%	2%	2%	3%	3%	4%
FN rate	41%	38%	38%	34%	30%	30%	30%
Precision positives	91%	90%	90%	90%	88%	88%	88%
Precision negatives	73%	75%	75%	78%	79%	79%	79%
Accuracy	79%	80%	80%	82%	83%	83%	83%
Usefulness	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Relative usefulness	53.3%	60.8%	68.3%	75.8%	84.1%	88.4%	88.1%

Logistic LASSO

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	49%	50%	57%	61%	71%	74%	74%
Recall negatives	99%	99%	98%	98%	97%	96%	95%
FP rate	1%	1%	2%	2%	3%	4%	5%
FN rate	51%	50%	43%	39%	29%	26%	26%
Precision positives	84%	84%	84%	84%	83%	82%	82%
Precision negatives	67%	68%	72%	74%	81%	84%	83%
Accuracy	74%	74%	78%	79%	84%	85%	85%
Usefulness	0.01	0.01	0.01	0.01	0.00	0.00	0.00
Relative usefulness	44.8%	52.5%	60.7%	69.3%	80.1%	85.7%	85.4%

Random forest

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	85%	85%	85%	85%	87%	88%	91%
Recall negatives	100%	100%	100%	100%	99%	99%	97%
FP rate	0%	0%	0%	0%	1%	1%	3%
FN rate	15%	15%	15%	15%	13%	12%	9%
Precision positives	100%	100%	100%	100%	99%	98%	96%
Precision negatives	89%	89%	89%	89%	90%	91%	94%
Accuracy	92%	92%	92%	92%	93%	93%	94%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.01
Relative usefulness	88.9%	89.3%	89.6%	90.1%	91.7%	92.7%	90.2%

XGBoost

	μ = 0.75	**μ = 0.8**	**μ = 0.85**	**μ = 0.9**	**μ = 0.95**	**μ = 0.97**	**μ = 0.98**
Recall positives	69%	71%	72%	79%	79%	83%	84%
Recall negatives	99%	99%	99%	99%	99%	99%	99%
FP rate	1%	1%	1%	1%	1%	1%	1%
FN rate	31%	29%	28%	21%	21%	17%	16%
Precision positives	99%	99%	98%	98%	98%	99%	98%
Precision negatives	79%	81%	82%	86%	86%	87%	88%
Accuracy	84%	85%	86%	89%	89%	91%	92%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Relative usefulness	94.0%	94.6%	95.4%	96.4%	97.5%	98.1%	98.0%

Table A10 – Second specification combined model evaluation measures changing μ

Combined model with weighted average

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	84%	84%	86%	90%	90%	90%	90%
Recall negatives	99%	99%	99%	99%	99%	99%	99%
FP rate	1%	1%	1%	1%	1%	1%	1%
FN rate	16%	16%	14%	10%	10%	10%	10%
Precision positives	99%	99%	99%	99%	99%	98%	98%
Precision negatives	88%	88%	90%	92%	92%	92%	92%
Accuracy	92%	92%	92%	94%	94%	94%	94%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Relative usefulness	89%	90%	92%	93%	95%	96%	95%

Combined model with simple average

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	81%	81%	81%	82%	86%	87%	90%
Recall negatives	100%	100%	99%	99%	99%	99%	99%
FP rate	0%	0%	1%	1%	1%	1%	1%
FN rate	19%	19%	19%	18%	14%	13%	10%
Precision positives	99%	99%	99%	99%	99%	99%	99%
Precision negatives	86%	86%	86%	87%	90%	91%	92%
Accuracy	90%	90%	90%	91%	92%	93%	94%
Usefulness	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Relative usefulness	93%	94%	94%	95%	96%	97%	96%

Note: The Table reports the following measures to evaluate the performance of the models: Recall positives (o TP rate) = TP/(TP+FN), Recall negatives (o TN rate) = TN/(TN+FP), Precision positives = TP/(TP+FP), Precision negatives = TN/(TN+FN), Accuracy = (TP+TN)/ (TP+TN+FP+FN), FP rate = FP/(FP+TN) e FN rate = FN/(FN+TP).

Table A11 – Third specification out-of-sample evaluation measures changing μ

Logit

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	9%	9%	11%	11%	11%	16%	16%
Recall negatives	99%	99%	99%	99%	98%	96%	96%
FP rate	1%	1%	1%	1%	2%	4%	4%
FN rate	91%	91%	89%	89%	89%	84%	84%
Precision positives	27%	27%	39%	39%	36%	46%	46%
Precision negatives	53%	53%	53%	53%	53%	54%	54%
Accuracy	54%	54%	55%	55%	54%	56%	56%
Usefulness	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Relative usefulness	23.2%	27.1%	31.0%	37.5%	48.7%	58.0%	64.4%

Logistic LASSO

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	28%	41%	45%	45%	49%	73%	75%
Recall negatives	70%	66%	63%	63%	58%	41%	40%
FP rate	30%	34%	37%	37%	42%	59%	60%
FN rate	72%	59%	55%	55%	51%	27%	25%
Precision positives			60%	60%	58%	55%	56%
Precision negatives	50%	54%	54%	54%	54%	65%	67%
Accuracy	49%	53%	54%	54%	53%	57%	58%
Usefulness	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Relative usefulness	21.2%	29.4%	38.8%	48.9%	63.7%	58.1%	50.9%

Random forest

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	68%	68%	86%	86%	86%	86%	92%
Recall negatives	100%	100%	100%	98%	94%	84%	65%
FP rate	0%	0%	0%	2%	6%	16%	35%
FN rate	32%	32%	14%	14%	14%	14%	8%
Precision positives	100%	100%	100%	96%	93%	89%	76%
Precision negatives	79%	79%	90%	89%	89%	89%	88%
Accuracy	84%	84%	93%	92%	90%	85%	78%
Usefulness	0.00	0.00	0.00	0.00	0.00	0.00	-0.01
Relative usefulness	22.2%	31.4%	43.8%	52.9%	70.7%	64.1%	58.9%

XGBoost

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	14%	14%	75%	75%	75%	83%	92%
Recall negatives	100%	99%	61%	61%	61%	60%	60%
FP rate	0%	1%	39%	39%	39%	40%	40%
FN rate	86%	86%	25%	25%	25%	17%	8%
Precision positives			74%	74%	74%	75%	75%
Precision negatives	54%	54%	82%	82%	82%	89%	94%
Accuracy	57%	57%	68%	68%	68%	72%	76%
Usefulness	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Relative usefulness	3.1%	6.1%	24.4%	49.7%	73.0%	82.1%	86.2%

Table A12 – Third specification combined model evaluation measures changing μ

Combined model with weighted average

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	28%	28%	36%	37%	71%	90%	90%
Recall negatives	90%	90%	79%	77%	98%	92%	79%
FP rate	10%	10%	21%	23%	2%	8%	21%
FN rate	72%	72%	64%	63%	29%	10%	10%
Precision positives	74%	74%	62%	62%	98%	92%	85%
Precision negatives	58%	58%	55%	55%	81%	93%	93%
Accuracy	59%	59%	57%	57%	84%	91%	85%
Usefulness	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Relative usefulness	50.1%	50.8%	51.5%	52.5%	54.7%	57.4%	56.5%

Combined model with simple average

	μ = 0.75	μ = 0.8	μ = 0.85	μ = 0.9	μ = 0.95	μ = 0.97	μ = 0.98
Recall positives	34%	34%	34%	51%	71%	71%	81%
Recall negatives	96%	96%	96%	95%	91%	94%	83%
FP rate	4%	4%	4%	5%	9%	6%	17%
FN rate	66%	66%	66%	49%	29%	29%	19%
Precision positives	91%	90%	90%	89%	85%	87%	82%
Precision negatives	61%	61%	61%	69%	82%	82%	85%
Accuracy	65%	65%	65%	73%	81%	83%	82%
Usefulness	0.01	0.01	0.01	0.01	0.00	0.00	0.00
Relative usefulness	46.3%	47.7%	49.3%	51.3%	61.1%	61.3%	57.6%

Note: The Table reports the following measures to evaluate the performance of the models: Recall positives (o TP rate) = TP/(TP+FN), Recall negatives (o TN rate) = TN/(TN+FP), Precision positives = TP/(TP+FP), Precision negatives = TN/(TN+FN), Accuracy = (TP+TN)/ (TP+TN+FP+FN), FP rate = FP/(FP+TN) e FN rate = FN/(FN+TP).

RECENTLY PUBLISHED "TEMI" (*)

N. 1314 – *Working horizon and labour supply: the effect of raising the full retirement age on middle-aged individuals*, by Francesca Carta and Marta De Philippis (February 2021).

N. 1315 – *Bank credit and market-based finance for corporations: the effects of minibond issuances*, by Steven Ongena, Sara Pinoli, Paola Rossi and Alessandro Scopelliti (February 2021).

N. 1316 – *Is inflation targeting a strategy past its sell-by date?*, by Alberto Locarno and Alessandra Locarno (February 2021).

N. 1317 – *Declining natural interest rate in the US: the pension system matters*, by Jacopo Bonchi and Giacomo Caracciolo (February 2021).

N. 1318 – *Can we measure inflation expectations using Twitter?*, by Cristina Angelico, Juri Marcucci, Marcello Miccoli and Filippo Quarta (February 2021).

N. 1319 – *Identifying deposits' outflows in real-time*, by Edoardo Rainone (February 2021).

N. 1320 – *Whatever it takes to save the planet? Central banks and unconventional green policy*, by Alessandro Ferrari and Valerio Nispi Landi (February 2021).

N. 1321 – *The power of text-based indicators in forecasting the Italian economic activity*, by Valentina Aprigliano, Simone Emiliozzi, Gabriele Guaitoli, Andrea Luciani, Juri Marcucci and Libero Monteforte (March 2021).

N. 1322 – *Judicial efficiency and bank credit to firms*, by Giacomo Rodano (March 2021).

N. 1323 – *Unconventional monetary policies and expectations on economic variables*, by Alessio Anzuini and Luca Rossi (March 2021).

N. 1324 – *Modeling and forecasting macroeconomic dowside risk*, by Davide Delle Monache, Andrea De Polis and Ivan Petrella (March 2021).

N. 1325 – *Foreclosures and house prices*, by Michele Loberto (March 2021).

N. 1326 – *inancial structure and bank relationships of Italian multinational firms*, by Raffaello Bronzini, Alessio D'Ignazio and Davide Revelli (March 2021).

N. 1327 – *Foreign investors and target firms' financial structure: cavalry or lucusts?*, by Lorenzo Bencivelli and Beniamino Pisicoli (April 2021).

N. 1328 – *Board composition and performance of state-owned enterprises: quasi experimental evidence*, by Audinga Baltrunaite, Mario Cannella, Sauro Mocetti and Giacomo Roma (April 2021).

N. 1329 – *Can internet banking affect households' participation in financial markets and financial awareness?*, by Valentina Michelangeli and Eliana Viviano (April 2021).

N. 1330 – *(In)Efficient separations, firing costs and temporary contracts*, by Andrea Gerali, Elisa Guglielminetti and Danilo Liberati (April 2021).

N. 1331 – *The catalytic role of IMF programs*, by Claudia Maurini and Alessandro Schiavone (April 2021).

N. 1332 – *Dating the euro area business cycle: an evaluation*, by Claudia Pacella (April 2021).

N. 1333 – *Population aging, relative prices and capital flows across the globe*, by Andrea Papetti (April 2021).

N. 1334 – *What drives investors to chase returns?*, by Jonathan Huntley, Valentina Michelangeli and Felix Reichling (April 2021).

N. 1335 – *Managerial talent and managerial practices: are they complements?*, by Audinga Baltrunaite, Giulia Bovini and Sauro Mocetti (April 2021).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2019

ALBANESE G., M. CIOFFI and P. TOMMASINO, *Legislators' behaviour and electoral rules: evidence from an Italian reform,* European Journal of Political Economy, v. 59, pp. 423-444, **WP 1135 (September 2017).**

APRIGLIANO V., G. ARDIZZI and L. MONTEFORTE, *Using the payment system data to forecast the economic activity,* International Journal of Central Banking, v. 15, 4, pp. 55-80, **WP 1098 (February 2017).**

ARNAUDO D., G. MICUCCI, M. RIGON and P. ROSSI, *Should I stay or should I go? Firms' mobility across banks in the aftermath of the financial crisis,* Italian Economic Journal / Rivista italiana degli economisti, v. 5, 1, pp. 17-37, **WP 1086 (October 2016).**

BASSO G., F. D'AMURI and G. PERI, *Immigrants, labor market dynamics and adjustment to shocks in the euro area,* IMF Economic Review, v. 67, 3, pp. 528-572, **WP 1195 (November 2018).**

BATINI N., G. MELINA and S. VILLA, *Fiscal buffers, private debt, and recession: the good, the bad and the ugly,* Journal of Macroeconomics, v. 62, **WP 1186 (July 2018).**

BURLON L., A. NOTARPIETRO and M. PISANI, *Macroeconomic effects of an open-ended asset purchase programme,* Journal of Policy Modeling, v. 41, 6, pp. 1144-1159, **WP 1185 (July 2018).**

BUSETTI F. and M. CAIVANO, *Low frequency drivers of the real interest rate: empirical evidence for advanced economies,* International Finance, v. 22, 2, pp. 171-185, **WP 1132 (September 2017).**

CAPPELLETTI G., G. GUAZZAROTTI and P. TOMMASINO, *Tax deferral and mutual fund inflows: evidence from a quasi-natural experiment,* Fiscal Studies, v. 40, 2, pp. 211-237, **WP 938 (November 2013).**

CARDANI R., A. PACCAGNINI and S. VILLA, *Forecasting with instabilities: an application to DSGE models with financial frictions,* Journal of Macroeconomics, v. 61, **WP 1234 (September 2019).**

CHIADES P., L. GRECO, V. MENGOTTO, L. MORETTI and P. VALBONESI, *Fiscal consolidation by intergovernmental transfers cuts? The unpleasant effect on expenditure arrears,* Economic Modelling, v. 77, pp. 266-275, **WP 1076 (July 2016).**

CIANI E., F. DAVID and G. DE BLASIO, *Local responses to labor demand shocks: a re-assessment of the case of Italy,* Regional Science and Urban Economics, v. 75, pp. 1-21, **WP 1112 (April 2017).**

CIANI E. and P. FISHER, *Dif-in-dif estimators of multiplicative treatment effects,* Journal of Econometric Methods, v. 8. 1, pp. 1-10, **WP 985 (November 2014).**

CIAPANNA E. and M. TABOGA, *Bayesian analysis of coefficient instability in dynamic regressions,* Econometrics, MDPI, Open Access Journal, v. 7, 3, pp.1-32, **WP 836 (November 2011).**

COLETTA M., R. DE BONIS and S. PIERMATTEI, *Household debt in OECD countries: the role of supply-side and demand-side factors,* Social Indicators Research, v. 143, 3, pp. 1185–1217, **WP 989 (November 2014).**

COVA P., P. PAGANO and M. PISANI, *Domestic and international effects of the Eurosystem Expanded Asset Purchase Programme,* IMF Economic Review, v. 67, 2, pp. 315-348, **WP 1036 (October 2015).**

ERCOLANI V. and J. VALLE E AZEVEDO, *How can the government spending multiplier be small at the zero lower bound?,* Macroeconomic Dynamics, v. 23, 8. pp. 3457-2482, **WP 1174 (April 2018).**

FERRERO G., M. GROSS and S. NERI, *On secular stagnation and low interest rates: demography matters,* International Finance, v. 22, 3, pp. 262-278, **WP 1137 (September 2017).**

FOA G., L. GAMBACORTA, L. GUISO and P. E. MISTRULLI, *The supply side of household finance,* Review of Financial Studies, v.32, 10, pp. 3762-3798, **WP 1044 (November 2015).**

GERALI A. and S. NERI, *Natural rates across the Atlantic,* Journal of Macroeconomics, v. 62, article 103019, **WP 1140 (September 2017).**

GIORDANO C., M. MARINUCCI and A. SILVESTRINI, *The macro determinants of firms' and households' investment: evidence from Italy,* Economic Modelling, v. 78, pp. 118-133, **WP 1167 (March 2018).**

GOMELLINI M., D. PELLEGRINO and F. GIFFONI, *Human capital and urban growth in Italy,1981-2001,* Review of Urban & Regional Development Studies, v. 31, 2, pp. 77-101, **WP 1127 (July 2017).**

LIBERATI D. and M. LOBERTO, *Taxation and housing markets with search frictions,* Journal of Housing Economics, v. 46, article 101632, **WP 1105 (March 2017).**

MAGRI S., *Are lenders using risk-based pricing in the Italian consumer loan market? The effect of the 2008 crisis,* Journal of Credit Risk, v. 15, 1, pp. 27-65, **WP 1164 (January 2018).**

MERCATANTI A., T. MAKINEN and A. SILVESTRINI, *The role of financial factors for european corporate investment,* Journal of International Money and Finance, v. 96, pp. 246-258, **WP 1148 (October 2017).**

MIGLIETTA A., C. PICILLO and M. PIETRUNTI, *The impact of margin policies on the Italian repo market,* The North American Journal of Economics and Finance, v. 50, **WP 1028 (October 2015).**

MONTEFORTE L. and V. RAPONI, *Short-term forecasts of economic activity: are fortnightly factors useful?,* Journal of Forecasting, v. 38, 3, pp. 207-221, **WP 1177 (June 2018).**

NERI S. and A. NOTARPIETRO, *Collateral constraints, the zero lower bound, and the debt–deflation mechanism,* Economics Letters, v. 174, pp. 144-148, **WP 1040 (November 2015).**

PANCRAZI R. and M. PIETRUNTI, *Natural expectations and home equity extraction,* Journal of Housing Economics, v. 46, 4, **WP 984 (November 2014).**

PEREDA FERNANDEZ S., *Teachers and cheaters. Just an anagram?*, Journal of Human Capital, v. 13, 4, pp. 635-669, **WP 1047 (January 2016).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment,* Macroeconomic Dynamics, v. 23, 2, pp. 590-624, **WP 871 (July 2012).**

2020

ALESSANDRI P. and M. BOTTERO, *Bank lending in uncertain times,* R European Economic Review, V. 128, **WP 1109 (April 2017).**

ANTUNES A. and V. ERCOLANI, *Public debt expansions and the dynamics of the household borrowing constraint,* Review of Economic Dynamics, v. 37, pp. 1-32, **WP 1268 (March 2020).**

ARDUINI T., E. PATACCHINI and E. RAINONE, *Treatment effects with heterogeneous externalities,* Journal of Business & Economic Statistics, , v. 38, 4, pp. 826-838, **WP 974 (October 2014).**

BOLOGNA P., A. MIGLIETTA and A. SEGURA, *Contagion in the CoCos market? A case study of two stress events,* International Journal of Central Banking, v. 16, 6, pp. 137-184, **WP 1201 (November 2018).**

BOTTERO M., F. MEZZANOTTI and S. LENZU, *Sovereign debt exposure and the Bank Lending Channel: impact on credit supply and the real economy,* Journal of International Economics, v. 126, article 103328, **WP 1032 (October 2015).**

BRIPI F., D. LOSCHIAVO and D. REVELLI, *Services trade and credit frictions: evidence with matched bank – firm data,* The World Economy, v. 43, 5, pp. 1216-1252, **WP 1110 (April 2017).**

BRONZINI R., G. CARAMELLINO and S. MAGRI, *Venture capitalists at work: a Diff-in-Diff approach at late-stages of the screening process,* Journal of Business Venturing, v. 35, 3, **WP 1131 (September 2017).**

BRONZINI R., S. MOCETTI and M. MONGARDINI, *The economic effects of big events: evidence from the Great Jubilee 2000 in Rome,* Journal of Regional Science, v. 60, 4, pp. 801-822, **WP 1208 (February 2019).**

COIBION O., Y. GORODNICHENKO and T. ROPELE, *Inflation expectations and firms' decisions: new causal evidence,* Quarterly Journal of Economics, v. 135, 1, pp. 165-219, **WP 1219 (April 2019).**

CORSELLO F. and V. NISPI LANDI, *Labor market and financial shocks: a time-varying analysis,* Journal of Money, Credit and Banking, v. 52, 4, pp. 777-801, **WP 1179 (June 2018).**

COVA P. and F. NATOLI, *The risk-taking channel of international financial flows,* Journal of International Money and Finance, v. 102, **WP 1152 (December 2017).**

D'ALESSIO G., *Measurement errors in survey data and the estimation of poverty and inequality indices,* Statistica Applicata - Italian Journal of Applied Statistics, v. 32, 3, **WP 1116 (June 2017).**

DEL PRETE S. and S. FEDERICO, *Do links between banks matter for bilateral trade? Evidence from financial crises,* Review of World Economic, v. 156, 4, pp. 859 - 885, **WP 1217 (April 2019).**

D'IGNAZIO A. and C. MENON, *The causal effect of credit Guarantees for SMEs: evidence from Italy,* The Scandinavian Journal of Economics, v. 122, 1, pp. 191-218, **WP 900 (February 2013).**

ERCOLANI V. and F. NATOLI, *Forecasting US recessions: the role of economic uncertainty,* Economics Letters, v. 193, **WP 1299 (October 2020).**

MAKINEN T., L. SARNO and G. ZINNA, *Risky bank guarantees,* Journal of Financial Economics, v. 136, 2, pp. 490-522, **WP 1232 (July 2019).**

MODENA F., E. RETTORE and G. M. TANZI, *The effect of grants on university dropout rates: evidence from the Italian case,* Journal of Human Capital, v. 14, 3, pp. 343-370, **WP 1193 (September 2018).**

NISPI LANDI V., *Capital controls spillovers,* Journal of International Money and Finance, v. 109, **WP 1184 (July 2018).**

PERICOLI M., *On risk factors of the stock–bond correlation,* International Finance, v. 23, 3, pp. 392-416, **WP 1198 (November 2018).**

RAINONE E., *The network nature of OTC interest rates,* Journal of Financial Markets, v.47, article 100525, **WP 1022 (July 2015).**

RAINONE E. and F. VACIRCA, *Estimating the money market microstructure with negative and zero interest rates,* Quantitative Finance, v. 20, 2, pp. 207-234, **WP 1059 (March 2016).**

RIZZICA L., *Raising aspirations and higher education. Evidence from the UK's widening participation policy,* Journal of Labor Economics, v. 38, 1, pp. 183-214, **WP 1188 (September 2018).**

SANTIONI, R., F. SCHIANTARELLI and P. STRAHAN, *Internal capital markets in times of crisis: the benefit of group affiliation*, Review of Finance, v. 24, 4, pp. 773-811, **WP 1146 (October 2017).**

SCHIANTARELLI F., M. STACCHINI and P. STRAHAN, *Bank Quality, judicial efficiency and loan repayment delays in Italy,* Journal of Finance , v. 75, 4, pp. 2139-2178, **WP 1072 (July 2016).**

2021

ALBANESE G., E. CIANI and G. DE BLASIO, *Anything new in town? The local effects of urban regeneration policies in Italy,* Regional Science and Urban Economics, v. 86, **WP 1214 (April 2019).**

FIDORA M., C. GIORDANO and M. SCHMITZ, *Real exchange rate misalignments in the Euro Area,* Open Economies Review, v. 32, 1, pp. 71-107, **WP 1162 (January 2018).**

LI F., A. MERCATANTI, T. MAKINEN and A. SILVESTRINI, *A regression discontinuity design for ordinal running variables: evaluating central bank purchases of corporate bonds,* The Annals of Applied Statistics, v. 15, 1, pp. 304-322, **WP 1213 (March 2019).**

LOSCHIAVO D., *Household debt and income inequality: evidence from Italian survey data,* Review of Income and Wealth. v. 67, 1, pp. 61-103, **WP 1095 (January 2017).**

NISPI LANDI V. and A. SCHIAVONE, *The effectiveness of capital controls,* Open Economies Review, v. 32, 1, pp. 183-211, **WP 1200 (November 2018).**

PEREDA FERNANDEZ S., *Copula-based random effects models for clustered data,* Journal of Business & Economic Statistics, v. 39, 2, pp. 575-588, **WP 1092 (January 2017).**

FORTHCOMING

ACCETTURO A., A. LAMORGESE, S. MOCETTI and D. PELLEGRINO, *Housing Price elasticity and growth: evidence from Italian cities,* Journal of Economic Geography, **WP 1267 (March 2020).**

ALBANESE G., G. DE BLASIO and A. LOCATELLI, *Does EU regional policy promote local TFP growth? Evidence from the Italian Mezzogiorno,* Papers in Regional Science, **WP 1253 (December 2019).**

ANZUINI A. and L. ROSSI, *Fiscal policy in the US: a new measure of uncertainty and its effects on the American economy*, Empirical Economics, **WP 1197 (November 2018).**

APRIGLIANO V. and D. LIBERATI, *Using credit variables to date business cycle and to estimate the probabilities of recession in real time*, The Manchester School, **WP 1229 (July 2019).**

BALTRUNAITE A., C. GIORGIANTONIO, S. MOCETTI and T. ORLANDO, Discretion and supplier selection in public procurement, Journal of Law, Economics, and Organization, **WP 1178 (June 2018)**

COVA P., P. PAGANO, A. NOTARPIETRO and M. PISANI, *Secular stagnation, R&D, public investment and monetary policy: a global-model perspective,* Macroeconomic Dynamics, **WP 1156 (December 2017).**

DE PHILIPPIS M., *Multitask agents and incentives: the case of teaching and research for university professors,* Economic Journal, **WP 1042 (December 2015).**

DEL PRETE S. and M. L. STEFANI, *Women as "Gold Dust": gender diversity in top boards and the performance of Italian banks,* Economic Notes, Monte dei Paschi di Siena, **WP 1014 (June 2015).**

HERTWECK M., V. LEWIS and S. VILLA, *Going the extra mile: effort by workers and job-seekers,* Journal of Money, Credit and Banking, **WP 1277 (June 2020).**

METELLI L. and F. NATOLI, *The international transmission of US tax shocks: a proxy-SVAR approach,* IMF Economic Review, **WP 1223 (June 2019).**

MOCETTI S., G. ROMA and E. RUBOLINO, *Knocking on parents' doors: regulation and intergenerational mobility,* Journal of Human Resources, **WP 1182 (July 2018).**

PERICOLI M. and M. TABOGA, *Nearly exact Bayesian estimation of non-linear no-arbitrage term-structure models,* Journal of Financial Econometrics, **WP 1189 (September 2018).**