BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

Corporate default forecasting with machine learning

by Mirko Moscatelli, Simone Narizzano, Fabio Parlapiano
and Gianluca Viggiano

Number 1256

December 2019

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

Corporate default forecasting with machine learning

by Mirko Moscatelli, Simone Narizzano, Fabio Parlapiano
and Gianluca Viggiano

Number 1256 - December 2019

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

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

CORPORATE DEFAULT FORECASTING WITH MACHINE LEARNING

by Mirko Moscatelli*, Simone Narizzano♦, Fabio Parlapiano* and Gianluca Viggiano♦

Abstract

We analyze the performance of a set of machine learning (ML) models in predicting default risk, using standard statistical models, such as the logistic regression, as a benchmark. When only a limited information set is available, for example in the case of financial indicators, we find that ML models provide substantial gains in discriminatory power and precision compared with statistical models. This advantage diminishes when high quality information, such as credit behavioral indicators obtained from the Credit Register, is also available, and becomes negligible when the dataset is small. We also evaluate the consequences of using an ML-based rating system on the supply of credit and the number of borrowers gaining access to credit. ML models channel a larger share of credit towards safer and larger borrowers and result in lower credit losses for lenders.

JEL Classification: G2, C52, C55, D83.
Keywords: credit scoring, machine learning, random forest, gradient boosting machine.
DOI: 10.32057/0.TD.2019.1256

Contents

---

* Bank of Italy, Research, Economics and Statistics Department.

♦ Bank of Italy, Markets and Payments Systems Department.

1. INTRODUCTION[1]

Default forecasting is of key importance for financial institutions and investors. For instance, banks use probabilities of default (PD) to screen potential borrowers, evaluate the terms of new loans and manage the risks stemming from lending activities. Investors also make extensive use of PD and probabilities of migration across different credit rating classes both for bond pricing and portfolio management. In addition, macroprudential authorities have an interest in the surveillance of default risk, as it represents a major source of risk for lenders.

Since the Basel II Accord, default forecasting methods based on a reduced-form regression approach have become popular in the banking industry. These methods consist of multivariate regression models which use firms' characteristics, such as economic and financial fundamentals, to predict the credit quality of a firm. Although the models are calibrated using statistical techniques, the selection of relevant predictors reflects a priori about the structural characteristics that explain the state of financial health of a firm (Resti and Sironi, 2007). More recently, owing to the availability of large dataset and unstructured information, a growing strand of research suggests that models based on machine learning algorithms (henceforth ML) also constitutea suitable alternative for modelling default risk. ML refers to a class of models that can perform complex forecasting tasks when the relationship between predictors and the outcome is complex or unknown. In turn, as established in a number of works (Baesens et al., 2003; Brown and Mues, 2012; Barboza et al., 2017), ML models can perform highly accurate out-of-sample forecasts without imposing strong assumptions on the structure of the data generating process.

This work aims to investigate how default forecasting can benefit from the use of ML algorithms. We contrast statistical models with ensemble decision trees, a class of ML models which can handle both complex relationships across different variables (e.g. non-linear or non-additive) and large datasets with correlated predictors. A decision tree algorithm is in essence a collection of rules that recursively partition the full set of firms into homogeneous subsets according to their characteristics and the outcome variable (i.e. default and non-default). Predictions are then obtained in the form of the odds of a given outcome in each subset. In this study, we use random forest (RDF) and gradient boosted tree (GBT) models, which combine a large number of predictions stemming from individual decision trees into a single (ensemble) highly accurate forecast. This method mitigates the tendency of decision trees to overfit the training dataset. The two models differ in the way in which individual trees are grown. In the RDF model, a random sample of the data and a random selection of variables are used for each tree in order to obtain less correlated individual predictions. The GBT model instead combines predictions obtained from trees that are tailored to deal sequentially with the forecasting errors of their predecessors.

[1] The authors wish to thank Francesco Columba, Filippo Giovannelli, Giorgio Gobbi, Aviram Levy and Franco Panfili for their comments, and express special thanks to Marco Orlandi for his technical contribution.

We estimate the models using a large dataset covering financial and credit behavioral indicators for Italian non-financial firms. We test the out-of-sample performance for these competing models comparing one-year-ahead PD estimates and observed default data for the 2011-17 period.

The analysis highlights the following main results:

(i) when a limited information set is used to train the models, for example financial information usually available to external credit analysts, ML models outperform statistical models both in terms of discriminatory power (the capacity to rank borrowers according to their riskiness) and precision (the ability to estimate PDs that deviate only marginally from actual default rates). When high quality information, such as confidential information available to lenders or information drawn from the Credit Register, is added to the training dataset, the gains from using ML are retained, albeit to a lesser extent. Finally, when only a small number of observations are available for training the models, as in the case of datasets available to lenders, the gains in forecasting performance from using ML are negligible;

(ii) in a comparative statics exercise, where credit is allocated to the banks' (current) clients on the basis of their PDs, we show that transition to a ML rating system would have a positive impact on the amount of credit (intensive margin) and moreover would not increase credit losses for lenders. In terms of the number of borrowers gaining access to credit (extensive margin), ML rating systems allocate credit to a more restricted pool of borrowers. These safer borrowers - which gain prioritized access to credit - are usually larger borrowers; and

(iii) we argue that improvements in forecasting performance from the use of ML are due to its capacity to exploit complex relationships between predictors and default outcomes: indicators presenting a non-linear relationship with the default outcome are more important for ML than for statistical models.

Our results contribute to the literature on corporate default forecasting in a number of ways. First, we document the extent to which there is a gain in accuracy when using ML compared with statistical models. The comparison covers different sets of data and different groups of borrowers. Second, contrary to previous studies, we document the precision of PD estimates using backtesting; these results complement the existing evidence on the discriminatory power of ML models and set out the practical implications for potential users of ML algorithms. Third, we study the impact of using ML for credit allocation, analyzing the changes in the supply of credit, the number of borrowers gaining access to credit and the default rate. To promote fair lending practices, the US banking regulator has recently introduced the obligation for creditors to provide notice of the adverse factors underlying a credit application.[2] This provision exacerbates the trade-off between using transparent statistical

[2] The Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) introduced the obligation for lenders to inform the borrower about the factors involved in taking actions that are adverse or unfavorable to the positive outcome of a credit application. In turn, these requirements provide actionable information to improve borrowers' credit standing. See 'What Are We Learning about Artificial Intelligence in Financial Services?', speech by Lael Brainard, Member of the US

models and adopting models that are more accurate but harder to explain. Nevertheless, our analysis sheds light on the circumstances in which the less transparent ML approach provides superior forecasts, warning lenders to cross-reference the PDs obtained using statistical models with those obtained using ML models.

The rest of the paper is organized as follows. Section 2 reviews the related literature on default forecasting; Section 3 provides a brief introduction to the credit risk models employed to estimate probabilities of default; Section 4 describes the dataset and the estimation process; Section 5 presents the results; and Section 6 concludes.

2. RELATED LITERATURE

The use of the ML approach in credit risk modelling has gained momentum and has recently been applied in the field of early warning systems for banking crises, predictions of household mortgage or consumer credit default, and corporate default.

Chakraborty and Joseph (2017) train a set of models to predict distress in financial institutions based on balance sheet items, finding that ML approaches generally outperform statistical models based on logistic regression. Specifically, the RDF model allows for a marked increase in discriminatory power of about 10 percentage points in terms of the Area under the Receiver Operating Characteristic (AuROC) compared with the logit model.

Using data on US household default on mortgages, Fuster et al. (2018) find that the RDF model generates more accurate predictions than the logit model, although the improvement is minimal and accounts for about 1.2 percentage points of AuROC. The authors argue that most of these gains result from the sophisticated functional form of the RDF model, which captures the complex relationships connecting different variables to default outcomes with greater discriminatory power.

The most popular application of ML algorithms in default forecasting is in modelling of consumer credit risk. Khandani et al. (2010) apply ML and statistical models to predict consumer defaults using a rich set of customer transactions and credit bureau data. Their work shows that RDF forecasts are an improvement over standard statistical techniques for measuring credit quality. Moreover, their use yields a substantial reduction in losses for lenders (from 6 to 23 per cent of banks' total assets). Albanesi and Vamossy (2019) develop a model to predict consumer default based on deep learning (i.e. a combination of forecasts from deep neural network and gradient boosting) in environments with high-dimensional data (over 200 variables). Deep learning models are shown to perform substantially better than logistic regression and adapt to the behavior of the aggregate default rate quite closely. The model is able to capture the sharp rise in credit risk in the run-up to the 2007-09 crisis. Other examples

---

Federal Reserve Board of Governors at the conference Fintech and the New Financial Landscape, Federal Reserve Bank of Philadelphia, Pennsylvania, 13 November 2018.

of their application to consumer credit risk (see, for example, Kruppa et al., 2013; Yuan, 2015) also confirm the superior performance of ML predictions.

There are also numerous applications of ML in corporate default forecasting. Using a large dataset covering the North American corporate sector for the period 1987-2013, Barboza et al. (2017) show that ML provides a substantial improvement of around 10 percentage points of AuROC over traditional models. In Bachman and Zhao (2017), the performance of ML models are compared to the Moody's proprietary algorithm based on a regression model using corporate data for the United States; this exercise shows that ML models deliver an AuROC of about 2-3 percentage points higher than those produced under the regression approach. However, their less transparent structure may lead to PDs that are difficult to relate to firms' underlying characteristics. Furthermore, the inclusion of credit behavioral variables within the set of predictors notably increases the discriminatory power of each model by over 10 percentage points (in terms of AuROC). Fantazzini and Figini (2009) compare RDF and logit model forecasts of the probability of survival using data on SMEs in Germany for the period 1996-2004. Their work documents a weak association between in-sample and out-of-sample forecasting performance, highlighting overfitting problems that can be associated with ML. In contrast to previous findings related to default probabilities, firms' out-of-sample survival forecasts obtained using the logit model outperform their RDF counterparts.

3. CREDIT RISK MODELS AT A GLANCE

We compare two different types of default forecasting models: statistical and ML models. The key differences between the two are as follows. On the one hand, statistical models are especially fit for the purpose of inference, and typically rely on many assumptions regarding the structural relationships between variables, the number of parameters that can be robustly estimated, and the distribution properties of the data generating process. On the other hand, ML models are mostly focused on prediction accuracy, and make very weak assumptions on the structure of the data generating process. This feature allows the detection of data-driven interactions and non-linear or non-monotonous relationships between predictors and the outcome variable. Moreover, ML approaches often involve the estimation of multiple models rather than a single model (this is the main reason why ML are more computationally intensive), and use only the most accurate model to perform prediction tasks. This feature of ML models is particularly relevant to credit risk applications, but it comes at a cost of less transparency compared with statistical models:[3] ML models do not provide estimates of the parameters that relate predictors to

[3] A recent line of research has put forward methods to increase the interpretability of ML models (see Guidotti et al., 2019). Moreover, the use of ML models for the purpose of inference is also considered in a number of works (see Chernozhukov et al., 2018 and Joseph, 2019).

the outcome variable (the models are non-parametric) and this ‘black box’ feature can make their rationale and forecasts difficult to explain.

In this section, we briefly review standard credit risk models (3.1) and introduce the ML models used in the empirical application, namely random forest and gradient boosted trees (3.2).

3.1 STATISTICAL MODELS

Statistical theory offers a variety of methods for estimating default probabilities, of which linear discriminant analysis and logistic regression are the most popular. The linear discriminant analysis (LDA) provides an assessment of corporates’ credit quality using a linear discriminant function that classifies corporate borrowers into groups (default and non-default) based on their characteristics. For more than three decades the discriminant analysis was used extensively by practitioners and performed reasonably well in predicting bankruptcy and other types of distress of privately and publicly held non-financial firms in the international context (Altman, 1968; Altman, 1983; Altman et al., 2017). However, criticism of its underlying assumptions (firms belonging to two different populations, the normal distribution of observables, and equal covariance matrices for the two populations) gradually opened the field to more flexible models such as logistic regression.

The logistic regression model (LOG) provides estimates of a continuous probability of default from firms’ observable characteristics using the two extreme values for the PD itself as the dependent variable: 0 for financially sound firms and 1 for defaulted firms. The model assumes that firms belong to the same population and that a known structural relationship (additive and linear) exists between the observable characteristics of the firm and the credit score. The parameters of the model are usually estimated using maximum likelihood. The LOG relaxes some of the assumptions needed for discriminant analysis (multivariate normality and equal covariance matrices), and has the significant advantage that its results can be easily interpreted, as it delivers a straightforward PD for individual borrowers.

These features attracted strong interest and prompted its diffusion as a scoring model amongst practitioners.[4] Nevertheless, a number of limitations still apply: the lack of consideration of non-linear or complex interactions between observables and defaults, the sensitivity to outliers or missing data, and the difficulty of fully exploiting large datasets. The penalized logistic regression model (PLR), also known as the elastic net regularization of the logistic regression, combines the linear nature of the previous models with a focus on prediction accuracy typical

[4] For example, the Bank of Italy’s In-house Credit Assessment System (BI-ICAS), which is used to assess credit claims posted as collateral in Eurosystem monetary policy operations, is based on a standard logit framework. BI-ICAS integrates two credit risk measures: the credit behaviour component that models monthly data sourced from the Italian Credit Register; and the financial component, based on yearly financial statement data reported in the company accounts data archive of the Bank of Italy. The statistical model of the BI-ICAS estimates at monthly frequency around 300,000 default probabilities for Italian non-financial limited companies. The forecasting performance of the model is assessed annually by the Eurosystem.

of ML models. The PLR has the same structural form as logistic regression, but it estimates the parameters using an objective function aimed at maximizing out-of-sample forecasting performance.

Previous evidence shows that PLR can outperform standard logistic regression in some prediction tasks (Zou and Hastie, 2005); however, the limitations of parametric models still apply.

Overall, statistical models are satisfactory forecasting devices that accommodate both accuracy and transparency requirements. This is owing to their plain functional form, which combines additively monotonic predictors of default into a probability with good out-of-sample performance.

However, the global financial crisis exposed some pitfalls in default forecasting based on commonly used rating systems approaches: i) their slow capacity to adapt to changes in the state of the economy, and ii) their limited ability to model complex non-linear interactions between economic, financial and credit variables. For example, credit ratings may overlook signals of a deteriorating economic and credit environment, such as a rapid increase in default rates or negative shocks to the supply of credit.

Availability of large datasets and unstructured information also pose a challenge to parsimonious rating systems based on a regression approach, since their predefined functional form may inhibit full exploitation of available information. Additionally, from a macroeconomic perspective, not all defaults entail equal consequences for the economy. For example, defaults or early stages of credit deterioration of sovereign borrowers are known to have a ripple effect through the economy. Similarly, the default of large corporate borrowers may spill over to other connected entities (partners, suppliers or customers) and impair their creditworthiness. Thus, being able to detect these infrequent, but critical, default events is a desirable property for a default forecasting model to have.

3.2 MACHINE LEARNING MODELS

We applied two powerful ML algorithms used extensively in credit risk applications: random forest (Breiman, 2001) and gradient boosted trees (Friedman, 2000). The building blocks for these models are classification trees, which are partition algorithms that recursively split the dataset into smaller sets (or branches) that best separate defaulters from non-defaulters.[5] At each iteration, the decision tree algorithm chooses from the covariates space **X** a variable and a value for that variable, so as to minimize a measure of heterogeneity (impurity index) in the resulting subgroups with respect to the classification variable.[6] The process continues within each branch until a

---

[5] A good introduction to trees, random forest and boosting algorithms can be found in Hastie, Tibshirani and Friedman (2001).

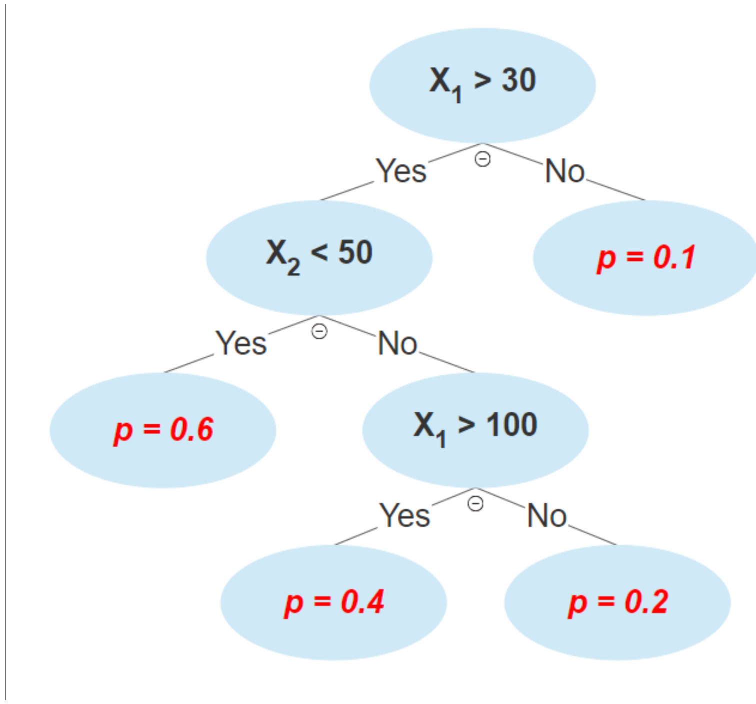
[6] An example of a measure of impurity, as well as the one that we use, is the Gini coefficient. Given a set of observations S with a binary value for each observation $y_i \in \{0,1\}$, the Gini impurity coefficient is defined as $GI(S) \coloneqq 2 \cdot \sum_i p_i (1-p_i)$, where p_i is the percentage of observations in S such that $y_i=1$. The Gini impurity coefficient ranges between 0, when $p_i = 0$ or $p_i = 1$, and 0.5 when $p_i = 0.5$.

stopping condition is reached, such as too few observations in the branch or no significant reduction in impurity. The final branches - called leaves - contain a constant estimate for the probability of default of firms in each leaf, computed as the proportion of defaulted firms over the total number of firms in that leaf in the estimation (or training) dataset.

A simplified classification tree containing only two splitting variables X_1 and X_2 and four leaves is shown in Figure 1. The first level of the tree splits the whole sample into two branches depending on the value of X_1. Firms with X_1 greater than 30 end up in a leaf and receive a default probability equal to0.1, whereas firms with X_1 less than or equal to 30 are classified in a middle branch and are divided further according to the value of X_2. Those firms with X_2 smaller than 50 end up in a leaf and receive a default probability equal to 0.6, while firms with X_2 greater or equal than 50 undergo a final split according to the value of X_1: if X_1 is greater than 100 their default probability equals 0.4; otherwise, they receive a default probability of 0.2.

Notice that the classification tree model: i) captures interactions between variables: the effect of X_2 on the default probability strongly depends on the value of X_1,for instance, if X_1 is greater than 30 the value of X_2 is irrelevant for the estimated default probability; ii) captures non-linear relationships, since the same variable can be used more than once with different values to split the tree, for example, X_1 is used twice to split the sample; and iii) the default probabilities are not a continuous function of the variables: a small increase from 100 to 101 in the value of X_1 results in the default probability of the firm going from 0.2 to 0.4.

Figure 1: Decision tree



Notes: A decision tree for predicting firms' default probabilities grown using two variables, namely X_1 and X_2. Decision rules according to which branches are split are reported in black, while estimated default probabilities are reported in red.

Classification trees have the desirable property of being low-bias, meaning that the leaves - which are defined using multiple variables simultaneously - can fit the data extremely well. However, this flexibility results in the model having an undesirable high-variance property, meaning that out-of-sample predictions are highly sensitive to small changes in the underlying estimation dataset, which often leads to the model having a low forecasting power.

This shortcoming is addressed by ensemble classifiers, such as those provided by the random forest and gradient boosted tree models. Instead of using a single classification tree, these models grow a large set of trees, and the final prediction is obtained as the average of the predictions stemming from the individual trees. In particular, the predictive function F takes the form of $F(x)=\frac{1}{k}\sum_{i=1}^{k} T_i(x)$, where $\{T_1, \ldots, T_k\}$ are the k different classification trees and $T_i(x)$ is the probability of default that the tree T_i associates to a borrower with x covariates. The two models differ according to how the different set of trees is constructed.

The random forest model grows the set of trees: i) using a different bootstrapped sample of the original dataset for each tree (i.e. a sample with replacement having the same number of observations of the original dataset); and ii) selecting at each branch the best split using only a randomly selected subset of the covariates. This procedure implies that the trees differ from one other, since the underlying information set is different.

The gradient boosted tree model grows the set of trees recursively, using an approach based on learning from mistakes, whereby at each step classification errors from the previous trees are used as the dependent variable to grow the next tree. The process proceeds as follows: the first tree T_1 is a standard classification tree as described above; from the second tree onwards, trees are grown using the same x covariate set, but a different dependent variable, computed as the difference between the 0/1 default outcome and the estimations of the previous trees.[7] In other words, the first tree T_1 will be trained on the model $y = T_1(x)$, the second tree T_2 will be trained on the model $y - T_1(x) = T_2(x)$, the third tree T_3 on the model $y - T_1(x) - T_2(x) = T_3(x)$, and so on, each time trying to predict the forecasting errors of the previous trees.[8] This procedure of continuously learning from previous forecasting errors can achieve very accurate predictions, but can also lead to overfitting, thus the number of trees is a very important feature of the model (hyper parameter) that is chosen using cross-validation.[9]

4. THE TRAINING DATASET

4.1 CORPORATE DEFAULTS

We use an extensive dataset of financial and credit behavioral indicators for Italian non-financial firms for the period 2011-17. Our dependent variable, namely financial default, is sourced from the Italian Credit Register and reflects a system-wide definition of the non-performing status of a borrower. A firm is classified as being in default in a given year if the ratio of non-performing credit to total credit drawn from the banking system is greater than 5 per cent for at least one month.[10] The default rate, i.e. the ratio of borrowers classified as non-performing in a given year to the total number of borrowers not in default at the beginning of the year, gives an aggregate measure of credit risk which we aim to model at firm level (Table 1).

Within the 2011-17 time period, credit risk stemming from the corporate sector peaked in 2014, in the aftermath of the European sovereign debt crisis and the associated slowdown of the Italian economy. Following the monetary policy measures adopted by the European Central Bank, the gradual improvement in the business cycle and the exit of vulnerable firms from the market, the aggregate default risk also decreased, with default rate levels approaching about 2.5 per cent in 2017.

[7] Regression trees differ from classification trees because the output variables are continuous rather than numerical. They are generated in the same way, the only difference being that the impurity function is the variance of the outcome variable instead of the Gini impurity coefficient.

[8] Residuals can be interpreted as negative gradients in F of the quadratic loss function $\frac{1}{2}\left(y\text{-}F(x)\right)^2$ from which the name "gradient boosting" is derived.

[9] Hyper parameters define the general characteristics of a model, such as its complexity, and can either be set a priori or learned from the data through optimizers such as grid search.

[10] The status of non-performing loans includes different stages of impairment: past-due 90 days, unlikely to pay and bad loans.

Table 1: Default rate

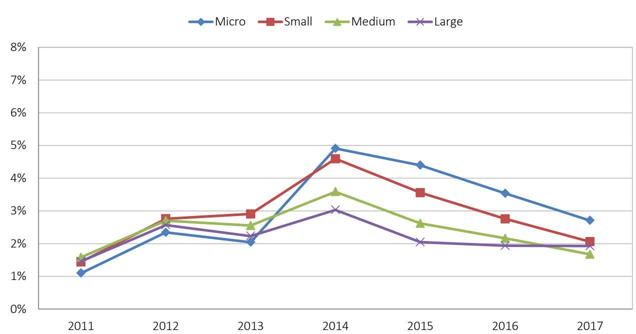
Year	N Firms	Default Rate
2011	222,879	1.20%
2012	233,157	2.45%
2013	259,166	2.23%
2014	249,566	4.76%
2015	252,059	4.12%
2016	260,156	3.31%
2017	269,657	2.53%

Notes: Own calculation based on National Credit Register data. N Firms refers to the number of firms included in our sample, while Default rate is the proportion of firms in default in a given year over the number of firms not in default at the beginning of the year.

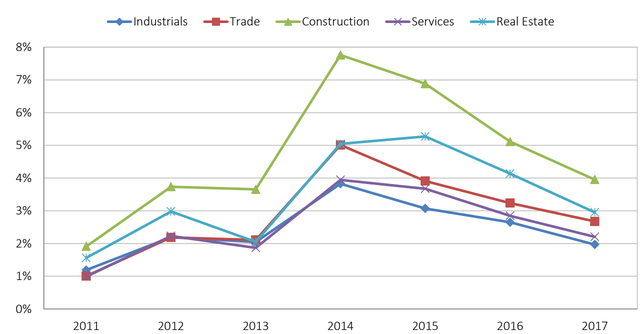
The sudden increase in the number of defaults in both 2012 and 2014, which more than doubled compared to the previous year, may pose a challenge to slow adapting credit risk models. From a qualitative point of view, the ability of models to forecast defaults in different stages of the credit cycle is a desirable property. This is covered in our remarks in the section on evaluation of the models' performance.

Figure 2: Default rate by size and sector

Panel A: Firm size



Panel B: Economic sector



Notes: Own calculation based on National Credit Register and Cerved data. Firm size classes are classified according to the following criteria defined by European Commission: micro, less than 10 employees and turnover (or total assets) up to EUR 2 million; small, less than 50 employees and turnover (or total assets) not exceeding EUR 10 million; medium, less than 250 employees and turnover less than EUR 50 million (or total assets not exceeding EUR 43 million); large, all remaining firms. A firm's economic sector affiliation is determined based on Cerved data.

Structural differences in default risk levels are usually associated with a firm's size class and sector (Figure 2); consistently, credit risk models relate a firm's PD to these factors (Altman et al., 2017; Wang and Dwyer, 2011).

Larger firms record lower default rates relative to SMEs (in our sample almost 3 per cent lower) and these differences are magnified in periods of economic turmoil (from 2014 onwards differences in default risk by size class are exacerbated). The lower riskiness of larger firms could be explained by their engagement in businesses with a longer history, a larger market share and less volatile revenues.

A firm's economic sector is another relevant indicator of its credit risk, with more cyclical sectors (usually investment goods or real estate) leading to volatile revenues and more fragile firms. Indeed, the construction and real estate sectors have significantly higher default rates than industrial sectors, with the latter usually including firms characterized by widely diversified businesses with greater exposure to non-domestic markets.

4.2 FINANCIAL AND CREDIT BEHAVIORAL INDICATORS

The majority of academic works on credit risk modelling and its industry applications use economic and financial ratios as potential indicators of corporate defaults. We add to those predictors a set of credit behavioral indicators on the firm-bank relationship.

Our dataset, drawn from the Company Accounts Data system (provided by Cerved) and the Italian Credit Register, contains a wide array of firm-level variables for Italian non-financial companies. Starting from financial ratios (time lag of one year), we compute 24 indicators covering: profitability, financing structure, debt sustainability and asset types. Credit behavioral indicators (with a time lag of two-months) include eight variables related to a firm's financial flexibility, that is the proportion of drawn to granted bank credit for different facilities, and the occurrence of delinquencies within a firm-bank credit relationship. After including firms' descriptive indicators, such as economic sector and geographical area, our set of default predictors contains 38 variables.

Variables were then selected to be included in the model using the following criteria:

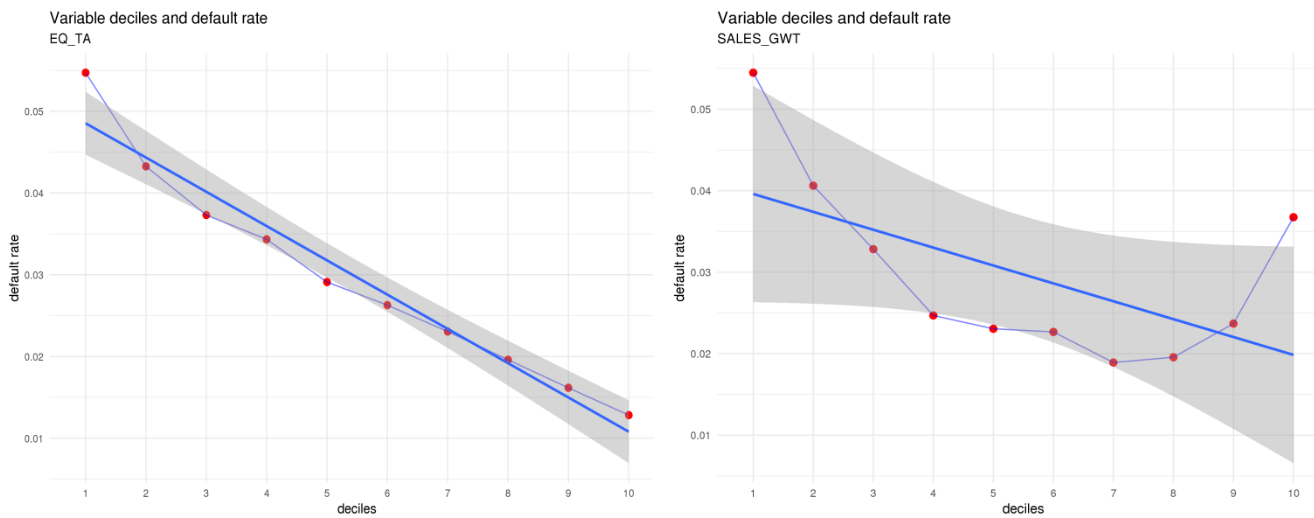
1. using univariate logit regressions for the probability of default, variables with an AuROC lower than 55 per cent were dropped;
2. using the Kolmogorov-Smirnov test, variables with insignificant differences in the distributions between the default and non-default groups were dropped;
3. from the list satisfying 1) and 2), only less correlated variables were retained (linear correlation < 0.7).

Our final set of default predictors consists of 26 variables (see Appendix 1 for a description of the variables), and the estimation dataset includes about 250,000 yearly observations for defaulted and non-defaulted firms.

The variables displaying a linear relationship with the default rate are suitable candidates to be default predictors in all models. For example, the financial leverage indicator (equity over total assets) correlates linearly with the default rate (Figure 3, Panel A). Indeed, the most capitalized firms are more resilient to financial distress or economic slowdown. ML models are expected to benefit more than statistical models from non-linear or non-monotonous indicators. For example, the sales growth rate indicator displays a non-monotonous relationship with default risk, with low and high sales growth firms being more risky than otherwise (Figure 3, Panel B).

Figure 3: Default rate and predicting variables deciles

Panel A: Equity over total assets | **Panel B: Sales growth**



Notes: Own calculation based on National Credit Register and Cerved data. Figure 3 plots (dots) the default rates associated with different deciles of two of the variables employed to predict defaults.

4.3 CALIBRATION OF THE MODELS AND SELECTION OF HYPER PARAMETERS

The estimation of credit risk models is subject to the rare-events problem, i.e. the dataset usually only contains a small proportion of default observations compared to non-default observations. In these circumstances, the discriminatory power of a forecasting model is weak: on the one hand, more importance is attributed to variables that identify sound firms (which are strongly represented in the estimation sample) rather than to variables that help distinguish distressed firms; on the other hand, the model tends to give very low PDs to all the firms.

Downsampling is commonly used to overcome this issue. In a first stage, the model is estimated on a balanced sample, namely a sample with an equal number of default and non-default observations, with the latter randomly sampled from the original dataset.[11] However, with a default rate equal to 50 per cent, the balanced sample does not reflect the real level of credit risk: estimated probabilities of default need to account for the economy's actual default rate. For this reason, during the second stage, a recalibration is performed via algebraic manipulation using Bayes correction (see Sugiyama et al., 2017; Dal Pozzolo et al., 2015a).[12]

[11] See Wallace and Dahabreh (2014).
[12] See Appendix 2 for a detailed description of the calibration procedure.

A second issue related to the estimation of ML models is the setting of hyper parameters, which define the general structure of the models. In RDF, we need to calibrate the number of variables selected at each split, whereas in GBT, we need to set the number of trees and leaves on each tree.

Using five-fold cross-validation, we choose the hyper parameters that maximize the out-of-sample performance of the models.[13] First, we randomly divide the training set into k disjoint subsets (folds) of equal size; then, we train the model k times, each time on a dataset composed by the union of different $k-1$ folds, and we use the remaining fold to compute an out-of-sample AuROC for the model. Finally, this gives us k out-of-sample accuracy measures for each combination of the parameters, and we select the parameters presenting the highest average AuROC.

5. RESULTS

In this section, we compare the credit risk models described in Section 2 based on their forecasting performance metrics: discriminatory power and precision. First, we evaluate their performance using a limited training dataset, which includes financial indicators and firms' characteristics only. This dataset resembles the case of a credit scoring model estimated using only publicly available financial information. Then we use the whole dataset, which covers both financial and credit behavioral information from the Italian Credit Register, with the latter set of indicators usually available only to lenders or supervisors. Second, we perform a static exercise to gain a picture of the allocation of credit that would result from a transition to an ML credit rating system. In particular, we set a credit assignment rule based on the probabilities of default and compare the amount of credit that each model would grant (intensive margin), the number of borrowers gaining access to credit (extensive margin), and the default rate a lender would record using a certain credit rating system. Finally, we evaluate the importance that each model gives to individual predictors in terms of their contribution to sort borrowers according to their credit standing. In particular, we compute the decrease in the AuROC that a model incurs if a variable is randomly permuted in the dataset before estimating the default probabilities.

5.1 DISCRIMINATORY POWER

We assess the discriminatory power of the one-year default probabilities estimated by the different models using the AuROC curve.[14] This is a measure of the ability of the model to assign higher default probabilities to firms that will default compared with financially sound firms: a random model that does not discriminate between sound and distressed firms has a 0.5 AuROC, while a perfect model has an AuROC of 1.

[13] For an introduction to cross-validation, see for example Geisser (1993).
[14] See Fawcett (2004), Chawla (2009) and Xu-Ying et al. (2009).

We first report the out-of-sample AuROC for the different models using a restricted set of information to train the model (Table 2). This dataset is limited to financial ratios and firm characteristics (such as geographic area, economic sector and firm size) that can be collected from publicly available sources. The AuROC score ranges from 72 to 77 per cent, a level of accuracy comparable to Wang and Dwyer (2011), Bacham and Zhao (2017), and Barboza et al., (2017).

We find that tree-based models outperform statistical models over the entire time span, with an average increase in discriminatory power over the LOG model of about 2.6 per cent. Linear discriminant analysis (LDA) and penalized logistic regression (PLR) display results very close to the LOG model, probably owing to similarities in their functional forms.

Table 2: Discriminatory power with restricted dataset (financial indicators)

Year	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
	LDA	LOG	PLR	RDF	GBT
2012	73,7%	73,9%	73,9%	76,6%	76,3%
2013	73,7%	73,9%	73,9%	77,2%	77,3%
2014	72,2%	72,3%	72,4%	74,4%	73,9%
2015	73,7%	73,7%	73,7%	76,1%	76,0%
2016	72,6%	72,6%	72,6%	75,3%	75,3%
2017	73,0%	73,0%	73,0%	75,7%	75,4%

Notes: Own calculation based on Cerved data. The AuROC score is computed using out-of-sample probabilities of defaults obtained from the various models and observed default data.

We then expand the information set used to train the different models in order to include credit behavioral indicators. The AuROC for these models are reported in Table 3. We observe a noticeable increase in overall discriminatory power when credit behavioral indicators are included: the increase accounts for about 10 percentage points in terms of AuROC. This finding is consistent with Bacham and Zhao (2017), for which credit behavioral indicators lead to similar increase in accuracy over a model based exclusively on financial ratios and firm characteristics. While still outperforming statistical models, tree-based models now provide a smaller increase in discriminatory power over the logistic model. We interpret this finding as the effect of the different information set utilized to train the models: with high quality data, the logistic regression already provides a very good forecasting performance which approaches the upper bound of feasible values for the discriminatory power. As a result, the improvement in default forecasting owing to the use of ML is necessarily less pronounced.

Table 3: Discriminatory power with complete dataset (financial and credit behavioral indicators)

Year	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
	LDA	LOG	PLR	RDF	GBT
2012	83,8%	84,0%	84,0%	84,6%	84,7%
2013	83,2%	83,3%	83,3%	84,2%	84,4%
2014	81,1%	81,6%	81,6%	82,5%	82,7%
2015	82,8%	82,9%	82,9%	84,4%	84,6%
2016	82,9%	83,0%	83,0%	84,1%	84,0%
2017	82,9%	83,1%	83,1%	84,1%	84,2%

Notes: Own calculation based on Cerved and National Credit Register data. The AuROC score is computed using out-of-sample probabilities of defaults obtained from the various models and observed default data.

Thanks to the dimension of our data set, confidence intervals for the AuROC statistics are very tight (less than 1 per cent for 95 per cent confidence intervals). In any case, we performed a DeLong significance test on the difference between the AuROC scores for the LOG and RDF models over all the available years both for the complete data set and for the financial data set: all of them strongly reject the null hypothesis of no difference in the AuROC scores at a 1 per cent confidence level.

Finally, we assess the sensitivity of forecasting performance of the different models to the size of the training dataset. To this end, we compile a smaller dataset which includes only 10 per cent of the original observations. Our choice is founded on the idea that datasets available to individual lenders usually include only a small fraction of the firms covered in our complete dataset. Contrary to previous results, we find that tree-based and logistic models achieve substantially similar levels of discriminatory power: ML models need many observations to produce highly accurate forecasts.

Table 4: Discriminatory power with a small dataset (financial and credit behavioral indicators)

Year	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
	LDA	LOG	PLR	RDF	GBT
2012	82,5%	82,6%	83,5%	83,9%	83,5%
2013	82,8%	82,9%	82,9%	83,0%	83,2%
2014	80,5%	80,8%	80,7%	81,1%	80,8%
2015	82,6%	82,7%	82,7%	83,3%	83,4%
2016	82,6%	82,7%	82,7%	82,7%	82,3%
2017	82,4%	82,6%	82,6%	82,5%	82,2%

Notes: Own calculation based on Cerved and National Credit Register data. The AuROC score is computed using out-of-sample probabilities of defaults obtained from the various models and observed default data.

In the following sections we concentrate on the default probabilities estimated using the dataset containing only balance sheet indicators and firm characteristics. This is the case where default forecasting can most benefit from the use of ML models. In Appendix 3, we report the results for the models estimated on the complete dataset also containing credit behavioral indicators from the Italian Credit Register.

5.2 MODELS DISCRIMINATORY POWER ACROSS CLUSTER OF FIRMS

We decompose the overall gain in discriminatory power attributable to ML models by computing the AuROC of PDs estimated for different clusters of borrowers: by firm sector and size (Table 5). The RDF increase in discriminatory power appears to be higher for those clusters that have lower AuROC scores under the LOG models, in particular micro and large firms, and the construction and agriculture sectors. This result suggests that the relative advantage of using ML models is greater when statistical models report weaker performances. This is probably due to a complex underlying relationship between variables and defaults that statistical models are not able to capture.

Table 5: AuROC by sector and size

Sector	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
Manufacturing	77,1%	77,2%	77,2%	79,8%	79,7%
Services	72,0%	72,1%	72,1%	74,9%	74,7%
Agriculture	70,8%	71,1%	71,1%	74,9%	73,6%
Energy	70,0%	70,2%	70,2%	72,4%	72,3%
Construction	69,3%	69,4%	69,4%	72,6%	72,2%
Real estate	67,9%	68,0%	68,0%	70,2%	70,3%

Size	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
Micro	71,6%	71,7%	71,7%	74,2%	74,1%
Small	77,2%	77,2%	77,3%	79,7%	79,6%
Medium	79,6%	79,6%	79,6%	81,4%	81,3%
Large	74,8%	74,6%	74,7%	77,2%	76,1%

Note: Own calculation based on National Credit Register data. Table 5 reports the AuROC score computed using out-of-sample probabilities of defaults from the various models, and observed default data. Results are grouped by firm sector and size.

5.3 BACKTESTING

To assess the degree to which probabilities of default match realized defaults, we perform a binomial-style test for different credit quality buckets, using the Credit Quality Steps (CQS) defined by the Eurosystem for the validation and annual monitoring of credit rating systems. In particular, we use a backtesting strategy outlined in Coppens et al. (2017) as the 'traffic light approach'. In essence, for each bucket, we test how often realized default rates are compatible with the PD forecasts and in the range of usual statistical deviations. A colour is assigned based on the p-value of the test, with green indicating that realized defaults are below the expected threshold, and yellow and red indicating that there is a positive or strongly positive discrepancy between expected and realized defaults respectively.

In Table 6 we report the results of the backtesting exercise: each cell indicates the realized default rate within a certain PD bucket, and the PD thresholds represent the upper limit for PD in each interval.

For the years 2013 and 2015-17, which are characterized by declining default rates (see Table 1), all of the rating systems report satisfactory performances. Structural models (LDA, LOG and PLR), however, show a lower capacity to correctly classify high credit quality borrowers, recording several red warnings in the CQSs 1-2 and 3. Tree-based models, on the other hand, do not report significant discrepancies between expected and realized default for these years.

In the years 2012 and 2014, which are characterized by a strong increase in default risk, the estimated probabilities of default often do not match realized defaults. Structural models tend to record weaker performances, presenting red warnings in all of the credit quality buckets.

Overall, these results show that the assessment of high credit quality borrowers and the adaptation to rapid deterioration in aggregate default risk is a common problem for structural models, while ML models can (partially) mitigate these issues. In particular, RDF predictions tend to be more precise compared with the other models in both economic upturns and downturns.

Table 6: Backtesting

		2012					2013				
CQS	*Threshold*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*
CQS1-2	*0,1%*	0,4%	0,5%	0,5%	0,0%	0,0%	1,0%	0,6%	0,6%	0,0%	0,0%
CQS3	*0,4%*	0,6%	0,7%	0,7%	0,4%	0,6%	0,4%	0,4%	0,5%	0,2%	0,2%
CQS4	*1%*	1,3%	1,4%	1,4%	1,1%	1,6%	0,6%	0,7%	0,7%	0,5%	0,6%
CQS5	*1,5%*	2,3%	2,5%	2,5%	2,3%	3,1%	0,9%	1,1%	1,0%	0,8%	1,2%
CQS6	*3%*	4,4%	4,5%	4,5%	4,5%	4,9%	1,7%	1,9%	1,8%	1,5%	2,0%
CQS7	*5%*	9,0%	9,0%	9,0%	8,9%	8,1%	3,4%	3,5%	3,5%	3,1%	3,6%
		2014					2015				
CQS	*Threshold*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*
CQS1-2	*0,1%*	0,5%	0,7%	0,7%	0,0%	0,0%	4,8%	2,7%	2,0%	0,5%	NA
CQS3	*0,4%*	1,0%	1,1%	1,0%	0,2%	0,5%	0,9%	0,8%	0,8%	0,1%	0,1%
CQS4	*1%*	1,4%	1,6%	1,6%	0,7%	1,4%	0,9%	0,9%	0,9%	0,3%	0,5%
CQS5	*1,5%*	2,1%	2,3%	2,3%	1,5%	2,7%	0,9%	1,1%	1,1%	0,6%	1,0%
CQS6	*3%*	3,3%	3,4%	3,3%	3,2%	4,1%	1,7%	1,8%	1,8%	1,4%	1,9%
CQS7	*5%*	5,4%	5,6%	5,6%	6,3%	6,4%	2,8%	2,8%	2,8%	3,1%	3,5%
		2016					2017				
CQS	*Threshold*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*
CQS1-2	*0,1%*	0,0%	6,4%	7,9%	0,0%	NA	12,5%	2,0%	1,8%	0,4%	0,0%
CQS3	*0,4%*	1,1%	1,0%	1,0%	0,0%	0,1%	0,5%	0,5%	0,4%	0,1%	0,2%
CQS4	*1%*	0,8%	0,8%	0,8%	0,3%	0,6%	0,7%	0,7%	0,7%	0,3%	0,6%
CQS5	*1,5%*	1,0%	1,1%	1,1%	0,8%	1,1%	1,0%	1,0%	1,1%	0,7%	1,2%
CQS6	*3%*	1,8%	1,9%	1,8%	1,6%	2,2%	1,7%	1,7%	1,6%	1,7%	2,2%
CQS7	*5%*	3,6%	3,5%	3,5%	3,5%	3,9%	3,7%	3,7%	3,7%	3,6%	3,7%

Notes: Own calculation. The Threshold column reports the upper limit of the CQS interval identified by the Euro Credit Assessment Framework (ECAF) scale. For instance, a firm is classified in CQS3 if the default probability is between 0.1 and 0.4 per cent. The percentages in the colored cells represent the realized default rate for each CQS in each year. The green, yellow and red colors denote the p-value of the "traffic light approach" test; where the H0 hypothesis is that true probabilities of defaults are less than or equal to the thresholds. The green shading indicates a p-value greater than 20 per cent, the yellow shading between 1 and 20 per cent, and the red shading less than 1 per cent.

5.4 CREDIT ALLOCATION

In this section we evaluate the impact of using ML rating systems on the allocation of credit. We run a static exercise, i.e. keeping other variables constant (the actual number of borrowers and the amount of drawn credit); hence, the results are driven by differences in PDs for the same borrower across different rating systems. In particular, we sort banks' (current) clients by their PDs and allocate credit from the safer to the riskier borrowers using two kind of constraints: first, we highlight the effect on the extensive margin (i.e. on the number and type of borrowers) by assuming that banks set the amount of credit that they are willing to issue; and second, we narrow down the effect on the intensive margin (i.e. on the credit granted) by setting the amount of credit risk that banks are willing to accept.

Table 7 shows the effects on the number of borrowers for different amounts of available credit, which is set as a proportion (1, 5, 10 and 20) of the total amount of credit drawn in our sample. While statistical models result in a broader allocation of credit, tree-based models focus credit on a smaller number of safer firms. Indeed, allocating credit to the latter pool of borrowers would imply a significantly smaller number of defaults and lower credit losses for lenders than the statistical models. This exercise gives insight into the distribution of PDs across different rating systems and accounting for the size of corporate borrowers. As the number of corporate borrowers drops, the average size of their loans increases. Thus, ML models tend to assess larger borrowers as less risky than statistical models. Moreover, within firm size classes, ML models are able to screen risky borrowers somewhat better than statistical models (as shown in Table 5).

Table 7: Credit allocation (fixed granted amount)

% Amount granted	Method	Number of firms	Defaulted amount	Default rate	Default rate - difference wrt LOG
1%	Linear discriminant analysis	8.247	29	0,68%	0,12%
	Logistic regression	7.724	23	0,56%	0,00%
	Penalized logistic regression	7.777	24	0,56%	0,00%
	Random forest	5.172	10	0,24%	-0,32%
	Gradient boosted trees	5.029	12	0,26%	-0,29%
5%	Linear discriminant analysis	19.741	147	0,69%	-0,12%
	Logistic regression	18.713	172	0,81%	0,00%
	Penalized logistic regression	18.951	169	0,79%	-0,01%
	Random forest	13.070	106	0,50%	-0,31%
	Gradient boosted trees	13.398	71	0,33%	-0,48%
10%	Linear discriminant analysis	29.702	593	1,38%	0,06%
	Logistic regression	28.349	564	1,33%	0,00%
	Penalized logistic regression	28.633	498	1,17%	-0,16%
	Random forest	20.022	232	0,54%	-0,79%
	Gradient boosted trees	20.479	349	0,81%	-0,52%
20%	Linear discriminant analysis	49.826	1.761	2,03%	-0,03%
	Logistic regression	48.745	1.788	2,06%	0,00%
	Penalized logistic regression	49.300	1.802	2,08%	0,02%
	Random forest	34.894	835	0,97%	-1,09%
	Gradient boosted trees	36.458	990	1,14%	-0,92%

Notes: Own calculation based on National Credit Register data. Allocated amount and defaulted amount in millions of euro.

Second, we set different thresholds for the default probabilities to decide whether a firm would be granted credit or not (with no restriction on the amount of credit available):

- PD ≤ 0.4% (BBB- investment grade);
- PD ≤ 1.5% (BB non-investment grade or high yield);
- PD ≤ 5% (B- low grade).

The PD thresholds correspond to the CQSs of the ECAF. Each CQS relates to a rating class used by international rating agencies to evaluate the credit quality of a certain issuer.

Table 8 shows the amount of credit that would be granted using different rating systems for different PD thresholds. The RDF and GBT models behave rather consistently across the different rating thresholds: both models would grant more credit than the statistical models, with the only exception being GBT in the first PD bucket, which allocates roughly the same amount. Although the two models would grant more credit, this would not result in higher credit losses for lenders: the default rate and defaulted amount of RDF and GBT is lower than that in other models.

The number of firms which would gain access to credit is always greater under the GBT model than under the LOG model. By contrast, the RDF model produces mixed evidence, cutting down the number of firms in the first threshold (compared with the LOG model), i.e. for investment grade borrowers, while providing a limited increase for high yield and low grade borrowers.

Overall, the results point to ML models having a consistently positive effect on the intensive margin and a reduction in the default rate and credit losses for lenders.

Table 8: Credit allocation (fixed PD thresholds)

Threshold	Method	Number of firms	Default rate	Allocated amount - % difference wrt LOG	Default rate - difference wrt LOG
0,40%	Linear discriminant analysis	8.060	0,76%	-46,0%	0,00%
	Logistic regression	12.637	0,75%	0,0%	0,00%
	Penalized logistic regression	12.154	0,74%	-5,6%	-0,02%
	Random forest	8.836	0,17%	11,3%	-0,59%
	Gradient boosted trees	15.338	0,28%	113,0%	-0,47%
1,50%	Linear discriminant analysis	49.157	1,08%	-9,9%	0,00%
	Logistic regression	53.133	1,12%	0,0%	0,00%
	Penalized logistic regression	52.290	1,10%	-2,8%	-0,01%
	Random forest	54.098	0,70%	46,1%	-0,42%
	Gradient boosted trees	74.532	1,02%	104,9%	-0,10%
5,00%	Linear discriminant analysis	151.450	2,62%	-0,4%	0,00%
	Logistic regression	151.822	2,62%	0,0%	0,00%
	Penalized logistic regression	151.730	2,62%	-1,1%	0,00%
	Random forest	152.906	2,41%	4,5%	-0,21%
	Gradient boosted trees	152.154	2,37%	-0,3%	-0,25%

Notes: Own calculation based on National Credit Register data. Allocated amounts in millions of euro.

5.5 VARIABLE IMPORTANCE AND MODEL ROBUSTNESS

We analyze the importance of individual predictors in each model in terms of the extent to which a single variable affects the overall discriminatory power of a model. In particular, for each variable, we run the following steps:

1) each firm receives a value for the selected variable that is randomly sampled from the distribution of that variable in the dataset (i.e. we permute the variable across all firms);
2) the AuROC of the model is computed on the permuted dataset; and
3) the measure of variable importance consists of the difference between the AuROC for the model applied to the original test dataset (having all the correct values for the selected variable) and the AuROC obtained for the permuted dataset.

If the difference computed in 3) results in large values, this implies that the variable contributes markedly to the discriminatory power of the model and also that the model is very sensitive to potential errors or noise attributable to this variable. Table 9 reports the variable importance (AuROC decrease) for each model for the set of explanatory variables. Predictors are ordered according to their average AuROC decrease across all models.

Comparing statistical models and tree-based models we highlight two main results:

- statistical models are more sensitive to the values assumed by single variables: the LOG model, for example, has six variables that induce a decrease in the AuROC of about 3 per cent or more when permuted, while both the RDF and GBT models have only one variable. Therefore, the overall gain in discriminatory power of ML models is probably attributable to their robustness to noise and their ability to use all of the information in the dataset;
- variables presenting non-linear and/or non-monotone relationships with default risk, such as Sales Growth, Payables Turnover and Receivables Turnover, are more important for tree-based models than for statistical ones. This result points to a connection between the increase in discriminatory power under ML approaches and their capacity to exploit complex relationships between predictors and the outcome variable.

Table 9: Variable importance

Variable	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
IE_CASHFLOW	4,5%	3,7%	3,7%	4,5%	5,4%
LOG_ASSETS	3,2%	3,1%	2,9%	1,7%	2,6%
DSCR	1,9%	2,9%	2,9%	2,2%	2,5%
EQ_TA	2,9%	3,0%	3,0%	1,2%	1,6%
TURNOVER	2,9%	2,9%	2,8%	1,4%	1,6%
CASH_ST_DEBT_S	2,4%	2,9%	2,8%	1,0%	2,3%
AREA_CVD	0,9%	0,9%	0,9%	0,3%	0,5%
ATECO_CVD	0,6%	0,6%	0,6%	0,4%	0,3%
PAYABLES_TURNOVER	0,4%	0,3%	0,3%	0,7%	0,6%
EBITDA_MARGIN	0,4%	0,3%	0,3%	0,3%	0,7%
DIM_CVD	0,6%	0,6%	0,5%	0,0%	0,0%
PFN_EBITDA	0,3%	0,3%	0,3%	0,3%	0,4%
SALES_GWT	0,0%	0,0%	0,0%	0,5%	0,9%
VA_TA	0,2%	0,2%	0,2%	0,4%	0,4%
FIN_MISMATCH	0,2%	0,1%	0,1%	0,5%	0,4%
CASH_TA	0,1%	0,2%	0,2%	0,2%	0,3%
RECEIVABLES_TURNOVER	0,1%	0,1%	0,1%	0,4%	0,3%
PFN_PN	0,0%	0,0%	0,0%	0,3%	0,2%

Notes: Own calculation based on National Credit Register data. See Appendix 1 for a description of the variables. Each column reports the decrease in the AuROC attributable to a permutation of the variable used in the model.

6. CONCLUSIONS

This work compares statistical models usually employed in credit risk modelling with ML models, namely random forest and gradient boosted tree models. We use a large dataset which includes financial ratios and credit behavioral indicators for about 300,000 Italian non-financial firms for the years 2011-17.

When the models are trained using only publicly available information, ML models have a more accurate forecasting performance, both in terms of discriminatory power and precision, compared with statistical models. This gain is reduced when high quality information, such as credit behavioral indicators, is added to the training dataset. However, when the size of the dataset is insufficient to make a robust estimate of the relationships between predictors and the outcome variable, the performance of ML models is similar to that of statistical models.

Moreover, the use of a credit allocation rule based on ML models would result in larger amounts of credit being granted and lower default rates. This is because ML models tend to channel credit towards safer and larger borrowers, resulting in lower credit losses for lenders.

We argue that the better forecasting performance of ML models is due to their ability to capture more precisely the complex relationship between the available firms' indicators and the default outcome. This is demonstrated by the fact that indicators presenting a non-linear or non-monotone relationship with the default are more effectively employed in ML models than in statistical ones. Our results suggest that the joint use of statistical and ML models by lenders or credit analysts may be beneficial for the accurate assessment of potential borrowers. For example, ML models, which are relatively non-transparent, may be used as a benchmark for probability of defaults obtained using more transparent statistical models. This is particularly useful in cases in which the estimates derived from the two models are particularly diverse, such as larger or riskier borrowers, allowing credit analysts to qualitatively adjust probabilities of default.

REFERENCES

Albanesi, S. and D.F. Vamossy, 'Predicting Consumer Default: A Deep Learning Approach', *NBER Working Paper,* No. 26165, 2019.

Altman, E.I., 'Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy', *Journal of Finance*, 4(23) 1968, 589-609.

Altman, E.I., M. Iwanicz-Drozdowska, E.K. Laitinen and A. Suvas, 'Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model', *Journal of International Financial Management and Accounting*, 28, 2017, 131-171.

Altman, E.I., *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding and Dealing With Bankruptcy*, Wiley Interscience, (John Wiley and Sons, 1983).

Bacham D. and J. Zhao, 'Machine Learning: Challenges, Lessons, and Opportunities in Credit Risk Modeling', *Moody's Analytics Risk Perspectives*, No. 9, July 2017.

Baesens, B., T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens and J. Vanthienen, 'Benchmarking state-of-the-art classification algorithms for credit scoring', *Journal of the Operational Research Society*, 54(6), 2003, 627-635.

Barboza, F., H. Kimura and E. Altman, 'Machine learning models and bankruptcy prediction', *Expert Systems with Applications: An International Journal*, 83(c), 2017, 405-417.

Breiman, L., 'Random Forests', *Machine Learning*, 45(1), 2001, 5-32.

Brown, I., and C. Mues, 'An experimental comparison of classification algorithms for imbalanced credit scoring data sets', *Expert Systems with Applications: An International Journal*, 39, 2012, 3446-3453.

Chakraborty, C. and A. Joseph, 'Machine learning at central banks', *Bank of England Staff Working Paper*, No. 674, September 2017.

Chawla, N.V., 'Data Mining for Imbalanced Datasets: An Overview', in O. Maimon and L. Rokach (eds.), *Data Mining and Knowledge Discovery Handbook* (Springer, Boston, MA, 2009), 875-886.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey and J. Robins, 'Double/debiased machine learning for treatment and structural parameters', *The Econometrics Journal*, 21(1), 2018, C1-C68.

Coppens, F., F. González and G. Winkler, 'The performance of credit rating systems in the assessment of collateral used in Eurosystem monetary policy operations', *Occasional Paper Series*, European Central Bank, 65, 2017, 1-40.

Dal Pozzolo, A., O. Caelen and G. Bontempi, 'When is Undersampling Effective in Unbalanced Classification Tasks?', Joint European Conference on Machine Learning and Knowledge Discovery in Databases , Springer, Cham, 2015a, 200-215

Dal Pozzolo, A., O. Caelen, R.A. Johnson and G. Bontempi, 'Calibrating Probability with Undersampling for Unbalanced Classification', *2015 IEEE Symposium Series on Computational Intelligence*, IEEE, 2015b, 159-166.

Fantazzini, D., and S. Figini, 'Random Survival Forests Models for SME Credit Risk Measurement', *Methodology and Computing in Applied Probability*, 11, 2009, 29-45.

Fawcett, T., 'ROC graphs: Notes and practical considerations for researchers', *Machine Learning*, 31(1), 2004, 1-38.

Friedman, J.H., 'Greedy function approximation: A gradient boosting machine', *The Annals of Statistics*, 29(5), 2000, 1189-1232.

Friedman, J., T. Hastie and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Interference, and Prediction* (New York: Springer Series in Statistics, 2001).

Fuster, A., P. Goldsmith-Pinkham, T. Ramodorai and A. Walther, 'Predictably Unequal? The Effects of Machine Learning on Credit Markets', working paper, 2018.

Geisser, S., *Predictive Inference: An Introduction*, Monograph on Statistics and Applied Probability, 55 (Routledge, 1993).

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi and F. Giannotti, 'A Survey of Methods for Explaining Black Box Models', *ACM computing surveys (CSUR)*, 51(5), 2019, 93.

Joseph, A., 'Shapley regressions: A framework for statistical inference on machine learning models', *Bank of England Working Paper*, No. 784, 2019.

Khandani, A.E., A.J. Kim and A.W. Lo, 'Consumer credit-risk models via machine-learning algorithms', *Journal of Banking & Finance*, 34(1), 2010, 2767-2787.

Kruppa, J., A. Schwarz, G. Arminger and A. Ziegler, "Consumer credit risk: Individual probability estimates using machine learning", *Expert Systems with Applications*, 40(13), 2013, 5125-5131.

Resti, A., and A. Sironi, *Risk Management and Shareholders' Value in Banking: From Risk Measurement Models to Capital Allocation Policies* (John Wiley & Sons Ltd, 2007).

Saerens, M., P. Latinne and C. Decaestecker, 'Adjusting the Outputs of a Classifier to New a Priori Probabilities: A Simple Procedure', *Neural Computation*, 14(1), 2002, 21-41.

Sugiyama, Masashi, Neil D. Lawrence, and Anton Schwaighofer. "Dataset shift in machine learning." (2017).

Wallace, B.C., and I.J. Dahabreh, 'Improving class probability estimates for imbalanced data', Knowledge and Information Systems, 41(1), 2014, 33-52.

Wang, J., and D.W. Dwyer, 'Moody's Analytics RiskCalc™ v3.1 Italy', *Moody's Analytics*, Inc., 2011, 1-33.

Xu-Ying, L., Wu, J., and Zhou, Z.H., 'Exploratory Undersampling for Class-Imbalance Learning', *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2), 2009, 539-550.

Yuan, D., 'Applications of machine learning: consumer credit risk analysis', Doctoral dissertation, Massachusetts Institute of Technology, 2015.

Zou, H., and T. Hastie, 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society*, 67, 2005, 301-320.

APPENDICES

APPENDIX 1

We describe the financial and credit behavioral indicators used to predict default of non-financial firms. By means of graphical inspection, those variables presenting a non-linear or non-monotonous relationship with respect to the default outcome are labelled "NL".

Variable	Description
TURNOVER (Asset Turnover Ratio)	Ratio between net sales and total assets. The asset turnover ratio is an efficiency ratio that measures a firm's ability to generate sales from its assets.
VA_TA (Value Added to Total Assets)	Ratio between economic value added and total assets. Operating profitability ratio that measures the firm's ability to generate value from its assets.
EBITDA_MARGIN (EBITDA to Net Sales) - ***NL***	Operating profitability ratio that measures how much earnings the firm is generating before interest, taxes, depreciation, and amortization, as a percentage of revenue.
PFN/PN (Net Debt to Equity)	Measure of a firm's financial leverage, calculated by dividing its net liabilities by stockholders' equity.
EQ_TA (Equity to Total Assets)	Ratio between equity and total assets. Used to assess a company's financial leverage
PFN/EBITDA (Net Debt to EBITDA)	Debt sustainability ratio gives an indication as to how long a firm would need to operate at its current level to pay off all its financial debt.
IE_CASHFLOW (Interest Expenses to Cash Flow)	Ratio that indicates the enterprise's ability to pay interest from generated cash flow.
DSCR (Debt Service Coverage Ratio)	Ratio of debt sustainability that refers to the amount of cash flow available to pay interest expenses and annual principal payments on financial debt.
FIN_MISMATCH (financial mismatch)	Ratio of the mismatch (difference) between short-term liabilities and short-term assets and total assets. Negative value of the ration (short-term liabilities > short-term assets) indicates that the firm has enough short-term assets to meet its short-term liabilities.
CASH_ST_DEBT_S (Current Assets to Short Term Debt)	Liquidity ratio that measures a firm's ability to pay off short-term debt obligations with cash and cash equivalents.
CASH_TA (Cash to Total Assets)	Ratio between cash and liquid assets to total assets. It measures a firm's liquidity and how easily it can service debt and short-term liabilities if the need arises.
RECEIVABLES_TURNOVER (Receivable Turnover Ratio) - ***NL***	Efficiency ratio that measures how efficiently a firm is using its assets. It measures the number of times over a given period (usually a year) that a firm collects its average accounts receivable.
PAYABLES_TURNOVER (Payable Turnover Ratio) - ***NL***	Efficiency and liquidity ratio that measures how many times a firm pays its creditors over an accounting period.
LOG_ASSETS (Natural Logarithm of Total Assets)	Measures the size of the firm.
SALES_GWT (Net Sales Growth) - ***NL***	Measures a firm's growth in a specific year. It also measures the stability of a firm's performance.
Variable	**Description**
DG_CR_TOT (Drawn amount to Granted Amount) - ***NL***	Financial flexibility ratio. It measures the percentage of available credit that the firm is actually using. It refers to all the different types of loans.

DG_CR_REV (Drawn Amount to Granted Amount of uncommitted short term loans) - ***NL***	Financial flexibility ratio. It measures the percentage of uncommitted short-term loans that the firm is actually using.
DG_CR_AUT (Drawn Amount/Granted Amount, short term loans) - ***NL***	Financial flexibility ratio. It measures the percentage of self-liquidating short-term loans that the firm is actually using.
DUMMY_SCONF (Overdrawns)	Dummy equal to 1 if the firm has an overdrawn amount greater than the granted amount, and 0 otherwise.
DEF_STORIA_CRED (Deteriorated loans)	Dummy equal to 1 if the firm has deteriorated loans, and 0 otherwise.
MORTGAGE (Mortgage)	Dummy variable equal to 1 if long-term loans are more than 90 per cent of total loans. It is used to mitigate the impact on PD estimation of a high drawn/granted ratio which is physiological for mortgages.
DUMMY_REV	Dummy equal to 1 if the firm has uncommitted short-term loans, and 0 otherwise.
DUMMY_AUT	Dummy equal to 1 if the firm has short-term loans, and 0 otherwise.
Variable	**Description**
AREA_CVD (geographical area)	Dummy variables identifying the geographical region where the firm operates (North-East, North-West, Center, South and Islands).
ATECO_CVD (economic sector)	Dummy variables identifying firms' economic sector.
DIM_CVD (size)	Dummy variables identifying firm size as defined by the European Commission.

APPENDIX 2

The performance of credit risk models typically suffers when they are estimated on a dataset that has many more non-defaults (healthy firms) than defaults (unhealthy firms). To improve the discriminatory power of the models, a new estimation dataset is derived from the original dataset using a downsampling strategy, by combining all the unhealthy firms with an equal number of randomly drawn healthy firms.

The in-sample default rate for this new estimation dataset is, by construction, 50 per cent, meaning that the dataset is balanced. However, while downsampling improves the discriminatory power of the model, it also creates an upward distortion of the estimated probabilities.

Let x be the vector of firm characteristics, D the event that the firm will go into default, and B the event that the firm is selected in the balanced estimation dataset. By training the model on the balanced training dataset we estimate the probabilities $P(D \mid B, x)$, and not $P(D \mid x)$, the true probabilities we are actually interested in. However, following a common procedure in the literature, we can link the two probabilities using the Bayes rule: [15]

$$P(D \mid B, x) = \frac{P(D, B, x)}{P(B, x)} = \frac{P(B \mid D, x)\cdot P(D \mid x)}{P(B \mid D, x)\cdot P(D \mid x) + P(B \mid \neg D, x)\cdot P(\neg D \mid x)}$$

$$= \frac{P(D \mid x)}{P(D \mid x) + P(B \mid \neg D, x)\cdot(1-P(D \mid x))}, \tag{A.1}$$

where A.1 follows from the fact that $P(B \mid D, x)=1$, since all defaulted firms are in the balanced dataset, and that $P(\neg D \mid x)=1-P(D \mid x)$. Notice that $P(B \mid \neg D, x)$, the probability that a healthy firm is in the balanced dataset, is known because it is the number of healthy firms in the balanced dataset out of the total number of healthy firms, which, by construction, is equal to the number of unhealthy firms over the number of healthy firms in the original dataset. If we solve the equation above for $P(D \mid x)$, we obtain:

$$P(D \mid x) = \frac{\beta\cdot P(D \mid B, x)}{\beta\cdot P(D \mid B, x) - P(D \mid B, x) + 1}, \tag{A.2}$$

where we have defined $\beta = P(B \mid \neg D, x) = \frac{\#\text{ unhealthy firms}}{\#\text{ healthy firms}}$. We use this formula to calibrate the probabilities estimated by the model from the balanced dataset. The calibration procedure is as follows:

1. We estimate the model on the balanced dataset.
2. We apply the model to the (unbalanced) test dataset, obtaining the probabilities $P(D \mid B, x)$.

[15] See Saerens et al. (2002) and Dal Pozzolo et al. (2015b).

3. We calibrate the probabilities from 2) using A.2, and obtain the final unbiased probabilities $P(D \mid x)$.

Appendix 3

We report the results for discriminatory power, backtesting, credit allocation and variable importance for the different models trained on the dataset also containing credit behavioral indicators from the Italian Credit Register.

Table A.1: AuROC by firm sector and size

Sector	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
Manufacturing	85,4%	85,5%	85,5%	86,2%	86,3%
Services	82,4%	82,5%	82,5%	83,7%	83,7%
Agriculture	81,3%	81,4%	81,4%	82,4%	82,4%
Energy	80,8%	81,0%	81,0%	82,7%	82,7%
Construction	80,7%	80,9%	80,9%	81,5%	81,8%
Real estate	79,8%	80,1%	80,1%	82,1%	82,2%

Size	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
Micro	81,7%	82,0%	82,0%	83,1%	83,2%
Small	85,7%	85,7%	85,7%	86,4%	86,5%
Medium	85,9%	86,0%	86,0%	86,6%	86,6%
Large	80,4%	80,4%	80,4%	82,1%	81,6%

Notes: Own calculation based on national Credit Register data. Table A.1 reports the AuROC score computed using out-of-sample PDs from the various models and observed default data. Results are grouped by firm sector and size.

Table A.2: Backtesting

				2012					2013		
CQS	*Threshold*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*
CQS1-2	*0,1%*	0,2%	0,2%	0,2%	0,1%	0,1%	0,0%	0,1%	0,1%	0,2%	0,0%
CQS3	*0,4%*	0,5%	0,4%	0,4%	0,3%	0,4%	0,3%	0,3%	0,3%	0,2%	0,2%
CQS4	*1%*	1,5%	1,4%	1,4%	1,0%	1,5%	0,6%	0,6%	0,5%	0,4%	0,6%
CQS5	*1,5%*	2,5%	2,5%	2,5%	2,1%	2,6%	1,1%	1,0%	1,0%	0,9%	1,2%
CQS6	*3%*	4,2%	4,2%	4,1%	3,9%	4,7%	1,9%	1,9%	1,9%	1,5%	2,2%
CQS7	*5%*	6,9%	7,7%	7,8%	7,8%	7,5%	3,4%	3,6%	3,5%	3,2%	3,8%
				2014					2015		
CQS	*Threshold*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*
CQS1-2	*0,1%*	0,2%	0,2%	0,2%	0,1%	0,1%	0,2%	0,2%	0,2%	0,0%	0,0%
CQS3	*0,4%*	0,4%	0,4%	0,4%	0,2%	0,3%	0,2%	0,2%	0,2%	0,2%	0,2%
CQS4	*1%*	1,5%	1,4%	1,4%	0,8%	1,4%	0,6%	0,6%	0,6%	0,4%	0,6%
CQS5	*1,5%*	2,9%	2,7%	2,7%	1,7%	2,7%	0,9%	1,0%	1,0%	0,6%	1,3%
CQS6	*3%*	4,0%	3,8%	3,8%	3,0%	4,2%	1,9%	1,9%	1,9%	1,5%	2,0%
CQS7	*5%*	6,5%	6,7%	6,6%	6,5%	6,7%	3,6%	3,5%	3,6%	2,9%	3,6%
				2016					2017		
CQS	*Threshold*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*	*LDA*	*LOG*	*PLR*	*RDF*	*GBM*
CQS1-2	*0,1%*	0,0%	0,2%	0,1%	0,0%	NA	0,0%	0,0%	0,0%	0,1%	0,0%
CQS3	*0,4%*	0,3%	0,3%	0,3%	0,1%	0,2%	0,2%	0,2%	0,2%	0,1%	0,1%
CQS4	*1%*	0,5%	0,6%	0,6%	0,3%	0,6%	0,5%	0,5%	0,5%	0,4%	0,5%
CQS5	*1,5%*	1,4%	1,3%	1,3%	0,8%	1,4%	1,0%	1,0%	0,9%	0,7%	1,1%
CQS6	*3%*	2,1%	2,3%	2,2%	1,8%	2,5%	1,6%	1,7%	1,7%	1,4%	1,8%
CQS7	*5%*	3,8%	4,0%	4,0%	3,9%	4,7%	3,1%	3,2%	3,2%	3,1%	3,1%

Notes: Own calculation. The Threshold column reports the upper limit of the CQS interval. For instance, a firm is in CQS3 if its default probability is greater than 0.1 per cent and lower than 0.4 per cent. The percentages represent the realized default rate for each CQS and year.

Table A.3: Credit allocation (fixed granted amount)

Percentage of the total amount granted	Method	Number of firms	Defaulted amount	Default rate	Default rate - difference wrt LOG
1%	Linear Discriminant Analysis	14.649	5	0,11%	0,00%
	Logistic Regression	13.242	5	0,11%	0,00%
	Penalized logistic regression	13.279	5	0,11%	0,00%
	Random Forest	7.887	7	0,16%	0,05%
	Gradient Boosted Trees	8.614	5	0,12%	0,02%
5%	Linear Discriminant Analysis	27.389	52	0,24%	0,00%
	Logistic Regression	26.643	51	0,24%	0,00%
	Penalized logistic regression	26.667	50	0,23%	-0,01%
	Random Forest	18.136	40	0,19%	-0,05%
	Gradient Boosted Trees	19.230	45	0,21%	-0,03%
10%	Linear Discriminant Analysis	36.145	154	0,36%	-0,01%
	Logistic Regression	35.805	160	0,37%	0,00%
	Penalized logistic regression	35.753	157	0,37%	-0,01%
	Random Forest	26.112	113	0,26%	-0,11%
	Gradient Boosted Trees	27.324	118	0,27%	-0,10%
20%	Linear Discriminant Analysis	52.271	427	0,50%	-0,06%
	Logistic Regression	52.823	482	0,56%	0,00%
	Penalized logistic regression	52.638	459	0,53%	-0,03%
	Random Forest	41.709	395	0,46%	-0,10%
	Gradient Boosted Trees	43.642	422	0,49%	-0,07%

Notes: Own calculation based on National Credit Register data. Allocated amount and defaulted amount in millions of euro.

Table A.4: Credit allocation (fixed PD thresholds)

Threshold	Method	Allocated amount	Number of firms	Default rate	Allocated amount - % difference wrt LOG	Default rate - difference wrt LOG
0,40%	Linear Discriminant Analysis	28.952	40.401	0,27%	-1,2%	0,00%
	Logistic Regression	29.306	41.336	0,27%	0,0%	0,00%
	Penalized logistic regression	28.403	40.439	0,26%	-3,1%	0,00%
	Random Forest	28.624	30.938	0,18%	-2,3%	-0,08%
	Gradient Boosted Trees	44.910	43.162	0,23%	53,2%	-0,04%
1,50%	Linear Discriminant Analysis	143.969	100.633	0,77%	5,4%	0,00%
	Logistic Regression	136.564	100.799	0,75%	0,0%	0,00%
	Penalized logistic regression	135.414	100.102	0,74%	-0,8%	-0,01%
	Random Forest	130.639	91.316	0,56%	-4,3%	-0,18%
	Gradient Boosted Trees	159.719	110.954	0,76%	17,0%	0,01%
5,00%	Linear Discriminant Analysis	268.245	160.791	1,74%	-0,1%	0,00%
	Logistic Regression	268.410	164.069	1,79%	0,0%	0,00%
	Penalized logistic regression	268.932	164.047	1,80%	0,2%	0,00%
	Random Forest	285.407	166.059	1,70%	6,3%	-0,09%
	Gradient Boosted Trees	277.776	167.218	1,71%	3,5%	-0,08%

Notes: Own calculation based on National Credit Register data. Allocated amount in millions of euro.

Table A.5: Variable importance

Excluded variable	Linear discriminant analysis	Logistic regression	Penalized logistic regression	Random forest	Gradient boosted trees
DG_CR_REV	10,7%	8,4%	8,4%	4,6%	5,6%
DG_CR_TOT	3,6%	4,8%	4,8%	3,9%	4,6%
DEF_STORIA_CRED	1,5%	1,7%	1,7%	1,3%	1,4%
LOG_ASSETS	1,3%	1,1%	1,1%	0,8%	1,3%
DUMMY_REV	1,6%	1,4%	1,4%	0,1%	0,2%
IE_CASHFLOW	0,8%	0,6%	0,6%	0,7%	0,8%
CASH_ST_DEBT_S	0,7%	0,7%	0,6%	0,2%	0,3%
DG_CR_AUT	0,4%	0,2%	0,2%	0,5%	0,8%
DSCR	0,1%	0,3%	0,3%	0,5%	0,6%
EQ_TA	0,4%	0,4%	0,4%	0,2%	0,2%
DUMMY_SCONF	0,2%	0,3%	0,3%	0,2%	0,1%
TURNOVER	0,2%	0,2%	0,2%	0,2%	0,3%
PAYABLES_TURNOVER	0,2%	0,1%	0,1%	0,2%	0,2%
AREA_CVD	0,2%	0,2%	0,2%	0,1%	0,1%
SALES_GWT	0,0%	0,0%	0,0%	0,3%	0,5%
ATECO_CVD	0,2%	0,2%	0,2%	0,1%	0,1%
DUMMY_AUT	0,1%	0,3%	0,3%	0,0%	0,0%
EBITDA_MARGIN	0,2%	0,1%	0,1%	0,1%	0,2%
CASH_TA	0,1%	0,2%	0,1%	0,1%	0,1%
VA_TA	0,1%	0,1%	0,1%	0,1%	0,1%
DIM_CVD	0,1%	0,1%	0,1%	0,0%	0,0%
PFN_EBITDA	0,0%	0,0%	0,0%	0,1%	0,1%
RECEIVABLES_TURNOVER	0,0%	0,0%	0,0%	0,1%	0,0%
PFN_PN	0,0%	0,0%	0,0%	0,0%	0,1%
MORTGAGE	0,0%	0,0%	0,0%	0,1%	0,0%
FIN_MISMATCH	0,0%	0,0%	0,0%	0,1%	0,0%

Notes: Own calculation based on National Credit Register data. See Appendix 1 for a description of the variables. Each column reports the decrease in the AuROC score attributable to a permutation of the variable used in the model.

RECENTLY PUBLISHED “TEMI” (*)

N. 1232 – *Risky bank guarantees*, by Taneli Mäkinen, Lucio Sarno and Gabriele Zinna (July 2019).

N. 1233 – *News and consumer card payments*, by Guerino Ardizzi, Simone Emiliozzi, Juri Marcucci and Libero Monteforte (October 2019).

N. 1234 – *Forecasting with instabilities: an application to DSGE models with Financial Frictions*, by Roberta Cardani, Alessia Paccagnini and Stefania Villa (October 2019).

N. 1235 – *The real effects of ‘ndrangheta: firm-level evidence*, by Litterio Mirenda, Sauro Mocetti and Lucia Rizzica (October 2019).

N. 1236 – *Forward-looking effective tax rates in the banking sector*, by Ernesto Zangari and Elena Pisano (October 2019).

N. 1237 – *A profit elasticity approach to measure banking competition in Italian credit markets*, by Michele Benvenuti and Silvia Del Prete (October 2019).

N. 1238 – *What do almost 20 years of micro data and two crises say about the relationship between central bank and interbank market liquidity? Evidence from Italy*, by Massimiliano Affinito (October 2019).

N. 1239 – *Bank credit, liquidity and firm-level investment: are recessions different?*, by Ines Buono and Sara Formai (October 2019).

N. 1240 – *Youth drain, entrepreneurship and innovation*, by Massimo Anelli, Gaetano Basso, Giuseppe Ippedico and Giovanni Peri (October 2019).

N. 1241 – *Fiscal devaluation and labour market frictions in a monetary union*, by Lorenzo Burlon, Alessandro Notarpietro and Massimiliano Pisani (October 2019).

N. 1242 – *Financial conditions and growth at risk in Italy*, by Piergiorgio Alessandri, Leonardo Del Vecchio and Arianna Miglietta (October 2019).

N. 1243 – *Cross-country differences in the size of venture capital financing rounds. A machine learning approach*, by Marco Taboga (November 2019).

N. 1244 – *Shifting taxes from labour to consumption: the efficiency-equity trade-off*, by Nicola Curci and Marco Savegnago (November 2019).

N. 1245 – *Credit supply, uncertainty and trust: the role of social capital*, by Maddalena Galardo, Maurizio Lozzi and Paolo Emilio Mistrulli (November 2019).

N. 1246 – *Financial development and growth in European regions*, by Paola Rossi and Diego Scalise (November 2019).

N. 1247 – *IMF programs and stigma in Emerging Market Economies*, by Claudia Maurini (November 2019).

N. 1248 – *Loss aversion in housing assessment among Italian homeowners*, by Andrea Lamorgese and Dario Pellegrino (November 2019).

N. 1249 – *Long-term unemployment and subsidies for permanent employment*, by Emanuele Ciani, Adele Grompone and Elisabetta Olivieri (November 2019).

N. 1250 – *Debt maturity and firm performance: evidence from a quasi-natural experiment*, by Antonio Accetturo, Giulia Canzian, Michele Cascarano and Maria Lucia Stefani (November 2019).

N. 1251 – *Non-standard monetary policy measures in the new normal*, by Anna Bartocci, Alessandro Notarpietro and Massimiliano Pisani (November 2019).

(*) 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.

2017

AABERGE, R., F. BOURGUIGNON, A. BRANDOLINI, F. FERREIRA, J. GORNICK, J. HILLS, M. JÄNTTI, S. JENKINS, J. MICKLEWRIGHT, E. MARLIER, B. NOLAN, T. PIKETTY, W. RADERMACHER, T. SMEEDING, N. STERN, J. STIGLITZ, H. SUTHERLAND, *Tony Atkinson and his legacy,* Review of Income and Wealth, v. 63, 3, pp. 411-444, **WP 1138 (September 2017).**

ACCETTURO A., M. BUGAMELLI and A. LAMORGESE, *Law enforcement and political participation: Italy 1861-65,* Journal of Economic Behavior & Organization, v. 140, pp. 224-245, **WP 1124 (July 2017).**

ADAMOPOULOU A. and G.M. TANZI, *Academic dropout and the great recession,* Journal of Human Capital, V. 11, 1, pp. 35–71, **WP 970 (October 2014).**

ALBERTAZZI U., M. BOTTERO and G. SENE, *Information externalities in the credit market and the spell of credit rationing,* Journal of Financial Intermediation, v. 30, pp. 61–70, **WP 980 (November 2014).**

ALESSANDRI P. and H. MUMTAZ, *Financial indicators and density forecasts for US output and inflation,* Review of Economic Dynamics, v. 24, pp. 66-78, **WP 977 (November 2014).**

BARBIERI G., C. ROSSETTI and P. SESTITO, *Teacher motivation and student learning,* Politica economica/Journal of Economic Policy, v. 33, 1, pp.59-72, **WP 761 (June 2010).**

BENTIVOGLI C. and M. LITTERIO, *Foreign ownership and performance: evidence from a panel of Italian firms,* International Journal of the Economics of Business, v. 24, 3, pp. 251-273, **WP 1085 (October 2016).**

BRONZINI R. and A. D'IGNAZIO, *Bank internationalisation and firm exports: evidence from matched firm-bank data,* Review of International Economics, v. 25, 3, pp. 476-499 **WP 1055 (March 2016).**

BRUCHE M. and A. SEGURA, *Debt maturity and the liquidity of secondary debt markets,* Journal of Financial Economics, v. 124, 3, pp. 599-613, **WP 1049 (January 2016).**

BURLON L., *Public expenditure distribution, voting, and growth,* Journal of Public Economic Theory,, v. 19, 4, pp. 789–810, **WP 961 (April 2014).**

BURLON L., A. GERALI, A. NOTARPIETRO and M. PISANI, *Macroeconomic effectiveness of non-standard monetary policy and early exit. a model-based evaluation,* International Finance, v. 20, 2, pp.155-173, **WP 1074 (July 2016).**

BUSETTI F., *Quantile aggregation of density forecasts,* Oxford Bulletin of Economics and Statistics, v. 79, 4, pp. 495-512, **WP 979 (November 2014).**

CESARONI T. and S. IEZZI, *The predictive content of business survey indicators: evidence from SIGE,* Journal of Business Cycle Research, v.13, 1, pp 75–104, **WP 1031 (October 2015).**

CONTI P., D. MARELLA and A. NERI, *Statistical matching and uncertainty analysis in combining household income and expenditure data,* Statistical Methods & Applications, v. 26, 3, pp 485–505, **WP 1018 (July 2015).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, v. 49, pp. 74-83, **WP 787 (January 2011).**

D'AMURI F. and J. MARCUCCI, *The predictive power of google searches in forecasting unemployment,* International Journal of Forecasting, v. 33, 4, pp. 801-816, **WP 891 (November 2012).**

DE BLASIO G. and S. POY, *The impact of local minimum wages on employment: evidence from Italy in the 1950s,* Journal of Regional Science, v. 57, 1, pp. 48-74, **WP 953 (March 2014).**

DEL GIOVANE P., A. NOBILI and F. M. SIGNORETTI, *Assessing the sources of credit supply tightening: was the sovereign debt crisis different from Lehman?,* International Journal of Central Banking, v. 13, 2, pp. 197-234, **WP 942 (November 2013).**

DEL PRETE S., M. PAGNINI, P. ROSSI and V. VACCA, *Lending organization and credit supply during the 2008–2009 crisis,* Economic Notes, v. 46, 2, pp. 207–236, **WP 1108 (April 2017).**

DELLE MONACHE D. and I. PETRELLA, *Adaptive models and heavy tails with an application to inflation forecasting,* International Journal of Forecasting, v. 33, 2, pp. 482-501, **WP 1052 (March 2016).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, v. 40, 10, pp. 2078-2096, **WP 877 (September 2012).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court,* Journal of Economic Geography, v. 17, 6, pp. 1251-1282, **WP 898 (January 2013).**

LOBERTO M. and C. PERRICONE, *Does trend inflation make a difference?,* Economic Modelling, v. 61, pp. 351–375, **WP 1033 (October 2015).**

MANCINI A.L., C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading,* Review of Economics of the Household, v. 15, 3, pp 965–993, **WP No. 958 (April 2014).**

MEEKS R., B. NELSON and P. ALESSANDRI, *Shadow banks and macroeconomic instability,* Journal of Money, Credit and Banking, v. 49, 7, pp. 1483–1516, **WP 939 (November 2013).**

MICUCCI G. and P. ROSSI, *Debt restructuring and the role of banks' organizational structure and lending technologies,* Journal of Financial Services Research, v. 51, 3, pp 339–361, **WP 763 (June 2010).**

MOCETTI S., M. PAGNINI and E. SETTE, *Information technology and banking organization,* Journal of Journal of Financial Services Research, v. 51, pp. 313-338, **WP 752 (March 2010).**

MOCETTI S. and E. VIVIANO, *Looking behind mortgage delinquencies,* Journal of Banking & Finance, v. 75, pp. 53-63, **WP 999 (January 2015).**

NOBILI A. and F. ZOLLINO, *A structural model for the housing and credit market in Italy,* Journal of Housing Economics, v. 36, pp. 73-87, **WP 887 (October 2012).**

PALAZZO F., *Search costs and the severity of adverse selection,* Research in Economics, v. 71, 1, pp. 171-197, **WP 1073 (July 2016).**

PATACCHINI E. and E. RAINONE, *Social ties and the demand for financial services,* Journal of Financial Services Research, v. 52, 1–2, pp 35–88, **WP 1115 (June 2017).**

PATACCHINI E., E. RAINONE and Y. ZENOU, *Heterogeneous peer effects in education,* Journal of Economic Behavior & Organization, v. 134, pp. 190–227, **WP 1048 (January 2016).**

SBRANA G., A. SILVESTRINI and F. VENDITTI, *Short-term inflation forecasting: the M.E.T.A. approach,* International Journal of Forecasting, v. 33, 4, pp. 1065-1081, **WP 1016 (June 2015).**

SEGURA A. and J. SUAREZ, *How excessive is banks' maturity transformation?,* Review of Financial Studies, v. 30, 10, pp. 3538–3580, **WP 1065 (April 2016).**

VACCA V., *An unexpected crisis? Looking at pricing effectiveness of heterogeneous banks,* Economic Notes, v. 46, 2, pp. 171–206, **WP 814 (July 2011).**

VERGARA CAFFARELI F., *One-way flow networks with decreasing returns to linking,* Dynamic Games and Applications, v. 7, 2, pp. 323-345, **WP 734 (November 2009).**

ZAGHINI A., *A Tale of fragmentation: corporate funding in the euro-area bond market,* International Review of Financial Analysis, v. 49, pp. 59-68, **WP 1104 (February 2017).**

2018

ACCETTURO A., V. DI GIACINTO, G. MICUCCI and M. PAGNINI, *Geography, productivity and trade: does selection explain why some locations are more productive than others?,* Journal of Regional Science, v. 58, 5, pp. 949-979, **WP 910 (April 2013).**

ADAMOPOULOU A. and E. KAYA, *Young adults living with their parents and the influence of peers,* Oxford Bulletin of Economics and Statistics,v. 80, pp. 689-713, **WP 1038 (November 2015).**

ANDINI M., E. CIANI, G. DE BLASIO, A. D'IGNAZIO and V. SILVESTRINI, *Targeting with machine learning: an application to a tax rebate program in Italy,* Journal of Economic Behavior & Organization, v. 156, pp. 86-102, **WP 1158 (December 2017).**

BARONE G., G. DE BLASIO and S. MOCETTI, *The real effects of credit crunch in the great recession: evidence from Italian provinces,* Regional Science and Urban Economics, v. 70, pp. 352-59, **WP 1057 (March 2016).**

BELOTTI F. and G. ILARDI *Consistent inference in fixed-effects stochastic frontier models,* Journal of Econometrics, v. 202, 2, pp. 161-177, **WP 1147 (October 2017).**

BERTON F., S. MOCETTI, A. PRESBITERO and M. RICHIARDI, *Banks, firms, and jobs,* Review of Financial Studies, v.31, 6, pp. 2113-2156, **WP 1097 (February 2017).**

BOFONDI M., L. CARPINELLI and E. SETTE, *Credit supply during a sovereign debt crisis,* Journal of the European Economic Association, v.16, 3, pp. 696-729, **WP 909 (April 2013).**

BOKAN N., A. GERALI, S. GOMES, P. JACQUINOT and M. PISANI, *EAGLE-FLI: a macroeconomic model of banking and financial interdependence in the euro area,* Economic Modelling, v. 69, C, pp. 249-280, **WP 1064 (April 2016).**

BRILLI Y. and M. TONELLO, *Does increasing compulsory education reduce or displace adolescent crime? New evidence from administrative and victimization data,* CESifo Economic Studies, v. 64, 1, pp. 15–4, **WP 1008 (April 2015).**

BUONO I. and S. FORMAI *The heterogeneous response of domestic sales and exports to bank credit shocks,* Journal of International Economics, v. 113, pp. 55-73, **WP 1066 (March 2018).**

BURLON L., A. GERALI, A. NOTARPIETRO and M. PISANI, *Non-standard monetary policy, asset prices and macroprudential policy in a monetary union,* Journal of International Money and Finance, v. 88, pp. 25-53, **WP 1089 (October 2016).**

CARTA F. and M. DE PHLIPPIS, *You've Come a long way, baby. Husbands' commuting time and family labour supply,* Regional Science and Urban Economics, v. 69, pp. 25-37, **WP 1003 (March 2015).**

CARTA F. and L. RIZZICA, *Early kindergarten, maternal labor supply and children's outcomes: evidence from Italy,* Journal of Public Economics, v. 158, pp. 79-102, **WP 1030 (October 2015).**

CASIRAGHI M., E. GAIOTTI, L. RODANO and A. SECCHI, *A "Reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households,* Journal of International Money and Finance, v. 85, pp. 215-235, **WP 1077 (July 2016).**

CECCHETTI S., F. NATOLI and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, v. 14, 1, pp. 35-71, **WP 1025 (July 2015).**

CIANI E. and C. DEIANA, *No Free lunch, buddy: housing transfers and informal care later in life,* Review of Economics of the Household, v.16, 4, pp. 971-1001, **WP 1117 (June 2017).**

CIPRIANI M., A. GUARINO, G. GUAZZAROTTI, F. TAGLIATI and S. FISHER, *Informational contagion in the laboratory,* Review of Finance, v. 22, 3, pp. 877-904, **WP 1063 (April 2016).**

DE BLASIO G, S. DE MITRI, S. D'IGNAZIO, P. FINALDI RUSSO and L. STOPPANI, *Public guarantees to SME borrowing. A RDD evaluation,* Journal of Banking & Finance, v. 96, pp. 73-86, **WP 1111 (April 2017).**

GERALI A., A. LOCARNO, A. NOTARPIETRO and M. PISANI, *The sovereign crisis and Italy's potential output,* Journal of Policy Modeling, v. 40, 2, pp. 418-433, **WP 1010 (June 2015).**

LIBERATI D., *An estimated DSGE model with search and matching frictions in the credit market,* International Journal of Monetary Economics and Finance (IJMEF), v. 11, 6, pp. 567-617, **WP 986 (November 2014).**

LINARELLO A., *Direct and indirect effects of trade liberalization: evidence from Chile,* Journal of Development Economics, v. 134, pp. 160-175, **WP 994 (December 2014).**

NUCCI F. and M. RIGGI, *Labor force participation, wage rigidities, and inflation,* Journal of Macroeconomics, v. 55, 3 pp. 274-292, **WP 1054 (March 2016).**

RIGON M. and F. ZANETTI, *Optimal monetary policy and fiscal policy interaction in a non_ricardian economy,* International Journal of Central Banking, v. 14 3, pp. 389-436, **WP 1155 (December 2017).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, v. 22, 2, pp. 661-697, **WP 1100 (February 2017).**

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).**

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).**

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).**

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).**

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 985 (July 2016).**

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).**

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).**

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).**

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).**

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).**

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).**

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).**

FORTHCOMING

ALBANESE G., G. DE BLASIO and P. SESTITO, *Trust, risk and time preferences: evidence from survey data,* International Review of Economics, **WP 911 (April 2013).**

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

ARDUINI T., E. PATACCHINI and E. RAINONE, *Treatment effects with heterogeneous externalities,* Journal of Business & Economic Statistics, **WP 974 (October 2014).**

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, **WP 1131 (September 2017).**

BELOTTI F. and G. ILARDI, *Consistent inference in fixed-effects stochastic frontier models,* Journal of Econometrics, **WP 1147 (October 2017).**

CIANI E. and G. DE BLASIO, *European structural funds during the crisis: evidence from Southern Italy,* IZA Journal of Labor Policy, **WP 1029 (October 2015).**

COIBION O., Y. GORODNICHENKO and T. ROPELE, *Inflation expectations and firms' decisions: new causal evidence,* Quarterly Journal of Economics, **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, **WP 1179 (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).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, **WP 787 (January 2011).**

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

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

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, **WP 877 (September 2012).**

FERRERO G., M. GROSS and S. NERI, *On secular stagnation and low interest rates: demography matters,* International Finance, **WP 1137 (September 2017).**

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

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court,* Journal of Economic Geography, **WP 898 (January 2013).**

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

LOSCHIAVO D., *Household debt and income inequality: evidence from italian survey data,* Review of Income and Wealth, **WP 1095 (January 2017).**

NATOLI F. and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, **WP 1025 (July 2015).**

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

PEREDA FERNANDEZ S., *Teachers and cheaters. Just an anagram?,* Journal of Human Capital, **WP 1047 (January 2016).**

RAINONE E., *The network nature of otc interest rates,* Journal of Financial Markets, **WP 1022 (July 2015).**

RIZZICA L., *Raising aspirations and higher education. evidence from the UK's widening participation policy,* Journal of Labor Economics, **WP 1188 (September 2018).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, **WP 1100 (February 2017).**