BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

Targeting policy-compliers with machine learning: an application to a tax rebate programme in Italy

by Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio and Viola Salvestrini

December 2017

Number 1158

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Targeting policy-compliers with machine learning: an application to a tax rebate programme in Italy

by Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio and Viola Salvestrini

Number 1158 - December 2017

The purpose of the Temi di discussione *series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.*

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

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

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

TARGETING POLICY-COMPLIERS WITH MACHINE LEARNING: AN APPLICATION TO A TAX REBATE PROGRAMME IN ITALY

by Monica Andini*, Emanuele Ciani*, Guido de Blasio*,
Alessio D'Ignazio* and Viola Salvestrini**

Abstract

Machine Learning (ML) can be a powerful tool to inform policy decisions. Those who are treated under a programme might have different propensities to put into practice the behaviour that the policymaker wants to incentivize. ML algorithms can be used to predict the policy-compliers; that is, those who are most likely to behave in the way desired by the policymaker. When the design of the programme is tailored to target the policy-compliers, the overall effectiveness of the policy is increased. This paper proposes an application of ML targeting that uses the massive tax rebate scheme introduced in Italy in 2014.

JEL Classification: C5, H3.
Keywords: machine learning, prediction, programme evaluation, fiscal stimulus.

Contents

* Bank of Italy, Directorate General for Economics, Statistics and Research, Structural Economic Analysis Directorate, Via Nazionale 91, 00184 Rome, Italy.

** London School of Economics and Political Science, Department of Economics, Houghton Street, London WC2A2AE, United Kingdom.

1. Introduction[1]

Machine Learning (ML) algorithms have been developed in the computer science and statistical literature. Differently from the econometrics literature, which usually points towards obtaining estimators that are unbiased or asymptotically consistent, the focus of ML algorithms is to minimize the out-of-sample prediction error (Athey and Imbens, 2017; Mullainathan and Spiess, 2017), which may require to tradeoff bias for variance. In this respect, ML algorithms do not propose new solutions to the fundamental identification issues when dealing with causal effects.[2] Nevertheless, the success of any public intervention depends on the actual implementation, which often requires a decision about who to target. This decision is a "prediction policy problem", a term coined by Kleinberg et al. (2015). The basic idea is that those who are treated under a programme might have different propensities to put in practice the behaviour that the policymaker wants to incentivize, or have different payoffs from a given treatment. In this framework, ML algorithms can help targeting the programme towards the policy-compliers, that is those who are most likely to behave in the desired way, or have the highest payoff, so to increase the overall effectiveness. The potential of ML predictions for policy decisions cannot go unnoticed. Evaluations might deliver a rigorous assessment of whether a programme worked or not, but are often less useful when it comes to advising the policymakers on how to ensure that a given programme will work. In particular, even if evaluation studies often propose rich heterogeneity analyses of the policy effect in different sub-groups, they rarely advise on how to build a better targeting rule. In this respect, ML algorithms provide us with some powerful techniques to predict which individuals are more likely to benefit from the policy using all the available information.

Early applications of ML targeting include: (i) predicting the riskiest patients for which a joint replacement would be futile (Kleinberg et al., 2015); (ii) improving over judges' decision on whether to detain or release arrestees as they await adjudication of their case (Kleinberg et al., 2017); (iii) targeting restaurant hygiene inspections (Kang et al., 2013); (iv) predicting highest risk youth for anti-violence interventions (Chandler et al., 2011); (v) predicting the effectiveness of teachers in terms of value added (Rockoff et al., 2011); (vi) hiring police officers who will not behave violently, as well as promoting the best teachers only (Chalfin et al., 2016); (vii) improve poverty targeting (McBride and Nichols, 2015). We extend this literature by focusing on a tax

[1] This paper was partly written while Viola Salvestrini was an intern at the Structural Economic Analysis Directorate of the Bank of Italy. The authors would like to thank Audinga Baltrunaite, Stefano Gagliarducci, Christian Hansen, Matthew Harding, Giuseppe Ilardi, Andrea Neri, Enrico Rettore, Paolo Sestito and the participants at the *IX ESPAnet Conference* (Macerata, September 2016), the *First Dondena Workshop on Public Policy* (Milan, November 2016), the *Bank of Italy Seminar Series* (Rome, November 2016), the *Royal Economic Society Annual Conference* (Bristol, April 2017), the *Bank of Italy Internal Workshop on "Big Data"* (Rome, June 2017) for suggestions and comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the institutions they are affiliated with.

[2] ML algorithms can, however, improve the predictive content of standard identification strategies. Varian (2014) suggests that counterfactuals could be better estimated by using ML algorithms instead of OLS regressions. Chernozhukov et al. (2017) discuss how to use a specific algorithm (the LASSO) to select controls in identification strategies based on the selection on observables assumption, or to choose instruments in an IV setting. In this paper we focus on the use of ML for targeting.

rebate scheme introduced in Italy in 2014 with Decree Law 66/2014. The programme[3] has been a centerpiece of the Italian government policy efforts to counterbalance the negative consequences of the Great Recession. The scheme was financially considerable: a tax credit was provided to all employees whose annual income ranged between 8,145 and 26,000 euro. According to government estimates (Ministry of Economics and Finance, 2015), it entailed a transfer of almost 7 billion euro in 2014, equivalent to 0.4% of Italian GDP.

The effect of the scheme on a given outcome depends on the allocation rule. In this work, we consider the hypothetical situation in which the only purpose of the policy was to increase consumption. If the bonus already targets the recipients that would have benefited the most in terms of increased consumption, then the policy effect is maximized, otherwise there is room for improvement. Our exercise is conducted on data from the Bank of Italy's Survey on Household Income and Wealth (SHIW). Descriptive evidence suggests that those households who received the bonus in 2014 belong to different points of the household income distribution; not all of them can be classified as needy. In line with previous results about the effectiveness of the policy (see below), we motivate our targeting exercise by showing that having difficulties in making ends meet, as reported by the household head, is associated with a larger impact of the bonus on consumption. We therefore use such a variable to proxy the consumption constrained status. Then, we turn to our core "prediction policy problem". We assume to be in the position of the Italian government at the beginning of 2014, when the scheme was designed. We use the SHIW waves (2010 and 2012) available at that time to build a prediction model to identify the households most likely to be consumption constrained on the basis of observable variables. We also limit the information set to variables that are not sensitive to privacy issues or ethical concerns. Consistently with the current policy, we maintain the focus on employees, hence we restrict our analysis and our proposed alternative allocation rule to households with at least one employee.

Our analysis is mostly based on decision tree algorithms (see Hastie et al., 2009). We find them to be ideal for our targeting purposes because they provide an assignment mechanism that can be transparently communicated to the general public by an accountable policymaker. However, we also present, for the sake of comparison, evidence resulting from non-ML methods, linear probability models, and black-box ML routines, such as clustering algorithms and random forest. We show that the gains obtained by using the targeting rule given by the decision tree algorithm can be substantial: our results suggest that 29% of the actual expenditure (about 2 billion euro, yearly) has been allocated to recipients that are not identified as the best targets. Very importantly, the decision tree provides us with an allocation rule that involves few variables related to income and financial wealth and it is easily interpretable. In principle, the ML selected variables may be used for an allocation rule of a generic measure that aims at helping the needy households. Nevertheless, if the policymaker has general welfare purposes, as in the case of the measures to contrast poverty, a categorical intervention like the one we simulate would be less appropriate than a universal minimum income measure (Sestito, 2016). The effects of such a

[3] Hereafter, the expressions *programme*, *tax credit*, *tax rebate*, and *bonus* are used interchangeably.

universal measure could be appropriately analyzed in a micro-simulation framework, which is not our tool.
The impact of the bonus on consumption is currently being studied by Neri et al. (2017), Gagliarducci and Guiso (2015) and Pinotti (2015). Neri et al. (2017) use a Difference-in-Differences approach with the longitudinal component of the SHIW survey, and suggest that households who received the bonus spent around 50-60% of it. Gagliarducci and Guiso (2015) use data from the Survey on Household Consumption of the National Statistical Office, matched with administrative data on labor income. They apply a Regression Discontinuity Design exploiting the fact that the bonus depends on labor income thresholds, and find a positive impact of the bonus on food and on mortgage payments, suggesting that the entire bonus goes to consumption. However, using similar data, Pinotti (2015) argues that the effect on total consumption cannot be estimated with precision, and it may even be zero. Differently from these papers, we do not directly focus on the ex-post impact of the bonus on consumption, but we rather study how the policy could be improved by changing the allocation rule. Our evidence on futile expenditures, however, provides boundaries for the magnitude of the ex-post average impact on the treated. For instance, if 29% of the recipients are unlikely to increase consumption out of the bonus, as our ML exercise suggests, then estimated average treatment effects close to unity are not consistent with our study and should be judged as impractical.
The remainder of the paper proceeds as follows. Section 2 provides the relevant details on the design and the implementation of the policy. Section 3 sketches the SHIW features and gathers information on how the tax credit is detected in the 2014 wave. Section 4 contains descriptive evidence on the effect of the income tax credit on consumption. Section 5 describes the empirical framework used for our targeting analysis and presents the results. Section 6 focuses on information requirements. Section 7 concludes.

2. The bonus: institutional details

During the second dip of the economic crisis, total consumption of Italian households dropped from nearly 970 billion euro in 2010 to 909 in 2013. While foreign demand kept supporting exports, the recession was prolonged by the stagnation of internal demand. A government crisis led to the appointment of a new Prime Minister in February 2014. One of the first announcements was the proposal of a new monetary transfer to households. This proposal was reiterated on several occasions during the early weeks of the new government, and formally announced to the press as a transfer on the 12th of March. The transfer was finally implemented with the Decree Law 66/2014, which was approved by the government on the 24th of April, and later ratified by the Parliament. There was no debate about this bonus earlier than the change of government in 2014, and therefore our analysis should not be affected by possible anticipation effects, as it focuses on the entire year 2014 without distinguishing between specific months.
The benefit was designed as a tax-credit. It targets employees and holders of similar income[4] with gross annual income between €8,145[5] and €26,000. In particular, the tax credit amounts at

[4] For instance, freelancers, priests, cooperative workers, recipients of unemployment or disability benefits, recipients of scholarships and unemployment insurance (*Cassa Integrazione*).

€640 per year if the gross annual income ranges between €8,145 and €24,000; for earnings between €24,000 and €26,000, the amount of the benefit is calculated as follows: $80 \times (26{,}000 - income)/2{,}000$. The tax credit is acknowledged automatically by the employer on the monthly salary paid without a specific request by the beneficiary, on the basis of the predicted annual income according to the current contract and as long as the gross (predicted) tax is greater than the tax deduction for employee income. An individual may receive the bonus on the basis of this predicted income, but will have to give it back if the actual income at the end of the year is outside the eligible range. This will happen at the moment of the annual tax return, which has to be submitted between April and the beginning of July 2015 for the 2014 annual income. Several factors contributed to determine the allocation rule.[6] First of all, the bonus was thought as a tax rebate: employees were therefore chosen as recipients since such tax credit could be automatically acknowledged by the withholding agent. This was clearly aimed at making the implementation of the programme much easier and faster. Following the same rationale, the "*incapienti*" (namely, those who earn so little that they do not pay taxes) were left out, so to avoid the withholding agent to pay the transfer out of pocket. Both mechanisms were introduced to speed up the allocation of the bonus, but they potentially reduced the ability to target households that are consumption constrained. On top of this, the focus on individual rather than household income and the exclusion of the *incapienti* make the measure not suitable to contrast poverty (Sestito, 2016). As last point, the tax credit should also be seen as a measure of remodeling the tax rates on labor income. As an in-work-benefit, it may have positive effects on labor supply but, at the same time, the intensive margin of labor supply from those who earn between 24,000 and 26,000 euro may be discouraged by the very sharp phasing out (the rapid decline in tax credit leads, indeed, to a sharp rise in effective marginal rates).

One year after its introduction as a temporary bonus, the measure became permanent. From an ex-post evaluation point of view, the stabilization of the benefit may have contributed to increase the positive impact of the measure on consumption; it may have also been a stimulus to the participation in the labor market for people with a low potential wage (Signorini, 2014). The impact on consumption might therefore be different if one estimates it in more recent years. Nevertheless, in our analysis we put ourselves in the ex-ante situation, earlier than 2014, in which the government had to decide the allocation rule. We therefore use the pre-intervention data (i.e. before 2014) to estimate a targeting rule aimed at addressing those who are consumption constrained. We cannot do it on the post-intervention data (i.e. 2014 and later years) because they are influenced by the bonus receipt itself. One could instead use such data to look at the ex-post impact of the measure, which is not our purpose. Furthermore, at the moment of writing, SHIW data for years after 2014 are not yet available as the field work for SHIW 2016 is ongoing.

---

[5] This threshold is €8,145 for individuals who worked the entire year, but it might be lower for workers who have been employed for less than 365 days. Indeed, the bonus was granted under the condition of a net tax greater than zero, which could happen for incomes below the tax area if the employment spell was less than 365 days.

[6] For a full description of the policy from a public finance point of view, and an assessment of its redistributive capacity, see Baldini et al. (2015).

3. The bonus in the SHIW

The Survey on Household Income and Wealth (SHIW) is a statistical survey conducted on a biennial basis by the Bank of Italy, to gather information on the economic behaviour of the Italian families at the microeconomic level. The SHIW (www.bancaditalia.it/statistiche) collects the following information: characteristics of the household and of its members (number of income earners, gender, age, education, job status, and characteristics of the dwelling); income (wage and salaries, income from self-employment, pensions and other financial transfers, income from financial assets and real estates); consumption and saving (food consumption, expenses for housing, health, insurance, spending on durable goods, and household saving); wealth in terms of real estate, financial assets, liabilities.

The 2014 wave contains some specific questions on the income tax credit: households were asked if they received the bonus, how many beneficiaries were present within the family and the overall amount received. The overall size of the sample for the 2014 wave is 8,156 households. Since a necessary condition to be eligible for the bonus is to be employed,[7] we consider only households with at least one employee, for a total of 3,646 observations. Since the survey data we rely on were collected between January and July 2015, that is when the incomes referring to the previous year were secured, we can reasonably assume that individuals knew at the moment of the interview if they were entitled to the bonus or they would have to give it back.

We also make use of previous waves to create the prediction model aimed at improving the targeting of the bonus. In order to keep information as homogeneous as possible, we only use the other two waves collected after the beginning of the recession (that is, 2010 and 2012). Although the dataset provides us with a large set of covariates, the actual income variable that determines the eligibility for the bonus is not observed. The reasons for this are two-fold. Firstly, the survey collects only income net of taxes and social contributions, while the bonus was determined on the basis of gross income. It is not possible to simply invert the tax formula, because it depends on a full set of deductions that are household-specific (and not reported in the survey). Secondly, the bonus was assigned according to the predicted income, which may differ from the actual annual income.

The SHIW collects both annual expenditure on durable goods and the average monthly non-durable consumption during the year. Households are also asked to report average monthly expenditure on food consumed at home and, separately, on food consumed outside home. The main limitation of these questions is that they are retrospective. The Survey on Household Consumption, carried out by the National Statistical Institute and used by Gagliarducci and Guiso (2015) and Pinotti (2015), is based instead on diaries filled in by a sample of households, who are asked to report detailed expenditures during a single month. Although this reduces the risk of misreporting, the fact that consumption refers to a single month of the year increases the volatility of the measure, thus reducing the ability to detect the effect of the bonus. In this respect, the nature of the consumption variables in SHIW, which are referred to the whole year (although reported as totals or monthly averages), may be more adequate for our purpose.

[7] We do not consider the other categories of bonus recipients, which however represent a small fraction of the beneficiaries.

4. Descriptive evidence

Using the 2014 SHIW wave, we first show that the households who received the bonus are not always the needy ones, and this signals serious problems of targeting. Figure 1 presents the distribution of the sample by bonus recipients: nearly 40% of the sample received the bonus. Figure 2 reports the difficulty that households face making ends meet by treatment status: there does not seem to be a relevant difference in the distribution of the sample between beneficiaries and non-beneficiaries. Figure 3 presents the distribution of beneficiaries and non-beneficiaries who reported to be liquidity-constrained, i.e. the household was at least partially rejected a request for a mortgage, or would have liked to apply for it but had not because of the assumption to be rejected. Among the recipients nearly 6% reported to be liquidity-constrained, against 2.6% among the non-recipients. Finally, Figure 4 reports the distribution of bonus recipients and non-recipients by income quartiles. Among the recipients, nearly 16% are in the top income quartile. This is due to the fact that we are considering household average income rather than individual income.

[Figure 1 to 4]

We can recover an estimate of the impact of the bonus on consumption by relying on selection on observables. Let $ammbonus_i$ be the amount in euro of the bonus received by household i. The main outcome of interest is consumption c_i (measured in euro) while x_i is a vector of household characteristics, including average income within the household, household size (and its square) and several other characteristics, plus a constant. We then estimate

$$c_i = \delta \; ammbonus_i + \beta_x x_i + \varepsilon_i \qquad (1)$$
$$E[\varepsilon_i \mid ammbonus_i, x_i] = 0 \qquad (2)$$

Given that both c_i and $ammbonus_i$ are in euro, δ can be interpreted as the fraction of the bonus spent in consumption, conditional on x_i. In principle, the assumption of selection on observables is appropriate for the policy under scrutiny because the bonus was automatically distributed to all eligible individuals on the basis of their tax-relevant information and therefore no self-selection occurred. For the reasons outlined in Section 3, our dataset does not provide information on the precise individual income variables involved in the allocation rule. Nevertheless, we observe a complete set of variables related to it and that, at the same time, have an impact on consumption (usually modeled as well as measured at the household level). It should be noted that, in terms of these observable characteristics, there is a large overlapping between the group of households who received the bonus and the others, and therefore we can compare households who are quite similar but differ by bonus receipt. Notice also that having an higher external validity, as it is the case for selection on observables, is desirable for our targeting analysis, which elaborates on the heterogeneity of the effect. More internally valid estimates - in particular, those obtained with a Regression Discontinuity Design - would provide us with "local" estimates, making it more difficult to estimate heterogeneity and to make predictions on the overall sample. Having said that, our estimates, as shown later in this section, are pretty much in line with those provided by Neri et al. (2017), which adopt a more rigorous identification framework.

We start by estimating the effect of the bonus on average monthly total non-durable consumption (Table 1; Table A.1 provides a description of the variables used in the baseline specifications). We focus only on non-durable consumption because, as discussed by Neri et al. (2017), durable consumption tends to be more volatile and therefore it is difficult to detect the impact of the bonus. The total consumption of non-durables excludes rents (imputed or actual), mortgages, and in-kind benefits from the employer. To begin with, the average monthly amount of the bonus perceived and household annual disposable income (net of the bonus) are considered as regressors (column (1)). Since current consumption depends - possibly not linearly - on the size of the family, we also control for the number of components and its square (column (2)). Then, we subsequently add a rich set of demographic characteristics such as age, education, gender, marital status (columns (3)-(6)). In column (7), regional fixed effects are included to capture specific factors that might affect all the people residing in the same area. Then, in line with the current debate on the destination of the bonus (Gagliarducci and Guiso, 2015; Pinotti, 2015), we repeat the above estimation on monthly spending for food eaten at home (Table 2). While the effect of the bonus on total consumption is generally not statistically significant and oscillates in magnitude, when considering food expenditure the effect of the bonus is quite consistent throughout the specifications: in particular, for every additional euro received as bonus, roughly 31.5 cents are spent in food consumed at home. This is in line, although slightly larger, with the results by Neri et al. (2017). In a nutshell, we find evidence that the bonus has an effect on food consumption, it is statistically significant even when we introduce additional controls and its economic magnitude is stable across specifications. The effect of the tax rebate on total consumption is not easily detectable in our data because total consumption is likely to be more volatile than food expenditure. In the remaining of the paper, we will mainly focus on food consumption in estimating the inefficiency in the current allocation of the bonus.

[Table 1]

[Table 2]

Using the specification in column (7) of the above tables, we then investigate whether the bonus has a heterogeneous effect on consumption according to different definitions of "needy" households (Table 3). For the estimation in Panel A and B, we construct two indicators taking value 1 for households who report facing some difficulty to make ends meet given the household's income. In particular, in Panel A difficulty =1 if the household makes ends meet with great difficulty, with difficulty or with some difficulty. Noting that a considerable number of households reported to face some difficulty getting through the month, in Panel B we slightly modify the definition of difficulty so to include only households making ends meet with great difficulty or with difficulty. In both specifications, total consumption does not seem to significantly react to the tax rebate. Nevertheless, the bonus seems to increase food consumption for all households and the effect is larger and statistically significant for families facing difficulties. In Panel C, we estimate the effect of the bonus among households that report to be liquidity-constrained and households that do not. Differently from the difficulty in making ends meet, this indicator provides an assessment of the overall households' wealth, rather than income only. The bonus has a significant and positive effect only on food consumption for both constrained and unconstrained households, and such increase is larger for constrained families. In our data, being consumption-constrained is a relevant phenomenon, characterizing about 60% of households; on the other hand, only 4% of households report themselves to be liquidity-constrained.

[Table 3]

In the remaining of the paper, our preferred indicator of needy households is the difficulty to make ends meet. First of all, this indicator comes from a questionnaire variable available for each wave in the same manner. SHIW 2010 and 2012 contain hypothetical questions about the propensity to consume out of an income shock that are possibly closer to our scope and that have been used also by Jappelli and Pistaferri (2014). However, the question changes significantly between the two waves and it is not available in 2014, making impossible to estimate the heterogeneity in the effect of the bonus according to the different answers. Secondly, we prefer to work with a reasonable number of observations of needy households. This excludes the liquidity constrained indicator that identifies few households as needy, as shown in Panel C of Table 3. Nevertheless, about 90% of the liquidity constrained households also report to have difficulties in making ends meet. Thirdly, we prefer to use a variable that allows us to estimate an effect of the bonus on consumption which is in line with the other papers focused on estimating the effect of the bonus. As one can see from Table 3, the results in Panel A obtained using our preferred indicator are the closest to the ones obtained by Neri et al. (2017). Fourthly, we prefer to use the unrestricted version of the difficulty to make ends meet indicator as the restricted one likely brings us to identify the poorest households, which is not our purpose ex ante.

5. Targeting analysis

5.1 *Variable selection*

For the targeting analysis, we want to be able to identify in the 2014 sample those households that are more likely to be needy. In this way, we can evaluate the efficiency of the current allocation and suggest possible alternatives. Although we observe needy households also in the 2014 sample, it is not useful for targeting: we need to rely on information available prior to the start of the policy, which could have been used by the policymaker. We therefore focus on a pooled dataset of 2010 and 2012 waves and estimate models that allow us to predict the needy status on the basis of a set of observable covariates.[8] We only maintain one rule of the current

[8] The sample includes a longitudinal component, but we ignore it because the sample size would be too small for a reliable targeting analysis. Therefore, i univocally identifies a household-year pair.

policy, that is the focus on employees. Hence we select only households with at least one employee.
In order to choose the covariates set, we consider variables that are recorded in both the pooled 2010-12 dataset and the 2014 one, so as to predict the needy status in 2014 using the prediction model estimated on the 2010-12 dataset. Among those variables, we select only those that are observable by the policymaker and not sensitive to privacy issues or ethical concerns so as to end up with a feasible targeting rule. That is, an assignment mechanism that can be transparently communicated to the general public by an accountable policymaker. Finally, we dismiss all the variables that are excluded for collinearity reasons by running a simple regression of our needy status proxy on the covariates set.[9] The complete list of variables used for prediction can be found in Table A.2. They essentially refer to household income, wealth, and demographic characteristics.
5.2 *Prediction*
ML techniques, which rely on highly flexible functional forms, aim at reaching out-of-sample predictive accuracy. In general, the optimal degree of flexibility of the model function f is the result of the minimization of a theoretical loss function, which can be usually broken down into three parts: variance, bias and irreducible error. The variance part pertains to how much our model function f will change if we estimate it on a different set of observations, while the bias is the error that we make when we approximate a complex function with a more simple one; the irreducible error cannot be minimized. The following trade-off always holds: as we allow for more flexibility in the function f, the bias decreases while the variance increases. For instance, a very simple function f will likely lead to high bias and low variance (and under-fitting of the estimation data), while a complex function will likely result in high variance and low bias (and over-fitting the estimation data). In order to achieve both low bias and low variance, and hence obtain a good out-of-sample performance, each ML algorithm comes with a regularization of the complexity level. In particular, ML algorithms rely on empirical tuning, where the model performance is evaluated over a small portion (randomly chosen) of the dataset. This procedure is repeated several times and the regularization parameter chosen is the one characterized by the best performance on average (cross validation). Given that the main purpose is out-of-sample prediction, we estimate and tune the models on a training subsample, composed of a randomly selected 2/3 of the 2010-12 pooled sample. The remaining 1/3 of such dataset constitutes the testing subsample.
Our main ML algorithm is the *decision tree* (Hastie et al., 2009; James et al., 2014), which allows in principle to reach a perfect in-sample fit by adding more and more leaves, while the regularization is made by pruning the tree. Decision trees are particularly appropriate for applications in which the assignment mechanism needs to be transparent; for instance, when the results need to be shared in order to facilitate decision making (Lantz, 2013). As it will be clear in Section 5.3, the output of a decision-tree algorithm can be easily described in a graph. On a more technical ground, decision trees are non-parametric learning algorithms that perform quite

[9] This step allows us to select only one variable among a set of variables that represent the same thing: for instance, age and year of birth; education represented with different levels of accuracy, and so on.

well in the case of non-linear relationships, being also robust to the presence of outliers. The algorithm divides data into progressively smaller subsets to identify patterns that can be used for predicting a specific output. In our case, the algorithm creates a decision rule which partitions the observations according to their (binary) needy status on the basis of the values of the observable covariates (z_i). Non-linearities and interactions are captured by the sequence of splits. Following a top-down approach, at each step the algorithm selects a variable z_{gi} from z_i and splits the observations into two groups according to a threshold z_g (or according to a subset of values in case of a multinomial discrete variable). Both the variable used to split and the threshold are chosen to obtain the largest possible reduction in heterogeneity (impurity) of the variable to be predicted (Siroky, 2009). In the decision tree algorithm that we use,[10] the degree of impurity at each node (leaves) is measured using a heterogeneity index. The algorithm then proceeds to the next step by further splitting the sub-samples at each terminal node. It stops when the degree of impurity of a terminal node is as low as possible. A high number of levels in a tree is likely to overfit the data. This could lead to a model which performs very well in the training sample, but gives highly imprecise predictions out-of-sample (Athey and Imbens, 2016; Lantz, 2013; Breiman et al., 1984). A solution to this problem is to reduce the complexity of the tree by setting a complexity parameter (cp) and use it to prune the tree. We choose the optimal cp by using a rule of thumb suggested in the literature (Hastie et al., 2009).[11]

We compare the findings obtained with the decision tree with those deriving from other ML algorithms, the k-Nearest Neighbours (k-NN) and random forest, as well as a standard Linear Probability Model (LPM). In the k-NN algorithm (Lantz, 2013), the trade-off between bias and variance is solved by choosing the optimal number of neighbours (i.e. the level of k). For each observation in the testing sample, the algorithm identifies the k closest observations from the training sample (the so-called nearest neighbours) and assigns a prediction on the basis of a majority rule, i.e. takes as prediction the most frequent outcome among those of the nearest neighbors. We chose the optimal number k of neighbours by using 10-fold cross-validation. The k-NN algorithm is the closest to the standard non-parametric analysis, and therefore it is useful for the reader who wants to compare the performance of less common ML algorithms with methods that an econometrician may be more familiar with. Random forest explores a richer set of possible models. Essentially, it estimates a large number of trees on a series of new samples generated by randomly drawing (with replication) from the original sample (i.e. bootstrapping), using for each tree only a randomly selected subset of the regressors. To obtain the final prediction for each observation, random forest takes the majority vote across the predictions generated by each tree. Intuitively, the algorithm works as a decision tree that moves around over lots of regressors. Although the random forest performance improves with strong and moderately important predictors, the algorithm is not free of the risk of averaging over noise as it may also select regressors that are highly correlated with predictors. Therefore, one should use random

[10] We use the R package "rpart" [https://cran.r-project.org/web/packages/rpart/rpart.pdf].

[11] First, the complexity parameter associated to the smallest cross-validation error (say $errmin$) is found. Then, the optimal cp is the one that has a cross-validated error which is the closest to $errmin + standard\_error(errmin)$. The rule of thumb leads to a simpler tree because the cross-validation error curve tends to be flat around its minimum, hence there is a small gain in picking exactly the minimum while there is a higher risk of over-fitting.

forest only if the number of regressors is really big, which is not our case. Following the work of Chandler et al. (2011), we also make use of a LPM prediction.[12] In order to make the latter more comparable with ML predictions, we include all the variables depicted in Table A.2, the squares and cubes of the continuous variables, plus all interactions between themselves and all interactions between them and the discrete covariates. In the case of LPM the prediction is continuous, so we consider the dummy needy having the value one if the predicted probability is larger than 0.5. k-NN, random forest and LPM are used essentially to probe the prediction quality of the decision tree. In our case, they cannot be considered as real alternatives, as we are looking for a transparent assignment mechanism.

5.3 *Empirical findings*

The decision tree leads to the assignment mechanism shown in Figure 5. It depends on few variables, essentially referring to household income and wealth. The targeted households would be: (a) those that have financial assets lower than 13,255 euro; among these ones, the needy are those that either perceive income lower than 36,040 euro yearly or those that earn more than 36,040 euro but the maximum income perceived within the household is lower than 34,500 euro; (b) those that have financial assets higher than 13,255 euro; among these ones, the needy are those that earn less than 52,591 euro yearly and have an income from financial assets lower than 432.9 euro together with a minimum income perceived within the household lower than 13,895 euro.[13] As for a comparison, using either LPM, random forest or k-NN to target households would be a much more challenging task. These methods do not select a subset of the variables, and therefore the actual allocation of the bonus would require acquiring a larger amount of information on each household. Furthermore: (i) these methods require cumbersome computations to obtain the actual index that is used for the allocation and (ii) they do not provide clear insights (or not at all, in the case of k-NN) on which characteristics of the household are pivotal in the selection rule.

[Figure 5]

Table 4 compares the performance of the three models in terms of correctly predicting the "needy" status. Notwithstanding its simplicity, the decision tree correctly identifies 74.1% of the observations, a share very close to that of its alternatives (respectively, 73%, 77% and 75% for the k-NN, the random forest and the LPM). Since we are using 2010-2012 information to predict 2014 needy households, we also investigate whether the association between the actual needy status and tree-selected predictors is stable. We run two separate LPM regressions for the 2010 and 2012 subsamples, using as dependent variable the dummy for difficulty in making ends meet and as covariates the variables selected by the tree. The relationship between the observables and

[12] The results are unchanged if we use a Probit model. We then decide to rely on a LPM since it is easier to interpret the coefficients.

[13] In principle, targeted households may also include the *incapienti* (see Section 2). Nevertheless, we cannot argue on the actual presence of *incapienti* among members of targeted households as the decision rule we suggest is based on household rather than individual income.

the needy status appears to be quite stable, as coefficients change only marginally. Results are presented in Table A.3.

[Table 4]

We proceed to estimate with the 2014 data the effect of the bonus between the households that, according to our decision-tree assignment, should have received the bonus, as they would be predicted to be needy, and those that should have not. Table 5 reports the results of the estimation. The effect of the bonus for food consumption is positive and significant for the households that would have been targeted with our assignment rule. The effect is instead neither statistically nor economically different from zero for households that received the bonus without being consumption constrained according to the decision tree rule. In particular, households predicted to be needy spend on average 36.9% of the bonus in food consumption. This share is very close to the one estimated by using 2014 data (see Table 2).[14]

[Table 5]

Table 6 provides the percentage of overlap between predicted status (i.e. being needy or not) and the receipt of the bonus. The overlap includes households that: (i) both receive the bonus and are predicted to be needy, and (ii) both do not receive the bonus and are predicted to be non-needy. This fraction is quite low, around 49%. This implies that several households received the bonus but would have not if the allocation rule was the one that we propose. Given that we find evidence of an impact on consumption only among those predicted to be needy, this implies that there were margins to improve the total effect.

[Table 6]

In order to capture this misallocation, we focus on a measure of spending inefficiency due to the actual allocation rule. As shown in Table 6, 70.9% of the households that receive the bonus are predicted to be needy by the decision tree algorithm. Our spending inefficiency measure refers to the remaining 29.1%. We look at the amount that was spent for the bonus recipients that the decision tree does not identify as needy households. The way we compute such a measure is as follows. Let A be the number of bonus recipients in our dataset, and B the subset of A made up of predicted needy households. The total expenditure for the tax rebate is given by

$$E_{total} = \sum_{i=1}^{A} ammbonus_i \qquad (3)$$

[14] One issue is that predictors and the needy status are both measured at the same time. In principle, one would predict the needy status with variables that have been already observed at the time the policy is implemented. Our data do not allow us to follow such a strategy. However, note that the selected predictors such as income and wealth are characterized by a high degree of persistence. In particular, we use the panel component of the dataset and regress each predictor measured in 2014 on its 2010-12 average value. Such an estimate is roughly 0.9 for the two main predictors (income and financial assets).

while the “efficient” expenditure (namely, the amount spent for the predicted consumption constrained households) is given by

$$E_{correct} = \sum_{i=1}^{B} ammbonus_i \qquad (4)$$

Therefore, the percentage of expenditure that has been allocated inefficiently can be computed as

$$\frac{E_{total} - E_{correct}}{E_{total}} \qquad (5)$$

This share turns out to be equal to 29% of the total expenditure. In order to maximize the coverage of the programme, this amount could be reallocated to those households that are predicted to be needy but did not receive the bonus. One possibility is to endow this group with a transfer which is set to be equal to the per capita transfer received by households belonging to B (i.e. roughly 57 euro). In such a case, keeping fixed the total public expenditure for this transfer, we could reach 30% of predicted needy households that did not receive the bonus. In this way, 60% of the households we predict as needy would be endowed with a bonus.

6. Data requirements

The decision tree rule is based on information at household level that, at least in principle, is observable by a policymaker. As a matter of example, the equivalent economic situation indicator (i.e., the so called “ISEE”) enables the policymakers to collect information on income and wealth at household level. We are aware that implementing the targeting rule we suggest may increase the costs of the policy in the short term because it would require, using the same example, to know the ISEE of all Italian households. However, the use of household-level information is also in line with other recent proposals to review some assistance benefits policies aiming at the use of eligibility criteria that approximate the ISEE or, more generally, the household economic condition. In short: data defined at the household level are going to be collected anyway to comply with a more efficient welfare system.

Note also that having only a subset of the (few) variables included in the decision tree assignment rule will deliver lower but still sizable benefits. Indeed, a useful feature of the decision tree algorithm is the possibility to compute the fraction of households that would be incorrectly identified as needy by observing only a subset of characteristics among those involved in the tree. For instance, let us assume that the policymaker can observe household financial assets and disposable income only. In this case, her decision rule to identify needy households could be based only on the financial assets and income thresholds given by the tree. In terms of Figure 5, needy households would be those that have financial assets lower than 13,255 and disposable income lower than 36,040 euro and non-needy households would be those that have financial assets at least equal to 13,255 and disposable income at least equal to 36,040 euro yearly. Clearly, these groups do not overlap with the groups of predicted needy and predicted non-needy households identified through the use of all the variables involved in the tree. Using a decision rule based on financial assets and disposable income only, 22% of the households would be allocated to a status that does not correspond to the one predicted by the use of all the variables (i.e., the entire tree). If the policymaker observes the maximum income perceived within the household too, and constructs a decision rule also based on this variable using the thresholds given by the tree, then the fraction of incorrectly allocated households decreases to 5%. Finally, the fraction of incorrectly identified needy households is obviously 0 in case both income from financial assets and minimum income perceived within the household are observable, because the decision rule now coincides with the entire tree.

As a last point, one may argue that people may react by manipulating the subset of endogenous variables selected by the tree when they are used by the policymaker to distribute the bonus. We believe manipulation is not likely in our context. In case of manipulation, the potential recipient should manipulate more than one single variable (say, both financial assets and income in the simplest case) and more than one single family member (given that we are considering mainly household level variables and that both the minimum and maximum income are, taken together with all the other variables, good predictors of the needy status). Manipulating one single variable, as the case of the actual measure, is instead more likely and even more realistic if it refers to individual features.

7. Conclusions

During economic downturns, well-designed programmes may contribute to the recovery. A key ingredient for an effective policy is an accurate targeting of beneficiaries, who should behave in the way the policymaker wants to incentivize. This ideal framework unlikely corresponds to the actual one because of a trade-off between ease and accuracy of the targeting rule. Machine Learning algorithms help addressing such a trade-off as they allow to target units that most likely behave in the desired way or to gain more from the policy.

In this paper, we focus on a massive tax rebate programme recently implemented in Italy. We assume that the only purpose of the policy was to increase consumption. We make use of ML techniques to identify the households that would have benefited the most from the programme in terms of increased consumption. To do so we use a decision tree and find that 29% of the actual expenditure has been allocated to recipients that are not the best target for this objective.

References

Athey, S. and Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27): 7355–7360.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2): 3–32.

Baldini M., Giarda E., Olivieri A., Pellegrino S., and Zanardi, A. (2015). Il "bonus" degli 80 euro: caratteristiche ed effetti redistributivi. *Rivista di Diritto Finanziario e Scienza delle Finanze*, 74(1): 3–22.

Breiman, L., Friedman, J., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Belmont: Wadsworth.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., and Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review: Papers & Proceedings*, 106(5): 124–127.

Chandler, D., Levitt, S. D., and List, J. A. (2011). Predicting and preventing shootings among at-risk youth. *American Economic Review: Papers & Proceedings*, 101(3): 288–292.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. (2017). Double/Debiased/Neyman machine learning of treatment effects. *American Economic Review: Papers & Proceedings*, 107(5): 261–265.

Gagliarducci, S. and Guiso, L. (2015). Gli 80 euro? Spesi al supermercato. *Lavoce.info*. Available at: http://www.lavoce.info/archives/36685/gli-80-euro-spesi-al-supermercato.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The elements of statistical learning – Data mining, inference, and prediction*. New York: Springer.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An introduction to statistical learning – With applications in R*. New York: Springer.

Jappelli, T. and Pistaferri, L. (2014). Fiscal policy and MPC heterogeneity. *American Economic Journal: Macroeconomics*, 6(4): 107–136.

Kang, J. S., Kuznetsova, P., Luca, M., and Choi, Y. (2013). Where not to eat? Improving public policy by predicting hygiene inspections using online reviews. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1443–1448. Available at: https://aclanthology.info/pdf/D/D13/D13-1150.pdf.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human decisions and machine predictions. NBER Working Paper No. 23180, February 2017, National Bureau of Economic Research.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review: Papers & Proceedings*, 105(5): 491–495.

Lantz, B. (2013). *Machine learning with R*. Birmingham: Packt Publishing Ltd.

McBride, L. and Nichols, A. (2015). Improved poverty targeting through machine learning: An application to the USAID poverty assessment tools. Unpublished manuscript. Available at: http://www.econthatmatters.com/wp-content/uploads/2015/01/improvedtargeting_21jan2015.pdf.

Ministry of Economics and Finance (2015). Statistiche sulle dichiarazioni fiscali. Analisi dei dati Irpef - Anno d'imposta 2014. Technical report. Available at: http://www1.finanze.gov.it/finanze2/analisi_stat/v_4_0_0/contenuti/analisi_dati_2014_irpef.pdf?d.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2): 87–106.

Neri, A., Rondinelli, C., and Scoccianti, F. (2017). Household spending out of a tax rebate: Italian "€80 tax bonus". Bank of Italy Occasional Paper No. 379, June 2017, Bank of Italy.

Pinotti, P. (2015). 80 euro rimasti nel portafoglio. *Lavoce.info*. Available at: http://www.lavoce.info/archives/37086/80-euro-rimasti-nel-portafoglio.

Rockoff, J. E., Jacob, B. A., Kane, T. J., and Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1): 43–74.

Sestito, P. (2016). Audizione preliminare sulla delega recante norme relative al contrasto della povertà, al riordino delle prestazioni e al sistema degli interventi e dei servizi sociali. Available at: https://www.bancaditalia.it/pubblicazioni/interventi-vari/int-var-2016/sestito-040416.PDF.

Signorini, L. F. (2014). Audizione preliminare all'esame dei documenti di bilancio per il triennio 2015-2017. Available at: https://www.bancaditalia.it/pubblicazioni/interventi-direttorio/int-dir-2014/signorini-03112014.pdf.

Siroky, D. S. (2009). Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys*, 3: 147–163.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2): 3–28.

Figure 1: Bonus Recipients

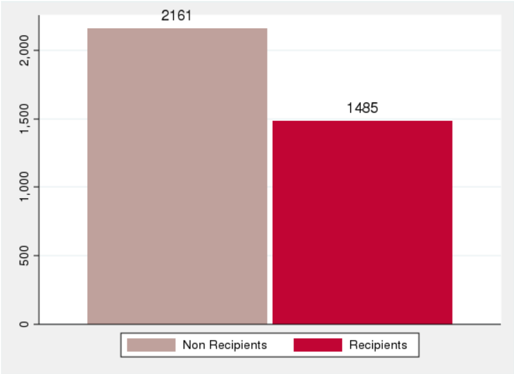


Figure 2: Difficulty making ends meet by treatment status

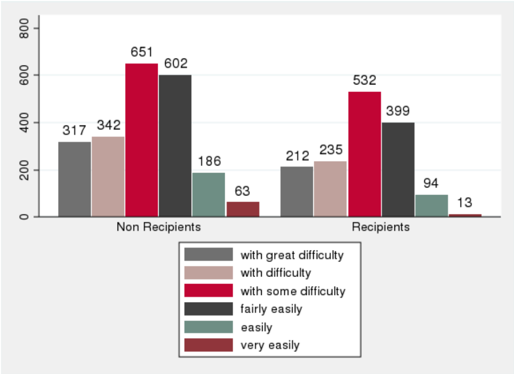


Figure 3: Liquidity constraints by treatment status

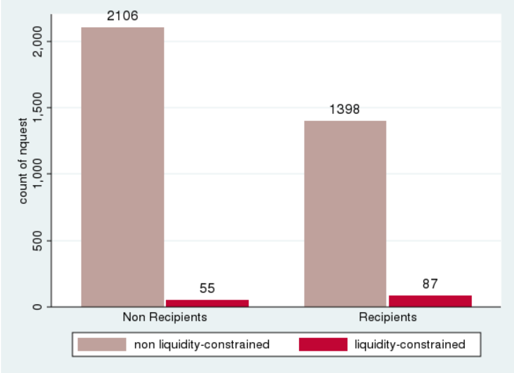


Figure 4: Income quartiles by treatment status

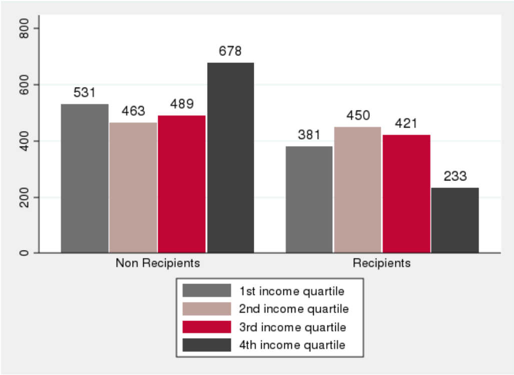
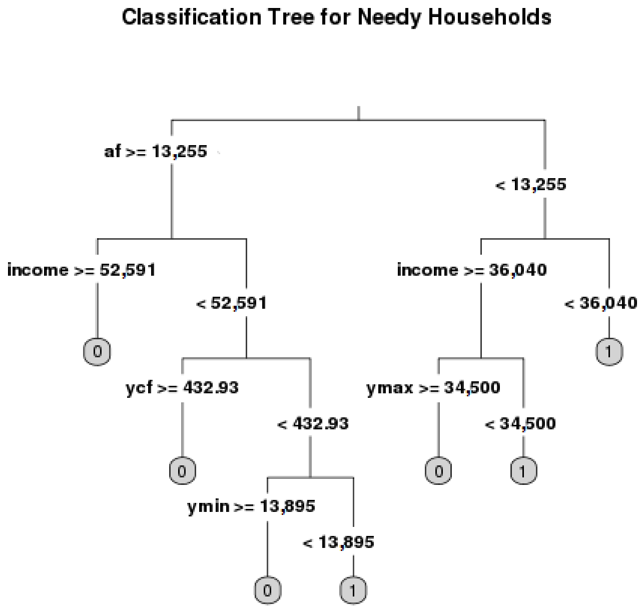


Figure 5: Decision tree output



Legend	
af	household yearly financial assets
income	income household yearly disposable
ycf	household income from financial assets
ymin	minimum individual labor income within the household
ymax	maximum individual labor income within the household

Table 1: Effect of the bonus on total non-durable consumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ammbonus	0.537*	0.114	0.134	0.217	0.214	0.208	0.373
	(0.281)	(0.273)	(0.275)	(0.275)	(0.276)	(0.276)	(0.276)
income	0.0231***	0.0217***	0.0216***	0.0205***	0.0204***	0.0204***	0.0202***
	(0.000859)	(0.000885)	(0.000922)	(0.000992)	(0.00100)	(0.00100)	(0.00106)
ncomp		159.4***	157.5***	176.8***	177.1***	159.3***	153.5***
		(27.04)	(26.74)	(27.13)	(27.15)	(28.92)	(28.45)
ncomp2		-10.20**	-9.796**	-11.73***	-11.78***	-9.947**	-8.189*
		-4.379	-4.335	-4.303	-4.309	-4.381	-4.200
age			0.797	2.042**	2.024**	1.972**	2.600***
			(0.764)	(0.866)	(0.866)	(0.862)	(0.894)
diploma				69.53***	69.59***	69.73***	68.24***
				(21.14)	(21.14)	(21.13)	(20.96)
degree				150.7***	151.5***	152.0***	148.7***
				(35.21)	(35.27)	(35.30)	(35.35)
male					6.974	2.076	9.219
					(18.58)	(18.93)	(19.06)
married						34.39	42.03*
						(22.45)	(22.35)
Constant	551.4***	247.1***	211.5***	94.08*	91.90*	109.4**	145.8**
	(27.72)	(34.22)	(44.67)	(55.16)	(55.26)	(55.47)	(74.23)
Regional FE	NO	NO	NO	NO	NO	NO	YES
N	3,646	3,646	3,646	3,646	3,646	3,646	3,646
$R2$	0.470	0.493	0.493	0.497	0.497	0.497	0.510

Notes: Estimation on the 2014 dataset. *-**-*** denotes statistical significance at 10%-5%-1%. Robust standard errors in parentheses. The unit of observation is the household and we restrict the sample to those with at least one employee among their members. See Table A.1 for a description of the covariates.

Table 2: Effect of the bonus on food consumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ammbonus	0.497***	0.207**	0.255***	0.263***	0.265***	0.262***	0.315***
	(0.101)	(0.0933)	(0.0944)	(0.0941)	(0.0940)	(0.0939)	(0.0960)
income	0.00527***	0.00426***	0.00400***	0.00387***	0.00389***	0.00386***	0.00375***
	(0.000271)	(0.000253)	(0.000253)	(0.000276)	(0.000278)	(0.000278)	(0.000286)
ncomp		114.2***	109.5***	111.5***	111.3***	102.3***	102.8***
		(11.19)	(10.62)	(10.72)	(10.74)	(11.38)	(11.07)
ncomp2		-7.850***	-6.868***	-7.040***	-7.007***	-6.072***	-5.951***
		(1.890)	(1.796)	(1.793)	(1.795)	(1.816)	(1.754)
age			1.936***	2.097***	2.109***	2.082***	2.237***
			(0.259)	(0.286)	(0.285)	(0.284)	(0.285)
diploma				11.98*	11.94*	12.01*	9.926
				(6.892)	(6.893)	(6.888)	(6.803)
degree				16.45	16.00	16.25	17.90
				(11.46)	(11.48)	(11.50)	(11.43)
male					-4.279	-6.773	-1.840
					(5.911)	(6.000)	(5.981)
married						17.51**	21.95***
						(7.633)	(7.635)
Constant	292.0***	78.00***	-8.416	-24.11	-22.77	-13.87	-30.81
	(9.319)	(13.03)	(15.83)	(18.50)	(18.73)	(19.15)	(25.05)
Regional FE	NO	NO	NO	NO	NO	NO	YES
N	3,646	3,646	3,646	3,646	3,646	3,646	3,646
R2	0.269	0.388	0.397	0.398	0.398	0.398	0.416

Notes: Estimation on the 2014 dataset. *-**-*** denotes statistical significance at 10%-5%-1%. Robust standard errors in parentheses.

Table 3: Effect of the bonus on consumption: heterogeneity analysis

	Total Consumption		Food Consumption	
	(1)	(2)	(3)	(4)
Panel A	No Difficulty	Difficulty	No Difficulty	Difficulty
ammbonus	-0.222	0.503	0.0518	0.357***
	(0.492)	(0.333)	(0.161)	(0.117)
N	1,357	2,289	1,357	2,289
*R*2	0.485	0.438	0.385	0.437
Panel B	No Difficulty (restricted)	Difficulty (restr.)	No Difficulty (restr.)	Difficulty (restr.)
ammbonus	0.0815	0.635	0.151	0.486***
	(0.340)	(0.444)	(0.116)	(0.166)
N	2,540	1,106	2,540	1,106
*R*2	0.478	0.474	0.398	0.450
Panel C	Liquidity unconstrained	Liquidity constrained	Liquidity unconstrained	Liquidity constrained
ammbonus	0.292	-0.721	0.294***	1.057***
	(0.273)	-2.040	(0.0987)	(0.393)
N	3,504	142	3,504	142
*R*2	0.522	0.365	0.413	0.613

Notes: Estimation on the 2014 dataset. All controls of specification (7) in Tables 1-2 included. In Panel A, *difficulty* =1 if the household reports making ends meet with great difficulty, with difficulty or with some difficulty. In Panel B, *difficulty* =1 if the household reports making ends meet with great difficulty or with difficulty. In Panel C, *liquidity constrained* =1 if the household was at least partially rejected a request for a mortgage, or would have liked to apply for it but had not because they thought they would have been rejected. *-**-*** denotes statistical significance at 10%-5%-1%. Robust standard errors in parentheses.

Table 4: Decision Tree, k-NN, LPM and random forest models performance compared

		Real status		
		Not Needy	Needy	Total
Panel A: decision tree				
	Not Needy	608	232	840
Predicted Status	Needy	447	1,334	1,781
	Total	1,055	1,566	2,621
	% Correctly Predicted	57.6%	85.1%	74.1%
Panel B: k-NN				
	Not Needy	593	244	837
Predicted Status	Needy	462	1,322	1,784
	Total	1,055	1,566	2,621
	% Correctly Predicted	56.2%	84.4%	73.0%
Panel C: random forest				
	Not Needy	680	218	898
Predicted Status	Needy	375	1,348	1,728
	Total	1,055	1,566	2,621
	% Correctly Predicted	64.4%	86.0%	77.3%
Panel D: LPM				
	Not Needy	608	208	816
Predicted Status	Needy	447	1,358	1,805
	Total	1,055	1,566	2,621
	% Correctly Predicted	57.6%	86.7%	75.0%

Notes: Out-of-sample estimation on the testing subsample of the 2010-2012 pooled dataset. *Needy* =1 if the household makes ends meet with great difficulty, with difficulty or with some difficulty.

Table 5: Effect of the bonus on consumption by predicted needy households (decision tree)

	Total Consumption		Food Consumption	
	(1)	(2)	(3)	(4)
	Not Needy	Needy	Not Needy	Needy
ammbonus	-0.527	0.710**	0.00907	0.369***
	(0.563)	(0.315)	(0.184)	(0.111)
N	1,146	2,500	1,146	2,500
*R*2	0.459	0.415	0.356	0.442

Notes: Estimation on the 2014 dataset. All controls of specification (7) in Tables 1-2 included. *Needy* =1 if according to the decision tree algorithm the household is predicted to make ends meet with great difficulty, with difficulty or with some difficulty. *-**-*** denotes statistical significance at 10%-5%-1%. Robust standard errors in parentheses.

Table 6: Decision Tree rule: predicted status and bonus recipient status

		Predicted status			
		Not Needy	Needy	Total	% Overlapping
Bonus recipient status	Not Recipient	715	1,446	2,161	33.0%
	Recipient	431	1,054	1,485	70.9%
	Total	1,146	2,500	3,646	
	% Overlapping	62.4%	42.1%		48.5%

Notes: Estimation on the 2014 dataset. *Needy* =1 if the household is predicted to make ends meet with great difficulty, with difficulty or with some difficulty.

Table A.1: Variables description: baseline regressions

total consumption	average monthly spending on all non-durable items
food consumption	average monthly spending on food eaten at home
ammbonus	overall amount of the bonus received monthly by the household
income	household annual disposable income (net of the bonus)
ncomp	number of household members
age	age of the head of the household
male	=1 if the head of the household is male
diploma	=1 if the head of the household has a upper secondary school diploma
degree	=1 if the head of the household has a university degree or more
married	=1 if the head of the household is married
liquidity constrained	=1 if the household is liquidity-constrained
difficulty	=1 if the household makes ends meet with great difficulty, difficulty or some difficulty
difficulty (restricted)	=1 if the household makes ends meet with great difficulty or with difficulty

Table A.2: Variables description: Machine Learning dataset

godabit	=1 if the household flat is property of household members
nfigli	number of household sons and daughters [used as discrete variable]
carta	=1 if some member of the household holds a credit card
bancomat	=1 if some member of the household holds a debit card
cartapre	=1 if some member of the household holds a prepaid card
altrab	=1 if some member of the household holds properties different than residence house
debita1	=1 if the household has house-related debts (acquisition or restructuring)
ncomp	number of household components [used as discrete variable]
income	household annual disposable income (net of the bonus)
yl	household income from employment
ytp	household income from retirement
ym	household income from self-employment
yca	household income from real estate
ycf	household income from financial assets
af	household financial assets
ymin	minimum individual labor income within the household
ymax	maximum individual labor income within the household
native	=1 if the head of the household is Italian
staciv	civil status of the head of the household
age	age of the head of the household
q	employment condition of the head household [used as discrete variable]
nperc	number of income perceivers within the household [used as discrete variable]
acom4c	dimensional class of the household municipality of residence [used as discrete variable]
degree	=1 if the head of the household has a university degree or more
diploma	=1 if the head of the household has a upper secondary school diploma
compulsory	=1 if the head of the household has a compulsory education
africa	=1 if the head of the household is African
asia	=1 if the head of the household is Asian
east europe	=1 if the head of the household is East-European
south america	=1 if the head of the household is South-American
south	=1 if the head of the household lives in the South of Italy

Table A.3: Probability of being a needy household

	(1)	(2)	(3)
af	0.0000149	-0.00000999	0.000107
	(0.000117)	(0.000174)	(0.000186)
income	-0.00873***	-0.00873***	-0.00874***
	(0.000426)	(0.000653)	(0.000535)
ycf	-0.00159	-0.0116**	-0.00131
	(0.00513)	(0.00483)	(0.00519)
ymax	0.000227	-0.000242	0.000424
	(0.00152)	(0.00126)	(0.00252)
ymin	-0.00653***	-0.00812***	-0.00518**
	(0.00140)	(0.00112)	(0.00227)
Constant	1.035***	1.040***	1.037***
	(0.0203)	(0.0175)	(0.0296)
N	7,802	3,939	3,863
R2	0.223	0.229	0.222

Notes: Columns (1), (2) and (3) are estimated on the 2010-2012 pooled dataset, the 2010 dataset and the 2012 dataset, respectively. We focus on a Linear Probability Model to ease the interpretation of the coefficients. All coefficients (excluding the constant term) have been multiplied by 1,000. *Needy* =1 if the household makes ends meet with great difficulty, with difficulty or with some difficulty. *-**-*** denotes statistical significance at 10%-5%-1%. Robust standard errors in parentheses.

RECENTLY PUBLISHED "TEMI" (*)

N. 1129 – *The effects of central bank's verbal guidance: evidence from the ECB*, by Maddalena Galardo and Cinzia Guerrieri (July 2017).

N. 1130 – *The Bank of Italy econometric model: an update of the main equations and model elasticities*, by Guido Bulligan, Fabio Busetti, Michele Caivano, Pietro Cova, Davide Fantino, Alberto Locarno, Lisa Rodano (July 2017).

N. 1131 – *Venture capitalists at work: what are the effects on the firms they finance?*, by Raffaello Bronzini, Giampaolo Caramellino and Silvia Magri (September 2017).

N. 1132 – *Low frequency drivers of the real interest rate: a band spectrum regression approach*, by Fabio Busetti and Michele Caivano (September 2017).

N. 1133 – *The real effects of relationship lending*, by Ryan Banerjee, Leonardo Gambacorta and Enrico Sette (September 2017).

N. 1134 – *Credit demand and supply: a two-way feedback relation*, by Ugo Albertazzi and Lucia Esposito (September 2017).

N. 1135 – *Legislators' behaviour and electoral rules: evidence from an Italian reform*, by Giuseppe Albanese, Marika Cioffi and Pietro Tommasino (September 2017).

N. 1136 – *Macroeconomic effects of non-standard monetary policy measures in the euro area: the role of corporate bond purchases*, by Anna Bartocci, Lorenzo Burlon, Alessandro Notarpietro and Massimiliano Pisani (September 2017).

N. 1137 – *On secular stagnation and low interest rates: demography matters*, by Giuseppe Ferrero, Marco Gross and Stefano Neri (September 2017).

N. 1138 – *Tony Atkinson and his legacy*, by Rolf Aaberge, François Bourguignon, Andrea Brandolini, Francisco H. G. Ferreira, Janet C. Gornick, John Hills, Markus Jäntti, Stephen P. Jenkins, Eric Marlier, John Micklewright, Brian Nolan, Thomas Piketty, Walter J. Radermacher, Timothy M. Smeeding, Nicholas H. Stern, Joseph Stiglitz and Holly Sutherland (September 2017).

N. 1139 – *Credit misallocation during the European financial crisis*, by Fabiano Schivardi, Enrico Sette and Guido Tabellini (September 2017).

N. 1140 – *Natural rates across the Atlantic*, by Andrea Gerali and Stefano Neri (September 2017).

N. 1141 – *A quantitative analysis of risk premia in the corporate bond market*, by Sara Cecchetti (October 2017).

N. 1142 – *Monetary policy in times of debt*, by Mario Pietrunti and Federico M. Signoretti (October 2017).

N. 1143 – *Capital misallocation and financial development: a sector-level analysis*, by Daniela Marconi and Christian Upper (October 2017).

N. 1144 – *Leaving your mamma: why so late in Italy?*, by Enrica Di Stefano (October 2017).

N. 1145 – *A Financial Conditions Index for the CEE economies*, by Simone Auer (October 2017).

N. 1146 – *Internal capital markets in times of crisis: the benefit of group affiliation in Italy*, by Raffaele Santioni, Fabio Schiantarelli and Philip E. Strahan (October 2017).

N. 1147 – *Consistent inference in fixed-effects stochastic frontier models*, by Federico Belotti and Giuseppe Ilardi (October 2017).

N. 1148 – *Investment decisions by European firms and financing constraints*, by Andrea Mercatanti, Taneli Mäkinen and Andrea Silvestrini (October 2017).

(*) 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.

2015

AABERGE R. and A. BRANDOLINI, *Multidimensional poverty and inequality,* in A. B. Atkinson and F. Bourguignon (eds.), Handbook of Income Distribution, Volume 2A, Amsterdam, Elsevier, **TD No. 976 (October 2014).**

ALBERTAZZI U., G. ERAMO, L. GAMBACORTA and C. SALLEO, *Asymmetric information in securitization: an empirical assessment*, Journal of Monetary Economics, v. 71, pp. 33-49, **TD No. 796 (February 2011).**

ALESSANDRI P. and B. NELSON, *Simple banking: profitability and the yield curve,* Journal of Money, Credit and Banking, v. 47, 1, pp. 143-175, **TD No. 945 (January 2014).**

ANTONIETTI R., R. BRONZINI and G. CAINELLI, *Inward greenfield FDI and innovation,* Economia e Politica Industriale, v. 42, 1, pp. 93-116, **TD No. 1006 (March 2015).**

BARONE G. and G. NARCISO, *Organized crime and business subsidies: Where does the money go?,* Journal of Urban Economics, v. 86, pp. 98-110, **TD No. 916 (June 2013).**

BRONZINI R., *The effects of extensive and intensive margins of FDI on domestic employment: microeconomic evidence from Italy*, B.E. Journal of Economic Analysis & Policy, v. 15, 4, pp. 2079-2109, **TD No. 769 (July 2010).**

BUGAMELLI M., S. FABIANI and E. SETTE, *The age of the dragon: the effect of imports from China on firm-level prices*, Journal of Money, Credit and Banking, v. 47, 6, pp. 1091-1118, **TD No. 737 (January 2010).**

BULLIGAN G., M. MARCELLINO and F. VENDITTI, *Forecasting economic activity with targeted predictors,* International Journal of Forecasting, v. 31, 1, pp. 188-206, **TD No. 847 (February 2012).**

BUSETTI F., *On detecting end-of-sample instabilities,* in S.J. Koopman, N. Shepard (eds.), Unobserved Components and Time Series Econometrics, Oxford, Oxford University Press, **TD No. 881 (September 2012).**

CESARONI T., *Procyclicality of credit rating systems: how to manage it,* Journal of Economics and Business, v. 82. pp. 62-83, **TD No. 1034 (October 2015).**

CIARLONE A., *House price cycles in emerging economies,* Studies in Economics and Finance, v. 32, 1, **TD No. 863 (May 2012).**

CUCINIELLO V. and F. M. SIGNORETTI, *Large bank, loan rate markup and monetary policy,* International Journal of Central Banking, v. 11, 3, pp. 141-177, **TD No. 987 (November 2014).**

DE BLASIO G., D. FANTINO and G. PELLEGRINI, *Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds,* Industrial and Corporate Change, v. 24, 6, pp. 1285-1314, **TD No. 792 (February 2011).**

DEPALO D., R. GIORDANO and E. PAPAPETROU, *Public-private wage differentials in euro area countries: evidence from quantile decomposition analysis,* Empirical Economics, v. 49, 3, pp. 985-1115, **TD No. 907 (April 2013).**

DI CESARE A., A. P. STORK and C. DE VRIES, *Risk measures for autocorrelated hedge fund returns,* Journal of Financial Econometrics, v. 13, 4, pp. 868-895, **TD No. 831 (October 2011).**

FANTINO D., A. MORI and D. SCALISE, *Collaboration between firms and universities in Italy: the role of a firm's proximity to top-rated departments,* Rivista Italiana degli economisti, v. 1, 2, pp. 219-251, **TD No. 884 (October 2012).**

FRATZSCHER M., D. RIMEC, L. SARNOB and G. ZINNA, *The scapegoat theory of exchange rates: the first tests,* Journal of Monetary Economics, v. 70, 1, pp. 1-21, **TD No. 991 (November 2014).**

NOTARPIETRO A. and S. SIVIERO, *Optimal monetary policy rules and house prices: the role of financial frictions,* Journal of Money, Credit and Banking, v. 47, S1, pp. 383-410, **TD No. 993 (November 2014).**

RIGGI M. and F. VENDITTI, *The time varying effect of oil price shocks on euro-area exports,* Journal of Economic Dynamics and Control, v. 59, pp. 75-94, **TD No. 1035 (October 2015).**

TANELI M. and B. OHL, *Information acquisition and learning from prices over the business cycle,* Journal of Economic Theory, 158 B, pp. 585–633, **TD No. 946 (January 2014).**

2016

ALBANESE G., G. DE BLASIO and P. SESTITO, *My parents taught me. evidence on the family transmission of values,* Journal of Population Economics, v. 29, 2, pp. 571-592, **TD No. 955 (March 2014).**

ANDINI M. and G. DE BLASIO, *Local development that money cannot buy: Italy's Contratti di Programma,* Journal of Economic Geography, v. 16, 2, pp. 365-393, **TD No. 915 (June 2013).**

BARONE G. and S. MOCETTI, *Inequality and trust: new evidence from panel data,* Economic Inquiry, v. 54, pp. 794-809, **TD No. 973 (October 2014).**

BELTRATTI A., B. BORTOLOTTI and M. CACCAVAIO, *Stock market efficiency in China: evidence from the split-share reform,* Quarterly Review of Economics and Finance, v. 60, pp. 125-137, **TD No. 969 (October 2014).**

BOLATTO S. and M. SBRACIA, *Deconstructing the gains from trade: selection of industries vs reallocation of workers,* Review of International Economics, v. 24, 2, pp. 344-363, **TD No. 1037 (November 2015).**

BOLTON P., X. FREIXAS, L. GAMBACORTA and P. E. MISTRULLI, *Relationship and transaction lending in a crisis*, Review of Financial Studies, v. 29, 10, pp. 2643-2676, **TD No. 917 (July 2013).**

BONACCORSI DI PATTI E. and E. SETTE, *Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register,* Journal of Financial Intermediation , v. 25, 1, pp. 54-76, **TD No. 848 (February 2012).**

BORIN A. and M. MANCINI, *Foreign direct investment and firm performance: an empirical analysis of Italian firms,* Review of World Economics, v. 152, 4, pp. 705-732, **TD No. 1011 (June 2015).**

BRAGOLI D., M. RIGON and F. ZANETTI, *Optimal inflation weights in the euro area,* International Journal of Central Banking, v. 12, 2, pp. 357-383, **TD No. 1045 (January 2016).**

BRANDOLINI A. and E. VIVIANO, *Behind and beyond the (headcount) employment rate,* Journal of the Royal Statistical Society: Series A, v. 179, 3, pp. 657-681, **TD No. 965 (July 2015).**

BRIPI F., *The role of regulation on entry: evidence from the Italian provinces,* World Bank Economic Review, v. 30, 2, pp. 383-411, **TD No. 932 (September 2013).**

BRONZINI R. and P. PISELLI, *The impact of R&D subsidies on firm innovation,* Research Policy, v. 45, 2, pp. 442-457, **TD No. 960 (April 2014).**

BURLON L. and M. VILALTA-BUFI, *A new look at technical progress and early retirement,* IZA Journal of Labor Policy, v. 5, **TD No. 963 (June 2014).**

BUSETTI F. and M. CAIVANO, *The trend–cycle decomposition of output and the Phillips Curve: bayesian estimates for Italy and the Euro Area,* Empirical Economics, V. 50, 4, pp. 1565-1587, **TD No. 941 (November 2013).**

CAIVANO M. and A. HARVEY, *Time-series models with an EGB2 conditional distribution,* Journal of Time Series Analysis, v. 35, 6, pp. 558-571, **TD No. 947 (January 2014).**

CALZA A. and A. ZAGHINI, *Shoe-leather costs in the euro area and the foreign demand for euro banknotes,* International Journal of Central Banking, v. 12, 1, pp. 231-246, **TD No. 1039 (December 2015).**

CESARONI T. and R. DE SANTIS, *Current account "core-periphery dualism" in the EMU,* The World Economy, v. 39, 10, pp. 1514-1538, **TD No. 996 (December 2014).**

CIANI E., *Retirement, Pension eligibility and home production,* Labour Economics, v. 38, pp. 106-120, **TD No. 1056 (March 2016).**

CIARLONE A. and V. MICELI, *Escaping financial crises? Macro evidence from sovereign wealth funds' investment behaviour,* Emerging Markets Review, v. 27, 2, pp. 169-196, **TD No. 972 (October 2014).**

CORNELI F. and E. TARANTINO, *Sovereign debt and reserves with liquidity and productivity crises,* Journal of International Money and Finance, v. 65, pp. 166-194, **TD No. 1012 (June 2015).**

D'AURIZIO L. and D. DEPALO, *An evaluation of the policies on repayment of government's trade debt in Italy,* Italian Economic Journal, v. 2, 2, pp. 167-196, **TD No. 1061 (April 2016).**

DE BLASIO G., G. MAGIO and C. MENON, *Down and out in Italian towns: measuring the impact of economic downturns on crime,* Economics Letters, 146, pp. 99-102, **TD No. 925 (July 2013).**

DOTTORI D. and M. MANNA, *Strategy and tactics in public debt management,* Journal of Policy Modeling, v. 38, 1, pp. 1-25, **TD No. 1005 (March 2015).**

ESPOSITO L., A. NOBILI and T. ROPELE, *The management of interest rate risk during the crisis: evidence from Italian banks,* Journal of Banking & Finance, v. 59, pp. 486-504, **TD No. 933 (September 2013).**

LIBERATI D., M. MARINUCCI and G. M. TANZI, *Science and technology parks in Italy: main features and analysis of their effects on hosted firms*, Journal of Technology Transfer, v. 41, 4, pp. 694-729, **TD No. 983 (November 2014).**

MARCELLINO M., M. PORQUEDDU and F. VENDITTI, *Short-Term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility,* Journal of Business & Economic Statistics , v. 34, 1, pp. 118-127, **TD No. 896 (January 2013).**

RODANO G., N. SERRANO-VELARDE and E. TARANTINO, *Bankruptcy law and bank financing,* Journal of Financial Economics, v. 120, 2, pp. 363-382, **TD No. 1013 (June 2015).**

ZINNA G., *Price pressures on UK real rates: an empirical investigation,* Review of Finance,v. 20, 4, pp. 1587-1630, **TD No. 968 (July 2014).**

2017

ADAMOPOULOU A. and G.M. TANZI, *Academic dropout and the great recession,* Journal of Human Capital, V. 11, 1, pp. 35–71, **TD No. 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, **TD No. 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, **TD No. 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, **TD No. 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, **TD No. 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 **TD No. 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, **TD No. 1049 (January 2016).**

BURLON L., *Public expenditure distribution, voting, and growth,* Journal of Public Economic Theory,, v. 19, 4, pp. 789–810, **TD No. 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, **TD No. 1074 (July 2016).**

BUSETTI F., *Quantile aggregation of density forecasts,* Oxford Bulletin of Economics and Statistics, v. 79, 4, pp. 495-512, **TD No. 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, **TD No. 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, **TD No. 1018 (July 2015).**

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, **TD No. 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, **TD No. 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, **TD No. 942 (November 2013).**

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, **TD No. 1052 (March 2016).**

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, **TD No. 1108 (April 2017).**

LOBERTO M. and C. PERRICONE, *Does trend inflation make a difference?,* Economic Modelling, v. 61, pp. 351–375, **TD No. 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, **D 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, **TD No. 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, **TD No. 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, **TD No. 752 (March 2010).**

MOCETTI S. and E. VIVIANO, *Looking behind mortgage delinquencies,* Journal of Banking & Finance, v. 75, pp. 53-63, **TD No. 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, **TD No. 887 (October 2012).**

PALAZZO F., *Search costs and the severity of adverse selection,* Research in Economics, v. 71, 1, pp. 171-197, **TD No. 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, **TD No. 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, **TD No. 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, **TD No. 1016 (June 2015).**

SEGURA A. and J. SUAREZ, *How excessive is banks' maturity transformation?,* Review of Financial Studies, v. 30, 10, pp. 3538–3580, **TD No. 1065 (April 2016).**

VACCA V., *An unexpected crisis? Looking at pricing effectiveness of heterogeneous banks,* Economic Notes, v. 46, 2, pp. 171–206, **TD No. 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, **TD No. 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, **TD No. 1104 (February 2017).**

FORTHCOMING

ADAMOPOULOU A. and E. KAYA, *Young Adults living with their parents and the influence of peers,* Oxford Bulletin of Economics and Statistics, **TD No. 1038 (November 2015).**

ALBANESE G., G. DE BLASIO and P. SESTITO, *Trust, risk and time preferences: evidence from survey data,* International Review of Economics, **TD No. 911 (April 2013).**

BOFONDI M., L. CARPINELLI and E. SETTE, *Credit supply during a sovereign debt crisis,* Journal of the European Economic Association, **TD No. 909 (April 2013).**

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, **TD No. 1077 (July 2016).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, **TD No. 787 (January 2011).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, **TD No. 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, **TD No. 898 (January 2013).**

NATOLI F. and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, **TD No. 1025 (July 2015).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment,* Macroeconomic Dynamics, **TD No. 871 (July 2012).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, **TD No. 1100 (February 2017).**